THE ORDERED WEIGHTED AVERAGING OPERATORS

Theory and Applications

THE ORDERED WEIGHTED AVERAGING OPERATORS

Theory and Applications

edited by

Ronald R. Yager
Iona College
New Rochelle, New York USA

and

Janusz Kacprzyk
Polish Academy of Sciences
Warsaw, Poland

Springer-Science+Business Media, LLC

Library of Congress Cataloging-in-Publication Data

A C.I.P. Catalogue record for this book is available
from the Library of Congress.

ISBN 978-1-4613-7806-8 ISBN 978-1-4615-6123-1 (eBook)
DOI 10.1007/978-1-4615-6123-1

Printed on acid-free paper.

Table of Contents

PREFACE

Aggregation plays a central role in many of the technological tasks we are faced with. The importance of this process will become even greater as we move more and more toward becoming an information-centered society, as is happening with the rapid growth of the Internet and the World Wide Web. Here we shall be faced with many issues related to the fusion of information.

One very pressing issue here is the development of mechanisms to help search for information, a problem that clearly has a strong aggregation-related component. More generally, in order to model the sophisticated ways in which human beings process information, as well as going beyond the human capabilities, we need provide a basket of aggregation tools.

The centrality of aggregation in human thought can be be very clearly seen by looking at neural networks, a technology motivated by modeling the human brain. One can see that the basic operations involved in these networks are learning and aggregation.

The Ordered Weighted Averaging (OWA) operators provide a parameterized family of aggregation operators which include many of the well-known operators such as the maximum, minimum and the simple average.

While the original insights leading to the creation of these operators owe a large debt to the rich framework provided by fuzzy sets theory, the use of these operators are not in any way restricted to fuzzy sets. For example, the simple process of taking the average of a collection of numbers in which we give preferential treatment to the higher numbers, as may be the case when we are averaging multiple estimates of future interest rates and want to be conservative, is an example of the use of OWA operators.

In this book we see papers from a number of domains where these operators have been applied: decision analysis under uncertainty, multi-criteria decision making, multi-person decision making and consensus formation, flexible database querying and information retrieval, learning and classification.

The editors would like to thank all the contributors without whose efforts this volume would be impossible.

Ronald R. Yager, *New York*
Janusz Kacprzyk, *Warsaw*

1

BASIC ISSUES IN AGGREGATION

KOLMOGOROV'S THEOREM AND ITS IMPACT ON SOFT COMPUTING

Hung T. Nguyen[1] and Vladik Kreinovich[2]
[1]Department of Mathematical Sciences,
New Mexico State University,
Las Cruces, NM 88003, USA,
e-mail hunguyen@nmsu.edu
[2]Department of Computer Science,
University of Texas at El Paso,
El Paso, TX 79968, USA,
e-mail vladik@cs.utep.edu

Abstract. In this chapter, we describe various applications of the Kolmogorov's theorem on representing continuous functions of several variables (as superpositions of functions of one and two variables) to soft computing. Kolmogorov's theorem leads to a theoretical justification, as well as to design methodologies, for *neural networks.* In the design of intelligent systems, Kolmogorov's theorem is used to show that general *logical operators* can be expressed in terms of basic fuzzy logic operations.

In the area of *reliable computing* (i.e., computing that takes into consideration the accuracy of the input data), an extended version of Kolmogorov's theorem justifies the need to use operations with three or more operands in soft computing. Such operations have already been actively used in soft computing; the simplest (and, so far, most used) of such operations are *ordered weighted averaging* (OWA) operators proposed by R. R. Yager.

Keywords: Neural networks, soft computing, Kolmogorov's theorem, reliable (and interval) computing, fuzzy logic, logical operations

1. INTRODUCTION

The area of Soft Computing, as defined by L. A. Zadeh, deals with methodologies and techniques for analyzing complex systems. These methodologies and

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

techniques include probability and statistics, neural networks, fuzzy logic, etc. Normally in scientific investigations, heuristics precede formal methods. The path from heuristics to formal methods is paved, in general, by some mathematical results, which may belong to another domain of pure mathematics. The story we are going to report in this chapter is no exception.

Among tremendous contributions to mathematics, ranging from Probability Theory to Analysis, A. Kolmogorov has left us with many fundamental theorems. His investigations into Hilbert's famous conjecture (known as 13th problem) has led to Kolmogorov's famous 1957 theorem on the possibility of representing continuous functions of several variables as superpositions of continuous functions of one and two variables. This theorem is the one we refer to in this chapter. While Kolmogorov's theorem is a purely mathematical, non-algorithmic result (disproving Hilbert's conjecture) that has applications in fields related to approximations of functions, its impact on Artificial Intelligence has not been noticed until recently.

In the next section, we will first describe Hilbert's 13th problem and Kolmogorov's theorem. In Sections 3–5, we will describe how Kolmogorov's theorem impacts soft computing. In general, *complex* systems are difficult to describe directly, so, a natural way to describe them is:

- to describe *simple* systems, and then
- to *combine* the descriptions of simple systems into a description of a complex system.

The simplest possible combination of systems is when the output of one subsystem becomes an input to another subsystem. In mathematical terms, such a combination corresponds to *superposition* of the corresponding functions. For example:

- In the *neural network* methodology, we describe an arbitrary complex system as a combination (superposition) of simple elements called *neurons*, in which:
  - first, input signals $x_1, \ldots, x_n$ undergo linear combination $x_1, \ldots, x_n \rightarrow x = w_1 x_1 + \ldots w_n x_n - w_0$;
  - second, a non-linear function $g(x)$ (called *activation function*) is applied to the result x of the linear combination, yielding the value $y = g(x)$.
- In *fuzzy systems* methodology, a general description of complex systems comes from combining such simpler descriptions as:
  - *membership functions* (that describe fuzzy properties);
  - *hedges* like "very" (that describe a degree to which a property is true), and
  - *t-norms* and *t-conorms* (fuzzy analogues of "and" and "or" logical operations), that *combine* the simpler descriptions.

In both cases, Kolmogorov's theorem shows that these methodologies can actually describe an arbitrary complex system. A similar result is proven for *logical* operators in fuzzy logic. Here, the impact of Kolmogorov's theorem is this: All (continuous) logical operations in fuzzy logic can be represented by basic logical connectives such as t-norms, t-conorms, and hedges (see Section 5).

At first glance, these results seem to indicate that operations with one or two arguments are sufficient to represent arbitrarily complex systems. This is, however, only true if we assume that the inputs $x_1, \dots, x_n$ are precisely known. If we take the input uncertainty into consideration (e.g., by assuming that the input belongs to an *interval*, or that the input is described by a *fuzzy* set), then, in contrast to numerical computations, there is a need to use operations with more than two variables on computers. In other words, Kolmogorov's theorem cannot be extended to the *interval-valued* and *fuzzy-valued* cases (see Section 6). In particular, this result justifies the necessity of using logical operations with three or more arguments, an example of which are Ordered Weighted Averaging (OWA) operations proposed by R. R. Yager (see, e.g., (Yager *et al* 1994)).

2. KOLMOGOROV'S THEOREM: EVERY FUNCTION OF SEVERAL VARIABLES CAN BE REPRESENTED AS A SUPERPOSITION OF FUNCTIONS OF ONE OR TWO VARIABLES

Let us first give a brief history of the mathematical developments that led to Kolmogorov's theorem (for a detailed history, see, e.g., (Boyer *et al.* 1991) and (Nakamura *et al.* 1993)).

2.1. Ideally, Computations Must Be Represented as a Sequence of Easy-to-Tabulate Steps

Since ancient times, people know how to solve linear equations $a_0 + a_1x = 0$: namely, if $a_1 \neq 0$, then the solution x can be obtained by simple arithmetic operations: $x = -a_0/a_1$. Even if we do not know a simple algorithm for division, the resulting function $x(a_0, a_1)$ (that inputs two real numbers a_0 and a_1 and returns the root x) can be actually *tabulated* as a 2-D table (similar to the multiplication table that is familiar to every schoolkid).

Next in complexity to linear equations are *quadratic* equations $a_0+a_1x+a_2x^2 = 0$. The desired root x is a function of *three* variables: a_0, a_1, and a_2. In principle, we could follow the example of linear equations and tabulate, for different values of a_0, a_1, and a_2, the corresponding values of x. However, this idea requires a 3-D table, and it is very difficult to represent a 3-D table on a 2-D sheet of paper (or, in ancient terms, on a 2-D papyrus :-). So, in order to solve quadratic equations, we must describe the transformation from (a_0, a_2, a_2) to x as a sequence of computation steps that correspond to *easy-to-tabulate* functions (i.e., functions of one or two variables).

In mathematical terms, a sequence of steps corresponds to a *superposition* of the corresponding functions. So, in these terms, we would like to represent the function $x(a_0, a_1, a_2)$ as a superposition of easy-to-tabulate functions.

of 5th and higher order resulted in a famous proof by Galois that there is no way to express such solutions in terms of basic algebraic operations ($+, -, \times, :, \sqrt[n]{\ }$). In other words, if $n \geq 5$, then a function f, that maps $(a_0, ..., a_n)$ into a solution of equation (2.1), cannot be represented as a superposition of these basic algebraic operations.

2.2. Hilbert's Problem and its Unexpected Solution: Kolmogorov's Theorem

The fact that the basic algebraic operations are not sufficient for solving equations of fifth and higher order does not necessarily mean that with other functions of one or two variables, like $\exp(x)$, $\sin(x)$, $\log(x)$, etc., we would not be more successful. However, numerous attempts to add these functions to the basic algebraic ones did not lead to any successful solution for fifth- and higher-order equations. The failure of these attempts led David Hilbert, the leading mathematician of the late 19th century, to formulate a natural hypothesis: that *not only one cannot express the solution of higher-order algebraic equations in terms of basic algebraic operations, but no matter what functions of one or two variables we add to these operations, we still won't be able to express the general solution.*

Hilbert even included this hypothesis (under No. 13) into the list of 23 major problems that he, on request from the mathematical community, formulated in 1900 as a challenge for the 20th century.

This problem remained a challenge until 1957, when (rather unexpectedly) Kolmogorov (Kolmogorov 1957) proved the following result:

THEOREM 2.1. (Kolmogorov 1957) *An arbitrary continuous function $f(x_1, ..., x_n)$ on an n-dimensional cube (of arbitrary dimension n) can be represented as a superposition of addition and functions of one variable.*

This result was later simplified and improved by D. Sprecher in (Sprecher 1965) (further improvements were later described in (Sprecher 1972)):

THEOREM 2.2. (Sprecher 1965) *For an arbitrary "box"*

$$B = [x_1^-, x_1^+] \times \ldots \times [x_n^-, x_n^+],$$

every real continuous function $f : B \to R$, can be represented as

$$f(x_1, ..., x_n) = \sum_{0 \leq q \leq 2n} \chi[\sum_{1 \leq p \leq n} \psi_{pq}(x_p)],$$

for some continuous functions χ and ψ_{pq}.

Such representation was known already to Babylonians; in modern terms, their solution for quadratic equations is

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

It is a superposition of arithmetic operations and a square root.

Unlike the solution of a linear equation, this solution is not a sequence of simple arithmetic operations: in addition to arithmetic operations, it includes an (algebraic) operation of taking a *square root*. The square root $z = \sqrt{u}$ is actually a solution of the quadratic equation $z^2 - u = 0$. So, if no specific algorithm for finding square roots is presented, this "solution" simply reduces the original quadratic equation $a_0 + a_1 x + a_2 x^2 = 0$ to the new one $z^2 - u = 0$. If we did not know how to solve the original equation, then we do not know how to solve the new one either. So, what did we gain by this reduction?

- Solving the original equation means, for every possible triple (a_0, a_1, a_2), to find a value x for which the desired equality is true. In terms of modern mathematics, we must describe a function $x(a_0, a_1, a_2)$ of *three* variables a_0, a_1, and a_2. We can compute the values of this function for different triples (i.e., solve different equations); however, it is *difficult to* tabulate or *store* the values of a function of three variables, because a natural tabulation would require a 3-D table.
- Solving the new equation $z^2 - u = 0$ means that we need, for every possible real number u, to find a value z for which this equality is true. The resulting function $z = \sqrt{u}$ is a function of *one* variable u and is, therefore, *easy to tabulate* (before computers, tables of square roots were actually widely used by engineers and scientists).

In short, the above formula for solving quadratic equations reduces the original function of three variables to functions of one and two variables $(+, -, \times, :, \sqrt{})$, and thus, reduces the original difficult-to-tabulate computational problem to a sequence of easy-to-tabulate computations.

A natural question is: can we do the same thing for cubic, quartic, and higher order equations? In the 16th century, Tartaglia, Cardan, and Ferrari succeeded in designing formulas for solving cubic equations $a_0 + a_1 x + a_2 x^2 + a_3 x^3 = 0$ and quartic equations $a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 = 0$. These formulas express the root $x(a_0, \ldots, a_3)$ (corr., $x(a_0, \ldots, x_4)$) of the corresponding equation as a superposition of functions of one or two variables: namely, of arithmetic operations $(+, -, \times, :)$ and of the corresponding root ($\sqrt[3]{\ }$ or $\sqrt[4]{\ }$). Thus, if we tabulate these functions, we can easily solve an arbitrary cubic (quartic) equation.

The success of these formulas led to the natural hypothesis that similar formulas are possible for higher-order equations. Numerous attempts to find such a general solution for algebraic equations

$$a_0 + a_1 x + \ldots + a_n x^n = 0 \tag{2.1}$$

3. NEURAL NETWORKS: EVERY DETERMINISTIC COMPLEX SYSTEM CAN BE SIMULATED BY A NETWORK OF LINEAR ELEMENTS AND SINGLE-INPUT NONLINEAR ELEMENTS

In this section, we will consider *deterministic* complex systems, i.e., systems in which the values of several quantities $x_1, \ldots, x_n$ uniquely determine the value of every other quantity y. A typical example of such a system is a *physical* system for which all equations are known.

In mathematical terms, the fact that y is uniquely determined by $x_1, \ldots, x_n$ means that y is a *function* of $x_1, \ldots, x_n$: $y = f(x_1, \ldots, x_n)$. This function must satisfy the following two natural requirements:

- First, each variable is normally *bounded*: e.g., possible values of velocity are bounded by the speed of light; possible values of coordinates are bounded by the area in which the system is operating, etc. Therefore, we can assume that each of the variables x_i can only take values from a certain interval $[x_i^-, x_i^+]$. So, the function f is only defined on a box $[x_1^-, x_1^+] \times \ldots \times [x_n^-, x_n^+]$.
- Second, in most real-life deterministic systems, small changes in x_i lead to small changes in y, i.e. in mathematical terms, the function f is *continuous*.

Deterministic complex systems can be very complicated. To simplify their analysis and simulation, it is therefore desirable to represent complex systems as compositions of simpler ones. Which systems are simple? The complexity of a system is determined by:

- the *number of quantities* n that determine the state of the system, and
- the *complexity* of the function $f(x_1, \ldots, x_n)$ (linear functions $f(x_1, \ldots, x_n) = w_1 x_1 + \ldots + w_n x_n - w_0$ are the simplest).

A natural question is: can we represent an arbitrary system as a composition of systems with $n = 1$ and of linear systems (i.e., systems that are characterized by linear functions f)? *Kolmogorov's theorem shows that such a representation is indeed possible.*

If we represent the component "simple" systems in hardware, then we can conclude that an arbitrary continuous function can be computed by a network of linear elements and 1-input-1-output elements (that compute functions of one variable). Namely, for an arbitrary continuous function $f(x_1, \ldots, x_n)$, we can set up a 4-layer network that computes this function:

- First, the input signals $x_1, \ldots, x_n$ go into the elements that implement functions $\psi_{pq}(x_p)$.
- Second, the signals generated by the first layer go into addition elements, generating the values $v_q = \sum \psi_{pq}(x_p)$.
- Third, the results v_q of the second layer go into the elements that implement the function $\chi(x)$, resulting in $\chi(v_q)$.
- Finally, the results $\xi(v_q)$ of the third layer go into the addition element, generating the desired value $f(x_1, \ldots, x_n) = \sum \chi(v_q)$.

Hardware elements that implement functions of one variable and addition are often called *neurons*, because biological neurons in our brain can be (within a good approximation) described as devices with n real-valued inputs $x_1, \ldots, x_n$ and an output $y = g(w_1x_1 + \ldots + \ldots w_nx_n - w_0)$, where w_i are constants called *weights* (w_0 is also called a *threshold*), and $g(z)$ is a function of one variable called an *activation function*. The network in which the output of some neurons is sent as input to others is called a *neural network*. Thus, Kolmogorov's theorem proves that in principle, *an arbitrary continuous function can be computed by a neural network.*

The question of whether neural networks are indeed *universal computational devices* (in this sense) was an open problem until 1987, when R. Hecht-Nielsen noticed (Hecht-Nielsen 1987) that the (positive) answer follows from Kolmogorov's 1957 result. So, *Kolmogorov's theorem justifies the neural networks methodology for simulating arbitrary deterministic (crisp) systems.*

In the following section, we will show that an appropriate modification of Kolmogorov's theorem justifies a similar approach to simulating arbitrary *fuzzy* systems.

4. FUZZY SYSTEMS: JUSTIFICATION OF FUZZY CONTROL METHODOLOGY

In the previous section, we considered *deterministic* complex systems, in which the knowledge of some quantities $x_1, \ldots, x_n$ uniquely determines the values of all other quantity y that characterize the system. In many real life situations, for given values $x_1, \ldots, x_n$ of input variables, we cannot predict the exact value of the output variables y; instead, for different possible y, we have different *truth values* (sometimes called *degrees of belief* or *degrees of certainty*) that characterize to what extent this number y can be the possible value of the desired quantity. So, for every $x_1, \ldots, x_n$, instead of a single value of y, we get different possible values of y, with different truth values attached to each value. In other words, for every $x_1, \ldots, x_n$, instead of a single value of y, we get a *fuzzy set* that characterizes possible values of y. Hence, systems represented in this manner can be naturally called *fuzzy systems.*

These truth values are usually characterized by numbers from the interval [0, 1]. Therefore:

- instead of a mapping $x_1, \ldots, x_n \to y$ (typical for deterministic systems), that determined y based on x_i,
- we have a mapping $x_1, \ldots, x_n, y \to [0, 1]$ that determines, for given values of x_i and y, to what extent y is possible for given x_i.

In the following text, we will denote this mapping by $F(x_1, \ldots, x_n, y)$.

Comment. Similarly to the continuous case, it is natural to make two assumptions:

- that the possible values of x_i and y are *bounded*, and
- that small changes in x_i and y leads to small changes in the resulting truth values, i.e., that the function F is a *continuous* function.

Similarly to deterministic systems, the simplest possible fuzzy systems are the ones in which the membership function F depends on a single variable only, i.e., systems described by membership functions $A_i(x_i)$ or $A(y)$, where $A_i : [x_i^-, x_i^+] \to [0, 1]$ and $A : [y^-, y^+] \to [0, 1]$. In logical terms, such simplest systems represent *fuzzy properties.* In terms of *expert knowledge*, such systems describe the simplest possible knowledge: the knowledge about the value of a single variable (input or output) of the type "x_1 is small" or "y is large".

The natural next step in representing knowledge is knowledge that relates the values of different variables. This knowledge is usually represented in terms of the *rules* of the type "if x_1 is small, and x_2 is medium, ..., then y is large", or, in general, rules r of the type $(A_1(x_1)\&\ldots\&A_n(x_n)) \to A(y)$, where A_i and A are fuzzy properties. Such rules form the basis of the *fuzzy control* methodology, one of the most successful applications of soft computing (Kandel *et al.* 1994, Nguyen *et al.* 1995, Hirota *et al.* 1996). In fuzzy control, the truth value R of the fuzzy rule r is usually defined as $R = f_\&(A_1(x_1), \ldots, A_n(x_n), A(y))$, where $f_\&(a, b)$ is a function called *t-norm.* This formula was originally proposed by Mamdani, the pioneer of fuzzy control applications and is, therefore, called *Mamdani rule.*

Usually, experts are pretty confident in their rules, but in principle, they can be confident only to a certain extent: they can be "very confident", or "slightly confident", etc. The words like "very", "slightly", etc., that describe a degree to which a rule is true, are called "hedges". In fuzzy logic, hedges are usually described by functions h from the set $[0, 1]$ (of all possible truth values) to itself: For example, "very" is usually interpreted as $h(x) = x^2$, so that if "John is young" has the the truth value 0.6, then "John is very young" has a truth value $0.6^2 = 0.36$. Another example: "slightly" is usually interpreted as $h(x) = \sqrt{x}$. If a rule r comes with a hedge $h(x)$, then its truth value is equal to $R = h(f_\&(A_1(x_1), \ldots, A_n(x_n), A(y)))$.

Hedges describe "fine structure" of control rules, added to better describe expert statements. Hedges are efficiently used in fuzzy *expert systems* (see, e.g., (Werbos 1993, Tseng *et al.* 1994), but right now, fuzzy *control* systems are mostly not yet at the level where this "fine structure" has to be taken into consideration. Hence, as of now, hedges are rarely used in fuzzy control. However, since we are talking about the *potential abilities* of this methodology, we have to consider hedges as well.

Usually, experts can formulate several rules $r_1, \ldots, r_m$; these rules describe the behavior of the system in different situations, and form the so-called *knowledge* *base.* The knowledge base is applicable to a certain situation iff one of its rules is applicable. Hence, the degree of belief that a knowledge base correctly describes the situation can be naturally determined as the expert's degree of belief that one of its rules is applicable, i.e., that the statement $r_1 \vee \ldots \vee r_m$ is true. The degree of belief in this statement can be described, by using a t-conorm $f_\vee$ (a fuzzy analogue of $\vee$), as $f_\vee(R_1, \ldots, R_m)$.

An expert may have some reservations about his/her knowledge base (just like he/she may have reservations about its rules). As a result, the formula for the true degree of belief in a knowledge base may include a hedge, i.e., it may be described as $h(f_\vee(R_1, \ldots, R_m))$ for some hedge $h(x)$.

By combining membership functions, t-norms, t-conorms, and hedges, we can describe reasonably complicated fuzzy systems. A natural question is: can we thus describe an arbitrary fuzzy system? A modification of Kolmogorov's result enables us to answer "yes". To formulate this result, let us introduce some formal definitions.

Definition 4.1. *Let $f_\&$ and $f_\vee$ be a t-norm and a t-conorm* (Klir *et al.* 1995, Nguyen *et al.* 1996).

- *By a fuzzy system, we mean a continuous function*
$$F : [x_1^-, x_1^+] \times \ldots \times [x_n^-, x_n^+] \times [y^-, y^+] \to [0, 1].$$
- *By a fuzzy property A, we mean a continuous function $A : R \to [0, 1]$.*
- *By a hedge $h(x)$, we mean a continuous function $h : [0, 1] \to [0, 1]$.*
- *By a fuzzy rule r, we mean a pair consisting of a hedge h and an expression of the type $(A_1(x_1)\&\ldots\&A_n(x_n)) \to A(y)$, where A_i and A are fuzzy properties. For given $x_1, \ldots, x_n, y$, the degree of belief $R(x_1, \ldots, x_n, y)$ in a rule r is defined as $h(f_\&(A_1(x_1), \ldots, A_n(x_n), A(y)))$.*
- *By a fuzzy knowledge base k, we mean a pair consisting of a hedge h and a finite set of fuzzy rules $r_1, \ldots, r_m$. For given $x_1, \ldots, x_n, y$, the degree of belief $K(x_1, \ldots, x_n, y)$ described by a knowledge base is defined as $h(f_\vee(R_1, \ldots, R_m))$, where R_i is a degree of belief in a rule r_i.*
- *We say that a fuzzy knowledge base k describes a fuzzy system F iff for all $x_i \in [x_i^-, x_i^+]$, and for all $y \in [y^-, y^+]$, the degree of belief $K(x_1, \ldots, x_n, y)$ described by this knowledge base is equal to $F(x_1, \ldots, x_n, y)$.*

THEOREM 4.1. *If the t-norm $f_\&$ and the t-conorm $f_\vee$ are strictly Archimedean, then every fuzzy system can be described by an appropriate knowledge base.*

This result shows that *the fuzzy control methodology can indeed describe an arbitrary fuzzy system.*

Idea of the proof. The main idea of the proof is to reformulate Kolmogorov's representation of a function $F(x_1, \ldots, x_n, y)$ in desired terms. The inner term in Kolmogorov's expression is the sum, while we want it to be the t-norm. It is known that every strictly Archimedean t-norm can be represented as $f_\&(a, b) =$

$\varphi^{-1}(\varphi(a)+\varphi(b))$ for some function φ. For $A=\varphi(a)$ and $B=\varphi(b)$, we can thus conclude that $A+B=\varphi(f_\&(\varphi^{-1}(A),\varphi^{-1}(B))$. Therefore, the expression $\chi(\sum \psi_{pq}(x_q))$ can be represented as

$$\chi(\varphi(f_\&(\varphi^{-1}(\psi_{1q}(x_1)),\ldots,\varphi^{-1}(\psi_{nq}(x_n)),\varphi^{-1}(\psi_{n+1,q}(y))),$$

i.e., as $h(f_\&(A_1(x_1),\ldots,A_n(x_n),A(y)))$ for $h(x)=\varphi(\chi(x))$, $A_i(x_i)=\varphi^{-1}(\psi_{iq}(x_i))$, and $A(y)=\varphi^{-1}(\psi_{n+1,q}(y))$.

A similar transformation describes the outer sum of the Kolmogorov's expression in terms of the strictly Archimedean t-conorm $f_\vee$. As a result, we get exactly the desired form for $F(x_1,\ldots,x_n,y)$. Q.E.D.

5. FUZZY LOGIC: ARBITRARY LOGICAL OPERATIONS CAN BE REPRESENTED IN TERMS OF "AND", "OR", AND HEDGES

In our description of the expert's knowledge, we used only the logical operations (connectives) "and", "or", and hedges (like "very") as a way of combining different statements. Experts actually use other logical connectives as well, such as "if – then". For each such combination $\circ(a_1,\ldots,a_n)$, we must be able to describe the degree of belief A in the resulting combined statement a in terms of the degrees of belief $A_1,\ldots,A_n$ in component statements $a_1,\ldots,a_n$. Therefore, we can define a *logical operation* as a function $f_\circ(A_1,\ldots,A_n)$ from $[0,1]^n$ to $[0,1]$ that transforms n degrees of belief $A_1,\ldots,A_n$ into a single value A. An important example of logical operators are Yager's Ordered Weighted Averaging Operators (OWA) (see, e.g., (Yager *et al.* 1994)).

In classical logic, where the set $\{0,1\}$ of possible truth values consists of only two values "true" ($=1$) and "false" ($=0$), we can, similarly, define a logical operation as a function $\{0,1\}^n \to \{0,1\}$. It is known that for two-valued logic, every logical operation can be represented as a composition of operations "and", "or", and "not". Moreover, every logical operation $f(A_1,\ldots,A_n)$ can be represented in each of the two special forms:

- *Conjunctive Normal Form* (CNF, for short), in which f is represented as $D_1\&\ldots\&D_k$, where each D_i is a *disjunction* $a\vee\ldots\vee b$, and $a,\ldots,b$ are *literals* (i.e., variables A_i or their negations);
- *Disjunctive Normal Form* (DNF, for short), in which f is represented as $C_1\vee\ldots\vee C_k$, where each C_i is a *conjunction* $a\&\ldots\&b$, and $a,\ldots,b$ are literals.

These representation enable us to represent an arbitrary function of n variables as a composition of functions of one or two variables. It is natural to ask whether similar representations are possible for fuzzy logic. Again, an appropriate modification of Kolmogorov's theorem leads to an answer "yes".

Definition 5.1. *Let $x_1, \ldots, x_n$ be fuzzy variables, i.e., variables that take values from the interval* $[0, 1]$. *Let operations $f_\&$ and $f_\vee$ be defined as functions from* $[0,1] \times [0,1]$ *to* $[0,1]$. *For simplicity, we will also use & and $\vee$ instead of $f_\&$ and $f_\vee$.*

- *By a fuzzy logical operation, we mean a continuous function $f(x_1, \ldots, x_n)$ from* $[0,1] \times \ldots \times [0,1]$ *to* $[0,1]$.
- *By a fuzzy literal, we mean either a variable x_i, or an expression of the type $h(x_i)$, where h is a hedge.*
- *By a fuzzy conjunction, we mean an expression of the type $h(a\& \ldots \&b)$, where h is a hedge and $a, \ldots, b$ are literals.*
- *By a fuzzy disjunction, we mean an expression of the type $h(a \vee \ldots \vee b)$, where h is a hedge, and $a, \ldots, b$ are literals.*
- *By a fuzzy conjunctive normal form (fuzzy CNF, for short), we mean an expression of the type $h(D_1\& \ldots \&D_k)$, where h is a hedge, and $D_1, \ldots, D_k$ are fuzzy disjunctions.*
- *By a fuzzy disjunctive normal form (fuzzy CNF, for short), we mean an expression of the type $h(C_1 \vee \ldots \vee C_k)$, where h is a hedge, and $C_1, \ldots, C_k$ are fuzzy conjunctions.*

THEOREM 5.1. (Kreinovich *et al* 1996) *If $f_\&$ and $f_\vee$ are strictly Archimedean operations, then every fuzzy logical operation can be represented both in fuzzy CNF and in fuzzy DNF.*

In particular, this theorem shows that *an arbitrarily complex fuzzy logical operation can be represented in terms of "and", "or", and hedges.*

Idea of the proof is similar to the proof of Theorem 4.1: actually, the idea from that proof leads to a DNF form (with & inside and $\vee$ outside). A similar transformation leads to a CNF form (with $\vee$ inside and & outside). Q.E.D.

Comment. As shown in (Perfilieva *et al.* 1995), if we do not use hedges (i.e., if we only use "and", "or", and "not"), then not all logical operations can be thus represented. Thus, the use of hedges is necessary.

6. KOLMOGOROV'S THEOREM CANNOT BE EXTENDED TO INTERVAL-VALUED AND FUZZY CASES, AND HENCE, OPERATIONS WITH THREE OR MORE OPERANDS ARE NECESSARY

6.1. Usually, Inputs Are Known Only With Interval or Fuzzy Uncertainty

At first glance, the results from Sections 3–5 seem to indicate that operations with one or two arguments are sufficient to represent arbitrarily complex systems.

This is, however, only true if we assume that the inputs $x_1, \ldots, x_n$ are precisely known. In reality, the values of the input variables x_i are only known with some uncertainty:

- These values may come from *measurements*, in which case, if a measurement result is $\tilde{x}_i$, then we can only guarantee that the actual value x_i belongs to an *interval* $[x_i^-, x_i^+]$, where:
  - $x_i^- = \tilde{x}_i - \Delta_i$,
  - $x_i^+ = \tilde{x}_i + \Delta_i$, and
  - Δ_i is the accuracy of the measuring device (that is guaranteed by the manufacturer of this device).
- These values can also come from *experts*. Experts can usually not only provide us with an interval $[x_i^-, x_i^+]$ of possible values of x_i, but they can also describe to what extent each value from this interval is possible. For each value $x \in [x_i^-, x_i^+]$, this "extent" can be characterized by a number from the interval $[0, 1]$. The expert's knowledge about x_i can be thus characterized by a function that assigns to every value $x \in [x^-, x^+]$, the expert's degree of belief $\in [0, 1]$ that this value x is a possible value of x_i. In other words, the expert's knowledge can be characterized by a function $[x_i^-, x_i^+] \to [0, 1]$, i.e., by a *fuzzy set*.

6.2. For Interval and Fuzzy Uncertainty, Operations With Three or More Operands are Necessary

If we take into consideration that the input is described either by an *interval*, or by a *fuzzy* set, then, in contrast to numerical computations, one can show that we cannot any longer represent an arbitrary operation with many operands in terms of operations with one or two operands (Kreinovich *et al.* 1995, Nguyen *et al.* 1997). (In other words, Kolmogorov's theorem cannot be extended to the *interval-valued* and *fuzzy-valued* cases.) Therefore, there is a need to use operations with more than two variables.

In particular, this result justified the necessity of using logical operations with three or more arguments, such as the Ordered Weighted Averaging (OWA) operations proposed by R. R. Yager (Yager *et al.* 1994).

6.3. Logical Operations: the Best Operations to Use are Operations of the Type $\sum w_i \cdot a_i \cdot b_i$, i.e., Generalized Yager's Ordered Weighted Averaging Operations

We have already mentioned that, according to (Nguyen *et al.* 1997), in order to represent arbitrary interval-valued and fuzzy-valued operations with three or more operands, it is not sufficient to use only operations with one or two operands, operations with three or more operands are also needed. In other words, not every complicated system can be represented as a combination of simplest systems (with one or two inputs).

This result does not mean that complicated systems cannot be reduced to simpler ones; it simply means that the original set of "simpler" systems (consisting of systems with one or two inputs), that was sufficient for the case of precise inputs, is no longer sufficient in case we take input uncertainty into consideration.

The natural question is:

> *How can we enlarge the original set so that it will be sufficient for representing arbitrary complicated systems not only for the case of precise inputs, but also for the case of inputs with uncertainty?*

In (Nguyen *et al.* 1997), we have shown that to get a representation within a certain order of approximation, all we need to add to this list of "basic" systems (simple functions/operations) are:

- operations of the type
$$\sum_{i=1}^{n} w_i \cdot a_i \cdot b_i,$$
where a_i and b_i are inputs (that can be interval-valued or fuzzy), and w_i are numerical coefficients; and
- operations that are obtained from the above formula by the process of *simplification*, i.e., by setting some of the variables a_i, b_i to be constants, and some others to be equal to each other.

In particular, for *fuzzy logical operations*, we must add operations of the type

$$a, \ldots, a_n, b_1, \ldots, b_n \to \sum_{i=1}^{n} w_i \cdot a_i \cdot b_i$$

and their simplifications:

- For $a_1 = \ldots = a_n = 1$, the resulting simplification is simply a linear combination $\sum w_i \cdot b_i$ of the input variables. Such linear combination (locally) coincides with one of the *Ordered Weighted Averaging* operations.

- In general, this operation coincides (locally) with the result of an appropriate OWA operation applied to the statements $a_1 \cdot b_1, \ldots, a_n \cdot b_n$. These statements, in their turn, represent (in case we use algebraic product as "and") the conjunctions $a_1 \& b_1, \ldots, a_n \& b_n$. Thus, the general operation can be represented as OWA$(a_1 \& b_1, \ldots, a_n \& b_n)$, i.e., as a *generalized* Yager's OWA.

7. CONCLUSIONS

In 1957, Kolmogorov has proven that an arbitrary continuous function $f(x_1, \ldots, x_n)$ of several real variables $x_1, \ldots, x_n$ can be represented as a composition of functions of one or two variables. This theoretical result (and its appropriate modifications) leads to a justification of several soft computing methodologies:

- Kolmogorov's theorem shows that an arbitrary *deterministic* system $f(x_1, \ldots, x_n)$ can be represented as a composition of linear elements and elements with one input. This result justifies the use of *neural networks* in describing and simulating deterministic systems.
- A modification of Kolmogorov's theorem shows that an arbitrary *fuzzy* system can be represented by a knowledge base of the type used in *fuzzy control*. Thus, this theorem justifies the use of fuzzy control methodology in describing fuzzy systems.
- Another modification of Kolmogorov's result shows that an arbitrary logical operation in fuzzy logic can be represented as a composition of basic fuzzy logical operations: "and", "or", and hedges ("very", "slightly", etc). This theorem explains why in traditional fuzzy control methodology, all expert's statements are reformulated in terms of these basic logical operations.

These "reduction" results are only true if we assume that the inputs $x_1, \ldots, x_n$ are known precisely. In real-life situations, these inputs are only known with some interval or fuzzy accuracy (i.e., the set of all possible values is described either by an interval, or by a fuzzy set).

For interval-valued and fuzzy-valued inputs, it is no longer possible to represent an arbitrary system with arbitrary number of inputs as a composition of systems with one or two inputs. To be able to represent arbitrarily complex systems, we must add to our set of basic systems (that previously contained only systems with one or two inputs) some systems with three or more inputs.

Similarly, to describe arbitrary *logical* operations, *we must add some operations with three or more operands* to our original list of basic logical operations. It turns out that *the best operations to add are* certain generalizations of *the Ordered Weighted Averaging operations.*

Acknowledgments. This work was partially supported by NSF Grant No. EEC-9322370 and by NASA Grant No. NAG 9-757. The authors are thankful to Ronald R. Yager and Janusz Kacprzyk for the encouragement, and to Elbert A. Walker for valuable discussions.

References

1. C. B. Boyer and U. C. Merzbach, *A History of Mathematics*, Wiley, N.Y., 1991.
2. R. Hecht-Nielsen, "Kolmogorov's Mapping Neural Network Existence Theorem", *IEEE International Conference on Neural Networks*, San Diego, SOS Printing, 1987, Vol. 2, pp. 11–14.

3. K. Hirota and M. Sugeno, *Industrial Applications of Fuzzy Technology in the World*, World Scientific, Singapore, 1996.
4. A. Kandel and G. Langholtz (Editors), *Fuzzy Control Systems*, CRC Press, Boca Raton, FL, 1994.
5. G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*, Prentice Hall, Upper Saddle River, NJ, 1995.
6. A. N. Kolmogorov, "On the Representation of Continuous Functions of Several Variables by Superposition of Continuous Functions of One Variable and Addition", *Dokl. Akad. Nauk SSSR*, 1957, Vol. 114, pp. 369–373.
7. V. Kreinovich and H. T. Nguyen, "On Hilbert's Thirteenth Problem for Soft Computing", *Proceedings of the Joint 4th IEEE Conference on Fuzzy Systems and 2nd IFES*, Yokohama, Japan, March 20–24, 1995, Vol. IV, pp. 2089–2094.
8. V. Kreinovich, H. T. Nguyen, and D. A. Sprecher, *Normal Forms for Fuzzy Logic – an Application of Kolmogorov's Theorem*, University of Texas at El Paso, Department of Computer Science, Technical Report UTEP-CS-96-8, January 1996; available by ftp from cs.utep.edu, login anonymous, file pub/reports/tr96-8.tex; submitted to *International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems.*
9. M. Nakamura, R. Mines, and V. Kreinovich, "Guaranteed intervals for Kolmogorov's theorem (and their possible relation to neural networks)", *Interval Computations*, 1993, No. 3, pp. 183–199.
10. H. T. Nguyen, V. Kreinovich, V. Nesterov, and M. Nakamura, "On hardware support for interval computations and for soft computing: a theorem", *IEEE Transactions on Fuzzy Systems*, 1997 (to appear).
11. H. T. Nguyen, M. Sugeno, R. Tong, and R. Yager (Eds.), *Theoretical aspects of fuzzy control*, J. Wiley, N.Y., 1995.
12. H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, CRC Press, Boca Raton, Florida, 1996 (to appear).
13. I. Perfilieva and A. Tonis, "Functional system in fuzzy logic formal theory", *Busefal*, 1995, Vol. 64, pp. 42–50.
14. D. A. Sprecher, "On the Structure of Continuous Functions of Several Variables", *Transactions Amer. Math. Soc.*, 1965, Vol. 115, No. 3, pp. 340–355.
15. D. A. Sprecher, "An Improvement in the Superposition Theorem of Kolmogorov", *Journal of Mathematical Analysis and Applications*, 1972, Vol. 38, pp. 208–213.
16. H. C. Tseng and D. W. Teo, "Medical expert system with elastic fuzzy logic", *Proc. IEEE Conference on Fuzzy Systems*, 1994, Vol. III, pp. 2067–2071.
17. P. J. Werbos, "Elastic fuzzy logic: a better way to combine neural and fuzzy capabilities", *World Congress on Neural Networks*, 1993, Vol. 2.
18. R. R. Yager and D. P. Filev, *Essentials of fuzzy modeling and control*, J. Wiley, N.Y., 1994.

POSSIBILITY AND NECESSITY IN WEIGHTED AGGREGATION

Christer Carlsson*, Robert Fullér** and Szvetlana Fullér**

* Institute for Advanced Management Systems Research, Åbo Akademi University, Åbo, Finland
*** Department of Computer Science, Eötvös Loránd University, Budapest, Hungary

Abstract

Yager [14] discussed the issue of weighted min and max aggregations and provided for a formalization of the process of importance weighted transformation. Generalizing Yager's principle we suggest the use of fuzzy implication operators for importance weighted transformation.

1 Definitions

In this section we provide the definitions of terms needed in the process of weighted aggregation.

Triangular norms were introduced by Schweizer and Sklar [8] to model the distances in probabilistic metric spaces. In fuzzy sets theory triangular norms are extensively used to model logical connective *and*. Triangular conorms are extensively used to model logical connective *or*.

Definition 1.1 *A mapping*

$$T\colon [0,1] \times [0,1] \to [0,1]$$

is a triangular norm (t-norm for short) iff it is symmetric, associative, non-decreasing in each argument and $T(a, 1) = a$, *for all* $a \in [0, 1]$.

Definition 1.2 *A mapping*

$$S\colon [0,1] \times [0,1] \to [0,1]$$

is a triangular co-norm (t-conorm for short) if it is symmetric, associative, non-decreasing in each argument and $S(a,0) = a$, *for all* $a \in [0,1]$.

If T is a t-norm then the equality

$$S(a,b) := 1 - T(1-a, 1-b) \tag{1}$$

defines a t-conorm and we say that S is derived from T. The basic t-norms and t-conorms pairs are

- minimum/maximum:

$$MIN(a,b) = \min\{a,b\} = a \wedge b, \quad MAX(a,b) = \max\{a,b\} = a \vee b$$

- Łukasiewicz:

$$LAND(a,b) = \max\{a+b-1, 0\}, \quad LOR(a,b) = \min\{a+b, 1\}$$

- probabilistic: $PAND(a,b) = ab, \quad POR(a,b) = a + b - ab$
- weak/strong:

$$WEAK(a,b) = \begin{cases} \min\{a,b\} & \text{if } \max\{a,b\} = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$STRONG(a,b) = \begin{cases} \max\{a,b\} & \text{if } \min\{a,b\} = 0 \\ 1 & \text{otherwise} \end{cases}$$

- Hamacher:

$$HAND_\gamma(a,b) = \frac{ab}{\gamma + (1-\gamma)(a+b-ab)},$$

$$HOR_\gamma(a,b) = \frac{a+b-(2-\gamma)ab}{1-(1-\gamma)ab}, \ \gamma \geq 0$$

- Yager:

$$YAND_p(a,b) = 1 - \min\{1, \sqrt[p]{(1-a)^p + (1-b)^p}\},$$

$$YOR_p(a,b) = \min\{1, \sqrt[p]{a^p + b^p}\}, \ p > 0$$

Definition 1.3 *Let* A *and* B *be two fuzzy predicates defined on the real line* $\mathbf{R}$. *Knowing that 'X is B' is true, the degree of possibility that the proposition 'X is A' is true,* $\Pi[A|B]$, *is given by*

$$\Pi[A|B] = \sup\{A(t) \wedge B(t) | t \in \mathbf{R}\}, \tag{2}$$

the degree of necessity that the proposition 'X is A' is true, $N[A|B]$, *is given by*

$$N[A|B] = 1 - \Pi[\neg A|B],$$

where A and B are the possibility distributions (for simplicity we write A instead of μ_A*) defined by the predicates A and B, respectively, and*

$$(\neg A)(t) = 1 - A(t)$$

for any t. We can use any t-norm T in (2) to model the logical connective and*:*

$$\Pi[A|B] = \sup\{T(A(t), B(t))|t \in \mathbf{R}\}.$$

There are three important classes of fuzzy implication operators:

- S-**implications**: defined by

  $$x \rightarrow y = S(n(x), y) \tag{3}$$

  where S is a t-conorm and n is a negation on $[0, 1]$. These implications arise from the Boolean formalism $p \rightarrow q = \neg p \vee q$. We shall use the following S-implications: $x \rightarrow y = \min\{1 - x + y, 1\}$ (Łukasiewitz) and $x \rightarrow y = \max\{1 - x, y\}$ (Kleene-Dienes).

- R-**implications**: obtained by residuation of continuous t-norm T, i.e.

  $$x \rightarrow y = \sup\{z \in [0, 1] \mid T(x, z) \leq y\}$$

  These implications arise from the Intutionistic Logic formalism. We shall use the following R-implication: $x \rightarrow y = 1$ if $x \leq y$ and $x \rightarrow y = y$ if $x > y$ (Gödel), $x \rightarrow y = \min\{1 - x + y, 1\}$ (Łukasiewitz)

- **t-norm implications**: if T is a t-norm then

  $$x \rightarrow y = T(x, y)$$

  Although these implications do not verify the properties of material implication they are used as model of implication in many applications of fuzzy logic. We shall use the minimum-norm as t-norm implication (Mamdani).

Consider again the definition of t-norm-based possibility

$$\Pi[A|B] = \sup\{T(A(t), B(t))|t \in \mathbf{R}\}, \tag{4}$$

where T is t-norm. Then for the measure of necessity of A, given B we get

$$N[A|B] = 1 - \Pi[\neg A|B] = 1 - \sup_t T(1 - A(t), B(t))$$

Let S be a t-conorm derived from T by (1), then

$$1 - \sup_t T(1 - A(t), B(t)) = \inf_t\{1 - T(1 - A(t), B(t))\} =$$

$$\inf_t \{S(1 - B(t), A(t))\} = \inf_t \{B(t) \to A(t)\}$$

where the implication operator is defined in the sense of (3). That is,

$$N[A|B] = \inf_t \{B(t) \to A(t)\}.$$

Let A and W be discrete fuzzy sets in the unit interval, such that

$$A = a_1/(1/n) + a_2/(2/n) + \cdots + a_n/1$$

and

$$W = w_1/(1/n) + w_2/(2/n) + \cdots + w_n/1$$

where $n > 1$, and the terms $a_j/(j/n)$ and $w_j/(j/n)$ signify that a_j and w_j are the grades of membership of j/n in A and W, respectively, i.e.

$$A(j/n) = a_j, \quad W(j/n) = w_j$$

for $j = 1, \dots, n$, and the plus sign represents the union. Then we get the following simple formula for the measure of necessity of A, given W

$$N[A|W] = \min_{j=1,\dots,n} \{W(j/n) \to A(j/n)\} = \min_{j=1,\dots,n} \{w_j \to a_j\} \tag{5}$$

and we use the notation

$$N[A|W] = N[(a_1, a_2, \dots, a_n)|(w_1, w_2, \dots, w_n)]$$

2 Weighted aggregations

A classical MADM problem can be expressed in a matrix format. A decision matrix is an $m \times n$ matrix whose element x_{ij} indicates the performance rating of the i-th alternative, x_i, with respect to the j-th attribute, c_j:

$$\begin{array}{ccccc}
 & c_1 & c_2 & \dots & c_n \\
x_1 & x_{11} & x_{12} & \dots & x_{1n} \\
x_2 & x_{21} & x_{22} & \dots & x_{2n} \\
\vdots & \vdots & \vdots & & \vdots \\
x_m & x_{m1} & x_{m2} & \dots & x_{mn}
\end{array}$$

In fuzzy case the values of the decision matrix are given as degrees of "how an alternative satisfies a certain attribute". Let x be an alternative such that for any criterion $C_j(x) \in [0, 1]$ indicates the degree to which this criterion is satisfied by x. So, in fuzzy case we have the following decision matrix

$$
\begin{array}{ccccc}
 & C_1 & C_2 & \dots & C_n \\
x_1 & a_{11} & a_{12} & \dots & a_{1n} \\
x_2 & a_{21} & a_{22} & \dots & a_{2n} \\
\vdots & \vdots & \vdots & & \vdots \\
x_m & a_{m1} & a_{m2} & \dots & a_{mn}
\end{array}
$$

where $a_{ij} = C_j(x_{ij})$, for $i = 1, \dots, m$ and $j = 1, \dots, n$.

Let x be an alternative and let

$$(a_1, a_2, \dots, a_n)$$

denote the degrees to which x satisfies the criteria, i.e.

$$a_j = C_j(x), \ i = 1, \dots, n.$$

In many applications of fuzzy sets as multi-criteria decision making, pattern recognition, diagnosis and fuzzy logic control one faces the problem of weighted aggregation. The issue of weighted aggregation has been studied by Carlsson and Fullér [1], Dubois and Prade [2, 3, 4], Fodor and Roubens [5, 6] and Yager [9, 10, 11, 12, 13, 14, 15, 16].

Assume associated with each fuzzy set C_j is a weight $w_j \in [0, 1]$ indicating its importance in the aggregation procedure, $j = 1, \dots, n$. The general process for the inclusion of importance in the aggregation involves the transformation of the fuzzy sets under the importance. Let **Agg** indicate an aggregation operator, max or min, to find the weighted aggregation. Yager [14] first transforms each of the membership grades using the weights

$$g(w_i, a_i) = \hat{a}_i,$$

for $i = 1, \dots, n$, and then obtain the weighted aggregate

$$\mathbf{Agg}\langle \hat{a}_1, \dots, \hat{a}_n \rangle.$$

The form of g depends upon the type of aggregation being performed, the operation **Agg**.

As discussed by Yager [14] in incorporating the effect of the importances in the min operation we are interested in reducing the effect of the elements which have low importance.

In performing the min aggregation it is the elements with low values that play the most significant role in this type of aggregation, one way to reduce the effect of elements with low importance is to transform them into big values, values closer to one. Yager introduced a class of functions which can be used for the inclusion of importances in the min aggregation

$$g(w_i, a_i) = S(1 - w_i, a_i)$$

where S is a t-conorm, and then obtain the weighted aggregate

$$\min\{\hat{a}_1, \dots, \hat{a}_n\} = \min\{S(1 - w_1, a_1), \dots S(1 - w_n, a_n)\} \tag{6}$$

We first note that if $w_i = 0$ then from the basic property of t-conorms it follows that

$$S(1 - w_i, a_i) = S(1, w_i) = 1$$

Thus, zero importance gives us one. Yager notes that the formula can be seen as a measure of the degree to which an alternative satisfies the following proposition:

All important criteria are satisfied

Example 1 *Let*

$$(0.3, 0.2, 0.7, 0.6)\cdot$$

be the vector of weights and let

$$(0.4, 0.6, 0.6, 0.4)$$

be the vector of aggregates. If $g(w_i, a_i) = \max\{1 - w_i, a_i\}$ *then we get*

$$g(w_1, a_1) = (1 - 0.3) \vee 0.4 = 0.7, \ \ g(w_2, a_2) = (1 - 0.2) \vee 0.6 = 0.8$$

$$g(w_3, a_3) = (1 - 0.7) \vee 0.6 = 0.6, \ \ g(w_4, a_4) = (1 - 0.6) \vee 0.4 = 0.4$$

That is

$$\min\{g(w_1, a_1), \dots, g(w_4, a_4)\} = \min\{0.7, 0.8, 0.6, 0.4\} = 0.4$$

As for the max aggregation operator: Since it is the large values that play the most important role in the aggregation we desire to transform the low importance elements into small values and thus have them not play a significant role in the max aggregation. Yager suggested a class of functions which can be used for importance transformation in max aggregation

$$g(w_i, a_i) = T(w_i, a_i)$$

where T is a t-norm. We see that if $w_i = 0$ then $T(w_i, a_i) = 0$ and the element plays no rule in the max.

Let **Agg** indicate any aggregation operator and let

$$(a_1, a_2, \dots, a_n)$$

denote the vector of aggregates. We define the weighted aggregation as

$$\mathbf{Agg}\,\langle g(w_1, a_1), \dots, g(w_n, a_n)\rangle.$$

where the function g satisfies the following properties

- if $a > b$ then $g(w, a) \geq g(w, b)$
- $g(w, a)$ is monotone in w
- $g(0, a) = id, \ \ g(1, a) = a$

where the identity element, *id*, is such that if we add it to our aggregates it doesn't change the aggregated value.

3 Extensions

Let us recall formula (5)

$$N[(a_1, a_2, \ldots, a_n)|(w_1, w_2, \ldots, w_n)] = \min\{w_1 \rightarrow a_1, \ldots, w_n \rightarrow a_n\} \qquad (7)$$

where

$$A = a_1/(1/n) + a_2/(2/n) + \cdots + a_n/1$$

is the fuzzy set of performances and

$$W = w_1/(1/n) + w_2/(2/n) + \cdots + w_n/1$$

is the fuzzy set of weights; and the formula for weighted aggregation by the minimum operator

$$\min\{\hat{a}_1, \ldots, \hat{a}_n\}$$

where

$$\hat{a}_i = g(w_i, a_i) = S(1 - w_i, a_i)$$

and S is a t-conorm.

It is easy to see that if the implication operator in (7) is an S-implication then from the equality

$$w_j \rightarrow a_j = S(1 - w_j, a_j)$$

it follows that the weighted aggregation of the a_i's is nothing else, but

$$N[(a_1, a_2, \ldots, a_n)|(w_1, w_2, \ldots, w_n)],$$

the necessity of performances, given weights.

This observation leads us to a new class of transfer functions (which contains Yager's functions as a subset):

$$\hat{a}_i = g(w_i, a_i) = w_i \rightarrow a_i \qquad (8)$$

where $\rightarrow$ is an arbitrary implication operator. Then we combine the $\hat{a}_i$'s with an appropriate aggregation operator **Agg**.

However, we first select the implication operator, and then the aggregation operator **Agg** to combine the $\hat{a}_i$'s. If we choose a t-norm implication in (8) then we will select the max operator, and if we choose an R- or S-implication then we will select the min operator to aggregate the $\hat{a}_i$'s.

Example 2 *Let*

$$(0.3, 0.2, 0.7, 0.6)$$

be the vector of weights and let

$$(0.4, 0.6, 0.6, 0.4)$$

be the vector of aggregates. If $g(w_i, a_i) = \min\{1, 1 - w_i + a_i\}$ *is the Łukasiewicz implication then we compute*

$$g(w_1, a_1) = 0.3 \to 0.4 = 1, \; g(w_2, a_2) = 0.2 \to 0.6 = 1$$

$$g(w_3, a_3) = 0.7 \to 0.6 = 0.9, \; g(w_4, a_4) = 0.6 \to 0.4 = 0.8$$

That is

$$\min\{g(w_1, a_1), \dots, g(w_4, a_4)\} = \min\{1, 1, 0.9, 0.8\} = 0.8$$

If $g(w_i, a_i)$ *is implemented by the Gödel implication then we get*

$$g(w_1, a_1) = 0.3 \to 0.4 = 1, \; g(w_2, a_2) = 0.2 \to 0.6 = 1$$

$$g(w_3, a_3) = 0.7 \to 0.6 = 0.6, \; g(w_4, a_4) = 0.6 \to 0.4 = 0.4$$

That is

$$\min\{g(w_1, a_1), \dots, g(w_4, a_4)\} = \min\{1, 1, 0.6, 0.4\} = 0.4$$

If $g(w_i, a_i) = w_i a_i$ *is the Larsen implication then we have*

$$g(w_1, a_1) = 0.3 \times 0.4 = 0.12, \; g(w_2, a_2) = 0.2 \times 0.6 = 0.12$$

$$g(w_3, a_3) = 0.7 \times 0.6 = 0.42, \; g(w_4, a_4) = 0.6 \times 0.4 = 0.24$$

That is

$$\max\{g(w_1, a_1), \dots, g(w_4, a_4)\} = \max\{0.12, 0.12, 0.42, 0.24\} = 0.42$$

It should be noted that if we choose an R-implication in (8) then the equation

$$\min\{w_1 \to a_1, \dots, w_n \to a_n\} = 1$$

holds iff $w_i \le a_i$ for all i, i.e. when each performance rating is at least as big as its associated weight. In other words, if a performance rating with respect to an attribute exceeds the value of the weight of this attribute then this rating does not matter in the overall rating. However, ratings which are well below of the corresponding weights play a significant role in the overall rating. Thus the formula (7) with an R-implication can be seen as a measure of the degree to which an alternative satisfies the following proposition:

All scores are bigger than or equal to the importances

It should be noted that the min aggregation operator does not allow any compensation, i.e. a higher degree of satisfaction of one of the criteria can not compensate for a lower degree of satisfaction of another criteria. Averaging operators realize *trade-offs* between objectives, by allowing a positive compensation between ratings.

Another possibility is to use an *andlike* or an *orlike* OWA-operator to aggregate the elements of the bag

$$\langle w_1 \to a_1, \dots, w_n \to a_n \rangle.$$

Let A and W be discrete fuzzy sets in $[0, 1]$, where $A(t)$ denotes the performance rating and $W(t)$ denotes the weight of a criterion labeled by t. Then the weighted aggregation of A can be defined by,

- a t-norm-based measure of necessity of A, given W:

$$N[A|W] = \min_t \{W(t) \to A(t)\}$$

  For example, the Kleene-Dienes implication opeartor,

$$w_i \to a_i = \max\{1 - w_i, a_i\},$$

  implements Yager's approach to fuzzy screening [11].

- a t-norm-based measure of possibility of A, given W:

$$\Pi[A|W] = \max_t \{T(A(t), W(t))\}$$

- an OWA-operator defined on the bag

$$\langle W(t) \to A(t) \mid t \rangle$$

Other possibility is to take the value

$$\frac{\int_0^1 A(t) \wedge W(t)\, dt}{\int_0^1 W(t)\, dt}$$

for the overall score of A. If $A(t) \geq W(t)$ for all $t \in [0, 1]$ then the overall score of A is equal to one. However, the bigger the set $\{t \in [0, 1] | A(t) \leq W(t)\}$ the smaller the overall rating of A.

4 Summary and Conclusions

Generalizing Yager's principles for weighted min and max aggregations we introduce fuzzy implication operators as a means for importance weighted transformation.

Weighted aggregations are important in decision problems where we have multiple attributes to consider and where the outcome is to be judged in terms of attributes which are not equally important for the decision maker. The importance is underscored if there is a group of decision makers with varying value judgments on the attributes and/or if this group has factions promoting some subset of attributes.

The results shown here can easily be implemented with a number of software tools.

Acknowledgements

The second author, who is presently a Donner Visiting Professor at *Institute for Advanced Management Systems Research*, Åbo Akademi University, has been partially supported by OTKA T 14144 and T 019455.

References

[1] C. Carlsson and R. Fullér, On fuzzy screening system, in: *Proceedings of EUFIT'95 Conference*, August 28-31, 1995 Aachen, Germany, Verlag Mainz, Aachen, 1995 1261-1264.

[2] D.Dubois, H.Prade, Criteria aggregation and ranking of alternatives in the framework of fuzzy set theory, *TIMS/Studies in the Management Sciences*, 20(1984) 209-240.

[3] D. Dubois and H. Prade, A review of fuzzy sets aggregation connectives, *Information Sciences*, 36(1985) 85-121.

[4] D. Dubois and H. Prade, Weighted minimum and maximum operations in fuzzy sets theory, *Information Sciences*, 39(1986) 205-210.

[5] J.C. Fodor and M. Roubens, Aggregation and scoring procedures in multicriteria decision-making methods, *Proceedings of the IEEE International Conference on Fuzzy Systems*, San Diego, 1992 1261–1267.

[6] J.C. Fodor and M. Roubens, *Fuzzy Preference Modelling and Multicriteria Decision Support*, Theory and Decision Library, Series D, system Theory, Knowledge Engineering and Problem Solving, Vol. 14, Kluwer, Dordrecht, 1994.

[7] J. Kacprzyk and R.R. Yager, Using fuzzy logic with linguistic quantifiers in multiobjective decision making and optimization: A step towards more human-consistent models, in: R.Slowinski and J.Teghem eds., *Stochastic versus Fuzzy Approaches to Multiobjective Mathematical Programming under Uncertainty*, Kluwer Academic Publishers, Dordrecht, 1990 331-350.

[8] B.Schweizer and A.Sklar, Associative functions and abstract semigroups, *Publ. Math. Debrecen*, 10(1963) 69-81.

[9] R.R.Yager, Fuzzy decision making using unequal objectives, *Fuzzy Sets and Systems*, 1(1978) 87-95.

[10] R.R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making, *IEEE Transactions on Systems, Man and Cybernetics* 18(1988) 183-190.

[11] R.R.Yager, Fuzzy Screaning Systems, in: R.Lowen and M.Roubens eds., *Fuzzy Logic: State of the Art*, Kluwer, Dordrecht, 1993 251-261.

[12] R.R.Yager, Families of OWA operators, *Fuzzy Sets and Systems*, 59(1993) 125-148.

[13] R.R.Yager, Aggregation operators and fuzzy systems modeling, *Fuzzy Sets and Systems*, 67(1994) 129-145.

[14] R.R.Yager, On weighted median aggregation, *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 1(1994) 101-113.

[15] R.R.Yager, Constrained OWA aggregation, Technical Report, #MII-1420, Mashine Intelligence Institute, Iona College, New York, 1994.

[16] R.R.Yager, Quantifier guided aggregation using OWA operators, Technical Report, #MII-1504, Mashine Intelligence Institute, Iona College, New York, 1994.

OWA operators and an extension of the contrast model

Bernadette Bouchon-Meunier

Maria Rifqi

LAFORIA-IBP
Université P. et M. Curie
Case 169 - 4, Place Jussieu
75252 Paris Cedex 05

1 Introduction

A framework was proposed (Bouchon-Meunier *et al.* 1996) to classify and to generate measures of comparison. It is based on concepts analogous to those developed by Tversky for his *contrast model* (Tversky 1977) (Suppes *et al.* 1990). This framework distinguishes, in the set of measures of comparison, measures of similitude and measures of dissimilarity. A measure of similitude can, in particular, depend on its use and thus on required properties.

Tversky suggests to compare objects globally even if he considers all the features of an object. In Tversky's axiomatic, a measure of similarity should take features of objects into account once for all. Thus, a comparison of objects yields one value : there does not exist any step of aggregation.

We prefer to consider that an object is described by means of fuzzy attributes. Two objects are compared attribute by attribute. Thus, a measure of comparison can be used for each pair of values of attributes. We obtain as many values of comparison as there exist attributes. Aggregation of those values is then necessary to compare the objects globally.

We study the properties of each kind of measures of comparison when a step of aggregation is added to the model.

2 Model of measures of comparison

Let us briefly expose the model of measures of comparison we consider.

For any set Ω of elements, let $F(\Omega)$ denote the set of fuzzy subsets of Ω, f_A the membership function of any description A in $F(\Omega)$. We use the classical definition of intersection: $f_{A\cap B} = \min(f_A, f_B)$ to describe the elements belonging to A and B.

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

We also use the definition of set-inclusion in the sense of Zadeh: $A \subseteq B$ if and only if $f_A(x) \leq f_B(x)$ $\forall x \in F(\Omega)$. We need to define an operation of difference between two fuzzy subsets.

Definition 1 *An operation on $F(\Omega)$ is called a* difference *and denoted by $-$, if it satisfies for every A and B in $F(\Omega)$:*

D1 *: if $A \subseteq B$, then $A - B = \emptyset$.*

D2 *: $B - A$ is monotonous with regard to B : $B \subseteq B'$ entails $B - A \subseteq B' - A$*

This definition is slightly different from the definition of difference given by A. Kaufmann (Kaufmann 1973), or by D. W. Roberts (Roberts 1987) or also D. Dubois and H. Prade (Dubois and Prade 1983), but it is compatible with the definition of a difference in the case of crisp subsets of Ω, where $A - B$ can be defined as the complement of $A \cap B$ in A or, equivalently to the intersection of A with the complement of B in Ω.

We suppose that we are given a means of evaluating the weight of the elements of the universe characterized by a fuzzy set through a fuzzy set measure M.

Definition 2 *A fuzzy set measure M is a mapping : $F(\Omega) \rightarrow I\!R^+$ such that, for every A and B in $F(\Omega)$:*

- $M(\emptyset) = 0$
- *if $B \subseteq A$, then $M(B) \leq M(A)$.*

Definition 3 *An M-*measure of comparison *on Ω is a mapping $S : F(\Omega) \times F(\Omega) \times F(\Omega) \rightarrow [0,1]$ such that $S(A,B) = F_S(M(A \cap B), M(B - A), M(A - B))$, for a given mapping $F_S : I\!R^+ \times I\!R^+ \times I\!R^+ \rightarrow [0,1]$ and a fuzzy set measure M on $F(\Omega)$.*

We consider the following properties for F_S:

- symmetry: $S(A,B) = S(B,A)$
- minimality: $S(A,A) = 0$
- reflexivity: $S(A,A) = 1$
- containment: if $B \subseteq A$ then $S(A,B) = 1$
- exclusiveness: if $A \cap B = \emptyset$ then $S(A,B) = 0$

In particular cases, we will restrict ourselves to continuous M-measures of comparison S, defined by continuous functions F_S.

Definition 4 *An M-*measure of similitude *S on Ω is an M-measure of comparison S such that $F_S(u,v,w)$ is non decreasing in u, non increasing in v and w.*

The measure of dissimilarity is not defined as the dual of a measure of similitude, but it has specific properties.

Definition 5 *An M-measure of dissimilarity S on Ω is an M-measure of comparison satisfying the minimality property and such that $F_S(u, v, w)$ is independent of u and non decreasing in v and w.*

In order to classify the existing measures more subtly, we distinguish three types of M-measures of similitude: satisfiability, inclusion and resemblance.

Definition 6 *An M-measure of resemblance on Ω is an M-measure of similitude S which satisfies reflexivity and symmetry properties.*

Definition 7 *An M-measure of inclusion S on Ω is an M-measure of similitude satisfying reflexivity and exclusiveness properties such that $F_S(u, v, w)$ is independent of v.*

Definition 8 *An M-measure of satisfiability S on Ω is an M-measure of similitude S satisfying containment and exclusiveness properties and such that $F_S(u, v, w)$ is independent of w.*

Interpretation

The five above-mentioned properties (minimality, reflexivity, symmetry, exclusiveness and containment) can be considered in a more restrictive way as:

- symmetry: $F_S(u, v, w) = F_S(u, w, v)$
- minimality: $F_S(u, 0, 0) = 0$
- reflexivity: $F_S(u, 0, 0) = 1$
- containment: $F_S(u, 0, w) = 1$ whatever $u \neq 0$ and w may be.
- exclusiveness: $F_S(0, v, w) = 0$ whatever v and w may be.

3 Aggregation

Let Θ be a set of objects described by means of attributes defined on sets Ω_1, $\Omega_2, ..., \Omega_n$. An object $O = (A_1, A_2, ..., A_n)$ is associated with values $A_1, A_2, ..., A_n$ of the attributes, respectively defined as fuzzy subsets of Ω_1, $\Omega_2, ..., \Omega_n$. We suppose that a fuzzy set measure M^i and a difference $-$ are defined for every Ω_i, $i \in [1, n]$. We consider an M-measure of comparison S_i on $\Omega_i, i \in [1, n]$. Let us consider another object $O' = (B_1, B_2, ..., B_n)$. We say that:

- an object O is included in an object O' if and only if each value describing the object O is included in the corresponding value describing the object O',

- the intersection of O and O' is defined as the object $O \cap O'$ with values of attributes $A_i \cap B_i$,
- the difference between objects O and O' is defined as the object $O - O'$ with values of attributes $A_i - B_i$.
- a measure $\mathcal{M}$ is defined on Θ as a mapping lying in $I\!R^{+n}$ such that, for every O in Θ :

$$\mathcal{M}(O) = (M^1(A_1), M^2(A_2), \ldots, M^n(A_n))$$

As a consequence, $\mathcal{M}$ is monotonous : if $O \subseteq O'$ then $M^i(A_i) \leq M^i(B_i)$ $\forall i \in [1, n]$ and we write: $\mathcal{M}(O) \preceq \mathcal{M}(O')$ with minimal element $\bigcirc = (0, ..., 0)$.

If the values of all attributes of two objects are disjoint ($A_i \cap B_i = \emptyset$ $\forall i = 1, \ldots, n$), then these two objects are disjoint ($O \cap O' = (\emptyset, \ldots, \emptyset)$).

3.1 Measures of comparison

We want to obtain a global measure of comparison of objects which satisfies the properties of the measure chosen to compare the values of attributes.

We consider an OWA-operator h_W, with $W = \{w_1, \ldots, w_n\}$ a set of coefficients in $[0, 1]$ such that $\sum_{i=1}^{i=n} w_i = 1$, $h_W(a_1, \ldots, a_n) = \sum_{i=1}^{i=n} w_i b_i$, with b_i the i-th largest value of $a_1, \ldots, a_n$.

It is interesting to see that the five properties (minimality, reflexivity, symmetry, exclusiveness and containment) are preserved by OWA operators (Yager 1988) in both representations of these properties (restrictive or not).

Definition 9 *An $\mathcal{M}$-measure of comparison S is a mapping $\Theta \times \Theta \to [0, 1]$ such that $S(O, O') = F_S(\mathcal{M}(O \cap O'), \mathcal{M}(O' - O), \mathcal{M}(O - O'))$, for any pair $O = (A_1, \ldots, A_n)$ and $O' = (A'_1, \ldots, A'_n)$ of objects of Θ, with $F_S : I\!R^{+n} \times I\!R^{+n} \times I\!R^{+n} \to [0, 1]$.*

We can extend the properties of symmetry, minimality, reflexivity, containment and exclusiveness defined for M-measures of comparison to $\mathcal{M}$-measures of comparison. We then define particular measures of comparison in a way analogous to definitions 4–8. Since $h_W(a_1, \ldots, a_n)$ is independent of the order of the elements $a_1, \ldots, a_n$ and h_W is idempotent (i.e. $h_W(a, \ldots, a) = a$ for any a), we can prove that the aggregation of M-measures of comparison by an OWA operator is also an $\mathcal{M}$-measure of comparison. More particularly :

Proposition 1 *Let S_i be an M-measure of similitude (resp. dissimilarity, resemblance, satisfiability, inclusion). For any OWA-operator h_W, the measure defined for any pair (O, O') of objects of Θ by :*

$$S(O, O') = h_W(S_1(A_1, B_1), \ldots, S_n(A_n, B_n))$$

is an $\mathcal{M}$-measure of similitude (resp. dissimilarity, resemblance, satisfiability, inclusion).

This means that every family of particular M-measures of comparison (similitude, dissimilarity, resemblance, satisfiability, inclusion) is closed under any aggregation by an OWA operator.

3.2 Extension of the contrast model

Tversky has pointed out a model of measure of similarity in psychology, considering objects described by a list of crisp characteristics. For the so-called contrast model, he considers the following major properties for a measure of similarity: matching, monotonicity, independence and solvability. We propose to extend these properties to fuzzy sets in such a way that the main M-measures of comparison used in fuzzy techniques are compatible with the contrast model. Let us remark another attempt to generalize the contrast model to fuzzy sets (Shiina 1988) not taking into consideration the major properties but only the mathematical expression of the measure exhibited by Tversky.

Let A and B be two fuzzy subsets of Ω, M a fuzzy set measure on Ω. We consider an M-measure of similitude $S(A,B)$ of B to A. Following Tversky's definition of agreement between descriptions, we consider that the pairs (A,B) and (C,D) *agree* on one, two or three components, respectively, when one, two or three of the following hold: $M(A\cap B) = M(C\cap D)$, $M(A-B) = M(C-D)$, $M(B-A) = M(D-C)$. Then, the four major properties of a similarity can be interpreted as follows for any continuous M-measure of similitude S:

1. *Matching*

   $$S(A,B) \text{ is a function of } M(A\cap B), M(A-B) \text{ and } M(B-A)$$

2. *Monotonicity*

   $$S(A,B) \geq S(A,C)$$

   whenever

   $$M(A\cap B) \geq M(A\cap C), \quad M(A-B) \leq M(A-C)$$
   $$\text{and} \quad M(B-A) \leq M(C-A)$$

3. *Solvability*

   For any A, B, B' such that $M(A\cap B) > M(A\cap B')$ and $M(B-A) < M(B'-A)$ and $M(A-B) < M(A-B')$, there exists a pair (P,Q) such that $M(P\cap Q) = M(A\cap B), M(Q-P) = M(B'-A)$, $M(P-Q) = M(A-B')$ and $S(A,B) > S(P,Q) > S(A,B')$ if S is strictly increasing in $M(A\cap B)$ and strictly decreasing in $M(B-A)$ and in $M(A-B)$.

If we consider M-measures of similitude that are linear (F_S is a linear function), then they satisfy the generalized property of independence expressed as:

4. *Independence*
   Suppose the pairs (A, B) and (C, D) as well as the pairs (A', B') and (C', D') agree on the same two components, while the pairs (A, B) and (A', B') as well as the pairs (C, D) and (C', D') agree on the remaining component. Then,

$$S(A, B) \geq S(A', B') \text{ iff } S(C, D) \geq S(C', D')$$

If we consider M-measures of satisfiability or M-measures of inclusion, the independence property restricted to two components is satisfied.

If the contrast model is generalized to a comparison attribute by attribute, it can be proven that those properties are still valid when the chosen operator of aggregation is an OWA operator and the representation of objects as follows:

Proposition 2 *Let S_i be a continuous M-measure of similitude defined on Ω_i, for $1 \leq i \leq n$. For any OWA-operator h_W, the measure defined for any pair (O, O') of objects of Θ by :*

$$S(O, O') = h_W(S_1(A_1, B_1), \ldots, S_n(A_n, B_n))$$

satisfies properties of matching, monotonicity and solvability. In the case of linear M-measures of similitude or in the case of M-measures of satisfiability or inclusion, $S(O, O')$ satisfies also the property of independence.

4 Conclusion

We have studied the global comparison of objects regarded as the aggregation of the elementary comparison of their attributes. Global and elementary measures of comparison of objects satisfy the same basic properties, as soon as the aggregation is performed by means of an OWA-operator. The basic properties are either functional (minimality, reflexivity, symmetry, exclusiveness and containment) or stemmed from Tversky's approach (matching, monotonicity, solvability, independence).

If we consider the compatibility between the contrast model and the model of measures of comparison we propose (Bouchon-Meunier *et al.* 1996), (Rifqi 1995) it can be proven that M-measures of satisfiability have the properties required by Tversky's model.

If the considered measures have three components: $M(A \cap B), M(B - A)$ and $M(A - B)$, such as M-measures of resemblance or similitude , then properties of independence and solvability are very restrictive. Solvability is the less restrictive:

if F_S is a continuous function, S satisfies the axiom of solvability. The independence property is inherited from measurement and decision theory. It is not a compulsory property for a general measure of comparison. For example, it does not allow a ratio function. Tversky himself introduces a *ratio model* in order to take into account all the existing measures expressed by means of a ratio function.

Let us finally remark that the OWA-operators are not the only aggregation operators preserving the studied properties.

References

Bouchon-Meunier, B., M. Rifqi and S. Bothorel (1996). 'Towards general measures of comparison of objects'. *Fuzzy Sets and Systems* **84**(2), december.

Dubois, D. and H. Prade (1983). 'Ranking fuzzy numbers in the setting of possibility theory'. *Information Sciences* **30**, 183–224.

Kaufmann, A. (1973). *Introduction à la théorie des sous-ensembles flous.* Vol. 3. Masson.

Rifqi, M. (1995). Mesures de similitude et leur agrégation. In 'Rencontres sur la Logique Floue et ses Applications, LFA'95'. Paris. pp. 80–87.

Roberts, D. W. (1987). 'An anticommutative difference operator for fuzzy sets and relations'. *Fuzzy Sets and Systems* **21**, 35–42.

Shiina, K. (1988). 'A fuzzy-set-theoretic feature model and its application to asymmetric data analysis'. *Japanese psychological research* **30**(3), 95–104.

Suppes, P., D. H Krantz, R. Duncan Luce and A. Tversky (1990). *Foundations of measurement.* Vol. 2. Academic Press, New-York.

Tversky, A. (1977). 'Features of similarity'. *Psychological Review* **84**, 327–352.

Yager, R. R. (1988). 'On ordered weighted averaging operators on multicriteria decision-making'. *IEEE Transactions on Systems, Man and Cybernetics* **18**, 183–190.

2

FUNDAMENTAL ASPECTS OF OWA OPERATORS

EQUIVALENCE OF CHANGES IN PROPORTIONS AT CROSSROADS OF MATHEMATICAL THEORIES

János Aczél*, Günter Rote**, Jens Schwaiger***

*Faculty of Mathematics
University of Waterloo
Waterloo, Ontario
Canada N2L 3G1

**Institut für Mathematik
Technische Universität
Steyrergasse 30
A-8010 Graz
Austria

***Institut für Mathematik
Karl-Franzens Universität
Heinrichstrasse 36/4
A-8010 Graz
Austria

Abstract
After restating certain 'reasonable' requirements for changes in proportions to be equivalent, we determine all relations which satisfy them (there are uncountably many). The proof makes use of the theories of webs, of iteration groups, of concave functions and of functional equations.

1. Introduction

Y.-K. Ng [4] formulated a set of "reasonable" conditions for equivalence of changes in proportions (equality of quotients or of differences would not do since they could produce proportions greater than one or smaller than zero). Solving two problems posed by him we determine all families of functions establishing equivalences which satisfy these conditions; it turns out that there are uncountably many of them. Surprisingly, the theories of webs and of iteration groups play significant roles in the proof, as well as do concave functions and (not so surprisingly) functional equations and the theories of measurement and of probability distributions. Our conditions are actually weaker but imply those in [4].

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

2. Conditons, webs and iteration groups

To begin with, it is required that, for every pair r, x and every s, there exist exactly one y such that the change from s to y is equivalent to that from r to x. It is convenient to hold r and s strictly between 0 and 1, while allowing x and y to be also 0 or 1 (or anything in between). However, from any s only the change to y = 0 or y = 1 is equivalent to the change from any r to x = 0 or x = 1, respectively (r and s strictly between 0 and 1).

When also x and y are temporarily restricted to the open interval]0,1[(as were r and s in the first place) then it is postulated that the correspondence between (r,s) and (x,y) and also that between (r,x) and (s,y) be equivalence relations: reflexive (each pair P is equivalent to itself), symmetric (if P is equivalent to Q then Q is equivalent to P) and transitive (if P is equivalent to Q and Q is equivalent to R then P is equivalent to R).

Considering y a function of x for every fixing of r and s, we get a family of functions. We suppose either that all these functions are strictly increasing on the closed interval [0,1] or that all are continuous there.

Under the above conditions the graphs of these functions and the (vertical and horizontal) coordinate lines form a web (geometric net) on $]0,1[^2$ (the interior of the unit square), which means here that the graph of exactly one of these functions goes through each point in $]0,1[^2$. Moreover, the requirement above, that the correspondence between (r,x) and (s,y) be an equivalence relation, implies that the so called Reidemeister condition of web theory is satisfied ([1], [2]) and, eventually, that the functions establishing this correspondence form a continuous iteration group, containing the identity, with every function its inverse and with every pair of functions their composition.

Such iteration groups can be described by a parameter going through all real numbers and by a continuous and strictly increasing function g mapping]0,1[onto the set of all reals. The first pair (r,s), which we have temporarily fixed, determines the parameter, and then the equivalence of (r,x) and (s,y) is described by

(*) $g(y) - g(s) = g(x) - g(r).$

(This equation plays an essential role in the theory of measurement, see e.g. [3]).

3. Two additional conditions: symmetry and concavity

A further condition in [4] is that replacing the proportions (say unemployment rates), r, s, x by 1-r, 1-s, 1-x (in this example by the proportion of those employed) transforms y into 1 - y. We can prove that this means that the graph of the above function g is symmetric with respect to the point (1/2, g(1/2)). Y.-K. Ng asked in [4] whether there exist essentially different families of functions establishing equivalent changes in proportions, which satisfy all these conditions. Since there are uncountably many essentially different functions g with the above properties, the answer to Ng's question is yes.

A second problem posed in [4] was whether adding to the conditions the concavity of those functions in the iteration group, whose graph lies above that of the identity, gives an essentially unique family of functions. The answer is no. The functions g yielding through (*) the correspondences satisfying all the requirements of Ng, including the last, concavity condition, are the threshold functions (the inverses of probability distribution functions) of continuous random variables whose support is the set of real numbers and whose probability density function is symmetric and log-concave (that is, its logarithm is concave), and only these. (One may take for "probability distribution function" a sufficiently smooth approximation of an empirical distribution function).

There are also uncountably many essentially different functions of this kind.

In [4] a one-parameter (though not the iteration parameter) family of functions have been presented which satisfies all the conditions required. While, as we just saw, it is not the only one by far, it has a nice interpretation (not in [4]): The "odds" r/(1-r), x/(1-x) and s/(1-s), y/(1-y) are proportional, that is,

$$\frac{y/(1-y)}{s/(1-s)} = \frac{x/(1-r)}{r/(1-r)}$$

Taking logarithms on both sides we get an equation of the form (*).

References

[1] Aczél, J. (1965), Quasigroups, nets and nomograms, Adv. Math., 1, 383-450.

[2] Aczél, J. (1966), Funktionskomposition, Iterationsgruppen und Gewebe, Arch. Math. (Basel), 17, 469-475.

[3] Krantz, D.H., Luce, R.D., Suppes, P., and Tversky, A. (1971), Foundations of Measurement 1, Additive and Polynomial Representations. Chapter 4, Academic Press, New York.

[4] Ng, Y.-K. (1991), Equivalent changes /differences for percentages and probabilities, Quart. Appl. Math., 49, 289-301.

ON THE INCLUSION OF IMPORTANCES IN OWA AGGREGATIONS

Ronald R. Yager
Machine Intelligence Institute, Iona College
New Rochelle, NY 10801
E-Mail: yager@panix.com

ABSTRACT
In this work we concentrate on the issue of the inclusion of importances in the OWA aggregation process. We first look at the ways in which importances have been included in some notable examples of OWA operators. We see that in these cases the importance is used to transform the scores into some effective score. Using this knowledge we build, using a fuzzy systems model, a general transform operator for the inclusion of importances. A second approach for the inclusion of importances is also discussed this approach requires the existence of a linguistic quantifier to generate the OWA weights modified by the importances.

KEYWORDS: Aggregation, Multi-Criteria, OWA,

1. INTRODUCTION

The OWA operators were original introduced in [1] to provide a means for aggregating scores associated with the satisfaction to multiple criteria. Subsequently they have proved to be a useful family of aggregation operators for many different types of problems, here we particularly note their application in generalizing decision making under uncertainty [2]. One issue of considerable interest related to the use of these operators is the development of an appropriate methodology for inclusion of importance weights in OWA aggregations. A first attempt at the solution of this problem was suggested in [1]. The difficulty associated with this problem rests with the large class of operators representable by the OWA operator. In particular the different members of the OWA family such as Max, Min and Average each have different ways of including importance. In this work we describe two approaches for the unified inclusion of importance in the OWA aggregation process. The first approach makes use of a transformation operator [3]. This transformation operator is obtained from a fuzzy systems model [4] which has a knowledge base built from the required transformations for known aggregation operations. The second approach which requires the knowledge of an underlying linguistic quantifier to guide the

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

aggregation [5, 6] and uses information about the importances to modify the OWA weights.

2. FUNDAMENTALS OF OWA OPERATORS

In [1] Yager introduced the ordered weighting averaging (OWA) operators which provide a general class of mean like aggregation operators. In a number of other works he extended ideas related to this operator [2, 6-12]. We define these operators in the following:

Definition: An OWA operator of dimension n is a mapping $F: R^n \rightarrow R$ which has an associated weighting vector

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ w_n \end{bmatrix}$$

in which

1. $w_j \in [0, 1]$
2. $\sum_{j=1}^{n} w_j = 1$

where

$$F(a_1, a_2,..., a_n) = \sum_{j=1}^{n} w_j b_j$$

with b_j being the j^{th} largest of the a_i.

The key feature of this operator is the ordering of the arguments by value, a process that introduces a nonlinearity into the operation. It can be shown that this operator is in the class of mean operators [13] as it is commutative, monotonic, and idempotent. It is also easy to show that for any weighting vector W, $\text{Min}(a_1) \leq F(a_1, a_2,..., a_n) \leq \text{Max}[a_i]$.

Its generality lies in the fact that by selecting the weights we can implement different aggregation operators. Specifically, by appropriately selecting the weights in W we can emphasize different arguments based upon their position in the ordering. If we place most of the weights near the top of W we can emphasize the higher scores while placing the weights near bottom of W emphasizes the lower scores in the aggregation. The following situation illustrates a prototypical use of this property. Assume we have n observations indicating the location of an enemy force. Our problem is to combine these readings to obtain a fused value. One approach is to simply take the average of these readings, however since the error of saying the enemy is farther away then closer is more costly then judging they are closer when they are far we may want to aggregate these values giving a preference to the lower scores, those saying the enemy is closer.

A number of special cases of these operators have been pointed out in the literature [8]. Each of these special cases is distinguished by the structure of the weighting vector W. Consider the situation where the weights are such that $w_1 = 1$ and $w_j = 0$ for all $j \neq 1$, this weighting vector is denoted as W^*, in this case we get

$$F(a_1, a_2, \ldots, a_n) = \mathrm{Max}_j[a_j].$$

Thus the Max operator is a special case of the OWA operator.

If the weights are such that $w_n = 1$ and $w_j = 0$ for $j \neq n$, denoted W_*, we get

$$F(a_1, a_2, \ldots, a_n) = \mathrm{Min}_j[a_j].$$

Thus the Min operator is a special case of of the OWA operator. As we noted above the Min and Max provide the extremes of this operator.

If the weights are such that $w_j = \frac{1}{n}$ for all j, denoted W_{ave}, then

$$F(a_1, a_2, \ldots, a_n) = \frac{1}{n} \sum_{j=1}^{n} a_j.$$

Thus we see that the simple average is also a special case of these operators.

If $w_k = 1$ and $w_j = 0$ for $j \neq k$, denoted $W_{(k)}$, then

$$F(a_1, a_2, \ldots, a_n) = b_k \quad \text{(the } k^{th} \text{ largest of the } a_i\text{)}$$

The median is also a special case of this family of operators. If n is odd we obtain the median by selecting $w_{\frac{n+1}{2}} = 1$ and letting $w_j = 0$ for $j \neq \frac{n+1}{2}$. If n is even we get the mode by selecting $w_{\frac{n}{2}} = w_{\frac{n}{2}+1} = \frac{1}{2}$ and letting $w_j = 0$ for all other terms.

An interesting class of these operators are the so-called olympic aggregators. The simplest example of this case is where we select $w_1 = w_n = 0$ and let $w_j = \frac{1}{n-2}$ for $j \neq 1$ or n. In this case we have eliminated the highest and lowest scores and taken the average of the rest. We note that this process is often used in obtaining aggregated scores from judges in olympic events such as gymnastics and diving.

In [1] we introduced two measures, dependent on the weighting vector, useful for characterizing the OWA operators. The first of these measures is

$$\alpha = \sum_{j=1}^{n} \frac{n-j}{n-1} w_j.$$

It can be shown $\alpha \in [0, 1]$. Furthermore it can be easily shown that

$$\alpha = 1 \qquad \text{if } W = W^*$$
$$\alpha = 0.5 \qquad \text{if } W = W_{ave}$$
$$\alpha = 0 \qquad \text{if } W_*.$$

Essentially α provides some indication of the inclination of the OWA operators for giving more weight to the higher scores or lower scores. The closer α is to one the more preference is given to the higher scores, the closer to zero the more preference is given to lower scores and a value close to 0.5 indicates an equal preference given to all scores. The actual semantics associated with α depends upon the application at hand. For example, in using the OWA operators to model logical connectives between the *and* and *or*, as was done in [1], α can be associated with a measure of the degree of *orness* associated with an aggregation. In [2], where the OWA operator was used to implement different procedures for modeling subjective attitudes in decision making under uncertainty, α was associated with the degree of optimism of a decision maker. Generically we can indicate α as a measure of *maxness* of the associated aggregation. We note that if we use $W_{(k)}$ then $\alpha = \frac{n - k}{n - 1}$ and we see that as k moves from one, Max, to n, Min, α gets smaller.

It can be shown that while $\alpha = 1$ only if $W = W^*$ and $\alpha = 0$ only if $W = W_*$ other values of α can be obtained for many different cases of W. A particularly interesting case is $\alpha = 0.5$. In [3] it was shown that for any OWA operator having a W with $w_{n-j+1} = w_j$ we get $\alpha = 0.5$. Thus we see any symmetric OWA operator has $\alpha = 0.5$, essentially these operators are in the same spirit as the simple average.

The second measure introduced in [1] was

$$\text{Disp}(W) = - \sum_{j=1}^{n} w_j \ln(w_j).$$

In [1] it was suggested that this measure can be used to measure the degree to which we use all the information in the argument. It can be shown that for all W

$$0 \leq \text{Disp}(w) \leq \ln(n).$$

We note $\text{Disp}(w) = 0$ iff $W = W_{(k)}$ and $\text{Disp}(w) = \ln(n)$ iff $W = \text{Wave}$. It can be shown that of all the symmetric implementations of W, those having $\alpha = 0.5$, W_{ave} has the largest measure of Disp.

3. OBTAINING THE OWA WEIGHTS

The great generality that the OWA aggregation function offers brings with it the problem of obtaining the weights to be used in a particular problem. Perhaps the most direct approach is to subjectively select the weights used in the vector W. This subjective selection is not to be done blindly, but should be guided by knowledge about the application at hand. Other than this direct approach there appears to be at least three categories of methods for obtaining the weights: learning of the weights from data, selection of a member from a given class of OWA operators and the use of linguistic imperatives to describe the aggregation. In the following we shall briefly describe each of these methods in turn.

As noted above, one approach to obtaining the weights is to learn the weights from available data. In the following we shall describe one methodology in this spent, more details about this approach can be found in [14, 15].

Assume a collection of m observations where each observation is comprised of an n-tuple of values $(a_{k1}, a_{k2}, \ldots, a_{kn})$, called the arguments, and an associated single value called the aggregated value, which we shall denote as d_k. Our goal is to find an OWA operator, a weighting vector W, that ***best*** models the process of aggregation used in this data set. Essentially we are interested in finding a weighting vector W such that for the entire collection of data we as faithfully as possible satisfy for every observation the condition

$$F(a_{k1}, a_{k2}, \ldots, a_{kn}) = d_k,$$

where F indicates the OWA aggregation of the arguments using W.

This problem can be simplified by taking advantage of the linearity of the OWA aggregation with respect to the ordered arguments. We denote the reordered objects of the k-th sample by b_{k1}, b_{k2}, ..., b_{kn} where b_{kj} is the j^{th} largest element of the argument collection a_{k1}, a_{k2},..., a_{kn}. Using these ordered arguments the problem becomes that of finding the vector of OWA weights $W = [w_1, w_2 ... w_n]^T$ that best satisfies

$$b_{k1} w_1 + b_{k2} w_2 + \dots + b_{kn} w_n = d_k$$

for every k = 1 to m.

Using the type of gradient descent techniques that have proved so successful in the learning algorithms used in neural networks we look for a weighting vector $W = [w_1, w_2, \dots, w_n]^T$ that minimizes the instantaneous errors e_k where,

$$e_k = \frac{1}{2}((b_{k1} w_1 + b_{k2} w_2 + \dots + b_{kn} w_n) - d_k)^2$$

with respects to weights w_i. However the solution to this problem is not the simple application of the Widrow-Hoff rule [16] because the situation is complicated by the fact that the minimization problem is a constrained optizimation problem, the w_i have to satisfy the properties

1. $\sum_{i=1}^{n} w_i = 1$
2. $w_i \in [0, 1]$, $i = (1, n)$.

To circumvent the constraints on the weights it was suggested in [14, 15] that we express the w_i's as

$$w_i = \frac{e^{\lambda_i}}{\sum_{j=1}^{n} e^{\lambda_j}}, \quad \text{for } i = 1 \text{ to } n.$$

We see that for any value of parameters λ_i the weights w_i will be positive and will sum to 1. Therefore the constrained minimization problem can be transformed into the unconstrained nonlinear programming problem of finding the λ_i that minimizes

$$e_k = \frac{1}{2}(b_{k1}\frac{e^{\lambda_1}}{\sum_{j=1}^{n} e^{\lambda_j}} + b_{k2}\frac{e^{\lambda_2}}{\sum_{j=1}^{n} e^{\lambda_j}} + \dots + b_{kn}\frac{e^{\lambda_n}}{\sum_{j=1}^{n} e^{\lambda_j}} - d_k)^2.$$

Using the gradient descent method we obtain the following rule for updating the parameters

$$\lambda_i(l + 1) = \lambda_i(l) - \beta\, w_i(l)\, (b_{ki} - \hat{d}_k)\, (\hat{d}_k - d_k)$$

where $\lambda_i(l + 1)$ is our new estimate for λ_i, β denotes the learning rate $(0 \le \beta \le 1)$, where for each i, $w_i(l) = \frac{e^{\lambda_i(l)}}{\sum_{j=1}^{n} e^{\lambda_j(l)}}$ is the estimate of w_i after the lth iteration and where $\hat{d}_k = b_{k1}\, w_1(l) + b_{k2}\, w_2(l) + \dots + b_{kn}\, w_n(l)$.

The process of updating the λ_i continues until we get small changes of parameter estimates $\Delta_i = |\lambda_i(l + 1) - \lambda_i(l)\,|$, $i = (1, n)$.

The second general class of approaches to the determination of the OWA weights is the selection of a weighting vector from a class of OWA operators. This generally simply involves the selection of one or perhaps two parameters to characterize an aggregation and thus reduces the difficulty of the problem. As an example of this approach consider the family of operators defined by

$$w_i = 0 \qquad \text{for } i = 1 \text{ to } q$$

$$w_i = 0 \qquad \text{for } i = n \text{ to } n - q + 1$$

$$w_i = \frac{1}{n - 2q} \qquad \text{for all other } i$$

where q is a parameter to be selected. Here we see that we essentially disregard the q highest and q lowest values and take the average of the remaining scores. In [8] we discuss a number of different families of OWA operator useful for this approach.

Perhaps the most often used approach in this parameterized spirit is the methodology suggested by O'Hagan [17] which makes use of the measure of maxness and dispersion introduced earlier. In this approach all that is required is that the user provide a value $\alpha \in [0, 1]$ corresponding to the degree of maxness. The ME-OWA operators, as they are called by O'Hagan, are obtained as the set of weights w_j, $j = 1$ to n, that satisfy the following mathematical programming problem:

$$\text{Maximize: } - \sum_{j=1}^{n} w_j \ln (w_j)$$

Subject to:

$$1.\ \alpha = \sum_{j=1}^{n} \frac{n - j}{n - 1}\, w_j$$

$$2.\ \sum_{j=1}^{n} w_j = 1$$

$$3.\ 0 \le w_j, \le 1 \qquad \text{for } j = 1 \text{ to } n.$$

Thus we see that these operators are those that have the maximum degree of dispersion for a given degree of maxness, α. As noted above a significant advantage of this approach is that we can specify an OWA operator by just providing the one parameter, α.

Another approach to the determination of the weights used in the OWA approach is based on the generation of the weights from some associated linguistic quantifier [18]. Let us first describe the spirit of this approach. Assume $a_i, \dots, a_n$ are a collection of scores we want to aggregate, however rather than considering all the scores in the aggregation we need only consider *most* of the scores. For example, in information retrieval we may have a collection of characteristics describing a desired document but rather than requiring the document to meet all these conditions it only need satisfy *most* of them. The term *most* is an example of what Zadeh called a linguistic quantifier. In addition to *most* linguistic quantifiers are exemplified by terms such as *almost all*, *many*, *at least half* and *some*. These terms can be viewed as fuzzy proportions. In [18] Zadeh suggested a formal representation of linguistic quantifiers in terms of fuzzy subsets. In particular, if $\underset{\sim}{Q}$ is a linguistic quantifier, such as *most*, we can represent this as a fuzzy subset Q over the unit interval where for any $r \in [0, 1]$ the membership grade Q(r) indicates the degree of compatibility of r with the concept $\underset{\sim}{Q}$. An important class of these quantifiers are what Yager [4] called Regular Increasing Monotone (RIM) quantifiers. These are characterized by the properties

$$Q(0) = 0$$
$$Q(1) = 1$$
$$Q(x) \geq Q(x) \quad \text{if } x > y.$$

Essentially these are quantifiers for which the degree of compatibility increases as the proportion of elements increases.

In [1] Yager suggested an approach to the generation of OWA weights from a linguistic quantifier. Assume we have a collection of n scores whose aggregation is guided by a RIM linguistic quantifier represented as a fuzzy subset Q. It is suggested that this aggregation can be implemented using an OWA aggregation in which the weights associated with the aggregation are obtained as

$$w_i = Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right) \qquad i = 1 \text{ to } n.$$

Thus if the quantifier *most* is represented as $Q(r) = r^2$ and m = 4, then

$$w_1 = Q\left(\frac{1}{4}\right) - Q\left(\frac{0}{4}\right) = \left(\frac{1}{4}\right)^2 - 0 = .063$$
$$w_2 = Q\left(\frac{2}{4}\right) - Q\left(\frac{1}{4}\right) = \left(\frac{2}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = .187$$
$$w_3 = Q\left(\frac{3}{4}\right) - Q\left(\frac{2}{4}\right) = \left(\frac{3}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = .313$$
$$w_4 = Q\left(\frac{4}{4}\right) - Q\left(\frac{3}{4}\right) = (1)^2 - \left(\frac{3}{4}\right)^2 = .437$$

4. MULTI-CRITERIA DECISION MAKING AND THE INCLUSION OF IMPORTANCES

An important class of aggregation problems are those we shall generically refer to as multi–criteria aggregation problems. Examples of these problems arise in many applications such as decision making, pattern recognition and information retrieval. In decision making the criteria generally refer to the goals, constraints and objectives of the decision maker. In pattern recognition the criteria generally refer to the characteristics of an object in a given class. In information retrieval the criteria are associated with the properties of the document one is looking for. Typically in these problems we have a collection criteria, A_i for $i = 1$ to n, and a collection X of alternative solutions. For a given alternative x we get a value $a_i \in [0, 1]$ indicating the degree to which alternative x satisfies the criteria A_i. Using these values, the a_i's, we now combine them to obtain an overall score for alternative x, this overall score is denoted

$$a^* = \mathrm{Agg}(a_1, a_2, \ldots, a_n).$$

The individual alternatives are then compared with respect to these overall scores.

As discussed in [1] a possible method for implementing the Agg operator is to use an OWA operator,

$$a^* = F_W(a_1, a_2, \ldots, a_n).$$

In the above F_W is a particular OWA operator. The selection of the OWA operator, choice of W, is a reflection of the relationship between the criteria being aggregated. If we require that *all* the criteria be satisfied then we use $W = W_*$, Min. If we are satisfied if *any* of the criteria are satisfied we use $W = W^*$, Max. If we require an average of the individual criteria satisfactions, we can use $W = W_{ave}$. As discussed in [1] the use of the OWA operators allows the representations of sophisticated relationships between the criteria.

In many of these multi-criteria problems we have a situation in which the individual criteria have different degrees of importance associated with them. For example, if we are involved in a decision problem to select a car we may have as our criteria that the car is inexpensive, comfortable and gives good gas mileage. However, in addition it may be indicated the price is a more important criteria then the others. The ability to use the OWA operators in these kinds of problems necessitates our being able to take into account in the OWA aggregation the different importances associated with the criteria. Thus we are forced with the problem of calculating

$$a^* = F_W\left((u_1, a_1,), (u_2, a_2)), \ldots, (u_n, a_n)\right)$$

where $a_i \in [0,1]$ is the score of the ith criteria and $u_i \in [0, 1]$ is the importance of the ith criteria. In [1] we suggested a number of approaches to addressing this problem. The difficulty in obtaining a fully satisfying approach to the inclusion of importances arises because of the different ways we must handle the importance for different types of OWA aggregation. In the following we shall first look at some of the classical approaches used for including importances in some of the distinctive OWA operators.

In the case of the average operator the prototypical approach to including importances is to use the weighted average,

$$a^* = \sum_{j=1}^{n} \frac{u_j}{T} a_j = \frac{1}{n} \sum_{j=1}^{n} \frac{n}{T} u_j a_j$$

where $T = \sum_{j=1}^{n} u_j$, the sum of the importance weights.

In the case of the Min operator, as discussed in [19], a general approach for the inclusion of importances is to use

$$a^* = \mathrm{Min}_j[S(\bar{u}_j, a_j)]$$

where $\bar{u}_j = 1 - u_j$, and S is any t-conorm [20]. Two notable cases are worth pointing out. When S = Max, we get

$$a^* = \mathrm{Min}_j[\mathrm{Max}(\bar{u}_j, a_j)].$$

When S(a, b) = a + b - ab, bounded sum, we get

$$a^* = \mathrm{Min}_j[(\bar{u}_j + u_j\, a_j)].$$

In the case of the Max operator a general approach to including importances [19] is to use

$$A^* = \mathrm{Max}_j[T(u_j, a_j)].$$

where T is any t-norm [20]. Among the notable cases are

$$a^* = \mathrm{Max}_j[u_j\, a_j] \quad \text{when } T(a,b) = a\, b$$

$$a^* = \mathrm{Max}_j[\mathrm{Min}(u_j, a_j)] \quad \text{when } T(a,b) = \mathrm{Min}(a, b).$$

While each of the three different cases, average, max and min use different methodologies for the inclusion of importance we can see some underlying regularity. Let us denote Agg as the basic operation used to combine the scores, Average, Max, or Min. Essentially Agg takes n scores in the unit interval and returns a value in the unit interval, Agg: $I^n \rightarrow I$.

In the case when we have importances associated with the criteria, instead of having single values we have types (u_j, a_j) and thus we can't directly use the Agg function. A careful look at the three procedures described above for incorporating importances indicates the following unifying process. We take each pair (u_j, a_j) and apply an importance transformation operation to get an efficient value, $b_j = G(u_j, a_j)$. We then apply the appropriate basic Agg operation on these transformed values,

$$a^* = \mathrm{Agg}(b_1, b_2, \ldots, b_n).$$

Table 1 summarizes these three cases.

Name	Agg Operator	Transformation of (u_j, a_j)
Max	$\mathrm{Max}_j\, [b_j]$	$b_j = T(u_j, a_j)$
Min	$\mathrm{Min}_j\, [b_j]$	$b_j = S(\bar{u}_j, a_j)$
Average	$\frac{1}{n}\sum_{j=1}^{n} b_j$	$b_j = \frac{n}{T}\, u_j\, a_j$

Table 1: Importance Inclusion for different aggregation operators

5. IMPORTANCES IN OWA AGGREGATIONS VIA FUZZY MODELING

We now describe an approach for including importances in the OWA aggregation method based on the idea of importance transformation [3]. As we shall see use is made of fuzzy systems modeling to obtain the form of the importance transformation operator. Assume F is an OWA operator of dimension n with weighting vector W. Furthermore, we assume that each of the arguments to be aggregated have an associated importance weight, thus we have a collection of n pairs, (u_i, a_i) where

$u_i \in [0, 1]$ is the importance weight and $a_i \in [0, 1]$ is the score. In the previous section we saw that one approach to aggregation when we have importance weighted scores was to introduce some transformation G that converts the importance weights and scores into some effective value, $b_j = G(u_j, a_j)$ and then aggregate these effective scores using our aggregation operator. It appears natural to try to use this approach in the case of the OWA aggregation. Using this idea we would calculate

$$a^* = F(b_1, b_2, \ldots, b_n)$$

where $b_j = G(u_j, a_j)$.

While this seems like a reasonable way to proceed one basic problem arises. As we noted in the previous section the form of G depends upon the type of aggregation being performed, F. That is we noted Max, Min, and Average had different forms for G. Since the OWA operator, depending upon the weighting vector W, can take on different types of aggregation it is not clear what form should be used for G. That is G depends upon the W. Since α, the degree of maxness of the aggregation, provides a simple indication of the type of aggregation being performed a natural approach is to try to construct G as a parameterized function of α. However, a further compounding factor arises due to the fact that we only know the form of G for some prototypical cases; Max, Min, and Average. One possible approach to dealing with this situation is to use fuzzy systems modeling [4] to construct G from our knowledge about the form of G for the known cases. We recall that fuzzy systems modeling allows the construction of complex relationship from partial knowledge about the relationship. In the following we use a particular form of fuzzy modeling introduced by Sugeno [21].

We recall if R and V are two variables taking their values in the spaces X and Y respectively then a fuzzy systems model of the Sugeno type is a collection of p rules of the form

if R is A_i then V is y_i

where A_i is a fuzzy subset of X and $y_i \in Y$.

In these models given an input value for R, x*,we calculate the output value y* as

$$y^* = \frac{\sum_{i=1}^{p} y_i A_i(x^*)}{\sum_{i=1}^{p} A_i(x^*)}$$

We see y* is the weighted average of the rule outputs where the weighting value is the normalized firing level of each rule. Thus denoting $\upsilon_i = \frac{A_i(x^*)}{\sum_{i=1}^{p} A_i(x^*)}$ we get

$$y^* = \sum_{i=1}^{p} v_i\, A_i(x^*).$$

Let us now apply this methodology to the problem of constructing G. In this environment we associate with R the value α indicating the type of aggregation

being performed. We associate with V the form of the corresponding importance transform function. In this situation our rules are of the type:
if **degree of orness** is α_i then **the importance transform function** is $G_i(u, a)$.

Let us first consider the construction of this generalized transformation operator just using our knowledge about the way of including importances in the Max and Min aggregation. Since we know that when $\alpha = 1$ we get a Max type aggregation and when $\alpha = 0$ we get a Min type aggregation we can consider the construction of G from the following simple fuzzy system model:

if **value of** α is *high* then **G(u, a)** is $G_{Max}(u, a)$

if **value of** α is *low* then **G(u, a)** is $G_{Min}(u, a)$

where $G_{Max}(u, a)$ and $G_{Min}(u, a)$ are the respective transformation for Max and Min type aggregations.

Using this model if for a given OWA aggregation we have a degree of orness equal α then we get as our form for G

$$G(u, a) = \frac{high(\alpha) G_{Max}(u, a) + low(\alpha) G_{Min}(u, a)}{high(\alpha) + low(\alpha)}$$

Using the following simple linear definitions for high and low

$$high(\alpha) = \alpha$$

$$low(\alpha) = 1 - high(\alpha) = 1 - \alpha,$$

we have $low(\alpha) + high(\alpha) = 1$ and we get

$$G(u, a) = \alpha\, G_{max}(u, a) + (1 - \alpha)\, G_{min}(u, a)$$

which is simply a weighted average of the Max and Min type importance transformation.

Using $G_{max}(u, a) = T(u, a)$ and $G_{min}(u, a) = S(\bar{u}, a)$ we get

$$G(u, a) = \alpha\, T(u, a) + (1 - \alpha)\, S(\bar{u}, a).$$

If we use the product and probabilistic sum, $T(u, a) = u\, a$ and $S(\bar{u}, a) = \bar{u} + a - \bar{u}\, a = \bar{u} + u\, a$ we get $G(u, a) = \alpha\, a\, u + (1 - \alpha)(\bar{u} + u\, a)$ after some algebraic manipulation of the above we get the following nice form

$$G(u, a) = \bar{u}\, \bar{\alpha} + u\, a.$$

Before we proceed to investigate this form in more detail let us look at an example of its application.

Example: Assume we desire to aggregate

$$(0.7, 0.8), (1, 0.7), (0.5, 1), (0.3, 0.9),$$

where the first term in these tuples is the importance weight and the second is the score. Assume our aggregation is an OWA operator F with $W = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.2 \\ 0.1 \end{bmatrix}$.

The following is a step by step illustration of the procedure

1. Calculate the degree of orness α of the OWA operator using $\alpha = \sum_{j=1}^{n} \frac{n-j}{n-1} w_j$. Since n = 4 we have

$$\alpha = \sum_{j=1}^{4} \frac{4-j}{4-1} w_j = \frac{1}{3}[(3)(0.4) + (2)(0.3) + (1)(0.2)] = 2/3 = 0.67$$

2. Transform each of the argument tuples using

$$b_j = G(u_j, a_j) = \bar{\alpha}\,\bar{u}_j + u_j\, a_j = \frac{1}{3}\bar{u}_j + u_j\, a_j$$

hence

$$b_1 = \frac{1}{3}(0.3) + (0.7)(0.8) = 0.66$$

$$b_2 = \frac{1}{3}(0) + (1)(0.7) = 0.7$$

$$b_3 = \frac{1}{3}(0.5) + (0.5)(1) = 0.17 + 0.5 = 0.67$$

$$b_4 = \frac{1}{3}(0.7) + (0.3)(0.9) = 0.23 + 0.22 = 0.5$$

3. Calculate $F_W(0.66, 0.7, 0.67, 0.6)$

   i. order the arguments: $c_1 = 0.7$; $c_2 = 0.67$; $c_3 = 0.66$; $c_4 = 0.5$

   ii. $a^* = w_1\, c_1 + w_2\, c_2 + w_3\, c_3 + w_4\, c_4$

   $a^* = (.04)(.07) + (0.3)(0.67) + (0.2)(0.66) + (0.1)(0.5) = 0.66$

Let us now look in more detail at this suggested transformation for including importances, $b_j = G_\alpha(u_j, a_j) = \bar{\alpha}\,\bar{u}_j + a_j\, u_j$. It is clear that this is monotonic with respect to the scores, if $a_j > \hat{a}_j$, then $G_\alpha(u_j, a_j) \geq G(u_j, \hat{a}_j)$.

Secondly we note that if the importance value is 1, $u_j = 1$, then $G_\alpha(u_j, a_j) = a_j$. We also see that if $u_j = 0$, then $G_\alpha(u_j, a_j) = \bar{\alpha}$, the complement of degree of maxness.

Let us investigate the functioning of this transformation for the distinguished cases, Max, Min and average. We first see that if $\alpha = 1$, Max type aggregation, $b_j = a_j\, u_j$ which is an appropriate transformation for this type of aggregation, it is a t-norm. If $\alpha = 0$, Min type aggregation, $b_j = \bar{u}_j + a_j\, u_j$ which is again an appropriate type transform.

Let us now look at the performance of this transformation in the case of the simple average. We recall that for the average the OWA weights are $w_j = \frac{1}{n}$ and $\alpha = 0.5$. In this case

$$b_j = \bar{\alpha}\,\bar{u}_j + a_j\, u_j = \frac{1}{2}\,\bar{u}_j + a_j\, u_j.$$

Since all the w_j are equal there is no need to reorder the transformed values to calculate the OWA aggregation,

$$a^* = F_{WA}(b_1, b_2, \ldots, b_n) = \frac{1}{n}\sum_{j=1}^{n} b_j = \frac{1}{n}\sum_{j=1}^{n} a_j u_j + \frac{1}{2}\left(1 - \frac{T}{n_j}\right)$$

where $T = \sum_{j=1}^{n} u_j$. We can furthermore express this as

$$a^* = \frac{T}{n}\frac{1}{T}\sum_{j=1}^{n} a_j u_j + \frac{1}{2}(1-\frac{T}{n}).$$

Denoting the usual weighted average, $\frac{1}{T}\sum_{j=1}^{n} a_j u_j$, as $\hat{a}$ and letting $\rho = \frac{T}{n}$ we get

$$a^* = \rho\, \hat{a} + \frac{1}{2}(1 - \rho).$$

We first note that if all the weights are equal one, $u_i = 1$ for all i, then $T = n$, $\rho = 1$, and we get $a^* = \hat{a}$. Thus for the case when the importances are all one we get the usual simple average.

Let us further look at this case. Consider two sets of arguments having the same associated importance weights:

$$(u_j, a_j) \qquad j = 1, 2, ..., n.$$
$$(u_j, b_j) \qquad j = 1, 2, ..., n.$$

Let us denote $\hat{a} = \frac{1}{T}\sum u_j a_j$ and $\hat{b} = \frac{1}{T}\sum u_j b_j$. We note that in this case since the importance weights are the same for both cases the ρ's are the same, $\rho = \frac{\sum u_i}{n}$. We see that $a^* - b^* = \rho\,(\hat{a} - \hat{b})$, from this it follows that the comparative ordering of a^* and b^* is the same as $\hat{a}$ and $\hat{b}$ that is if $\hat{a} > \hat{b}$ then $a^* > b^*$.

Thus we see that this approach, gives the same ordering as the weighted average. Thus if we have two alternatives in a multi-criteria decision problem comparing them using the weighted average or the above method would lead to the same ordering. Thus we have shown that while the use of this model to obtain the importance transformation operator doesn't exactly reproduce the weighted average operation it provides an ordering that is equivalent to the use of the weighted average. We should note that the simplicity of the form of G makes this a very appealing method of calculating the importance transformation.

In an attempt to improve the fidelity with respect to the average operation we can consider the construction of G from a knowledge base that includes a rule describing the transformation to be used when we have an OWA aggregation corresponding to a simple average.

Consider the following rule base

if **value of** α is *high* then **G(u, a)** is $G_{Max}(u, a)$

if **value of** α is *high* then G(u, a) is $G_{Max}(u, a)$

if **value of** α is *medium* then **G(u, a)** is $G_{Avg}(u, a)$

if **value of** α is *low* then **G(u, a)** is $G_{Min}(u, a)$

In this case we get as our transformation operator for a given value α

$$G(u, a) = \frac{high(\alpha)\, G_{Max}(u, a) + medium(\alpha)\, G_{Max}(u, a) + low(\alpha) G_{Min}(u, a)}{high(\alpha) + medium(\alpha) + low(\alpha)}$$

In figure #1 we show one partitioning of the the α value on the unit interval

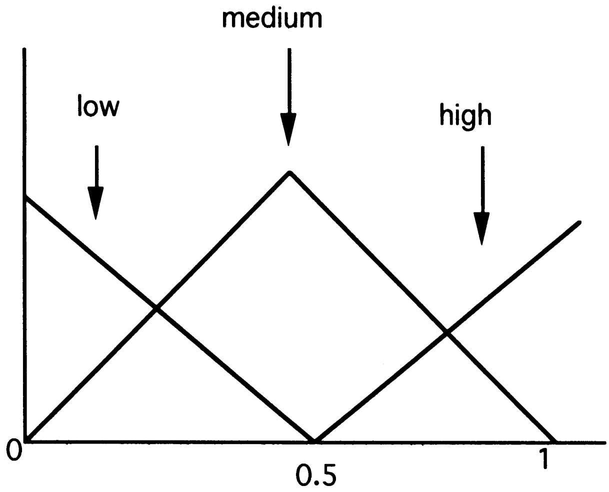


Figure #1 A partitioning of the α values

In this case we have following definitions for the associated fuzzy subsets:

High:

$$High(\alpha) = 0 \quad \text{for } \alpha \le 0.5$$
$$High(\alpha) = 2\alpha - 1 \quad \text{for } \alpha \ge 0.5$$

Low:

$$Low(\alpha) = -2\alpha + 1 \quad \text{for } \alpha \le 0.5$$
$$Low(\alpha) = 0 \quad \text{for } \alpha \ge 0.5$$

Medium:

$$Medium(\alpha) = 2\alpha \quad \text{for } \alpha \le 0.5$$
$$Medium(\alpha) = 2 - 2\alpha \quad \text{for } \alpha \ge 0.5$$

We should note that this partitioning has the property that for all α

$$Low(\alpha) + Medium(\alpha) + High(\alpha) = 1.$$

Using these definitions for the partitioning of α interval we get the following formulation for G(u, a):

$$G(u, a) = (-2\alpha + 1)\, G_{Min}(u, a) + 2\alpha\, G_{Av}(u, a) \qquad \text{for } \alpha \le 0.5$$
$$G(u, a) = (2\alpha - 1)\, G_{Max}(u, a) + (2 - 2\alpha)\, G_{Av}(u, a) \qquad \text{for } \alpha \ge 0.5$$

We see that G(u, a) is a piecewise linear function of α, where $\alpha = 0.5$ is the point where we change from one function to the next. Using as our prime transformations

$$G_{Max} = \underline{u}\, a$$
$$G_{Min} = \bar{u} + u\, a$$
$$G_{Av} = \frac{n}{T}\, u\, a$$

we get

$$G_1(u, a) = (-2\alpha + 1)\, (\bar{u} + u\, a) + 2\alpha\left(\frac{n}{T}\, u\, a\right) \qquad \text{for } \alpha \le 0.5$$
$$G_2(u, a) = (2\alpha - 1)\, u\, a + (2 - 2\alpha)\left(\frac{n}{T}\, u\, a\right) \qquad \text{for } \alpha \le 0.5$$

We can somewhat simplify these formulas as follows:

$$G_1(u, a) = u\,a\,(2\alpha - 1 + \frac{2n}{T} - 2\alpha\,\frac{n}{T})$$

$$G_1(u, a) = u\,a\,(2\bar{\alpha}\,(\frac{n}{T} - 1) + 1) = \frac{ua}{T}(2\bar{\alpha}(n - T) + T)$$

We see that for $\alpha = 1$, $G(u, a) = u\,a$, for $\alpha = 0$ we get $G(u, a) = \bar{u} + u\,a$ and for $\alpha = 0.5$ $G(u, a) = \frac{n}{T}\,u\,a$ as desired. This form provides complete fidelity for Min, Max, and Average it does so at the expense of a more complex form for the importance transformation operation.

Example: Assume we desire to aggregate (0.7, 0.8), (1, 0.7), (0.5, 1) and (0.3, 0.9) where the first terms are the importance weights and the second are the scores. We note in this situation $n = 4$ and $T = 2.5$ and $\frac{n}{T} = 1.6$.

a. Consider first the aggregation using

$$W = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.2 \\ 0.1 \end{bmatrix}.$$

In this case $\alpha = 0.67$ and therefore we must use $G_2(u_j, a_j) = u_j\,a_j\,(2\bar{\alpha}\,(\frac{n}{T} - 1) + 1)$

Hence we get

$$G_2(u_j, a_j) = u_j\,a_j\,(1 + (0.67\,)(0.6)) = u_j\,a_j\,(1.4)$$

From this we get: $b_1 = 0.78$, $b_2 = 0.98$, $b_3 = 0.7$, $b_4 = 0.38$.

Recording these values we get: $c_1 = 0.98$, $c_2 = 0.78$, $c_3 = 0.7$, $c_4 = 0.38$

Calculating the OWA aggregation we get

$$a^* = (0.98)(0.4) + (0.78)(0.3) + (0.7)(0.2) + (0.38)(0.1) = 0.8$$

b. Consider the aggregation using

$$W = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{bmatrix}.$$

In this case $\alpha = \frac{1}{3}$ and therefore we must use $G_1(u_j, a_j) = \frac{1}{3}\,\bar{u}_j + 1.4\,u_i\,a_j$. From this we get: $b_1 = 0.88$, $b_2 = 0.98$, $b_3 = 0.87$, $b_4 = 0.6$.

Recording these values we get: $c_1 = 0.98$, $c_2 = 0.88$, $c_3 = 0.87$, $c_4 = 0.6$.

Calculating the OWA aggregated value we get

$$a^* = (0.1)(0.98) + (0.2)(0.88) + (0.3)(0.87) + (0.4)(0.6) = 0.77.$$

6. IMPORTANCES WITH LINGUISTIC QUANTIFIERS

In the preceding we have suggested an approach to the inclusion of importances in the OWA aggregation process using a transformation operation based upon the score and the importance of the object. In this approach no change is made to the

OWA weights used in the aggregation but the scores are transformed based upon the importance weight. In [6] Yager suggested an alternative approach to the inclusion of importances in the OWA aggregation process, see also Torra [22]. In this approach, rather than modifying the scores with respect to importance we use the importances to modify the weighting vector. This approach works most naturally in environments in which we have some linguistic quantifier $\mathcal{Q}$ guiding the aggregation process.

Assume we have a collection of n pairs (u_i, a_j) where $u_j \in [0,1]$ is the importance weight and $a_j \in [0,1]$ indicates the score. Let Q be a fuzzy subset on the unit interval corresponding to some linguistic quantifier. In the framework of multi-criteria decision making where u_j is the importance of the jth criteria and a_j is the satisfaction of a given alternative to the jth criteria our aggregation imperative would be.

$\mathcal{Q}$ important criteria are satisfied by a satisfactory alternative.

The procedure suggested by Yager in [5, 6] is as follows.

The first step is to order the a_j's in descending order, here we shall denote b_i be the ith largest of the a_j. Furthermore. we let v_i denote the importance associated with the score that has the ith largest value. Thus if a_5 is the largest of a_i then $b_1 = a_5$ and $v_1 = u_5$. At this point we can consider our information to be aggregated to be a collection of n pairs (v_i, b_i) where the b_i's are in descending order. Our next step is to obtain the OWA weights associated with this aggregation. We obtain these weights as

$$w_i = Q\left(\frac{S_i}{T}\right) - Q\left(\frac{S_{i-1}}{T}\right) \qquad \text{for } i = 1 \text{ to } n.$$

where $S_i = \sum_{k=1}^{i} v_k$ and $T = S_n = \sum_{k=1}^{n} v_k$. Thus T is the sum of all importances and S_i is the sum of the importances for the i highest scores

Finally we calculate the aggregated value a* as

$$a^* = \sum_{i=1}^{n} b_i w_i.$$

The following example illustrates the application of the above method.

Example: Assume we have two alternative x and y. We shall assume four criteria A_1, A_2, A_3, A_4. The importances associated with these criteria are: $u_1 = 1$, $u_2 = 0.6$, 0.5 and $u_4 = 0.9$. We note $T = \sum_{i=1}^{4} u_j = 3$. In addition we shall assume that the satisfactions to each of the criteria by the alternatives is given by the following:

$$A_1(x) = 0.7,\ A_2(x) = 1,\ A_3(x) = 0.5,\ A_4(x) = 0.6$$

$$A_1(y) = 0.6,\ A_2(y) = 0.3,\ A_3(y) = 0.9,\ A_4(y) = 1.$$

We shall assume the quantifier guiding this aggregation to be *most* which is defined by $Q(r) = r^2$.

We first consider the aggregation for x. In this case the ordering of the criteria satisfaction gives us:

	b_j	v_j
A_4	1	0.6
A_3	0.7	1
A_1	0.6	0.9
A_2	0.5	0.5

Calculating the weights associated with x, which we denoted $w_i(x)$, we get

$$w_1(x) = Q\left(\frac{0.6}{3}\right) - Q\left(\frac{0}{3}\right) = (0.2)^2 - 0 = 0.04$$

$$w_2(x) = Q\left(\frac{1.6}{3}\right) - Q\left(\frac{0.6}{3}\right) = .28 - .04 = .24$$

$$w_3(x) = Q\left(\frac{2.5}{3}\right) - Q\left(\frac{1.6}{3}\right) = .69 - .28 = .41$$

$$w_4(x) = Q\left(\frac{3}{3}\right) - Q\left(\frac{2.5}{3}\right) = 1 - .69 = .31$$

To obtain D(x) we calculate

$$D(x) = \sum_{i=1}^{4} w_i(x)\, b_1 = (.04)(1) + (.24)(.7) + (.41)(.6) + (.31)(.5) = 0.609$$

To calculate the evaluation for y we proceed as follows. In this case the ordering of the criteria satisfaction is

	b_j	v_j
A_4	1	0.9
A_3	0.9	0.5
A_1	0.6	1
A_2	0.3	0.6

The weights associated with the aggregation are:

$$w_1(y) = Q\left(\frac{.9}{3}\right) - Q\left(\frac{0}{3}\right) = .09 - 0 =.09$$

$$w_2(y) = Q\left(\frac{1.4}{3}\right) - Q\left(\frac{.9}{3}\right) = .22 - .09 =.13$$

$$w_3(y) = Q\left(\frac{2.4}{3}\right) - Q\left(\frac{1.4}{3}\right) = .64 - .22 =.42$$

$$w_4(y) = Q(1) - Q\left(\frac{2.4}{3}\right) =1 - .64 =.36$$

To obtain D(y) we calculate

$$D(y) = \sum_{i=1}^{4} w_i(y)\, b_i = (.09)(1) + (.13)(.9) + (.42)(.6) + (.36)(.3) = 0.567$$

Hence in this example x is the preferred alternative.

It is important to observe that the weights are different for the two aggregations this is due to the fact that the ordering of the satisfactions to the A_i's are different for x and y which lead to a different ordering of the v_j's resulting in a different weighting vector.

More details with respect to the properties of this methodology can be found in

[5, 6].

7. CONCLUSION

We have introduced two approaches for the inclusion of importance weights in the OWA aggregation method. The first approach makes use of a transformation operator on the scores and importances to obtain a transformed score. This transformation operator is obtained from a fuzzy systems model which has a knowledge base built from the required transformations for known aggregation operations. The second approach which requires the knowledge of an underlying linguistic quantifier to guide the aggregation and uses information about the importances to modify the OWA weights.

8. REFERENCES

[1]. Yager, R. R., "On ordered weighted averaging aggregation operators in multi-criteria decision making," IEEE Transactions on Systems, Man and Cybernetics 18, 183-190, 1988.

[2]. Yager, R. R., "Decision making under Dempster-Shafer uncertainties," International Journal of General Systems 20, 233-245, 1992.

[3]. Yager, R. R., "Including Importances in OWA Aggregations Using Fuzzy Systems Modeling," Technical Report# MII-1625 Machine Intelligence Institute, Iona College, New Rochelle, NY, 1996.

[4]. Yager, R. R. and Filev, D. P., Essentials of Fuzzy Modeling and Control, John Wiley: New York, 1994.

[5]. Yager, R. R., "Multicriteria decision making using fuzzy quantifiers," Proceedings of the IEEE Conference on Computational Intelligence for Financial Engineering, New York, 42-47, 1995.

[6]. Yager, R. R., "Quantifier guided aggregation using OWA operators," International Journal of Intelligent Systems 11, 49-73, 1996.

[7]. Yager, R. R., "Applications and extensions of OWA aggregations," International Journal of Man-Machine Studies 37, 103-132, 1992.

[8]. Yager, R. R., "Families of OWA operators," Fuzzy Sets and Systems 59, 125-148, 1993.

[9]. Yager, R. R. and Filev, D. P., "Fuzzy logic controllers with flexible structures," Proceedings Second International Conference on Fuzzy Sets and Neural Networks, Iizuka, Japan, 317-320, 1992.

[10]. Yager, R. R., "Interpreting linguistically quantified propositions," International Journal of Intelligent Systems 9, 541-569, 1994.

[11]. Yager, R. R. and Filev, D. P., "Parameterized "andlike" and "orlike" OWA operators," International Journal of General Systems 22, 297-316, 1994.

[12]. Yager, R. R., Filev, D. P. and Sadeghi, T., "Analysis of flexible structured fuzzy logic controllers," IEEE Transactions on Systems, Man and Cybernetics 24, 1035-1043, 1994.

[13]. Dubois, D. and Prade, H., "A review of fuzzy sets aggregation connectives," Information Sciences 36, 85 - 121, 1985.

[14]. Filev, D. P. and Yager, R. R., "Learning OWA operator weights from data," Proceedings of the Third IEEE International Conference on Fuzzy Systems, Orlando, 468-473, 1994.

[15]. Filev, D. P. and Yager, R. R., "On the issue of obtaining OWA operator weights," Fuzzy Sets and Systems, (To Appear).

[16]. Widrow, B. and Hoff, M. E., "Adaptive switching circuits," IRE WESCON Convention Record, 96-104, 1960.

[17]. O'Hagan, M., "Aggregating template or rule antecedents in real-time expert systems with fuzzy set logic," Proceedings 22nd Annual IEEE Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, Ca., 681-689, 1988.

[18]. Zadeh, L. A., "A computational approach to fuzzy quantifiers in natural languages," Computing and Mathematics with Applications 9, 149-184, 1983.

[19]. Yager, R. R., "A note on weighted queries in information retrieval systems," J. of the American Society of Information Sciences 38, 23-24, 1987.

[20]. Alsina, C., Trillas, E. and Valverde, L., "On some logical connectives for fuzzy set theory," J. Math Anal. & Appl. 93, 15-26, 1983.

[21]. Sugeno, M., "An introductory survey of fuzzy control," Information Sciences 36, 59-83, 1985.

[22]. Torra, V., "The weighted OWA operator," International Journal of Intelligent Systems, (To Appear).

ON THE LINGUISTIC OWA OPERATOR AND EXTENSIONS

F. Herrera and E. Herrera-Viedma

Department of Computer Science and Artificial Intelligence
University of Granada
18071–Granada, Spain
E-mail: herrera@decsai.ugr.es

Abstract

A summary on the linguistic OWA operators existing in the literature is presented. To deal with linguistic information with equal importance the LOWA and ordinal OWA operators are studied. To deal with weighted linguistic information two extensions of the LOWA operator are analyzed, the LWA operator when the weights have linguistic nature and the L-WOWA operator if they have numerical nature.

1 Introduction

Usually, experts express their opinions by means of numerical values. Sometimes, however an expert could have a vague knowledge about this preference valuations, for example the preference degree of any alternative over any other alternative, and cannot estimate his preference with an exact numerical value. Then, a more realistic approach may be to use linguistic assessments instead of numerical values, that is, to suppose that the variables which participate in the problem are assessed by means of linguistic terms [10, 16, 24, 28].

The linguistic approach considers the variables which participate in the problem assessed by means of linguistic terms instead of numerical values. A linguistic variable differs from a numerical one in that its values are not numbers, but words or sentences in a natural or artificial language [28]. Since words, in general, are less precise than numbers, the concept of a linguistic variable serves the purpose of providing a means of approximated characterization of

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

phenomena, which are too complex, or too ill-defined to be amenable its description in conventional quantitative terms. The linguistic approach has been applied to different areas, for instance, we can find applications on information retrieval [2], medical diagnostic [4], control systems [14], decision making [7, 16, 24], etc.

In contexts with multiple information sources an aggregation operator of linguistic labels is needed. Assuming a linguistic context, two main different approaches can be found in order to aggregate linguistic values: the first acts by direct computation on labels [5, 6, 7, 12, 22, 24, 25], and the second uses the associated membership functions [1, 4, 16]. Most available applications use techniques belonging to the latter kind, however, the final results of those methods are fuzzy sets which do not correspond to any label in the original term set. If one finally wants to have a label, then a "linguistic approximation" is needed [1, 4, 16].

In this work, we study aggregation operators of linguistic information, which work by means of direct computation on labels. These operators summary the linguistic approaches to the Ordered Weighting Averaging (OWA) operator [21]. To combine non-weighted linguistic information we study the *Linguistic OWA (LOWA)* operator [7, 12] and the *Ordinal OWA* operator [22]. To deal with weighted linguistic information we present two extensions of the LOWA operator: the *Linguistic Weighted Averaging* (LWA) [6] and the *Linguistic Weighted OWA* (L-WOWA) [17] operators. We complete our study presenting the properties and axioms that these operators verify.

To do so, the paper is structured as follows: Section 2 presents the label set used to provide the opinions; Section 3 shows the LOWA and ordinal OWA operators; Section 4 analyzes two extensions of the LOWA operator to combine weighted linguistic information; and finally, some concluding remarks are pointed out.

2 Preliminaries

We use label sets with an odd cardinal, representing the middle term an assessment of "approximately 0.5", with the rest of the terms being placed symmetrically around it and the limit of granularity 11 or no more than 13. The semantic of the elements in the label set is given by fuzzy numbers defined on the [0,1] interval, which are described by membership functions. Because the linguistic assessments are just approximate ones given by the experts, we can consider that linear trapezoidal membership functions are good enough to capture the vagueness of those linguistic assessments, since it may be impossible or unnecessary to obtain more accurate values. This representation is achieved by the 4-tuple, $(a_i, b_i, \alpha_i, \beta_i)$, the first two paremeters indicate the interval in which the membership value is 1; the third and fourth parameters indicate the left and right widths. Moreover, the term set, $S = \{s_0, \ldots, s_T\}$, must have the following characteristics:

1) The set is ordered: $s_i \geq s_j$ if $i \geq$ j.

2) There is the negation operator: Neg(s_i) = s_j such that $j = T$-i.
3) Maximization operator: Max(s_i, s_j) = s_i if $s_i \geq s_j$.
4) Minimization operator: Min(s_i, s_j) = s_i if $s_i \leq s_j$.

Example 1. For example, this is the case of the following term set [1]:

VH	$Very\_High$	$(1, 1, 0, 0)$
H	$High$	$(.98, .99, .05, .01)$
MH	$Moreorless\_High$	$(.78, .92, .06, .05)$
$FFMH$	$From\_Fair\_to\_Moreorless\_High$	$(.63, .80, .05, .06)$
F	$Fair$	$(.41, .58, .09, .07)$
$FFML$	$From\_Fair\_to\_Moreorless\_Low$	$(.22, .36, .05, .06)$
ML	$Moreorless\_Low$	$(.1, .18, .06, .05)$
L	Low	$(.01, .02, .01, .05)$
VL	$Very\_Low$	$(0, 0, 0, 0)$

In the following we shall use this set of nine labels in all examples.

3 Linguistic Versions of OWA Operators

Here, we present the two linguistic versions of the OWA operator and their properties.

3.1 The LOWA Operator

The *LOWA operator* [7, 12] is based on the *OWA operator* defined by Yager [21] and the *convex combination of linguistic labels* defined by Delgado et al. [5].

Definition 1. *Let $A = \{a_1, \ldots, a_m\}$ be a set of labels to be aggregated, then the LOWA operator, ϕ, is defined as*

$$\phi(a_1, \ldots, a_m) = W \cdot B^T = \mathcal{C}^m\{w_k, b_k, k = 1, \ldots, m\} =$$

$$= w_1 \odot b_1 \oplus (1 - w_1) \odot \mathcal{C}^{m-1}\{\beta_h, b_h, h = 2, \ldots, m\}$$

where $W = [w_1, \ldots, w_m]$, is a weighting vector, such that, (i) $w_i \in [0, 1]$ and, (ii) $\Sigma_i w_i = 1$,. $\beta_h = w_h / \Sigma_2^m w_k, h = 2, \ldots, m$, and $B = \{b_1, \ldots, b_m\}$ is a vector associated to A, such that,

$$B = \sigma(A) = \{a_{\sigma(1)}, \ldots, a_{\sigma(n)}\}$$

in which,$a_{\sigma(j)} \leq a_{\sigma(i)}$ $\forall\ i \leq j$, with σ being a permutation over the set of labels A. $\mathcal{C}^m$ is the convex combination operator of m labels and if m=2, then it is defined as

$$\mathcal{C}^2\{w_i, b_i, i = 1, 2\} = w_1 \odot s_j \oplus (1 - w_1) \odot s_i = s_k, \quad s_j,\ s_i\ \in\ S,\ (j \geq i)$$

such that, $k = min\{T, i + round(w_1 \cdot (j - i))\}$, where "round" is the usual round operation, and $b_1 = s_j,\ b_2 = s_i$.

If $w_j = 1$ and $w_i = 0$ with $i \neq j$ $\forall i$, then the convex combination is defined as:

$$\mathcal{C}^m\{w_i, b_i, i = 1, \ldots, m\} = b_j.$$

Next, we present an extension of the LOWA operator, and Inverse LOWA operator, that will be used in the definition of some weighted operators [6].
Definition 2. *An Inverse-Linguistic Ordered Weighted Averaging (I-LOWA) operator, ϕ^I, is a type of LOWA operator, in which*

$$B = \sigma^I(A) = \{a_{\sigma(1)}, \ldots, a_{\sigma(n)}\}$$

where,

$$a_{\sigma(i)} \leq a_{\sigma(j)} \ \forall\, i \leq j.$$

If m=2, then it is defined as

$$\mathcal{C}^2\{w_i, b_i, i = 1, 2\} = w_1 \odot s_j \oplus (1 - w_1) \odot s_i = s_k, \quad s_j,\ s_i \ \in \ S, \ (j \leq i)$$

such that

$$k \ = \ min\{T, i \ + \ round(w_1 \cdot (j - i))\}.$$

Example 2. Suppose that we want to aggregate by means of the LOWA operator the following four labels, $\{L, ML, H, VH\}$. Assuming the weighting vector $W = [0.3, 0.2, 0.4, 0.1]$ the general expression of the aggregation of labels is:

$$\begin{aligned} \phi(L, ML, H, VH) &= [0.3, 0.2, 0.4, 0.1](VH, H, ML, L) \\ &= \mathcal{C}^4\{(0.3, VH), (0.2, H), (0.4, ML), (0.1, L)\}. \end{aligned}$$

Then, we obtain the final result applying the recursive definition of the convex combination, $\mathcal{C}^4$, as follows. Firstly, we develop $\mathcal{C}^4$ until its most simple expression in the following steps:

1. For $m = 4$,
$$\begin{aligned} &\mathcal{C}^4\{(0.3, VH)(0.2, H)(0.4, ML)(0.1, L)\} \\ &= 0.3 \odot VH \oplus \mathcal{C}^3\{(0.29, H), (0.57, ML), (0.14, L)\}. \end{aligned}$$

2. For $m = 3$,
$$\mathcal{C}^3\{(0.29, H), (0.57, ML), (0.14, L)\} = 0.29 \odot H \oplus \mathcal{C}^2\{(0.57, ML), (0.14, L)\}.$$

Now, we are going to go back solving the most simple cases until to obtain the final result:

1. For $m = 2$,
$$\mathcal{C}^2\{(0.57, ML), (0.14, L)\} == 0.57 \odot ML \oplus (1 - 0.57) \odot L = ML \ (s_2),$$
since as $ML = s_2$ and $L = s_1$ then
$$\begin{aligned} &min\{9, 1 + round(0.57 \cdot (2 - 1))\} = \\ &= \ min\{8, 1 + round(0.57)\} = G22min\{8, 2\} = 2. \end{aligned}$$

2. For $m = 3$,

$$\begin{aligned} & \mathcal{C}^3\{(0.29, H), (0.57, ML), (0.14, L)\} = \\ & = \quad 0.29 \odot H \oplus \mathcal{C}^2\{(0.57, ML), (0.14, L)\} = FFML \ (s_3) \end{aligned}$$

since as $H = s_7$ and $ML = s_2$ then

$$min\{8, 2+round(0.29\cdot(7-2))\} = min\{8, 2+round(1.45)\} = min\{8, 3\} = 3.$$

3. Finally, we obtain the final result for $m = 4$,

$$\mathcal{C}^4\{(0.3, VH)(0.2, H)(0.4, ML)(0.1, L)\} =$$

$$0.3 \odot VH \oplus \mathcal{C}^3\{(0.29, H), (0.57, ML), (0.14, L)\} = FFMH \ (s_5),$$

since as $VH = s_8$ and $FFML = s_3$ then

$$min\{8, 3+round(0.3\cdot(8-3))\} = min\{8, 3+round(1.5)\} = min\{8, 5\} = 5.$$

How to calculate the weighting vector of LOWA operator, W, is a basic question to be solved. A possible solution is that the weights represent the concept of fuzzy majority in the aggregation of LOWA operator using fuzzy linguistic quantifiers [29]. Yager proposed an interesting way to compute the weights of the OWA aggregation operator, which, in the case of a non-decreasing proportional fuzzy linguistic quantifier, Q, is given by this expression [21, 23]:

$$w_i = Q(i/n) - Q((i-1)/n), i = 1, \ldots, n,$$

being the membership function of Q, as follows:

$$Q(r) = \begin{cases} 0 & \text{if } r < a \\ \frac{r-a}{b-a} & \text{if } a \leq r \leq b \\ 1 & \text{if } r > b \end{cases}$$

with $a, b, r \in [0, 1]$. Some examples of non-decreasing proportional fuzzy linguistic quantifiers are: "most" $(0.3, 0.8)$, "at least half" $(0, 0.5)$ and "as many as possible" $(0.5, 1)$. When a fuzzy linguistic quantifier, Q, is used to compute the weights of LOWA operator, ϕ, it is symbolized by ϕ_Q. Similarly happens for the I-LOWA operator, i.e., in this case it is symbolized by ϕ^I_Q.

Example 3. Suppose that we want to aggregate by means of the LOWA operator the same four above labels, $\{L, ML, H, VH\}$. If we use the fuzzy linguistic quantifier "as many as possible" to calculate the weighting vector, i.e., $W = \{0, 0, 0.5, 0.5\}$ and the aggregation of the labels, working in the similar way, is the following:

$$\phi_Q(L, ML, H, VH) = [0, 0, 0.5, 0.5](VH, H, ML, L)$$

$$= \mathcal{C}^4\{(0, VH), (0, H), (0.5, ML), (0.5, L)\} = ML.$$

The LOWA operator is a rational aggregation operator because it verifies some properties and axioms of any acceptable aggregation operator [9].

3.1.1 Properties of LOWA Operator

Property 1.- The LOWA operator is **increasing monotonous** respect to the argument values, in the following sense: let $A = [a_1, a_2, \ldots, a_n]$ be an ordered argument vector, let $B = [b_1, b_2, \ldots, b_n]$ be a second ordered argument vector, such that $\forall\ j,\ a_j \geq bj$ then $\phi(A)\ \geq\ \phi(B)$.

Property 2.- The LOWA operator is **commutative**, i.e.,

$$\phi(a_1, a_2, \ldots, a_n) \;=\; \phi(\pi(a_1), \pi(a_2), \ldots, \pi(a_n)),$$

where π is a permutation over the set of arguments.
Property 3.- The LOWA operator is an **"orand" operator**. That is, for any weighting vector W and ordered labels vector A=$[a_1, a_2, \ldots, a_n]$, then

$$\text{Min}(A)\ \leq\ \phi(A)\ \leq\ \text{Max}(A).$$

3.1.2 Axiomatic of the LOWA Operator

In what follows we are going to study some of the proposed axioms in fuzzy setting considering the LOWA operator working with linguistically valued preferences. Before this, we include the following linguistic notation that we shall use.

Let $A = \{x_1, \ldots, x_n\}$ be a finite non-empty set of alternatives.
Let $E = \{e_1, \ldots, e_m\}$ be a panel of experts.
Let $S = \{s_i,\ i = 0..T\}$ be a label set to voice experts' opinions.
Let $x_{ij} \in S$ be the linguistic rating of alternative x_i by expert e_j.
Let F_j be the linguistic rating set over alternatives by expert e_j.
Let μ_{F_j} be the linguistic membership function of F_j such that $x_{ij} = \mu_{F_j}(x_i)$.
Let F be the linguistic rating set such that $F = \phi(F_1, \ldots, F_n)$.
The LOWA operator satisfies these axioms:

Axiom I: *Unrestricted domain.* For any set of individual preference patterns $\{F_j,\ j = 1, \ldots, m\}$ there is a social preference pattern F, which may be constructed,

$$\forall F_1, \ldots, F_m \in S^n, \exists F \in S^n \text{ such that } F = \phi(F_1, \ldots, F_m).$$

Axiom II: *Unanimity or Idempotence.* If everyone agrees on a preference pattern, it must be 3sen as the social 3ice pattern,i.e.,

$$F_j = F, \forall j \Rightarrow F = \phi(F, F, \ldots, F).$$

Axiom III: *Positive association of social and individual values.* If an individual increases his linguistic preference intensity for x_i then the social linguistic preference for x_i cannot decrease. This means that if F_j' and F_j are such that $\mu_{F_j} \leq \mu_{F_j'}$, then if $\phi(F_1, \ldots, F_j, \ldots, F_m) = F$ and $\phi(F_1, \ldots, F_j', \ldots, F_m) = F'$, then $\mu_F \leq \mu_{F'}$.

Axiom IV: *Independence of irrelevant alternatives.* The social preference intensity for x_i only depends on the individual preference intensity for x_i, and not for x_k, $k \neq i$, $\mu_{\phi(F_1,\ldots,F_m)}(x_i) = \varphi(x_{i1}, \ldots, x_{im})$. Clearly this axiom does not extend strictly speaking, since for preference relations the independence of irrelevant alternatives deals with pairs of alternatives.

Axiom V: *Citizen sovereignty.* It means that any social preference pattern can be expressed by the society of individuals; in other words

$$\forall F, \exists F_1, \ldots, F_m \text{ such that } F = \phi(F_1, \ldots, F_m).$$

Axiom VI: *Neutrality.* The neutrality axiom refers to the invariance properties of the voting procedure. There are several types:

1. *Neutrality with respect to alternatives.* If x_i and x_k are such that $x_{ij} = x_{kj}$, $\forall j$, then $\mu_{\phi(F_1,\ldots,F_m)}(x_i) = \mu_{\phi(F_1,\ldots,F_m)}(x_k)$
2. *Neutrality with respect to voters.* In a homogeneous group, this is the anonymity property, i.e., the commutativity of ϕ.

Some applications of the LOWA operator guided by fuzzy majority in group decision making problems can be found in [8, 10, 11, 12].

3.2 The Ordinal OWA Operator

The ordinal OWA operator was introduced in [22].
Definition 3. *Let $A = \{a_1, \ldots, a_m\}$ $(a_i \in S)$ be a set of labels to be aggregated, then, the ordinal OWA operator, ω, is defined as*

$$\omega(a_1, \ldots, a_m) = Max_j[w_j \min b_j],$$

where $W = [w_1, \ldots, w_m]$, is a weighting vector, such that, (i) $w_i \in S$, (ii) $w_j \geq w_i$ if $j > i$, and (iii) $Max_j[w_j] = s_T$; and B according to definition 1.

Clearly, we may see that two important differences between this operator and the LOWA operator are: a) the weights in this operator have linguistic nature and in the LOWA operator numerical one, and b) the different requeriments on the weighting vectors.

3.2.1 Properties of the Ordinal OWA

The ordinal OWA operator can be shown to have the following properties [22]:
Property 1.- The ordinal OWA operator is **increasing monotonous** respect to the argument values, in the following sense: let $A = [a_1, a_2, \ldots, a_n]$ be an argument vector, let $B = [b_1, b_2, \ldots, b_n]$ be a second argument vector, such that $\forall\ j,\ a_j \geq b_j$ then $\omega(A) \geq \omega(B)$.
Property 2.- The ordinal OWA operator is **commutative**, i.e.,

$$\omega(a_1, a_2, \ldots, a_n) = \omega(\pi(a_1), \pi(a_2), \ldots, \pi(a_n)).$$

Property 3.- The ordinal OWA operator is **an "orand" operator.** That is, for any weighting vector W and ordered labels vector A=$[a_1, a_2, \ldots, a_n]$, then

$$\mathrm{Min}(A) \leq \omega(A) \leq \mathrm{Max}(A).$$

Property 4.- The ordinal OWA operator present the idempotency property. That is, if we have an ordered vector of labels, A=$[a_1, a_2, \ldots, a_n]$, such that $a_i = a \; \forall i$, then for any weighting vector, W, $\omega(A) = a$.

Some applications of ordinal OWA operator in decision making problems can be found in [26].

4 Extensions of LOWA Operator

In the OWA operators the weights measure the importance of a value (in relation to other values) with independence of the information source. In other operators the weights measure the importance of an information source with independence of the value. Different aggregation operators of linguistic information that combines the advantages of both operators types have been proposed in [6, 17]. These operators are based on the LOWA operator and are designed to aggregate weighted linguistic information. In the first, called Linguistic Weighted Averaging (LWA) operator [6], the weights of information sources have linguistic nature, and in the second, called the Linguistic Weighted OWA (L-WOWA) [17], they have numerical nature. In what follows we shall analyze them.

4.1 The LWA Operator

Let $\{(c_1, a_1), \ldots, (c_m, a_m)\}$ be a set of weighted opinions such that a_i shows the opinion of an expert e_i, assessed linguistically on the label set, S, $a_i \in S$, and c_i the relevance degree of expert e_i, assessed linguistically on the same label set S, $c_i \in S$.

Following Cholewa's studies [3] and Montero's aggregation model [15], if we want to aggregate weighted information we have to define two aggregations:

- the aggregation of importance degrees (weights) of information, and
- the aggregation of weighted information (to combine information with weights).

The first aspect consists of obtaining a collective importance degree from individual importance degrees that characterizes the final result of aggregation operator. In the LWA operator, as the importance degrees are linguistic values, this is solved using the LOWA operator guided by the concept of fuzzy majority.

The aggregation of weighted information involves the transformation of the weighted information under the importance degrees. The transformation form depends upon the type of aggregation of weighted information being performed

[25]. In [19, 20] Yager discussed the effect of the importance degrees in the types of aggregation "MAX" and "MIN" and suggested a class of functions for importance transformation in both types of aggregation. For MIN type aggregation he suggested a family of t-conorms acting on the weighted information and the negation of the weights, which presents the non-increasing monotonic property in the weights. For MAX type aggregation he suggested a family of t-norms acting on weighted information and the weight, which presents the non-decreasing monotonic property in the weights. In [25] Yager proposed a general specification of the requirements that any *importance transformation function*, g, must satisfy for any type of the aggregation operator. The function, g, must have the following properties:

1. if $a > b$ then $g(w, a) \geq g(w, b)$
2. $g(w, a)$ is monotone in w
3. $g(0, a) = \text{ID}$
4. $g(1, a) = a$.

with $a,\ b \in [0, 1]$ expressing the satisfaction with regards to a criterion, $w \in [0, 1]$ the weight associated to the criterion, and "ID" an identity element, which is such that if we add it to our aggregations it doesn't change the aggregated value. Condition one means that the function g is monotonically non-decreasing in the second argument, that is, if the satisfaction with regards to the criteria is increased the overall satisfaction shouldn't decrease. The second condition may be viewed as a requirement that the effect of the importance be consistent. It doesn't specify whether g is monotonically non-increasing or non-decreasing in the first argument, but must be one of these. It should be denoted that conditions three and four actually determine the type of monotonicity obtained from two. If $a > \text{ID}$, the $g(w, a)$ is monotonically non-decreasing in w, while if $a < \text{ID}$, then it is monotonically non-increasing. The third condition is a manifestation of the imperative that zero importance items don't effect the aggregation process. The final condition is essentially a boundary condition which states that the assumption of all importances equal to one effectively is like not including importances at all [25].

Considering the aforementioned ideas and assuming a linguistic framework, that is a label set, S, to express the information and a label set, L, to express the weights, in [6] we proposed the LWA operator, with its respective aggregation operators and transformation functions:

1. Aggregation operator: $LOWA$ or $I - LOWA$.
2. Transformation function: $g_{(\text{LOWA})} = LC^{\rightarrow}(w, a)$ or $g_{(\text{I-LOWA})} = LI^{\rightarrow}(w, a)$.

Itis based on the combination of the LOWA and I-LOWA operator with several *linguistic conjunction functions* ($LC^{\rightarrow}$) and several *linguistic implication functions* ($LI^{\rightarrow}$), respectively. Therefore, the LWA operator is a type of fuzzy majority guided weighted aggregation operator.

Before defining the LWA operator, we let us present the following two families of connectives:

1. Linguistic conjunction functions ($LC^{\rightarrow}$).

   These linguistic conjunction functions, presented in [6], are monotonically non-decreasing t-norms in the weights:

   (a) *The classical MIN operator:*

   $$LC_1^{\rightarrow}(c,a) = MIN(c,a).$$

   (b) *The nilpotent MIN operator:*

   $$LC_2^{\rightarrow}(c,a) = \begin{cases} MIN(c,a) & \text{if } c > Neg(a) \\ s_0 & \text{otherwise.} \end{cases}$$

   (c) *The weakest conjunction:*

   $$LC_3^{\rightarrow}(c,a) = \begin{cases} MIN(c,a) & \text{if } MAX(c,a) = s_T \\ s_0 & \text{otherwise.} \end{cases}$$

2. Linguistic implication functions ($LI^{\rightarrow}$).

   These linguistic implication functions are monotonically non-increasing in the weights:

   (a) *Kleene-Dienes's implication function:*

   $$LI_1^{\rightarrow}(c,a) = MAX(Neg(c),a).$$

   (b) *Gödel's implication function:*

   $$LI_2^{\rightarrow}(c,a) = \begin{cases} s_T & \text{if } c \leq a \\ a & \text{otherwise.} \end{cases}$$

   (c) *Fodor's implication function:*

   $$LI_3^{\rightarrow}(c,a) = \begin{cases} s_T & \text{if } c \leq a \\ MAX(Neg(c),a) & \text{otherwise.} \end{cases}$$

Definition 4. *The aggregation of the set of weighted individual opinions, $\{(c_1,a_1),\ldots,(c_m,a_m)\}$, according to the LWA operator, Ω, is defined as a weighted collective opinion, (c_E,a_E), such that*

$$(c_E,a_E) = \Omega[(c_1,a_1),\ldots,(c_m,a_m)],$$

where the importance degree of the group opinion, c_E, is obtained as

$$c_E = \phi_Q(c_1,\ldots,c_m),$$

and, the opnion of the group, a_E, is obtained as

$$a_E = f[g(c_1, a_1), \ldots, g(c_m, a_m)],$$

where $f \in \{\phi_Q, \phi_Q^I\}$ is an linguistic aggregation operator of transformated information and g is a importance transformation function, such that, $g \in \{LC_1^{\rightarrow}, LC_2^{\rightarrow}, LC_3^{\rightarrow}\}$ if $f = \phi_Q$, and $g \in \{LI_1^{\rightarrow}, LI_2^{\rightarrow}, LI_3^{\rightarrow}\}$ if $f = \phi_Q^I$.

As it was commented in [6], when the aggregation operator, f, is the I-LOWA operator, ϕ_Q^I, and given that ϕ_Q^I is an aggregation operator with characteristics of a MIN type aggregation operator, then we have decided to use the linguistic implications functions, $LI^{\rightarrow}$, as the transformation function type. Something similar happens when f is the LOWA operator ϕ_Q.

In [6] we presented some evidence of its rational aggregation way. It aggregation has been checked examining some of the axioms that an acceptable weighted aggregation operator must verify.

Some applications of the LWA operator in multi-criteria and group decision making can be found in [6, 13].

4.2 The L-WOWA Operator

Let $\{(c_1, a_1), \ldots, (c_m, a_m)\}$ be a set of weighted opinions such that a_i shows the opinion of an expert e_i, assessed linguistically on the label set, S, $a_i \in S$, and c_i the relevance degree of expert e_i, such that, (i) $c_i \in [0, 1]$ and (ii) $\sum_i c_i = 1$.

Definition 5. *The aggregation of the set of weighted individual opinions, $\{(c_1, a_1), \ldots, (c_m, a_m)\}$, according to the L-WOWA operator is defined as*

$$\varphi[(c_1, a_1), \ldots, (c_m, a_m)] = [\lambda_1, \ldots, \lambda_m]B^T = \mathcal{C}^m\{\lambda_i, b_i,\ i = 1, \ldots, m\} =$$

$$= \lambda_1 \odot b_1 \oplus (1 - \lambda_1) \odot \mathcal{C}^{m-1}\{\beta_h, b_h, h = 2, \ldots, m\},$$

B according to definition 1 and the weights λ_i obtained as

$$\lambda_i = w^*(\sum_{i \geq j} c_{\sigma(j)}) - w^*(\sum_{j<i} c_{\sigma(j)})$$

with w^* being a monotonic increasing function that interpolates the points $(i/m, \sum_{i \geq j} c_j)$ together with the point (0,0).

Since the difference between the LOWA and L-WOWA operator is the calculation of the weights, the L-WOWA operator satisfies the same properties and axioms of the LOWA operator.

Some applications of the L-WOWA operator are shown in [18]

5 Concluding Remarks

In this paper, we have presented a summary of linguistic OWA operators provided in the literature. We have shown that there are OWA operators to deal with non-weighted and weighted linguistic information. Specifically we have presented the LOWA and ordinal OWA operators for management non-weighted linguistic information and the LWA and L-WOWA operators for weighted one.

References

P.P. Bonissone and K.S. Decker, Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-off Precision and Complexity, in: L.H. Kanal and J.F. Lemmer, Eds., *Uncertainty in Artificial Intelligence* (North-Holland, 1986) 217-247.

G. Bordogna and G. Pasi, A Fuzzy Linguistic Approach Generalizing Boolean Information Retrieval; a Model and Its Evaluation, *Journal of the American Society for Information Science* **44** (1993) 70-82.

W. Cholewa, "Aggregation of Fuzzy Opinions: An Axiomatic Approach," *Fuzzy Sets and Systems* **17** (1985) 249-259.

R. Degani and G. Bortolan, The Problem of Linguistic Approximation in Clinical Decision Making, *Inter. J. of Approximate Reasoning* **2** (1988) 143-162.

M. Delgado, J.L. Verdegay and M.A. Vila, On Aggregation Operations of Linguistic Labels, *Int. Journal of Intelligent Systems* **8** (1993) 351-370.

F. Herrera and E. Herrera-Viedma, Aggregation Operators for Linguistic Weighted Information, *IEEE Transactions on Systems, Man, and Cybernetics* (1996). To appear.

F. Herrera and J.L. Verdegay, Linguistic Assessments in group decision, *Proc. 1th European Congress on Fuzzy and Intelligent Technologies*, Aachen, 1993, 941-948.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, A Sequential Selection Process in Group Decision Making with Linguistic Assessment, *Information Sciences* **85** (1995) 223-239.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, Aggregating Linguistic Preferences: Properties of LOWA Operator, *Proc. 5th IFSA World Congress*, Sao Paulo, 1995, 153-156.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, A Linguistic Decision Process in Group Decision Making, *Group Decision and Negotiation* **5** (1996) 165-176.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, A Model of Consensus in Group Decision Making under Linguistic Assessments, *Fuzzy Sets and Systems* **78** (1996) 73-87.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, Direct Approach Processes in Group Decision Making Using Linguistic OWA Operators, *Fuzzy Sets and Systems* **79** (1996) 175-190.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, Choice Processes for Non-Homogeneous Group Decision Making in Linguistic Setting, *Fuzzy Sets and Systems* (1996). To appear.

E.H. Mamdani and S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, *Inter. Journal of Man Machine Studies* **7** (1975) 1-13.

J. Montero, Aggregation of Fuzzy Opinions in a Non-Homogeneous Group," *Fuzzy Sets and Systems* **25** (1988) 15-20.

M. Tong and P. P. Bonissone, A Linguistic Approach to Decision Making with Fuzzy Sets, *IEEE Transactions on Systems, Man and Cybernetics* **11** (1980) 716-723.

V. Torra, The Weighted OWA Operator, *Int. Journal of Intelligent Systems* (1996). To appear.

V. Torra, Weighted OWA Operators for Synthesis of Information, Proc. 5th IEEE Inter. Conference on Fuzzy Systems, New Orleans, USA, 1996, 966-971.

R.R. Yager, A New Methodology for Ordinal Multiple Aspect Decision Based on Fuzzy Sets, *Decision Sciences* **12** (1981) 589-600.

R.R. Yager, A Note on Weighted Queries in Information Retrieval Systems, *J. of the American Society of Information Sciences* **38** (1987) 23-24.

R.R. Yager, On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making, *IEEE Transactions on Systems, Man, and Cybernetics* **18** (1988) 183-190.

R.R. Yager, Applications and Extension of OWA Aggregation, *Int. Journal of Man-Machine Studies* **37** (1992) 103-132.

R.R. Yager, Families of OWA Operators, *Fuzzy Sets and Systems* **59** (1993) 125-148.

R.R. Yager, Fuzzy Screening Systems, in *Fuzzy Logic: State of the Art*, R. Lowen, Ed. Kluwer Academic Publishers, 1993, pp. 251-261.

R.R. Yager, On Weighted Median Aggregation, *Int. J. Uncertainty, Fuzzines and Knowledge-Based Systems* **2** (1994) 101-113.

R.R. Yager, An Approach to Ordinal Decision Making, *Int. Journal of Approximate Reasoning* **12** (1995) 237-261.

R.R. Yager, Fusion of Ordinal Information Using Weighted Median Aggregation, Machine Intelligence Institute, Technical Report #MII-1520.

L. A. Zadeh, The Concept of a Linguistic Variable and its Applications to Approximate Reasoning. Part I, *Information Sciences* **8** (1975) 199-249, Part II, *Information Sciences* **8** (1975) 301-357, Part III, *Information Sciences* **9** (1975) 43-80.

L. A. Zadeh, A Computational Approach to Fuzzy Quantifiers in Natural Languages, *Computers and Mathematics with Applications* **9** (1983) 149-184.

Alternative Representations of OWA Operators

Michel GRABISCH
Thomson-CSF, Central Research Laboratory
Domaine de Corbeville, 91404 Orsay Cedex, France
email grabisch@thomson-lcr.fr

Abstract

It is known that OWA operators appear to be particular cases of Choquet integral with respect to a suitable fuzzy measure. Recently, it has been shown that fuzzy measures can be expressed in three different, completely equivalent forms (called representations), among which the so-called interaction representation. It has been shown that only the interaction representation makes sense for a decision maker in multicriteria or multiattribute or multiperson problems, with close links to the Shapley value in cooperative game theory. In this paper we will give the three representations of an OWA operator, with emphasis on the interaction representation, in order to give new insights into its meaning in aggregation. We will also give best linear and 2nd order approximations of OWA operators.

Keywords: Choquet integral, Möbius transform, Shapley value, interaction, linear approximation

1 Introduction

It is a well known fact (see e.g. [15]) that OWA operators are a particular case of discrete Choquet integrals with respect to a fuzzy measure depending only on the cardinal of subsets. In fact, viewing Choquet integrals as aggregation operators, the class of OWA operators coincides with the class of commutative Choquet integrals [4]. As Choquet integrals have been thoroughly studied in the context of multicriteria decision problems [4, 6, 8, 14], the OWA operator can benefit from these studies.

A recent issue in fuzzy measures and Choquet integral is the introduction of alternative ways of representing fuzzy measures [7, 5, 3], by means of the Möbius transform and the notion of interaction. This has shed light on the interpretation of Choquet integral, and on its behavioral properties in a context of multicriteria decision making. In this paper, we will examine OWA operators under the point of view of these different representations, and get a new intepretation of these operators. In the last section, the issue concerning the best first and second order approximation of OWA will be addressed.

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

2 Fuzzy measures and their representation

This section will give the necessary mathematical background for the analysis of OWA operators. As operators deal with a finite number of arguments, all the following presentation will be done in a finite setting, thus avoiding measure theoretical intricacies. For a more complete treatment of fuzzy measures and Choquet integral, the reader is referred to the monographs of Denneberg [2] and Grabisch *et al.* [10].

2.1 Fuzzy measure and the Choquet integral

Let $X = \{x_1, \ldots, x_n\}$ be a finite set, which could be criteria in a multicriteria decision problem, persons in a voting problem, players in a game, etc., and let us denote $\mathcal{P}(X)$ the power set of X, i.e. the set of all subsets of X.

Definition 1 *A* fuzzy measure *on X is a set function $\mu : \mathcal{P}(X) \longrightarrow [0,1]$, satisfying the following axioms.*

(i) $\mu(\emptyset) = 0$, $\mu(X) = 1$.

(ii) $A \subset B$ *implies* $\mu(A) \leq \mu(B)$, *for* $A, B \in \mathcal{P}(X)$.

$\mu(A)$ can be viewed as the weight of importance of the set of elements A. The *dual fuzzy measure* of μ is defined by $\mu^*(A) := 1 - \mu(A^c)$.

An *additive measure* is such that $\mu(A \cup B) = \mu(A) + \mu(B)$, for every pair of disjoint $A, B \subset X$. If the fuzzy measure has no particular additional property, one needs to define the $2^n - 2$ coefficients corresponding to the 2^n subsets of X, except $\emptyset$ and X itself.

We introduce now the concept of discrete Choquet integral, viewed as an aggregation operator. For this reason, we will adopt a connective-like notation instead of the usual integral form, and the integrand will be a set of n values $a_1, \ldots, a_n$ in $[0,1]$.

Definition 2 *Let μ be a fuzzy measure on X. The* discrete Choquet integral *of $a_1, \ldots, a_n$ with respect to μ is defined by*

$$\mathcal{C}_\mu(a_1, \ldots, a_n) := \sum_{i=1}^{n} (a_{(i)} - a_{(i-1)})\mu(A_{(i)}),$$

with $A_{(i)} := \{x_{(i)}, \ldots, x_{(n)}\}$, and $a_{(0)} = 0$. Parentheses around the indices denote a permutation such that $a_{(1)} \leq a_{(2)} \leq \cdots \leq a_{(n)}$.

2.2 Shapley value and the interaction index of a pair of elements

From now on, the elements of X are simply denoted $1, \ldots, n$. In the context of cooperative game theory, Shapley has introduced the following index, representing the importance of a player in the game, based on axiomatic considerations.

Definition 3 *(Shapley [18]) Let μ be a fuzzy measure on X. The* Shapley index *for every $i \in X$ is defined by*

$$v_i := \sum_{k=0}^{n-1} \gamma_k \sum_{K \subset X \setminus i, |K|=k} (\mu(K \cup \{i\}) - \mu(K)), \tag{1}$$

with $\gamma_k := \frac{(n-k-1)!k!}{n!}$, and $|K|$ being the cardinal of K. The Shapley value *of μ is the vector $v(\mu) := [v_1 \cdots v_n]^t$.*

The Shapley index v_i can be interpreted as a kind of weighted average value of the contribution of element i alone in all coalitions. It satisfies $\sum_{i=1}^n v_i = 1$, so that the Shapley value is a kind of sharing out of the total value of the game μ to the individual players. Murofushi proposed in [12] its use in the context of multicriteria decision making.

Murofushi and Soneda, relying on concepts of multiattribute utility theory, have defined an interaction index between two elements.

Definition 4 *(Murofushi and Soneda [13]) Let μ be a fuzzy measure on X. The* interaction index *of elements i, j is defined by*

$$I_{ij} := \sum_{k=0}^{n-2} \xi_k \sum_{K \subset X \setminus \{i,j\}, |K|=k} (\mu(K \cup \{i,j\}) - \mu(K \cup \{i\}) - \mu(K \cup \{j\}) + \mu(K)), \tag{2}$$

with $\xi_k := \frac{(n-k-2)!k!}{(n-1)!}$.

The interaction index I_{ij} can be interpreted as a kind of average value of the *added value* given by putting i and j together, all coalitions being considered. The following cases can happen:

- $I_{ij} > 0$: i and j are said to be *complementary*, or to have a *positive synergy*, since put together they become very important.
- $I_{ij} < 0$: i and j are said to be *substitutive*, or to have a *negative synergy* since adding one or the either or both to a coalition is not very significant.
- $I_{ij} = 0$: i and j have no interaction, they are said to be *independent.*

We will come back to this interpretation later.

Relying on the analogy between (1) and (2), let us define interaction indexes for more than two elements.

Definition 5 *(Grabisch [7]) Let μ be a fuzzy measure on X. The* interaction index *of elements among $A \subset X$, $A \neq \emptyset$, is defined by:*

$$I(A) := \sum_{k=0}^{n-|A|} \Xi_k^{|A|} \sum_{K \subset X \setminus A, |K|=k} \sum_{L \subset A} (-1)^{|A|-|L|} \mu(L \cup K), \tag{3}$$

with $\Xi_k^p := \frac{(n-k-p)!k!}{(n-p+1)!}$.

Clearly, $v_i = I(\{i\}), i = 1, \ldots, n$, and we recover (2) substituting A by $\{i, j\}$ in (3).

Since (3) is defined for every $A \subset X$, $A \neq \emptyset$, putting $I_\emptyset = 0$, we have defined a set function on X, constructed from μ. In fact, as this will be explained in the next section, I is another representation of a fuzzy measure, very useful for the decision maker. This is made clear in the next section.

2.3 Three Representations of a Fuzzy Measure

Including the usual measure form of a fuzzy measure, there are three possible representations of a fuzzy measure :

the (usual) measure representation: the set function $\mu : \mathcal{P}(X) \longrightarrow [0, 1]$, determined by $\mu(A)$, for all $A \subset X$. These are 2^n coefficients in $[0, 1]$, which can be put into a lattice isomorphic to the lattice of subsets ordered by inclusion.

the Möbius representation: the Möbius inversion formula is known in the evidence theory of Shafer [17], to compute the basic probability assignment from a belief function. This formula is in fact known in combinatorics since a long time (see e.g. Rota [16]), and can be applied to any set function. Let us denote m the set function from $\mathcal{P}(X)$ to $\mathbb{R}$ which is the transformation of μ by the Möbius transform. It is defined by:

$$m(A) := \sum_{B \subset A} (-1)^{|A \setminus B|} \mu(B), \tag{4}$$

for any $A \subset X$. This is again a set of 2^n coefficients, but in $\mathbb{R}$ and without any lattice structure.

the interaction representation: it is given by the formula (3), for any $\emptyset \neq A \subset X$, and $I(\emptyset) := 0$. This is also a set function, whose values range in $\mathbb{R}$, and without any lattice structure.

Of course, these are not the only possible alternative representations of a fuzzy measure, but these ones present a particular interest in a decision making framework.

We give below relations between these different representations.

- passage from m to μ (this is known as the zeta transform):

$$\mu(A) = \sum_{B \subset A} m(B), \forall A \subset X. \tag{5}$$

- passage from m to I:

$$I(A) = \sum_{B \subset X \setminus A, |B| \leq n - |A|} \frac{1}{|B| + 1} m(A \cup B), \quad \forall A \subset X, A \neq \emptyset. \tag{6}$$

- passage from I to m:

$$m(A) = \sum_{j=0}^{n-|A|} \alpha_j \sum_{B\subset X\setminus A, |B|=j} I(A\cup B), \quad \forall A \subset X, A \neq \emptyset, \tag{7}$$

where the coefficients α_j are obtained in a recursive way by the formula

$$\alpha_j := -\sum_{l=0}^{j-1} \frac{\alpha_l}{j-l+1}\binom{j}{l},$$

and $\alpha_0 := 1$. First values are $\alpha_1 = -1/2, \alpha_2 = 1/6, \alpha_3 = 0, \alpha_4 = -1/30$, etc.

- passage from I to μ:

$$\mu(A) = \sum_{B\subset X} \beta^{|B|}_{|B\cap A|} I(B), \tag{8}$$

with β^l_k defined by:

$$\beta^l_k = \sum_{i=0}^{k} \binom{k}{i} \alpha_{l-i}.$$

The set of equations (3), (4), (5), (6), (7) and (8) allows us to pass freely from one representation to another.

The reader is referred to [3] for full details on this part, including proofs of the above relations, and properties of the different representations. Concerning properties, we just mention the following.

Property 1 *Let μ, μ^* be a pair of dual measures, and denote by I and I^* their respective interaction representation. Then*

$$I^*(A) = (-1)^{|A|+1} I(A).$$

Property 2 *(Chateauneuf and Jaffray [1]) Let μ be a fuzzy measure, with m its Möbius representation, and $a_1, \ldots, a_n$ a set of positive real numbers. Then the expression of the Choquet integral in terms of m is given by:*

$$\mathcal{C}_\mu(a_1, \ldots, a_n) = \sum_{A\subset X} m(A) \inf_{i\in A} a_i.$$

There is yet no equivalent simple expression of the Choquet integral in terms of the interaction representation. However, an interesting formula has been obtained in a particular case, namely, the case of *2-additive measures*. First, we need the following definition.

Definition 6 *A fuzzy measure on X is said to be a k*-order additive fuzzy measure *or k*-additive measure *for short if $m(A) = 0$ for any A such that $|A| > k$, and there is at least one $A \subset X$ of exactly k elements for wich $m(A) \neq 0$.*

Due to (6), we can replace in the above definition m by I without any change.

Property 3 *Let μ be a 2-additive measure, and $a_1, \ldots, a_n$ a set of positive real numbers. Then the expression of the Choquet integral in terms of I is given by:*

$$\begin{aligned} \mathcal{C}_\mu(a_1, \ldots, a_n) &= \sum_{I(\{i,j\})>0} (a_i \wedge a_j) I(\{i,j\}) + \sum_{I(\{i,j\})<0} (a_i \vee a_j) |I(\{i,j\})| \\ &+ \sum_{i=1}^{n} a_i \left[I(\{i\}) - \frac{1}{2} \sum_{j \neq i} |I(\{i,j\})| \right], \end{aligned} \tag{9}$$

and $I(\{i\}) - \frac{1}{2} \sum_{j \neq i} |I(\{i,j\})| \geq 0$, $i = 1, \ldots, n$.

Remark that this expression is in terms of the Shapley indices v_i and the interaction indices I_{ij} of pairs of elements. Since the Choquet integral appears as a sum of three terms, this makes clear the intrinsic meaning of I_{ij} and v_i:

- a positive I_{ij} implies a conjunctive behavior between i and j. In multicriteria decision making, this means that the simultaneous satisfaction of criteria i and j is significant for the global score, but a unilateral satisfaction has no effect.
- a negative I_{ij} implies a disjunctive behavior, which means that the satisfaction of either i or j is sufficient to have a significant effect on the global score.
- the Shapley value acts as a weight vector in a weighted arithmetic mean. This represents the linear part of Choquet integral.

3 Representation of OWA in Möbius and in interaction form

We now address the main aim of the paper. We want to apply the preceding concepts and results to OWA, since OWA can be put under the form of a Choquet integral. In order to fix notations, we define OWA operators by the following formula:

$$\mathrm{OWA}_{w_1, \ldots, w_n}(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i a_{\sigma(i)}, \quad a_1, \ldots, a_n \in [0,1]^n, \tag{10}$$

where $w_1, \ldots, w_n$ are weights in $[0,1]$ so that $\sum_{i=1}^{n} w_i = 1$, and σ is a permutation on $\{1, \ldots, n\}$ so that $a_{\sigma(1)} \geq a_{\sigma(2)} \geq \cdots \geq a_{\sigma(n)}$.

We have the following property:

Property 4 *Any OWA operator with weights $w_1, \ldots, w_n$ is a Choquet integral, whose fuzzy measure μ is defined by*

$$\mu(A) = \sum_{j=1}^{i} w_j, \quad \forall A \text{ such that } |A| = i,$$

where $|A|$ denotes the cardinal of A. Reciprocally, any commutative Choquet integral is such that $\mu(A)$ depends only on $|A|$, and coincides with an OWA operator, whose weights are $w_i = \mu(A_i) - \mu(A_{i-1}), i = 2, \ldots, n$, and $w_1 = 1 - \sum_{i=2}^{n} w_i$. A_i denotes any subset such that $|A_i| = i$.

Let us now express the fuzzy measure associated with an OWA in the Möbius and interaction representation. The following can be shown.

Theorem 1 *Let $w_1, \ldots, w_n$ define an OWA operator, and let μ be the equivalent fuzzy measure. Then its Möbius and interaction representations are given by:*

$$m(A) = \sum_{i=1}^{|A|} (-1)^{|A|-i} w_i \binom{|A|-1}{i-1}, \tag{11}$$

$$I(A) = \frac{1}{n - |A| + 1} \sum_{j=0}^{|A|-2} (-1)^j \binom{|A|-2}{j} [w_{n-j} - w_{|A|-j-1}], \tag{12}$$

for all $A \subset X$. In particular, the Shapley index and the interaction index for a pair of elements are:

$$v_i = \frac{1}{n}, \quad i = 1, \ldots, n, \tag{13}$$

$$I_{ij} = \frac{w_n - w_1}{n-1}, \quad i, j = 1, \ldots, n, i \neq j. \tag{14}$$

Proof: we compute the Möbius representation. We obtain easily:

$$\mu(A) = \sum_{k=1}^{|A|} \binom{|A|}{k} (-1)^{|A|-k} \mu(B_k)$$

with B_k any subset in X containing k elements. We can equivalently express the above formula using $w_1, \ldots, w_n$. We obtain:

$$\begin{aligned} \mu(A) &= \sum_{k=1}^{|A|} \binom{|A|}{k} (-1)^{|A|-k} \sum_{i=1}^{k} w_i \\ &= \sum_{i=1}^{|A|} w_i \sum_{k=i}^{|A|} \binom{|A|}{k} (-1)^{|A|-k} \\ &= \sum_{i=1}^{|A|} (-1)^{|A|-i} w_i \binom{|A|-1}{i-1} \end{aligned}$$

using the fact that $\binom{n}{k} = \binom{n}{n-k}$ and

$$\sum_{k=0}^{m} (-1)^k \binom{n}{k} = (-1)^m \binom{n-1}{m}.$$

We compute now the interaction representation, beginning by the Shapley value. We have:

$$\begin{aligned}
v_i &= \sum_{A \subset \{i\}^c} \frac{(n-|A|-1)!|A|!}{n!}(\mu(A \cup \{i\}) - \mu(A)) \\
&= \sum_{k=0}^{n-1} \frac{(n-|A|-1)!|A|!}{n!}\binom{n-1}{k} w_{k+1} \\
&= \sum_{k=0}^{n-1} \frac{1}{n} w_{k+1} \\
&= \frac{1}{n}.
\end{aligned}$$

We compute the interaction index for two elements.

$$\begin{aligned}
I_{ij} &= \sum_{A \subset \{i,j\}^c} \frac{(n-|A|-2)!|A|!}{(n-1)!}(\mu(A \cup \{i,j\}) - \mu(A \cup \{i\}) - \mu(A \cup \{j\}) + \mu(A)) \\
&= \sum_{k=0}^{n-2} \frac{(n-|A|-2)!|A|!}{(n-1)!}\binom{n-2}{k}(w_{k+2} - w_{k+1}) \\
&= \sum_{k=0}^{n-2} \frac{1}{n-1}(w_{k+2} - w_{k+1}) \\
&= \frac{w_n - w_1}{n-1}.
\end{aligned}$$

Doing similarly for three elements, we get

$$\iota(\{i,j,k\}) = \frac{(w_n - w_{n-1}) - (w_2 - w_1)}{n-2}.$$

We compute now the general expression, following the same technique.

$$\begin{aligned}
\iota(A) &= \frac{1}{n-|A|+1} \sum_{k=0}^{n-|A|} \sum_{l=0}^{|A|} (-1)^l \binom{|A|}{|A|-l} \mu(B_{|A|+k-l}) \\
&= \frac{1}{n-|A|+1} \sum_{j=0}^{|A|-2} (-1)^j \binom{|A|-2}{j} [w_{n-j} - w_{|A|-j-1}]
\end{aligned}$$

dropping some tedious computations. □

Some comments are noteworthy here.

- as the equivalent measure of an OWA depends only on the cardinal of the sets, the Möbius and the interaction representation too depend only on the cardinal of the sets.

- taking the Möbius representation, we see that $m(\{i\}) = w_1, m(\{i,j\}) = -w_1 + w_2, m(\{i,j,k\}) = w_1 - 2w_2 + w_3$ and so on, for every $i, j, k \in X$. Clearly, the sign of m may be positive for some A or negative for others, depending on the weight vector. This shows that in general an OWA is not a Choquet integral with respect to a belief function —for which m is nonnegative—, although this can happen with suitable weights (take for example $n = 3$ and $w_1 = 0.1, w_2 = 0.2$, and $w_3 = 0.7$).

- the Shapley index is the same for all elements, as expected since an OWA operator is commutative and does not favor any element. Thus, the global importance of an element in coalitions is the same for all elements.

- similarly, I_{ij} is the same for all pairs of elements, so that the interaction effect between elements is constant on all pairs of elements. It is then impossible to represent by OWA a positive synergy on a pair of elements. and simultaneously a negative synergy on another pair. OWA considers all elements *globally*.

- the sign of I_{ij} is decided solely by the sign of $w_n - w_1$. This means that if $w_n > w_1$, the interaction is positive, leading to a conjunctive behavior of the OWA operator. This can be interpreted by saying that w_n is the weight on the lowest value. If w_n is high (or at least higher than w_1 which is the weight on the highest value, then clearly OWA behaves like a minimum. Conversely, if $w_1 > w_n$, then we have a disjunctive behavior.

- we remark that the weights are involved in the interaction indices from the extremities towards the middle, i.e. considering first (w_1, w_n), then (w_2, w_{n-1}), etc. and finally $w_{n/2}$, when ranging from $|A| = 2$ to $|A| = n$. We will interpret this fact in the next section.

- using property 2 allows us to get an expression of OWA operators in terms of the Möbius representation m. However, since m is non zero in general, this expression will not be of interest in practice.

- OWA in general does not correspond to k-additive measures.

The following result concerning duality can be shown.

Property 5 *Let* $\mathrm{OWA}_{w_1,\dots,w_n}$ *be an OWA operator, whose associated fuzzy measure is* μ*. Then the dual fuzzy measure* μ^* *corresponds also to an OWA operator, which is* $\mathrm{OWA}_{w_n,\dots,w_1}$*, i.e. the OWA operator with the reversed weight vector. Moreover,*

$$\mathrm{OWA}_{w_n,\dots,w_1}(a_1,\dots,a_n) = 1 - \mathrm{OWA}_{w_1,\dots,w_n}(1-a_1,\dots,1-a_n).$$

Proof: consider the term $(-1)^j\binom{|A|-2}{j}[w_{n-j} - w_{|A|-j-1}]$ in $I(A)$, and replace the weight vector by its reversed version, thus obtaining $(-1)^j\binom{|A|-2}{j}[w_{j+1} - w_{n-|A|+j+2}]$. Letting $l := |A| - 2 - j$, we obtain $(-1)^{|A|-2-l}\binom{|A|-2}{|A|-2-l}[w_{|A|-l-1} - w_{n-l}]$, which can be written as $(-1)^{|A|+1}(-1)^l\binom{|A|-2}{l}[w_{n-l} - w_{|A|-l-1}]$. Due to property 1, this is a term of $I^*(A)$. As this remains true for any term, this proves that the two OWAs are dual. Now the relation between the OWA operators comes from the property of duality of Choquet integrals [9]:

$$\mathcal{C}_\mu(a_1, \ldots, a_n) = 1 - \mathcal{C}_{\mu^*}(1 - a_1, \ldots, 1 - a_n).$$

□

4 Approximation of OWA operators

The computation of OWA involves a sorting of elements, which is a time consuming task, so one can think of its approximation by simpler operators. We don't intend here to address this problem in its full generality, but merely to use existing results in the k-additive measure domain, which are related to this issue.

The fundamental idea is the following. In the field of complexity analysis and cooperative game theory, the problem of approximating a fuzzy measure μ by an additive measure ν, or by a k-additive fuzzy measure $\nu^{[k]}$, has been solved [11]. If the approximation criterion is to minimize $\sum_{A\subset X}(\mu(A) - \nu(A))^2$ —we call this the *squared error criterion* (SEC)—, then the answer is given by the *Banzhaf index* b_i:

$$\nu(\{i\}) = b_i := \frac{1}{2^{n-1}} \sum_{A\subset\{i\}^c} (\mu(A\cup\{i\}) - \mu(A)). \tag{15}$$

Remark the similarity with the Shapley index. Contrary to the Shapley index, the Banzhaf index has not the property $\sum_{i=1}^n b_i = 1$, so that $\nu(X) \neq \mu(X) = 1$. Keeping the same criterion under the additional constraint $\nu(X) = \mu(X) = 1$, the best *faithful* approximation of μ is:

$$\nu(\{i\}) = b_i + \frac{1}{n}(\mu(X) - \sum_{j=1}^n b_j). \tag{16}$$

Concerning the Shapley value, it has been shown that it is the best approximation of a fuzzy measure for a *weighted* criterion, which is $\sum_{A\subsetneq X}\frac{1}{\binom{n-2}{|A|-1}}(\mu(A) - \nu(A))^2$, under the constraint $\nu(X) = \mu(X) = 1$. We call this the *weighted squared error criterion* (WSEC). Notice that the weighted criterion put less importance on subsets of cardinal around $n/2$. This can be explained by saying that WSEC cares about elements in subsets, trying to avoid duplicacies. Since subsets of around $n/2$ elements are the most numerous, a single element is involved very often in such subsets and their contribution to the sum must be diminished. By contrast, the non weighted criterion considers subsets as individual and separated entities, without considering elements contained in them. In short, we could say that WSEC approximates on X, and SEC on $\mathcal{P}(X)$.

Another interesting fact is the following, shown by Murofushi [12].

Theorem 2 *(Murofushi [12]) Let $a_1, \dots, a_n$ be random variables in $[0, 1]$ such that the different orderings $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \cdots \leq a_{\sigma(n)}$ have same probability, which is $1/n!$. σ denotes any permutation on $\{1, \dots, n\}$. Then for any fuzzy measure μ,*

$$E[\mathcal{C}_\mu(a_1, \dots, a_n)] = \sum_{i=1}^{n} v_i a_i.$$

where $E[\cdot]$ denotes the expected value, and v_i the Shapley index of i.

Such a probability on orderings can be obtained simply by taking independent uniform random variables a_i, $i = 1, \dots, n$. The meaning of this theorem is the following: if one want to approximate $\mathcal{C}_\mu$ by a weighted sum, so that the expected value (taken on the whole hypercube $[0, 1]^n$ with uniform distribution) of the error is zero —we call this the *expected value criterion* (EVC)—, then the weight vector is the Shapley value.

Let us sum up the discussion at this point, concerning the approximation at the first order, i.e. by $\sum_{i=1}^{n} \nu_i a_i$, denoting $\nu(\{i\})$ by ν_i for short.

- using the squared error criterion (SEC), we find that ν_i is the Banzhaf index. For OWA, it is straightforward to get from its definition:

$$\nu_i = b_i = \frac{1}{2^{n-1}} \sum_{k=0}^{n-1} \binom{n-1}{k} w_{k+1}. \tag{17}$$

  Remark that, despite its simple definition compared to the Shapley index, b_i is much more complex that v_i for an OWA.

- using WSEC, we get $\nu_i = v_i = 1/n$.

- using EVC, we also get $\nu_i = 1/n$.

- the criterion which should be actually used here is rather the sum on the unit hypercube $[0, 1]^n$ of squared differences between $\text{OWA}_{w_1,\dots,w_n}(a_1, \dots, a_n)$ and $\sum_{i=1}^{n} \nu_i a_i$, but this criterion has still to be studied and expressed under a suitable form. The nearest in spirit is EVC, suggesting that the Shapley value should be chosen as the best linear approximation.

Obviously, this is a rather crude approximation, so we turn to the second order approximation, i.e. using a 2-additive measure $\nu^{[2]}$. Let us denote by $m^{[2]}$ its Möbius transform, and $m_i^{[2]}, m_{i,j}^{[2]}$ a short-hand for $m^{[2]}(\{i\}), m^{[2]}(\{i, j\})$. In [11], it is shown that the solution to SEC is:

$$\begin{aligned} m_i^{[2]} &= b_i := \frac{1}{2^{n-1}} \sum_{A \subset \{i\}^c} (\mu(A \cup \{i\}) - \mu(A)) \\ m_{i,j}^{[2]} &= b_{ij} := \frac{1}{2^{n-2}} \sum_{A \subset \{i,j\}^c} (\mu(A \cup \{i, j\}) - \mu(A \cup \{i\}) - \mu(A \cup \{j\}) + \mu(A)). \end{aligned}$$

Again remark the striking analogy between v_i, I_{ij} and b_i, b_{ij}. The computation of b_{ij} for OWA gives the following.

$$\begin{aligned} b_{ij} &= \frac{w_n - w_1}{2^{n-2}} + \frac{1}{2^{n-2}} \sum_{k=1}^{[n/2]-1} \left[\binom{n-2}{k} - \binom{n-2}{k-1} \right] (w_{n-k} - w_{k+1}) \\ &+ \begin{cases} -\left[\binom{n-2}{[n/2]} - \binom{n-2}{[n/2]-1} \right] w_{[n/2]+1}, & \text{if } n \text{ odd} \\ 0, & \text{if } n \text{ even} \end{cases} \end{aligned}$$

where $[a]$ is the integer part of a. Again the expression of b_{ij}, despite its simpler definition, is much more complex than the one of I_{ij}.

Unfortunately, there is yet no result about a possible weighted criterion leading to v_i and I_{ij} as solution, but we guess that it could be obtained following the same approach as for v_i and WSEC. Finally, turning to the probabilistic criterion, the argument goes as follows. One has to find a 2-additive measure $\nu^{[2]}$ satisfying:

$$E[\mathcal{C}_\mu(a_1, \ldots, a_n)] = E[\mathcal{C}_{\nu^{[2]}}(a_1, \ldots, a_n)].$$

Using theorem 2, this reduces to $v_i = v_i^{[2]}$, $i = 1, \ldots, n$, with $v^{[2]}$ the Shapley value of $\nu^{[2]}$. Clearly, a whole class of 2-additive measures satisfy this equality, since any monotonic 2-additive set function with $I(\{i\}) = v_i$ for all i and arbitrary $I(\{i,j\})$ values, provided they do not violate monotonicity, is a candidate (see [7, 3] for a complete treatment of this question). In particular, the 2-additive measure with $m_i^{[2]} = v_i$, $m_{ij}^{[2]} = I_{ij}$, for all i, j, is a candidate, and the 2-additive measure obtained by SEC is not. In fact, the expected value criterion is too loose to obtain a significant approximation , and variance has also to be taken into account. But this problem has to still to be studied.

The above facts seem to suggest that the solution with v_i, I_{ij} is a good solution for approximation, but further study is yet necessary to establish firmly this, as explained above. Under this hypothesis, the best 2nd order approximation of OWA is given by the following expression (use property 2):

$$\mathrm{OWA}_{w_1,\ldots,w_n}(a_1, \ldots, a_n) \approx \frac{1}{n} \sum_{i=1}^{n} a_i + \frac{w_n - w_1}{n-1} \sum_{\{i,j\} \subset X} (a_i \wedge a_j). \tag{18}$$

An important fact to stress here is that only w_1 and w_n are involved in the 2nd approximation. It is easy to see that, doing similarly with 3-additive, 4-additive, etc. measures, w_{n-1} and w_2 will be used, then w_{n-2} and w_3, etc. The important conclusion we can draw is that most of the information is contained in the extremities of the weight vector.

References

[1] A. Chateauneuf and J.Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17:263–283, 1989.

[2] D. Denneberg. *Non-Additive Measure and Integral.* Kluwer Academic, 1994.

[3] M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems.*

[4] M. Grabisch. Fuzzy integral in multicriteria decision making. *Fuzzy Sets & Systems*, 69:279–298, 1995.

[5] M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. In *4th Int. Conf. on Soft Computing*, Iizuka, Japan, september 1996.

[6] M. Grabisch. The application of fuzzy integrals in multicriteria decision making. *European J. of Operational Research*, 89:445–456, 1996.

[7] M. Grabisch. k-order additive fuzzy measures. In *6th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 1345–1350, Granada, Spain, july 1996.

[8] M. Grabisch. Fuzzy integral as a flexible and interpretable tool of aggregation. In B. Bouchon-Meunier, editor, *Aggregation of evidence under fuzziness*. Physica Verlag, to appear.

[9] M. Grabisch, T. Murofushi, and M. Sugeno. Fuzzy measures of fuzzy events defined by fuzzy integrals. *Fuzzy Sets and Systems*, 50:293–313, 1992.

[10] M. Grabisch, H.T. Nguyen, and E.A. Walker. *Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference.* Kluwer Academic, 1995.

[11] P.L. Hammer and R. Holzman. On approximations of pseudo-Boolean functions. *ZOR - Methods and Models of Operations Research*, 36:3–21, 1992.

[12] T. Murofushi. A technique for reading fuzzy measures (i): the Shapley value with respect to a fuzzy measure. In *2nd Fuzzy Workshop*, pages 39–48, Nagaoka, Japan, october 1992. In Japanese.

[13] T. Murofushi and S. Soneda. Techniques for reading fuzzy measures (iii): interaction index. In *9th Fuzzy System Symposium*, pages 693–696, Sapporo, Japan, may 1993. In Japanese.

[14] T. Murofushi and M. Sugeno. Non-additivity of fuzzy measures representing preferential dependence. In *2nd Int. Conf. on Fuzzy Systems and Neural Networks*, pages 617–620, Iizuka, Japan, july 1992.

[15] T. Murofushi and M. Sugeno. Some quantities represented by the Choquet integral. *Fuzzy Sets & Systems*, 56:229–235, 1993.

[16] G.C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2:340–368, 1964.

[17] G. Shafer. *A Mathematical Theory of Evidence.* Princeton Univ. Press, 1976.

[18] L.S. Shapley. A value for n-person games. In H.W. Kuhn and A.W. Tucker, editors, *Contributions to the Theory of Games, Vol. II*, number 28 in Annals of Mathematics Studies, pages 307–317. Princeton University Press, 1953.

3

MATHEMATICAL ISSUES AND OWA OPERATORS

USEFUL TOOLS FOR AGGREGATION PROCEDURES: SOME CONSEQUENCES AND APPLICATIONS OF STRASSEN'S MEASURABLE HAHN-BANACH-THEOREM

Heinz J. Skala, University of Paderborn
E-mail: hoexter@STABIF.wiwi.uni-paderborn.de

We state a celebrated theorem due to Strassen (1965) and derive from it Meyer's (1966) characterization of dilation kernels. It is shown that the latter result and thus Strassen's theorem provides a useful tool for deriving characterizations of certain orderings. As an example we prove a famous result due to Hardy/ Littlewood/Pólya (1934,1952). Finally we state some applications in the field of OWA-operators.

Key words: OWA-operators, ordering of weight systems, a measurable Hahn-Banach theorem, dilations.

1. INTRODUCTION AND MOTIVATION.

Some years ago Yager (1988) studied aggregation functions (which he called OWA operators) of the following type:

(1) $F(x) = \sum_{i=1}^{n} c_i \, x_i^*, \; x \in \mathbb{R}^n$

where the "weights" c_i are nonnegative, $\sum_{i=1}^{n} c_i = 1$ and $(x_1^*,\ldots,x_n^*)$ denotes the decreasing rearrangement of $(x_1,\ldots,x_n)$.

Soon it was observed that functions of the type (1) are similarly ordered additive (comonotonic additive as some authors prefer) and thus can be represented by a Choquet integral. See in particular the interesting papers by Fodor/Marichal/Roubens (199?), Ralescu/Ralescu (199?) and the literature cited therein. In simple situations such as the one considered by Schmeidler (1986) a respresentation is simply obtained by integrating the (available) distribution functions. In more complicated situations other techniques are needed which lead to a generalization of the Choquet integral. Such representations obtained by integrating the closure of certain martingales have been introduced by Skala (199?).

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

Corresponding aggregation procedures may be of practical importance but this is another story.

In my (1991) paper I compared aggregation procedures of type (1). As main technical tool I used Meyer's (1966) generalization of the Cartier/Fell/Meyer theorem (theorem 5 of the present paper). Although my main theorem there has a more elementary proof I did this with regard to generalizations to infinite situations if they should turn out to be of interest. In the meantime I received preprints of stimulating papers by Basile/D'Apuzzo (1996) and Ralescu/Ralescu (199?) which convinced me that techniques originating from Hardy, Littlewood and Pólya (1934/1952) should be of considerable interest to scientists working in the field of aggregation procedures. Techniques of the mentioned type can be found in books focussing different areas of mathematics. By lack of space we therefore refrain from giving an overview and prefer to present a very useful theorem due to Strassen (1965) and some of its consequences together with applications to OWA operators.

2. STRASSEN'S DISINTEGRATION THEOREM AND TWO REMARKABLE CONSEQUENCES

As Strassen's (1965) disintegration theorem may in a certain sense be viewed as measurable version of the most fundamental theorem in functional analysis, namely the Hahn-Banach theorem, its wide applicability is not surprising.

Let E be a Banach space with norm $\|\cdot\|$. We recall that a functional $p:E\to\mathbb{R}$ is said to be sublinear if

$$p(\alpha x) = \alpha\, p(x) \qquad \text{for all } \alpha \geq 0 \text{ and } x \in E$$

and

$$p(x+y) \leq p(x)+p(y).$$

We denote the family of sublinear functionals on E by S_E. As usual E' denotes the dual of E.

<u>Theorem 1.</u> (Strassen) Let E be a separable Banach space and let (Ω,F,λ) be a measure space within positive finite λ. Let p: $\Omega\to S_E$ be a bounded weakly measurable map (i.e. the real valued function $\omega \to p_\omega(x)$ is measurable for every $x\in E$). Observe that the integral

$$s(x) = \int_\Omega p_\omega(x)\, \lambda(d\omega) \quad , \ x \in E$$

defines a sublinear functional on E.

Let $x' \in E'$, then the following are equivalent:

(i) x' is dominated by s, i.e., $\langle x', x\rangle \leq s(x)$ for every $x \in E$,

(ii) there exists a bounded, weakly measurable map $\omega \to x'_\omega$ from Ω into E' such that

$$x'_\omega(x) \leq p_\omega(x) \qquad \text{for all } x \text{ and } \omega$$

and

$$\langle x', x\rangle = \int_\Omega \langle x'_\omega, x\rangle \lambda(d\omega) \qquad \text{for all } x \in E.$$

For proofs of this result we refer the interested reader to Strassen (1965) or Meyer (1966).

In order to convince the reader of the powerfullness of Strassen's theorem and to make the present paper as selfcontained as possible we derive from it a generalization of results associated with the names Blackwell, Cartier, Fell, Hardy/Littelwood/Pólya, Mokobotzki, Sherman, Stein, and Strassen.

The next two theorems and their proofs are taken from Meyer (1966). They have immediate applications to the theory of aggregation procedures.

We consider two compact metrizable spaces X and Y where X carries a positive measure λ. (By a measure we always understand a regular Borel measure.) Let K_Y denote the positive measures on Y with mass at most equal to 1. It is well known that K_Y is compact and metrizable in the weak topology. Let $x \to M_x \subset K_Y$ be a map such that M_x is nonempty and compact. We suppose that the set of pairs $(x,m) \in XxK_Y$ such that $m \in M_x$ is closed. Let C(Y) denote the continuous functions on Y, then for every $f \in C(Y)$ we define a bounded upper semicontinuous function on X by

$$\hat{f}(x) = \sup_{m \in M_x} \langle m, f\rangle$$

and put

$$p_\lambda(f) = \langle \lambda, \hat{f}\rangle.$$

(If ε_X denotes the unit mass at x we have $\hat{f}(x) = p_{\varepsilon_x}(f)$ for which we simply write $p_X(f)$.)

Theorem 2. Let μ be a positive measure on Y. The following are equivalent:

(i') $\mu(f) = \int_Y f\mu(dy) \leq p_\lambda(f)$ for every $f \in C(Y)$.

(ii') There exists a family $T = (T_x)_{x\in X}$ of positive measures on Y such that $T_x(A)$ is Borel measurable in $x\in X$ for every measurable $A\subset Y$ and $\mu = \int_X T_x \, \lambda(dx)$ and T_x belongs to M_x for λ-almost every $x\in X$.

Proof. Only the implication (i')→(ii') needs a proof. For this we apply theorem 1 and put $\Omega=X$, $\mathbb{F}$=σ-field of λ-measurable sets and E=C(Y). By the Riesz representation theorem we may and will identify C(Y)' with measures on Y. A sublinear functional on C(Y) is now defined by

$$s(f) = \int_Y \; p_*(f)\lambda(dx) = p_\lambda(f) \;\; = \;\; \langle \lambda, \hat{f} \rangle \;\; , \;\; f \in C(Y).$$

Clearly (i') implies that condition (i) of theorem 1 is satisfied. Therefore we obtain a bounded weakly measurable map $t{:}x \to t_x$ from X into C(Y)' such that

$$\langle t_x, f \rangle \le p_x(f) \qquad \text{for all } f \in C(Y), x \in X$$

and

$$\mu(f) = \int_X \langle t_x, f \rangle \lambda(dx).$$

Let $(f_n)_{n\in\mathbb{N}}$ be a dense subset in C(Y), then the λ-measurable functions $x \to \langle t_x, f_n \rangle$ are almost everywhere equal to Borel functions and therefore we can find a map $T{:}X \to K_Y$ which is equal almost everywhere to $t{:}X\to C(Y)'$ such that the functions $x \to \langle T_x, f_n \rangle$ are Borel for every n. ♦

Definition 3. Let C(X) be the family of continuous functions on some compact set X and let $S\subset C(X)$ be a convex cone which contains the positive constants and is closed under the operation $\wedge$ ($f\wedge g=\max(f,g)$). Let μ,λ be two Borel measures on X, then we write $\lambda \succsim_S \mu$ if $\lambda(f) \ge \mu(f)$ for every $f\in S$. (Observe that Meyer (1966) found it more convenient to reverse the ordering defined on the Borel measures.)

Example. Let S denote the cone of continuous convex functions on some finite dimensional space, then the corresponding ordering is well known from the comparison of sampling procedures (Blackwell 1953). For the one dimensional case the ordering has been characterized by Hardy/Littlewood/Pólya in terms of what is nowadays known as dilation.

Definition 4. Let X be a compact metrizable space and let S be a convex cone of continuous functions on X containing the constants and closed under $\wedge$. An S-dilation is any Borel kernel T on X such that $\varepsilon_x \succsim_S \varepsilon_x T$. (Remember that ε_x denotes the unit mass concentrated on x.

Theorem 5. Let X be a compact metrizable space and let λ,μ be two positive Borel measures on X. Assume that S is a convex cone of continuous functions on S containing the constants and closed under $\wedge$.
Equivalent are:

(a) $\mu(f) \leq p_\lambda(f)$ for all $f \in C(X)$.

(b) $\lambda \succsim_S \mu$, i.e. $\lambda(f) \geq \mu(f)$ for all $f \in S$.

(c) There exists an S-dilation T on X such that $\mu=\lambda T$.

Proof. We simply put X=Y in theorem 2 and let $M_x = \{m : m \in K_X, \varepsilon_x \succsim_S m\}$.

This proves the equivalence of (a) and (c). The equivalence of (a) and (b) has been shown in Meyer (1966). ♦

3. SOME APPLICATIONS WITH SPEZIAL ATTENTION TO AGGREGATION PROCEDURES

As the main ideas discussed in this paper may be traced back to Hardy/ Littlewood/Pólya (1934/1952) we want to derive one of their celebrated results from theorem 5. Actually we derive a somewhat more general result due to Blackwell (1953).

Let $x_1,...,x_n, y_1,...,y_n \in \mathbb{R}^m$, then we may form the mxn matrix $X = (x_1,...,x_n)$ and $Y = (y_1,...,y_n)$.

Theorem 6. Assume that $x_1,...,x_n, y_1,...,y_n \in \mathbb{R}^m$, then the following are equivalent:

(i) $\sum_{i=1}^{n} f(y_i) \geq \sum_{i=1}^{n} f(x_i)$ for every convex continuous function defined on $\mathbb{R}^m$

(ii) X=YT where the nxn matrix T is doubly stochastic.

Proof. Only the implication (i) → (ii) needs to be shown. Define μ as the sum of the n unit masses ε_{x_i} concentrated on the elements $x_i, i=1,\dots,n$ and correspondingly λ as the sum of the unitmasses concentrated on the $y_i, i=1,\dots,n$. Let S be the cone of convex continuous functions on $\mathbb{R}^m$, then $\lambda \succsim_S \mu$ iff (i) holds. Thus theorem 5 guarantees the existence of an S-dilation. We put $T_{ij} = T_{x_j}(\{y_i\})$. Observe that the functions $f(x)=x_i$, $i=1,\dots,n$ as well as their negative conterparts belong to S. This and the properties of a dilation imply that

$$x_j = \sum_{i=1}^{m} y_i T_{ij} \quad , \quad T_{ij} \text{ doubly stochastic.} \quad \blacklozenge$$

Let there be n criteria in a multiple criteria problem. To every object we associate a function $f: I=\{1,\dots,n\} \to \mathbb{R}$ where $f(i)=x_i$ denotes the degree to which the object satisfies the corresponding criterion. An aggregation procedure is simply a function F defined on the n-tuples $(x_1,\dots x_n)$. The most important aggregation procedures are defined by using weight systems. Skala (1991) proposed to characterize weight systems by means of convex cones of functions and obtained

<u>Theorem 7.</u> Let $I=\{1,\dots,n\}$ and let S be a convex cone of functions on I containing the positive constants and closed under $\wedge$. For any two nonnegative weight systems $c=(c_1,\dots,c_n)$ and $c'=(c_1',\dots,c_n')$ the following are equivalent:

(i) $\sum_{i=1}^{n} c_i\, f(i) \geq \sum c_i'\, f(i)$ for all $f \in S$.

(ii) There exists a substochastic matrix $T=(T_{ij})$ such that cT=c' and moreover $f(i) \geq \sum T_{ij}\, f(j)$ for all $i=1,\dots,n$; $f \in S$.

Proof. This is an easy consequence of theorem 5 (see Skala (1991)).

Remarks.

(1) Theorems 5 and 7 have obvious counterparts for convex cones closed under $\vee$.

(2) If the convex cone contains all constants, then T is stochastic, i.e. a Markov kernel or Markov matrix.

Let us now introduce the following extreme aggregation procedure

$$F_{\wedge}(x_1,\dots,x_n) = x_n^* = \min\{x_1,\dots,x_n\}$$

and

$$F_{\vee}(x_1,\dots,x_n) = x_1^* = \max\{x_1,\dots,x_n\}$$

with the corresponding weight systems $c_{\wedge} = (0,\dots0,1)$ and $c_{\vee} = (1,0\dots,0)$.
In order to measure the closeness of some weight related aggregation procedure F_c to the extremes $F_{\vee}$ and $F_{\wedge}$ respectively Yager (1988) used the special decreasing function $f(i) = x_i = (n-i)$ in order to obtain his measure

$$\text{orness}\,(c) = \sum_{i=1}^{n} (n-i)c_i.$$

A simple example from Skala (1991) shows that this is not a satisfying measure.
Let c=(1/3,1/3,1/3) and c'=(1/2,0,1/2), then orness (c)=orness(c')=1/2
but

$$F_c(2,1/2,0) = 5/6 < F_{c'}(2,1/2,0) = 1$$

and

$$F_c(2,3/2,0) = 7/6 > F_{c'}(2,3/2,0) = 1.$$

Thus the closeness of F_c to $F_{\vee}$ and the closeness of $F_{c'}$ to $F_{\vee}$ are not compatible although orness (c)=orness(c'). Skala (1991) therefore proposed to use an appropriate cone of functions for the comparison of OWA-operators and gave the following

<u>Definition 8.</u> An aggregation procedure (OWA-operator) with weight system c, $\Sigma c_i = 1$ is more or-like than one with weight system c', $\Sigma c_i' = 1$, shortly $c \succ c'$, iff there exists a stochastic matrix T such that c'=cT and $f(i) \geq \sum_{j=1}^{n} T_{ij}\, f(j)$ for all i=1,...,n and all decreasing functions $f\colon \{1,\dots,n\} \to \mathbb{R}_+$.

It is easy to prove that the relation $\succ$ defined above is reflexive, transitive and antisymmetric and that $c \succ c'$ implies orness (c)≥orness(c'). In addition weight systems of OWA-operators comparable by $\succ$ are related by upper triangular matrices and we have

<u>Theorem 9.</u> Let c and c' be weight systems of OWA-operators summing to 1, then the following are equivalent:

(i) c'=cT for some upper triangular stochastic matrix T.

(ii) $\sum_{i=1}^{k} c'_i \le \sum_{i=1}^{k} c_i,\ k = 1,\dots, n-1.$

A proof of theorem 9 and additional results may be found in Skala (1991).

The infinite counterparts of the results just discussed are easily available and we shall not discuss them here. However, we want to shortly comment on some proposals as discussed by Basile/D'Apuzzo (1996).
We start with a measure space of criteria which, for simplicity, is assumed to be the unit interval [0,1] endowed with the Lebesgue measure. To every alternative x we assign a function $f_x:[0,1] \to \mathbb{R}_+$ where $f_x(t)$ denotes the degree to which alternative x satisfies criterion t. It is supposed that the $f_x(\cdot)$ are integrable. For such functions Hardy/Littlewood/Pólya (1934,1952) have defined a unique (up to countably many exceptions) decreasing function $f_x^*(\cdot)$ such that f and f_x^* are equimeasurable. This clearly implies that $\int_0^1 \varphi(f_x(t))dt = \int_0^1 \varphi(f_x^*(t))dt$ for every measurable φ for which the integral exists. Thus it makes sense to consider aggregation procedures of the form

$$F_w(f_x) = \int_0^1 c(t)f_x^*(t)dt$$

where $c(\cdot)$ is a weight function such that $\int_0^1 w(t)dt = 1$. More generally Basile/D'Apuzzo propose procedures depending also on some strictly monotone continuous φ, i.e.

$$F_{w\varphi}(f_x) = \varphi^{-1}(\int_0^1 c(t)\varphi(f_x^*(t))dt).$$

Not surprisingly they show that the natural counterpart of Yager's measure of orness is also not satisfying. Motivated by theorem 4 in Skala (1991) the authors propose the following

<u>Definition 10.</u> Let c and c′ be weight functions such that $\int_0^1 c(t)dt = \int_0^1 c'(t)dt = 1$. We put

$$c' \succ c \quad \text{iff} \quad \int_0^x c'(t)dt \ge \int_0^x c(t)dt$$

for every $x \in [0,1]$.

Again it is easy to show that the relation $\succ$ is reflexive, transitive and, moreover, c' $\succ$ c and c $\succ$ c' imply c=c' almost everywhere in [0,1]. An easy calculation proves

Theorem 11. Let c, c' be two weight functions such that $c' \succ c$, then $F_{c'\varphi}(f_x(t)) \geq F_{c\varphi}(f_x(t))$ for every integrable $f_x:[0,1] \to \mathbb{R}_+$ and every strictly monotone and continuous function φ.

For further results and the proof of theorem 11 see Basile/D'Apuzzo (1996).

We would like to close this tutorial paper by observing that there is another result due to Strassen (1965) which should be of considerable interest to researchers working in the theory of aggregation procedures. Strassen's result has been generalized by Skala (1993) where the interested reader may also find a characterization of the ordering of Radon measures on ordered Hausdorff spaces.

References

Basile,L., L.D'Apuzzo, *Orderings for classes of aggregation operators.* Uncertainty, Fuzziness, and Knowledge-Based Systems 1996,4, 145-156.

Blackwell,D., *Equivalent comparison of experiments.* Ann. Math. Stat. 1953, 24, 265-272.

Fodor,J., J.L.Marichal, M.Roubens, *Characterization of the ordered weighted averaging operators.* To appear (199?).

Hardy,G.H., J.E.Littlewood, G.Pólya, *Inequalities.* Cambridge University Press, 1934/52.

Meyer,P.R., *Probability and Potential.* Blaisdell, 1966.

Ralescu,A.L., D.A.Ralescu, *Extensions of fuzzy aggregation.* To appear (199?).

Schmeidler,D., *Integral representation without additivity.* Proc. AMS 1986, 97, 255-261.

Skala,H.J., *Concerning ordered weighted averaging aggregation operators.* Statistical Papers 1991, 32, 35-44.

Skala,H.J., *The existence of probability measures with given marginals.* Ann. Prob. 1993, 21, 136-142.

Skala,H.J., *On a representation theorem of Schmeidler.* To appear in Statistical Papers (199?).

Strassen,V., *The existence of probability measures with given marginals.* Ann. Math. Stat. 1965, 36, 423-439.

Yager,R.R., *On ordered weighted averaging aggregation operators in multicriteria decision making.* I.E.E.E., Transactions on Systems, Man, and Cybernetics 1988, 18,1, 183-190.

OWA SPECIFICITY

Arthur Ramer

Knowledge Systems Group, Computer Science
University of New South Wales
Sydney 2052, Australia
E-mail: ramer@cse.unsw.edu.au

Abstract

A comprehensive model for evaluating *specificity* of fuzzy sets is presented. It is designed in terms of possibility values, independent of the domain of discourse. For a discrete distribution $\pi = (p_1 \geq p_2 \geq \cdots)$ two measures are defined

exponential $Sp(\pi) = p_1 - \sum_{i\geq 2}(\omega^{i-1} - \omega^i)p_i$

logarithmic $I(\pi) = \sum(p_i - p_{i+1}) \log i$

Measure $Sp(\pi)$ is derived from a few intuitively plausible properties of specificity; measure $I(\pi)$ is dual to *nonspecificity* in Dempster-Shafer theory.

The resulting model has a *natural* OWA structure, which follows necessarily from the basic assumptions. This leads to an *inverse* problem, one of developing, within the general OWA framework, the features successfully employed in specificity and uncertainty models. We suggest some directions in the concluding section.

1 Introduction

Fuzzy set (X, π) represents a form of a likelihood function on the domain of discourse X; we use π rather than the customary μ to stress the possibility interpretation. Elements of X where π reaches maximum play privileged role, and selecting one for which $\pi(x_0) = \max \pi(x)$ becomes a form of the maximum

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

likelihood decision. It is then important to ask how *definite* has been such decision, and whether another element would offer a close choice. The values $\pi(x)$ can be considered *possibilities* associated with the elements of X. In this interpretation, *specificity* becomes an attribute of the complete set of possibilities, the attribute assuming either numeric or linguistic values. Our paper develops a comprehensive model of specificity, expressed as a *numerical* function of a possibility assignment. We base our motivations and intuitions on fuzzy sets. However, our formal model deals exclusively with possibility values, gaining the flexibility of independence of a specific domain of discourse. Furthermore, it shows to advantage in decision analysis by permitting *specificity* as one of the 'objective' criteria. It is also reflected in our terminology and notation, where we speak of *possibility distributions* in preference to fuzzy sets.

Specificity as a numerical function was considered twice before in the literature on possibility theory; it was also considered on several occasions within Dempster-Shafer evidence theory. The explicit definition was given by Yager [16]. Indirectly, specificity was introduced as the possibilistic U-uncertainty [5]. It is closely related to the *nonspecificity* [1, 4, 11] in Dempster-Shafer theory, as for the possibility assignments the latter becomes precisely the U-uncertainty.

In the published literature, specificity and uncertainty are defined primarily for the discrete finite domains. In the companion chapter the continuous model is analyzed extensively. It is shown there how to define specificity measures for *arbitrary measurable* sets as domains of discourse. They can be discrete, finite or infinite or, as a measurable set X, have $\mu(X) < \infty$ or $\mu(X) = \infty$. Here we remain within the context of standard, discrete domain discourse making all the arguments and definitions completely elementary. Most of presentation can be viewed as defined simply for finite sets only. On occasion it is convenient to permit infinite sequences of possible choices, hence also infinite sequences of possibility values. They have a very attractive, practical interpretation. One can treat them as potentiall *unbound* listings of options; although we fully well know that we will stop at some point, we do not commit in advance to stopping at any specific point.

2 Notations and definitions

We adopt [3] convention of keeping subscripts and indices of summation to the minimum. We avoid stating the base of logarithms, as its change would amount to multiplying all the formulas by the same constant. However, to simplify such constant we assume a binary logarithm log_2 for the discrete and a natural logarithm ln for the continuous cases.

A discrete distribution is defined as a pairing of a finite or a countable infinite domain of discourse X and a possibility assignment $\pi : X \to [0, 1]$. As the domain X plays somewhat secondary role, we often refer to the multiset of values $\{\pi(x)|x \in X\}$ as the distribution. We *do not* require normalization, permitting $\sup \pi(x) < 1$; however, we require that the supremum is actually attained. For a discrete X, we enumerate its elements as $x_1, x_2, \ldots$, denoting possibility values $p_i = \pi(x_i)$. We use a special symbol $1^{(n)}$ for the distribution consisting of n ones.

3 Rearrangements of sequences

We denote $(a_i), i = 1, \ldots, n$ as (a_i) or even (a), if no ambiguity is possible, and use $(\tilde{a})$ to denote its decreasing rearrangement. For finite sequences, rearrangements are permutations of their elements. For infinite sequences we need to define rearrangements which may *no longer* have this property. As we need mostly descending rearrangements, we formalize their definition only.

Given $(a_i), i = 1, 2, \ldots$, we put as $\tilde{a}_i$ the supremum of the elements $a \in (a_i)$ for which there is at least i greater or equal elements in (a). In other words, for any α in $[\inf_i a_i, \sup_i a_i]$ we want to have the same number of elements above α both in (a) and $(\tilde{a})$. For a finite sequence this definition is equivalent to forming a decreasing permutation. However, the infinite case is different. For (a) without accumulation points from below, such $\tilde{a}_i$ will be always an element of the original sequence. If, in addition, all a_i are distinct we can ask for exactly i greater or equal elements. And if this number is infinite? Then we have $\tilde{a}_i = \tilde{a}_{i+1} = \tilde{a}_{i+2} = \ldots$ This has the effect of 'loosing' all the elements of (a) less than the greatest concentration point of the multiset $\{a_i\}$; in particular, if some a_0 occurs infinitely often, all $a_i < a_0$ are lost.

This construction can be formalized in terms of *cuts*. For any nonnegative real r a cut at that level is a multiset $c_r = \{a_j | a_j \geq r\}$; now $\tilde{a}_i = \sup\{r \| c_r| \geq i\}$. This method was first used as a basis of the rearrangements of *functions* [3, 6] where for f defined on X, we want $\tilde{f}$ be a descending equivalent of f. We require that all $\alpha-$cuts of f to be of the same measure as $\alpha-$cuts of $\tilde{f}$. The construction is now simple

$$P(y) = \mu(\{x : f(x) \geq y\}),$$

$$\tilde{f}(x) = P^{-1}(x).$$

As an example we define anew discrete rearrangements. We associate with the sequence $(a) = (a_1, \ldots, a_n, \ldots)$ a step function $f : x \mapsto a_{\lceil x \rceil}$, where $\lceil x \rceil$ denotes the greatest integer no less than x. Then the descending rearrangement $\tilde{f}$ corresponds to $(\tilde{a})$.

4 Principles of specificity

We use as a discrete model a countable infinite distribution $(p_i), i = 1, 2, \ldots$; finite distribution can be identified with an initial segment $(p_i), i = 1, \ldots n$, for some n. Our objective is to capture formally the informal intuition about specificity. The main premise is the principle of *juxtaposition:*

$Sp(\pi)$ expresses the preference for a certain maximal p_0 over any and all the remaining p_i.

Now let us consider how, having selected $p_0 = \max(p)$, its *informal specificity* is estimated. We look first for the next largest p_i and estimate how its presence diminishes the specificity. The process is then iterated in the order of decreasing values of p_i, every next value lowering the estimated specificity. We can picture it as a sequential process, its input the decreasing rearrangement $(\tilde{p}_i)$. We may also surmise that, for a given i, the drop in specificity caused by $\tilde{p}_i$ will not depend on the earlier inputs $\tilde{p}_1, \ldots, \tilde{p}_{i-1}$. This assumption of *independent influence* is consistent with the *juxtaposition* interpretation of specificity.

Let us consider the effect of a uniform modification of (p). For a scaling $\alpha\pi = (\alpha p_1, \ldots, \alpha p_n, \ldots), 0 \leq \alpha \leq 1$, we may assume that for any two distributions π and ϖ, their relative specificities remain unchanged

$$\frac{Sp(\alpha\pi)}{Sp(\alpha\varpi)} = \frac{Sp(\pi)}{Sp(\varpi)}$$

If the change is a shift of values $\pi - \beta = (p_1 - \beta, \ldots, p_n - \beta, \ldots)$ for $\beta \leq \inf_i p_i$, then no change of specificity should occur. This again results from *juxtaposition*, noting that for maximal p_0 and any given p_i, both their sequential placement and the difference in magnitude remain the same after the shift. We remark that such shift can be performed only for the infinite distributions.

Last item considered will be the effect of offering yet another choice, identical in value to several choices already provided. Consider distribution $1^{(n-1)}$, when another value 1 ‘arrives’, transforming it into $1^{(n)}$. The common perception of *specificity* is that the change due to such n-th choice will be ever less as n increases—a *diminishing return*. This view is also strongly supported by the analysis presented in the next section. We demonstrate there that, in the numerical terms, the decrease of specificity due to the n-th value 1 has limit zero as n goes to infinity. The statement above makes that limit to be reached monotonically.

Although in the absolute terms the influence of each additional 1 decreases, it does not say much about its *relative* effect. With rather less assurance, we can postulate that the arrival of each consecutive 1 takes away the same proportion of specificity $Sp(1^{(n-1)})$ still available. After all, we consider yet another *identical* choice; only we consider it at stage n and not sooner.

So far, we did not mandate any fixed values of the specificity for any special distributions. For definiteness we choose $[0, 1]$ range of values, assigning $Sp(1, 0, \ldots) = 1$. The least specific distributions are of the form $(c, \ldots, c, \ldots)$, in particular $(1, \ldots, 1, \ldots)$; we put $Sp(1, \ldots, 1, \ldots) = 0$.

5 Fuzzy specificity

We proceed to extract an analytical representation from the rules elaborated in the previous section. On their basis we assume *specificity* to be invariant under the permutations of (p) and, more significantly, under the descending rearrangements of (p). Informal specificity is estimated by selecting an element of the maximum value, then considering the alternatives, one by one in the descending order of values. For an infinite (p) certain values could be indefinitely delayed, thus allowing us to ignore considering those values altogether. This is exactly equivalent to replacing (p) by $(\tilde{p})$ and computing $Sp(\tilde{p})$.

Form the assumption of *independent influence* we have a sum form

$$Sp(\pi) = \sum_{i=1}^{\infty} f_i(\tilde{p}_i).$$

The rule about scaling leads to

$$\frac{Sp(\alpha\pi)}{Sp(\pi)} = g(\alpha),$$

with $g(\alpha\beta) = g(\alpha)g(\beta)$ for arbitrary α, β. Therefore $g(\alpha) = \alpha^k$ for some constant k and

$$Sp(\pi) = w_1\tilde{p}_1^k - \sum_{i\geq 2} w_i\tilde{p}_i^k.$$

To determine k we compute $Sp(1, \frac{1}{2}, \ldots, \frac{1}{2}, \ldots)$ in two different ways. First

$$1 - \sum_{i\geq 2} w_i \frac{1}{2^k} = 1 - \frac{1}{2^k}.$$

Next, applying additive shift gives $(\frac{1}{2}, 0, \ldots)$ and the specificity $\frac{1}{2^k}$. Equating those values yields $k = 1$ and a linear representation

$$Sp(\pi) = \tilde{p}_1 - \sum_{i\geq 2} w_i\tilde{p}_i$$

with $\sum_{i\geq 2} w_i = 1$. ¿From here we can conclude that $\lim_{i\to\infty} w_i = 0$, and 'diminishing returns' suggest $1 > w_2 > w_3 > \cdots$.

We shall consider the linear form of $Sp(\pi)$ with the monotonically decreasing coefficients to be a *general* form of the specificity function. It is general enough to fit most applications and, if w_i are supplied, it offers a comparison scale among the distributions.

Coefficients w_i can be established precisely if we assume the rule of constant influence of equal choices. After more calculations

$$w_i = \omega^{i-1} - \omega^i.$$

for some ω, $0 < \omega < 1$, producing a *definite* form of specificity

$$Sp(\pi) = \tilde{p}_1 - \sum_{i\geq 2} (\omega^{i-1} - \omega^i)\tilde{p}_i.$$

In many applications selecting the actual value of ω is not critical; its role, is similar to that of the base of logarithms. Choosing $\omega = \frac{1}{2}$, which has certain correspondence to binary logarithms, gives

$$Sp(\pi) = \tilde{p}_1 - \sum_{i\geq 2} \frac{1}{2^{i-1}} \tilde{p}_i.$$

In the above formulas the role of $\tilde{p}_1$ is manifestly different from that of $\tilde{p}_i, i \geq 2$. A more symmetric expression can be obtained defining $W_i = 1 - w_2 - \cdots - w_i$, resulting in a *general expression*

$$Sp(\pi) = \sum_{i\geq 1} W_i(\tilde{p}_i - \tilde{p}_{i+1})$$

and the *definite one*

$$Sp(\pi) = \sum_{i\geq 1} \omega^{i-1}(\tilde{p}_i - \tilde{p}_{i-1}).$$

They will play a very important role in the development of the *continuous* model of specificity.

6 Evidence specificity

Design of a specificity function can be also approached from the perspective of Dempster-Shafer theory. It is a very general framework for capturing numerically notions of evidence in support of assertions about the domain of discourse. The model we use applies to a finite domain of discourse X, where evidence is given as a mapping on a powerset of X

$$\mathbf{m} : \mathcal{P}(X) \to [0, 1].$$

Mapping $\mathbf{m}$ is called the basic assignment and the subsets $A \subset X$ where $\mathbf{m}(A) \neq 0$ are called focal; enumerating them $A_1, A_2, \ldots$ we put $m_i = \mathbf{m}(A_i)$. We require that $\sum m_i = 1$ and that the empty set is not focal, ie. $\mathbf{m}(\emptyset) = 0$.

For such structures several measures of *nonspecificity* have been proposed [15] among which

$$N(\mathbf{m}) = \sum m_i \log |A_i|$$

is usually preferred, being both additivity and subadditive [1, 11].

This model can be applied to fuzzy sets and possibility distributions. Given (p_i), hence also $(\tilde{p}_i)$, we construct a descending chain of $n = |X|$ subsets

$$A_1 = X \supset A_2 \supset \cdots \supset A_n.$$

We require that $\max_{x \in A_i} \pi(x) = \tilde{p}_i$, which determines subsets A_i uniquely if p_i are all different. Taking these A_i as focal sets with weights $m_i = \tilde{p}_i - \tilde{p}_{i+1}$ defines nonspecificity

$$U(\pi) = \sum (\tilde{p}_i - \tilde{p}_{i+1}) \log i.$$

We are interested in a *specificity* function, and an appropriate expression would be a complement of $U(\pi)$ with respect to the most nonspecific distribution $1^{(n)}$

$$I(\pi) = U(1^{(n)}) - U(\pi) = \log n - \sum(\tilde{p}_i - \tilde{p}_{i+1}) \log i.$$

Although we defined these functions starting from the evidence model, they both share the property of an extremely structured behavior under the *possibilistic operations.* They are both, as fuzzy operations, additive, symmetric, continuous and $U(\pi)$ is subadditive, while $I(\pi)$ is superadditive. These properties made them first recognized as uncertainty measures, but the earlier discussion suggests that $I(\pi)$ has merits of a specificity measure. We pursue this consideration in two ways. Here we show that $I(\pi)$ fits within the broad framework of $Sp(\pi)$. In the section on continuous distributions we show that $I(\pi)$ offers a natural model for the domains of finite measure ($Sp(\pi)$ works best for the infinite case).

Let us rewrite $I(\pi)$, remembering that $\tilde{p}_1 = 1$

$$I(\pi) = \tilde{p}_1 \log n - \sum_{i=2}^{n} \tilde{p}_i \log \frac{i}{i-1}.$$

Writing $w_1 = \log n,\ w_i \log \frac{i}{i-1},\ i = 2, \ldots, n$ we find that $w_1 > w_2 > \cdots > w_n$ and that $w_1 = \sum_2^n w_i$, in agreement with properties of $Sp(\pi)$. We shall not force the normalization $w_1 = 1$; it could be achieved by dividing by $\log n$, resulting in an amusing formula

$$\tilde{p}_1 - \sum_{i=2}^{n} \tilde{p}_i \log_n \frac{i}{i-1}.$$

However, we do not know of any natural interpretation of logarithms in base n; also, the original formula is better suited for continuous generalizations.

7 Concluding remarks

Our analysis can be extended to *pairs* of distributions defined on a common domain of discourse. A promising line of investgation would deal with extending the construction of information distances to, firstly, specificity measures and, secondly, to general OWA. The basic construction can be carried out quite simply and is described in our companion chapter and references therein [6]. Given two possibility assignments on the same domain $\pi : X \to [0,1]$ and $\varpi : X \to [0,1]$, such that for every $x \in X$ holds $\pi(x) \leq \varpi(x)$, define

$$G(\pi, \varpi) = U(\varpi) - U(\pi).$$

Here $U(\pi)$ and $U(\varpi)$ stand for U-uncertainty values. For general case of arbitrary π and ϖ we define

$$G(\pi, \varpi) = G(\pi, \pi \vee \varpi) + G(\varpi, \pi \vee \varpi).$$

This two-argument function $G(\pi, \varpi)$ becomes a *distance* on the space of all possibility assignments which are defined of the same basic domain X. We recall that U-uncertainty is itself an example of an OWA. It is now fairly straightforward defining similar distances, whether based on specificity or even on arbitrary OWA operators. Te more difficult part is verifying and proving the metric properties of these distances. For example, $G(\pi, \varpi)$ is a *metric*, but this property requires a nontrivial proof. A similar metric property can be expected for a large class of the proposed distances.

Another attractive direction is to analyze the use of generating functions in defining specificity and similar functionals. This is a large topic, and only the links with uncertainty measures have been investigated [7, 8]. In a different direction, we plan to adopt our first specificity measure $Sp(\pi)$ to a general Dempster-Shafer model. It would extend the relationship between the evidence nonspecificity $N(\mathbf{m})$ and fuzzy information $I(\pi)$.

References

[1] D. Dubois, H. Prade. *Properties of measures of information in evidence and possibility theories.* Fuzzy Sets and Systems 24(1988), 2:161–182.

[2] D. Dubois,H. Prade. *Possibility Theory.* Plenum Press, New York 1988.

[3] G. Hardy, J. Littlewood, G. Polya. *Inequalities, 2nd ed.* Cambridge University Press, Cambridge, 1964.

[4] M. Higashi. G. Klir. *On the notion of distance representing information closeness,* Int. J. General Systems, 9(1983), 2:103–115.

[5] M. Higashi, G. Klir. *Measures of uncertainty and information based on possibility distributions.* Int. J. General Systems, 9(1982), 1:43–58.

[6] A. Ramer. *Ordered continuous means and information.* Chapter in this book.

[7] A. Ramer. *Combinatorial interpretation of uncertainty and conditioning.* Chapter in: *Learning and Reasoning with Complex Representations,* ed: G. Antoniou, Lecture Notes in Computer Science, Springer-Verlag, New York 1997 (to appear).

[8] A. Ramer. *Uncertainty generating functions,* Chapter in: *Foundations and Applications of Possibility Theory,* ed: W. Chiang, J. Lee, World Scientific, Singapore, 1996.

[9] A. Ramer. *Structure of possibilistic information metrics and distances. Part 1: Properties.* Int. J. General Systems 17(1990), 1:21–32. Part 2: *Characterization. ibid* 18(1990), 1:1–10.

[10] A. Ramer. *Conditional possibility measures*, Cybernetics and Systems, 20(1989), 233–247. Reprinted in: *Readings in Fuzzy Sets for Intelligent Systems*, ed: D. Dubois, H. Prade, R. Yager. Morgan Kaufmann, San Mateo, CA 1993.

[11] A. Ramer. *Uniqueness of information measure in the theory of evidence.* Fuzzy Sets and Systems, 24(1987), 2:183–196.

[12] A. Ramer, L. Lander. *Classification of possibilistic measures of uncertainty and information.* Fuzzy Sets and Systems, 24(1987), 2:221–230.

[13] A. Ramer, C. Padet. *Possibility, probability, and specificity—decisions in a fuzzy world*, IFSICC'93 - Int. Fuzzy Systems and Intelligent Control Conf., Louisville, KY, March 1993.

[14] A. Ramer, R. Yager. *Specificity and information in fuzzy systems*, FUZZ-IEEE' 92, San Diego, CA, March 1992.

[15] R. Yager *Entropy and specificity in a mathematical theory of evidence.* Int. J. General Systems 9(1983).

[16] R. Yager. *On measuring specificity.* Tech. Rep., Iona College, New Rochelle, NY 1990.

ORDERED CONTINUOUS MEANS AND INFORMATION

Arthur Ramer

Knowledge Systems Group, Computer Science
University of New South Wales
Sydney 2052, Australia
E-mail: ramer@cse.unsw.edu.au

Abstract

Possibilistic measures of information have been characterized as a form of OWA operators on the possibility values. This correspondence can be carried on to the continuous domains and distributions.

Continuous information measures have been previously discussed only once in the open literature [12], while continuous OWA's were only hinted at in conference papers [17, 18]. This chapter presents a method of defining such functions on a continuous universe of discourse - a domain which is a measurable space of measure 1. The method is based on the concept of *rearangement* [5] of a function, used in lieu of sorting for the discrete possibility values.

For a continuous distribution, represented by a measurable function $f(x)$ on the domain of discourse X, first a decreasing *rearrangement*—$\tilde{f}(x)$ on $[0, \mu(X)]$—is constructed. Then, depending on $\mu(X)$, one of two definitions is appropriate

$\mu(\mathbf{X}) < \infty \quad I(f) = \int_0^{\mu(X)} \frac{1-\tilde{f}(x)}{x} dx$

$\mu(\mathbf{X}) = \infty \quad Sp(f) = k \int_0^{\infty} \tilde{f}(x) e^{-kx} dx$

For technical reasons the quantification of uncertainty must be in the form of information distance [6, 13] measuring the *departure* from the most 'uninformed' distribution (constant possibility 1). The final form of the information content for possibility distribution f, defined on

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

domain $X, \mu(X) = 1$, is given by the continuous OWA operator

$$I(f) = \int_0^1 \frac{1 - \tilde{f}(x)}{x} dx$$

Relationship with the discrete OWA's, and especially the discrete uncertainty measures, is discussed and various limit properties and approximations are established. Lastly, an investigation of continuous OWA as an *integral transform* is indicated.

KEYWORDS Function rearrangement, fuzzy set, fuzzy membership, information distance, information metric, information measure, information theory, ordered average, OWA, possibility distribution, weighted average.

1 Technical note

This chapter discusses basic concepts and motivations that led first to continuous fuzzy information functions and then, more generally, to the continuous OWA's. However, as the impetus was given by the study of quantitative uncertainty expressions suitable for possibilistic assignments [20], the presentation will commence with results and and examples in the information context. The general continuous OWA arose in the study of *specificity* of possibility distributions [17, 18]. Their general discussion is but a straighforward extension of that latter case. As all the motivating examples are based on the concept of quantitative specificity, we present all the OWA formulae only in this context.

While the discussion of the case $\mu(X) < \infty$ parallels that of the continuous uncertainty, the infinite measure domain of discourse $\mu(X) = \infty$ has to be given a seaprate treatment. Both cases are treated jointly in the second part of the chapter. The reader will notice a close relationship of the finite measure case and the information functions. The relevant formulae can be linked to Hilbert transform, while the infinite measure case has links to Laplace transform.

We state our main results as complete theorems, but we only present illustrative examples, rather than complete proofs. Most formulas we develop depend on some logarithmic function. In the discrete theory it is customary to use base 2 logarithm, as it corresponds to the usage in the Shannon's theory. Continuous theory is much better handled using natural logarithms, thus avoiding the nuisance of such constants as $\ln 2$. As all the information functions could be defined 'up to a multiplicative constant', we adopt this convention. We usually denote possibility values as $p_1 \leq \ldots \leq p_n = 1$ or, in a descending order, as $\tilde{p}_n = 1 \geq \ldots \geq \tilde{p}_1$. On occasion we assume tacitly that we have also defined $p_O = \tilde{p}_{n+l} = O$. We adopt Hardy *et al.* [5] style of writing summation symbols without explicit limits if the full range of sumation is meant. Such range is usually 1 to n, but it may become 1 to $n-1$ or 2 to n to avoid terms like $\log 0$ or $\frac{1}{0}$.

2 Introduction to uncertainty

A domain of discourse X becomes a fuzzy set whenever there is defined a fuzzy membership function $\pi : X \to [O, 1]$. Its values $\mu(x)$ are commonly interpreted as degrees of certainty or confidence or belief that element $x \in X$ has some property—equivalently, a degree of membership of x in a fuzzy set corresponding to that property. Zadeh [21] proposed that the values $\mu(x)$ be viewed as possibility assignments (possibilities for short)—numerical expressions of likelihood of some property being satisfied. He extended this measure of possibility to arbitrary subsets by putting $\pi(Y) = \sup_{x \in y} \pi(x)$. We can also interpret values $\pi(x)$ as likelihoods of occurrence of event x, given that domain X represents the set of possible outcomes. This perspective allows us to consider uncertainty associated with the domain and possibility defined thereon. Further, it leads to the definitions of possibilities of joint events and possibilities of groups of events (marginal possibilities). As implicit in Ramer [14], there is essentially a unique way of constructing these values, consistent with both combining and conditioning of possibilities. Given two domains X and Y and two independent possibility assignments $\pi_1 : X \to [0, 1], \pi_2 : Y \to [0, 1]$, we define a joint distribution $\pi_1 \otimes \pi_2$ as

$$\pi_1 \otimes \pi_2 : (x, y) \mapsto \min(\pi_1(x), \pi_2(y)).$$

Given an arbitrary assignment π on a product space $X \times Y$ we define the projected assignments π' on X and π'' on Y as

$$\begin{aligned} \pi'(x) &= \max_{y \in Y} \pi(x, y), \\ \pi''(x) &= \max_{x \in X} \pi(x, y). \end{aligned}$$

It is also convenient to define an *extension* of π from its domain X to larger set $Y \supset X$. We put

$$\begin{aligned} \pi^(y) &= \pi(y), \quad y \in X, \\ \pi^Y(y) &= 0, \qquad \text{otherwise}. \end{aligned}$$

Lastly, given a permutation s of $\{1, \ldots, n\}$ we can define a permuted possibility assignment $s(\pi)$ by putting

$$s(\pi)(x_i) = \pi(x_{s(i)}).$$

This is a standard structure of possibility assignments; by analogy with probability we often term π as *possibility distribution.* Although simple in structure, it already allows for introducing the concepts of information theory [7, 16]. We shall view the value $\pi(x)$ as expressing the degree of certainty of the event x (the possibility of x) and attempt to assign an overall value to the uncertainty inherent in the complete distribution π. We consider such (yet to be defined) *uncertainty value* as equivalent to the information that can be gained by selecting a specific event x from the total domain X.

We thus intend to define an information function I which would assign a nonnegative real value to an arbitrary distribution π. Following the established principles of information theory, we stipulate that such information function satisfies certain standard properties [1, 3]. Specifically, we require

- *additivity*
$$I(\pi_1 \otimes \pi_2) = I(\pi_1) + I(\pi_2)$$
- *subadditivity*
$$I(\pi) \leq I(\pi') + I(\pi'')$$
- *symmetry*
$$I(s(\pi)) = I(\pi)$$
- *expansibility*
$$I(\pi^Y) = I(\pi)$$
- *continuity*
$$\lim_{k\to\infty} I(\pi_k) = I(\pi), \quad \pi_k \to \pi$$

The subadditivity is taken with its sign reversed when $I(\pi)$ represents the information distance (from the uniform distribution) rather than the uncertainty value. The continuity, for the distributions π_k on a discrete domain, is meant as a coordinate-wise convergence to π on the same domain. For the continuous case, it is meant as function convergence; as the underlying domain is assumed compact, all the usual topologies coincide. For domain of infinite measures (which arise for general specificity) we assume the topology of uniform convergence of π_k.

Now our discussion must take into account the structure of the underlying domain X. The case of a finite domain of discourse is well understood and has been developed extensively [7, 9, 13, 16]. In the next section we summarize the results pertinent to the case of a finite set X; then the next two sections are devoted to the case of an infinite domain of discourse, represented by a space X endowed with a finite measure μ.

3 Information functions for finite domains

Given $\pi : X \to [0,1]$ let $\tilde{p}_1 \geq \ldots \geq_n$ be a descending rearrangement of the values $\pi(x), x \in X$. Then every information function (which satisfies *information conditions*) is of the form

$$I(\pi) = \sum (\tau(p_i) - \tau(p_{i+1})) \log i,$$

where τ is a nondecreasing mapping of [0, 1] onto itself [16]. If we also require that I satisfies some form of a linear property, like the branching condition [9], we arrive at a particularly simple expression, where τ is the identity function. This information function was first introduced by Higashsi and Klir [7] under the name of U-uncertainty

$$U(\pi) = \sum (p_i - p_{i+1}) \log i.$$

We note that this expression can be written in terms of the first backward differences of $\log i$

$$U(\pi) = \sum p_i \nabla \log i$$

where $\nabla \log i = \log i - \log(i-1)$. U-uncertainty serves to define an information distance [6, 13] between two distributions π and ρ defined on the same domain X. If $\pi(x) \leq \rho(x)$ for all $x \in X$, we put

$$g(\pi, \rho) = U(\rho) - U(\pi).$$

For the general case, given arbitrary π and ρ, we first define their lattice meet and join

$$\begin{aligned} \pi \wedge \rho \quad &: x \mapsto \quad \min(\pi(x), \rho(x)), \\ \pi \vee \rho \quad &: x \mapsto \quad \max(\pi(x), \rho(x)). \end{aligned}$$

We then put

$$\begin{aligned} G(\pi, \rho) &= g(\pi, \pi \vee \rho) + g(\rho, \pi \vee \rho), \\ H(\pi, \rho) &= g(\pi \wedge \rho, \pi) + g(\pi \wedge \rho, \rho). \end{aligned}$$

These functions have several attractive properties; for example G is a metric [6], while H is additive in both arguments [13].

4 Approches for the continuous universe of discourse

We shall attempt to extend previous definitions to arbitrary *continuous* domains. We develop our approach first for a special, albeit typical case where X is the unit interval. Now a fuzzy structure (possibility distribution) is given as a function

$$f : [0, 1] \rightarrow [0, 1]$$

such that $\sup_{x \in [0,1]} f(x) = 1$. Although in a variety of practical situations it is sufficient to consider only continuous functions, a much more general case of an arbitrary measurable function can be treated through the same means.

As the first step the discrete formula

$$U(\pi) = \sum p_i \nabla \log i$$

suggests forming an expression like

$$\int_0^1 \tilde{f}(x) d \ln x,$$

where $\tilde{f}$ would be some 'descending sorted' equivalent of f, while $d \ln x$ substitutes for $\nabla \log x$. The latter part is well known and simply represents $x^{-1}dx$; however, a proper definition of $\tilde{f}$ requires some work. We clearly want $\tilde{f}$ to be decreasing, or at least nonincreasing. We also would like it to 'stay' above any given value $\alpha, 0 \leq \alpha \leq 1$, over about the same space as the original function f does. This would assure us that all the α-cuts of f are of the same size (have the same measure) as the α-cuts of $\tilde{f}$. Such construction is well known in the literature; we follow here a classical description of Hardy *et al.* [5]. We define

$$L(y) = \mu(\{x : f(X) \geq y\}),$$

where is the standard Lebesgue measure (length) on $[0, 1]$. We then put $\tilde{f}(x) = L_{-1}(y)$. As an illustration let us consider a couple of examples.

Example 1

$$f(x) = \begin{cases} 2x, & \text{if } 0 \leq x \leq 0.5, \\ 2 - 2x, & \text{if } 0.5 \leq x \leq 1. \end{cases}$$

Then $L(y)$ represents the combined length of the intervals where $f(x) \geq y$. In our case

$$L(y) = 1 - y$$

and

$$\tilde{f}(x) = 1 - x.$$

It is immediate that $\tilde{f}(x) \geq \alpha$ over the set of the same measure as the set where $\tilde{f}(x) \geq \alpha$.

Example 2

$$f(x) = 3(x - \frac{1}{2})^2 = 4x^2 - 4x + 1.$$

Now $L(y) = 1 - \sqrt{y}$, as

$$\{x : f(x) \geq y\} = \left[0, \frac{1 - \sqrt{y}}{2}\right] \cup \left[\frac{1 + \sqrt{y}}{2}, 1\right].$$

Therefore $y = (1 - L(y))^2$ and $\tilde{f}(x) = (1 - x)^2$.

These and many other examples suggest strongly that using $\tilde{f}(x)dx$ to replace $p_i \nabla \log i$ would lead to very reasonable results. However, there remains one technical difficulty. The use of this definition would require considering

$$\int_0^1 \frac{\tilde{f}(x)}{x} dx$$

as the candidate expression for the value of information. Unfortunately, $\tilde{f}(x)$ is equal to 1 at 0, and the integral diverges. A solution can be found through a technique that has been used in probability theory. Shannon entropy [19]—a customary information function for discrete probability distributions, could be generalized to Boltzmann's entropy

$$-\int_0^1 f(x) log f(x) dx$$

for a continuous probability density $f(x)$. However the latter expression is not very satisfactory as an information function, as it cannot be obtained as a limit of the discrete approximations. The solution [3] is to use instead an information distance between the given density and the uniform one.

We shall use the same approach here. In possibility theory we consider a constant function $f(x) \equiv 1$ as representing a uniform distribution. This because it is the most 'uninformed' distribution—its discrete form clearly attains the maximum uncertainty value. Our final formula becomes

$$I(f) = \int_0^1 \frac{1 - \tilde{f}(x)}{x} dx.$$

This integral is well defined and avoids the annoying singularity at 0. We demonstrate its use on a class of polynomial functions.

Example 3
Let us consider possibility distributions represented by $f(x) = x^n, n = 0, 1, \ldots$ First a few complete integrations. Denoting $J_n = I(x^n)$ and remembering that $\widetilde{x^n} = (1-x)^n$, we find

$$J_0 = I(1) = \int_0^1 \frac{1-1}{x} dx = 0$$

$$J_1 = I(x) = \int_0^1 \frac{1-(1-x)}{x} dx = 1$$

$$J_2 = I(x^2) = \int_0^1 \frac{1-(1-x)^2}{x} dx = \int_0^1 (2-x) dx = 1\frac{1}{2}$$

To find a general expression for J_n, let us first compute $J_n - J_{n-1}$

$$\int_0^1 \frac{(1-(1-x)^n) - (1-(1-x)^{n-1})}{x} dx =$$
$$\int_0^1 \frac{(1-x)^{n-1} - (1-x)^n}{x} dx = \int_0^1 (1-x)^{n-1} dx = \frac{1}{n}$$

As $J_0 = 0$ we find that $J_n = 1 + \frac{1}{2} + \ldots + \frac{1}{n} = H_n$, the nth harmonic number.

5 Properties of continuous information

It requires precise proofs that $I(f)$ satisfies all the required properties of an information function. It is additive, subadditive, expansible, symmetric and continuous. Here we shall demonstrate the additivity using the example of a certain independent interest.

Example 4
Let $f(x) = x^2, g(y) = y^2$. We find, putting $h = f \otimes g, h(x,y) = \min(x^2, y^2)$, that

$$L(z) = \mu(\{(x,y) : h(x,y) \geq z\}) = \mu(\{(x,y) : x^2 \geq z, y^2 \geq z\}) = (1 - \sqrt{z})^2$$

and therefore

$$\tilde{h}(t) = L^{-l}(z) = (1 - \sqrt{t})^2.$$

Now

$$\int_0^1 \frac{1 - (1 - \sqrt{t})^2}{t} dt = \int_0^1 \left(\frac{2}{\sqrt{t}} - 1 \right) = -1 + 2 \int_0^1 t - \frac{1}{2} dt = 3.$$

This agrees with $I(f) = I(g) = 1\frac{1}{2}$.

Using$I(f)$ we can define continuous extensions of information distances g, G, and H. Specifically, if $f_1(x) \leq f_2(x), x \in [0,1]$

$$g(f_1, f_2) = \int_0^1 \frac{\tilde{f}_2(x) - \tilde{f}_1(x)}{x} dx.$$

For a general case we put

$$\begin{aligned} f_1 \wedge f_2 : x &\mapsto \min(f_1(x), f_2(x)), \\ f_1 \vee f_2 : x &\mapsto \max(f_1(x), f_2(x)), \end{aligned}$$

Now

$$G(f_1, f_2) = g(f_1, f_1 \vee f_2) + g(f_2, f_1 \vee f_2)$$

and

$$H(f_1, f_2) = g(f_1 \wedge f_2, f_1) + g(f_1 \wedge f_2, f_2).$$

As in the discrete case, G is a metric, while H is additive in both arguments.

Finally, we remark that $I(f)$ can be approximated as a limit of $U(\pi^{(n)})$, where $\pi^{(i)}$ form a sequence of discrete distributions approximating f. Already a nontrivial example is offered by a linear function.

Example 5
We select $f(x) = 1 - x$ and approximate it using the values at $\frac{1}{n}{}'\frac{2}{n}, \ldots, 1$. The

approximating distributions are $\pi^{(n)} = (\frac{n-1}{n}, \frac{n-2}{n}, \dots, 1)$ and have as their information measures

$$U(\pi^{(n)}) = \sum (p_i - p_{i+1} \ln i = \sum \left(\frac{n-i+1}{n} - \frac{n-i}{n}\right) \ln i = \frac{1}{n} \sum \ln i = \frac{1}{n} \ln n!$$

Invoking Stirling's formula, we find

$$\ln n! = n \ln n - n + O(\ln n),$$

and therefore

$$U(\pi^{(n)}) \sim \ln n - 1, \quad n \to \infty.$$

This gives

$$I(\pi^{(n)}) = \ln n - U(\pi^{(n)}) = 1$$

which agrees with $I(f)$.

6 Design of continuous specificity

For the definitions and terminology concerning the discrete specificity measures we refer to our companion chapter. Here we intend to replace the discrete distribution $\pi = (p_i)$ by a continuous function $f(x)$, the rearrangement $(\tilde{p}_i)$ by $\tilde{f}(i)$, and the coefficients w_i by a continuous 'coefficient' function fitted to their values. Then we replace the difference operator by a differential and summation by integration. A key step is defining the *coefficient function.* Let us analyze it in some detail. First, let us suppose that we have a finite sequence $w_1, \dots, w_n$. We can either interpret it as the values of a function $g(x)$ at points $\frac{1}{n}, \frac{2}{n}, \dots, 1$ or as the values of a function $h(x)$ at $1, 2, \dots, n$. In either case we would strive to have such putative function quite regular, thus presumably more useful for defining specificity operators.

We want $g(\frac{i}{n})$ to correspond to W_i and, perhaps more important, $g(\frac{i-1}{n}) - g(\frac{i}{n})$ to correspond to $w_i = W_{i-1} - W_i$. For the logarithmic coefficients it works well; we put $g(x) = \ln x$ and find $g(\frac{i}{n}) - g(\frac{i-1}{n}) = \ln \frac{i}{i-1}$ as required. For the exponential coefficients it is less successful. We want $h(x) = e^{-kx}$ for some constant k, so that $h(\frac{i}{n}) - h(\frac{i-1}{n}) = e^{-i} - e^{-(i-1)}$. Equivalently, we could posit $h(\frac{i}{n}) = e^{-i}$, suggesting $h(x) = e^{-nx}$. The first sign of difficulty is the dependence of $h(x)$ on n—actually we get a family of functions $h_n(x) = e^{-nx}$. A further difficulty lies in the limiting behavior of the coefficient function. As the interval $[0, 1]$ is subdivided into n parts, its end-point at 1 corresponds to the n-th position in the sequence of the discrete coefficients. Accordingly, we would like both $g(\frac{n}{n}) = g(1)$ and $h(\frac{n}{n}) = h(1)$ be approximately like $\ln \frac{n}{n-1}$ and e^{-n} as n goes to infinity. We have $g(1) = 0$, but $h(1)$ is a positive constant, again forcing the use of a sequence $h_n(x)$.

We now return to our main objective of correlating the discrete and the continuous distributions. We do it separately for the domains of finite measure and for the infinite domains.

First we suppose, without loss of generality, that our domain is $[0, 1]$ and the distribution is given by $f(x), 0 \leq f(x) \leq 1$. After the rearrangement we have $\tilde{f}(x)$ and can form specificity expressions

$$Sp_n(f) = 1 - \int_0^1 \tilde{f}(x)(-d(e^{-nx})) = 1 - n\int_0^1 \tilde{f}(x)e^{-nx}dx.$$

We can now define $Sp(f) = \lim_{n\to\infty} Sp_n(f)$.

For the logarithmic coefficients we define

$$I(f) = \int_0^1 (1 - \tilde{f}(x))(d\ln x) = \int_0^1 \frac{1-\tilde{f}(x)}{x}dx.$$

Now we can investigate the limiting behavior of discrete approximations. For $I(f)$ all is well, as the specificity values for the discrete distributions are simply Riemann sums for $I(f)$. For $Sp(f)$ there is nothing to approximate. We recall that we construct directly only the sequence $Sp_n(f)$ and are forced to define $Sp(f)$ as its limit. However, it is a very simple exercise to verify for every continuous $f(x)$

$$\lim_{n\to\infty} n\int_0^1 \tilde{f}(x)e^{-nx}dx = 1$$

resulting in $Sp(f) = 0$. This would result in viewing every continuous function (as well as most 'reasonable' measurable functions) as completely nonspecific, hence making such operator of little value.

Another form of approximation is through forming a distribution $f(x)$ over the infinite positive axis, taking care that $f(i) = p_i$ and, after the rearrangement, $\tilde{f}(i) = \tilde{p}_i$. Reasoning as before we derive two operators

$$Sp(f) = 1 - \int_0^\infty \tilde{f}(x)e^{-x}dx,$$

$$I(f) = \int_0^\infty \frac{1-\tilde{f}(x)}{x}dx.$$

(We observe that in the infinite case we can define $Sp(f)$ directly.)

Now the first operator gives useful values for most of the ordinary functions, eg. any polynomial over a finite interval $[0, a]$ and zero for $x > a$, or for polynomials in e^{-x}. The second operator, however, returns infinity as its value for all functions for which $\tilde{f}(x)$ is not identically 1. If for some $a > 0$, $\tilde{f}(a) = 1 - \varepsilon$, however small ε, then

$$I(\tilde{f}) \geq \int_a^\infty \frac{1-\tilde{f}(x)}{x}dx \geq \int_a^\infty \frac{1-(1-\varepsilon)}{x}dx$$

which diverges. Now the evidence based function assigns infinite specificity to practically every distribution.

7 Continuous specificity model

We propose a two-part structure, depending on the measure of the domain of discourse.

If $\mu(X)$ is finite we rearrange it to form $\tilde{f}(x)$ on $[0, \mu(X)]$. Then we propose as the basic measure

$$I(f) = \int_0^{\mu(X)} \frac{1-\tilde{f}(x)}{x} dx.$$

It is equivalent to the familiar integral over $[0,1]$ if we perform *scaling*, replacing $\tilde{f}(x)$ with $\tilde{f}'(x) = \tilde{f}(\mu(X)x)$ defined over $[0,1]$. We have

$$\begin{aligned} I(f) &= \int_0^{\mu(X)} \frac{1-\tilde{f}(x)}{x} dx \\ &= \int_0^1 \frac{1-\tilde{f}(\mu(X)x)}{\mu(X)x} d(\mu(X)x) \\ &= \int_0^1 \frac{1-f'(x)}{x} dx \end{aligned}$$

For X of infinite measure we propose using

$$Sp(f) = k\int_0^\infty \tilde{f}(x)e^{-kx} dx$$

or, in general

$$Sp(f) = \int_0^\infty \tilde{f}(x)W'(x)dx$$

for $W(x)$—a monotonically decreasing function satisfying

$$W(0) = 1, \quad \lim_{x\to\infty} W(x) = 0.$$

8 Further research

We presented a definite framework for constructing continuous OWA operators. It handles both the compact domains of discourse, with finite Lebesgue measure, and the domains of infinite measure. The resulting structures are different, as might be expected from the different nature of averaging over such domains.

Domains of the finite measure are the basis of continouos possibility assignments. For those assignments we gave a complete structure of information functions and distances. The same domains and assignments can be studied from the point of view of formal specificity. The use of the continuous OWA operators permits computing specificity values for all integrable, and in particular all continuous assignments.

For the domains of infinite measures our approach generalized the OWA of the infinite sequences. Again, the described framework handles all integrable possibility assignments.

We do not pursue the link of the continuous averaging operators we defined with some known, classical operators like Hilbert or Laplace transforms. This is a separate research topic, with several possible ramifications.

Another important direction is the derivation of OWA's as functionals on the space of 'information generating functions'. The concept of functions was introduced for probability (Shannon) entropy by Golomb [4] and for evidence nospecificity by Ramer [15]. These functions can be given a continuous form of integral transforms, applicable to possibility assignments [10, 11]. Their use permits applying analytic functions to reasoning about OWA properties.

We expect our result to find several applications, both in economic decision theory and in fuzzy control. In both these fields, great many problems are essentially continuous in nature, while often studied through the discrete approximations. A direct use of continuous quantitative methods should prove to advantage.

References

[1] J. Aczel, Z. Daroczy. *On Measures of Information and their Characterization.* Academic Press, New York, 1975.

[2] D. Dubois, H. Prade. *Possibility Theory.* Plenum Press, New York, 1988.

[3] S. Guiasu. *Information Theory and Applications.* McGraw-Hill, New York, 1976.

[4] S.W. Golomb, R.E. Peile, R.A. Scholtz. *Basic Cncepts in Information Theory and Coding.* Plenum Press, New York, 1994.

[5] G. Hardy, J. Littlewood, G. Polya. *Inequalities, 2nd ed.* Cambridge University Press, Cambridge, 1964.

[6] M. Higashi. G. Klir. *On the notion of distance representing information closeness,* Int. J. General Systems, 9(1983), 2:103–115.

[7] M. Higashi, G. Klir. *Measures of uncertainty and information based on possibility distributions.* Int. J. General Systems, 9(1982), 1:43–58.

[8] G. Klir, T. Folger. *Fuzzy Sets, Uncertainty and Information.* Prentice Hall, Englewood Cliffs, NJ, 1988.

[9] G. Klir, M. Mariano. *On the uniqueness of possibilistic measure of uncertainty and information.* Fuzzy Sets and Systems, 24(1987), 2:197-219.

[10] A. Ramer. *Combinatorial interpretation of uncertainty and conditioning.* Chapter in: *Learning and Reasoning with Complex Representations*, ed: G. Antoniou, Lecture Notes in Computer Science, Springer-Verlag, New York 1997 (to appear).

[11] A. Ramer. *Uncertainty generating functions*, Chapter in: *Foundations and Applications of Possibility Theory*, ed: W. Chiang, J. Lee, World Scientific, Singapore, 1996.

[12] A. Ramer. *Concepts of fuzzy information measures on continuous domains.* Int. J. General Systems 17(1990), 2&3:241–248.

[13] A. Ramer. *Structure of possibilistic information metrics and distances.* Part 1: *Properties.* Int. J. General Systems 17(1990), 1:21–32. Part 2: *Characterization. ibid* 18(1990), 1:1–10.

[14] A. Ramer. *Conditional possibility measures*, Cybernetics and Systems, 20(1989), 233–247. Reprinted in: *Readings in Fuzzy Sets for Intelligent Systems*, ed: D. Dubois, H. Prade, R. Yager. Morgan Kaufmann, San Mateo, CA 1993.

[15] A. Ramer. *Uniqueness of information measure in the theory of evidence.* Fuzzy Sets and Systems, 24(1987), 2:183–196.

[16] A. Ramer, L. Lander. *Classification of possibilistic measures of uncertainty and information.* Fuzzy Sets and Systems, 24(1987), 2:221–230.

[17] A. Ramer, C. Padet. *Possibility, probability, and specificity—decisions in a fuzzy world*, IFSICC'93 - Int. Fuzzy Systems and Intelligent Control Conf., Louisville, KY, March 1993.

[18] A. Ramer, R. Yager. *Specificity and information in fuzzy systems*, FUZZ-IEEE' 92, San Diego, CA, March 1992.

[19] C. Shannon. *The mathematical theory of communication.* Bell System Tech. J. 27(1948), 379–423, 623–656.

[20] R. Yager. *Aspects of possibilistic uncertainty.* Int. J. Man-Machine Studies, 12(1980), 283–298.

[21] L. Zadeh. *Fuzzy sets as a basis for a theory possibility.* Fuzzy Sets and Systems 1(1978), 3–28.

4

OWA OPERATORS IN DECISION ANALYSIS

OWA OPERATORS IN DECISION MAKING WITH UNCERTAINTY AND NONNUMERIC PAYOFFS

Ronald R. Yager
Machine Intelligence Institute.Iona College
New Rochelle, NY 10801, USA
E-mail Yager@Panix.Com

Maria Teresa Lamata
Departamento de Ciencias de la Computación e I.A
18071-Granada. España
e-mail mlamata@goliat.ugr.es

Abstract

We consider the problem of decision making in environments in which there exists some uncertainty about the state of nature. A general approach to the representation of uncertainty using the Dempster-Shafer belief structure is presented. A comprehensive methodology for evaluating the worth of each of the alternatives using the OWA operators to model the decision makers attitude is described. We then consider the situation in which the payoffs are nonnumeric values. Here we consider only the existence of a linear ordering on the allowable values. It is pointed out that in these nonnumeric environments a need arises for an operation to replace the weighted average. We show that in the case of only a linear ordering on the payoffs we can use a weighted median operation to replace the weighted average.

1 Introduction

Decision making in environments in which there exist some uncertainty with respect to a factor effecting the decision, generically called the state of nature, have been long considered in the literature [10] [13]. The methodologies for solving these problems have been strongly dependent upon the assumption of the type of uncertainty associated with the state of nature. Two assumptions about the knowledge of the uncertainty have dominated the literature.

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

The first case is where one assumes a probabilistic knowledge and the second is where one assumes a set of values which includes the actual value, however no probabilistic information is assumed. In the case of probabilistic knowledge considerable use is made of the expected value as a tool for obtaining the optimal alternative.

In the second case, sometimes called decision making under ignorance [13], the introduction of a decision making attitude, optimistic, pessimistic etc, is used to help provide an optimal answer. In [2,18] the authors suggested that by using the Dempster-Shafer belief structure we can provide a framework which unifies these two assumptions about our knowledge of the uncertainty and in addition provides a structure for representing other more sophisticated types of information about the uncertainty in the environment. Strat [17] and Nguyen and Walker [12] have also investigated decision making in the Dempster-Shafer environment. In [18] it was also suggested that with the aid of the OWA operators we can provide a unification of types of decision making attitude used in solving these problems.

In [18] it was assumed that the payoffs associated with the problem were numeric values. In many real problems the requirement of specific numeric values for the payoffs is not realistic. In this work we consider a relaxation of this restriction by allowing the nonnumeric values for the payoffs such as those carried by linguistic values [24],*big, small, about thirty thousand.* Considerations of nonnumeric values in decision making have also been considered in other works [20]. The relaxation of this requirements. however, comes at the price of not allowing the use of some operations, particularly the weighted average, needed to implement the process of determining the optimal solution.

In the following we suggest two operations which can be used in place of the weighted average in the decision making technique. The first operation, called the weighted median [21,22], only requires that the payoffs can be ordered. the second operation, called the weighted average on a uniform scale based on work in [4], requires a slightly more sophisticate scale but still doesn't require numbers.

2 Decision Under Various Kinds of Uncertainty

A classic paradigm of decision making can be captured by the following matrix. In this matrix the A_i indicate alternative actions that are possible for a decision maker to take. The S_j indicate possible states of nature and the r_{ij} indicate the payoff to the decision maker for selecting alternative i when the state of nature is S_j. The problem of concern is to select one of the alternatives as being the optimal. The solution to this problem can be decomposed into a two step.

Algorithm

1. For each alternative evaluate the effective value of selecting that alternative, we shall denote this as V_i.

2. Select the alternative with the highest V_i value.

	S_1	S_2	$\cdots$	S_j	$\cdots$	S_n
A_1	r_{11}	r_{12}	$\cdots$	r_{1j}	$\cdots$	r_{1n}
A_2	r_{21}	r_{22}	$\cdots$	r_{2j}	$\cdots$	r_{2n}
				$\vdots$		
A_i	r_{i1}	r_{i2}	$\cdots$	r_{ij}	$\cdots$	r_{in}
				$\vdots$		
A_q	r_{q1}	r_{q2}	$\cdots$	r_{qj}	$\cdots$	r_{qn}

The interest and difficulty associated with this problem arises because in many cases we don't know the state of nature and thus the calculation of the valuation of each alternative, V_i, is not straight forward. In the classic literature three different situations with respect to our knowledge of state of nature have been considered and means for finding V_i in these cases have been suggested.

1. **Decision Making Under Certainty** In this case the state of nature is exactly known. If we know that the state of nature is S_j, for example, we than simple calculate
$$V_i = r_{ij}$$

2. **Decision Making Under Risk** In this case it is assumed that we have a probability distribution on the set $S = \{S_1, ..., S_n\}$ such that P_j is the probability associated with state S_j. The valuation of any alternative A_i is obtained by taking the expected value, hence
$$V_i = \sum\nolimits_{j=1}^{n} P_j r_{ij}$$
We should note that the decision making under certainty can be considered as a special case in which $P_j = 1$ for alternative S_j.

3. **Decision Making Under Ignorance** In this case we have no information about the state of nature except that its some element in the set S. In this environment the decision maker replaces knowledge about the state of nature by assuming some particular decision making attitude. Among the attitudes considered in the literature are the following:

   - *Pessimistic Attitude* Using this attitude the decision maker selects as their valuation for an alternative the worst possible outcome under that alternative.
   - *Optimistic Attitude* Under this strategy the decision maker selects as their valuation for an alternative the best possible outcome under that alternative.
   - *Hurwicz Approach* In this approach the decision maker selects some value degree of optimism, $\alpha \in [0, 1]$, and then for each alternative

calculates V_i as the weighted average of the pessimistic and optimistic value,

$$V_i = \alpha \times Opt + (1-\alpha) \times Pess$$

- *Normative Approach* In this approach the decision maker uses average of the possible payoffs under an alternative,

$$V_i = 1/n \sum_{j=1}^{n} r_{ij}$$

More generally we can express the strategy for decision making under ignorance as

$$V_i = F(r_{i1}, r_{i2}, .., r_{im})$$

where the function F depends upon the attitude selected.

In [2,18] the authors extended and generalized the possible representation for our knowledge about the state of nature with the use of the Dempster-Shafer belief structure [5,15]. Assume S is a set of elements, in this environment the set of possible states of nature, a belief structure has associated with it a collection of non-null subsets of S, $B_1, B_2, ..., B_q$, called the focal elements and a function m called the probability density function, defined on the collection of focal elements such that

1. $m(B_i) > 0$
2. $\sum_i m(B_i) = 1$

Essentially $m(B_i)$ is a probability associated with each of focal elements. Using the concept of a belief structure the classic situations with respect to our knowledge about the state of nature can be easily represented in this framework.

1. **Certainty about the state of nature** :
Assume we know the state of nature is S_j. In this case we have just one focal element
$$B_1 = \{S_j\}$$
and
$$m(B_1) = 1$$

2. **Probabilistic knowledge about the state of nature**:
Assume we have P_j is the probability of S_j. In this case we have n focal elements
$$B_j = \{S_j\} \qquad j = 1, .., n$$
and
$$m(B_j) = P_j$$

3. **Ignorance about the state of nature**:
   Here we have one focal element

$$B = S$$

   where

$$m(B) = 1$$

In addition to unifying our representation regarding our knowledge of the states the use of belief structures allows for the representation of many other different kinds of knowledge about the state of nature. In [18] an approach to finding the valuation for a n alternative under the situation in which our knowledge about the state of nature is known in terms of a belief structure was suggested. Assume our knowledge about the state of nature is captured be the belief structure with focal elements $B_1, B_2, .., B_q$ and probability density function m. To obtain the valuation of alternative A_i we proceed as follows:

1. For each focal element obtain the collection of associated payoffs, denote these as E_k, $k = 1, .., q$.

2. Using a selected decision making attitude find the evaluation of each of the E_k, we shall denote these as $V_i(E_k)$

3. Calculate overall evaluation associated with A_i, V_i, as

$$V_i = \sum_{k=1}^{q} m(B_k) \times V_i(E_k)$$

Example: Consider the following abbreviated decision making matrix

	S_1	S_2	S_3	S_4	S_5
A_1	10	15	25	6	13

Assume our knowledge about the state of nature is

$$\begin{aligned} B_1 &= \{S_1, S_2, S_3\} \\ B_2 &= \{S_5\} \\ B_3 &= \{S_1, S_2, S_3, S_4, S_5\} \end{aligned}$$

where

$$m(B_1) = 0.5 \qquad m(B_2) = 0.3 \qquad m(B_3) = 0.2$$

Furthermore assume our decision attitude is a pessimistic one. We first calculate the E_k,

$$\begin{aligned} E_1 &= \{10, 15, 25\} \\ E_2 &= \{13\} \\ E_3 &= \{10, 15, 25, 6, 13\} \end{aligned}$$

Using the pessimistic attitude we calculate

$$V_1(E_k) = \text{Min}(E_k)$$

hence

$$V_1(E_1) = 10, \qquad V_1(E_2) = 13, \qquad V_1(E_3) = 6$$

Finally our overall valuation for this alternative is

$$V_1 = (0.5) \times 10 + (0.3) \times 13 + (0.2) \times 6 = 10.1$$

In the above we have suggested a very general approach to decision making under uncertainty. This approach, with the aid of the belief structure, allows for the inclusion very sophisticated information about the knowledge of the state of nature

3 Generalization of Attitudinal Evaluation

In the preceding when face with a collection $E = \{r_1, .., r_n\}$ of possible payoffs and no information about the probabilities associated with these we selected a decision making attitude and used this to obtain the valuation of E. The decision making attitude was selected from among the four previously describe attitudes. In [20] Yager suggested a generalization of this process. This generalization is based upon the use of the ordinal weighted averaging (OWA) operators introduced in [19]. In addition to providing a generalization of the process of decision making under ignorance this extension will provide a very useful semantic for the problem. We first introduce the OWA operator.

Definition (Yager) An ordered weighted averaging operator (OWA) of dimension n is a function

$$F : R^n \to R$$

that has associated with it a weighting vector W

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

or

$$W = [w_1, w_2, ..., w_n]^T$$

such that

1. $w_i \in [0, 1]$
2. $\sum_i w_i = 1$

where for any set of values $a_1, .., a_n$

$$F(a_1, .., a_n) = \sum_i w_i \star b_i$$

where b_i is the i-largest of the $a_1, .., a_n$.

Using the OWA operator with the appropriate choice of weighting vector we are able to obtain the valuation method used in the classic decision making attitudes. Furthermore this approach provides a large collection of other decision making attitudes.

- Pessimistic Attitude: If we select w where

  $$w_* = [0, 0, ..., 1]^T$$

  then $F(a_1, a_2, .., a_m) = Min_j[a_j] = u_{*j} \in U$, which is the aggregation rule used in the pessimistic strategy.

- Optimistic Attitude: If we select w^* where

  $$w^* = [1, 0, ..., 0]^T$$

  then $F(a_1, a_2, .., a_m) = Max_j[a_j] = l_j^* \in U$, which is what is used in the optimistic strategy.

- Hurwicz Strategy: If we select

  $$w_H = [\alpha, 0, ..., 0, (1 - \alpha)]^T$$

  then $F(a_1, a_2, .., a_m) = \alpha \times Max[a_j] + (1 - \alpha) \times Min[a_j] = \alpha \times l_j^* + (1 - \alpha) \times u_{*j} = u_j \in U$. This operation is the same that in subsection 4.2. This is the formulation used in the Hurwicz strategy.

Thus with the use of the OWA operators to aggregate the payoffs for a given alternative we are able to implement different types of decision making attitudes by the appropriate selection of the weighting vector W. It should be noted that if most of the weights are near the top of the vector we are essentially using an optimistic approach while if most of the weights are near the bottom we are using a pessimistic approach.

A view of this OWA approach can be introduced which somewhat reconciles the decision making under ignorance with that of decision making under uncertainty. In this view we can interpret the OWA weights as a kind of probability, we note the OWA weights have the properties of a probability distribution; they are in the unit interval and sum to one. In particular we can view w_j as the probability that the j^{th} best outcome will happen. In this spirit the OWA aggregation

$$F(a_1, a_2, ..., a_n) = \sum w_j b_j$$

can be seen as an expected value with this kind of probability distribution. Thus the OWA vector, the decision making attitude, can be seen as a probability distribution of this kind.

The use of the OWA approach raises the question of how we can select the weighting vector. A simple approach which requires only the selection of one parameter, the decision makers degree of optimism. This approach can be seen as a generalized Hurwicz strategy. In this approach we first select a degree of optimism β and use this to obtain the OWA weights. In particular we select $[w_1.w_2, ..., w_n]$ by solving a mathematical programming problem proposed by O'Hagan [11].

Maximize

$$\sum_j w_j ln(w_j) \qquad \text{(entropy)}$$

Subject to:

$$\sum_j (\frac{n-j}{n-1} \times w_j) = \beta \qquad \text{(degree of optimism)}$$

$$\sum_j w_j = 1$$

$$w_j \geq 0 \quad \text{for all} \quad j = 1, .., m$$

The basic approach using this OWA method can then be summarized as follows

1. Obtain the weighting vector W corresponding to the decision makers attitude.
2. For each alternative calculate $F(a_1, a_2, ..., a_n)$.
3. Select as optimum that achieves the maximun valuation.

- Normative Approach: If we select

$$w_N = [1/n, 1/n, ..., 1/n]^T$$

then $F(a_1, a_2, .., a_m) = 1/n \sum_j a_j = u_j \in U$. This function is essentially the normative strategy.

4 Decision with Ordinal Payoffs

In the preceding we have suggested a very general approach to decision making under uncertainty. This approach allows the representation of our knowledge about the state of nature in terms of a belief structure and uses the

OWA operator to find valuations of collections of potential payoffs. Let us now summarize this method.

Assume we have a belief structure on the state of nature with focal elements $B_1, B_2, ..., B_q$ and m. Let α be our degree of optimism. In order to find the valuation of a given alternative A_i, we proceed as follows:

1. For each B_k calculate the associated possible payoffs under alternative A_i denote these as $E_k, k = 1, 2, .., q$.

2. For each E_k, using α and the method suggested by O'Hagan calculate an associated weighting vector W_k. (We note the dimension of W_k depends upon the dimension of E_k.)

3. For each E_k calculate

$$V_i(E_k) = F_{W_k}(E_k) = \sum_{j=1}^{Dim(k)} w_{jk} \times b_{jk}$$

b_{jk} is the j^{th} largest element in E_k and $Dim(k)$ is the dimension of W_k.

4. Calculate

$$V_i = \sum_{k=1}^{q} m(B_k) \times V_i(E_k)$$

In the preceding we have implicitly assumed that the payoff, the r_{ij} 's, are numeric values. One can easily envision situations in which there may exist some difficulty in obtaining such crisp information regarding these payoffs. For example, we may have payoff values such as: *very high, low, about \$30,000, etc.* These can be seen as kinds of linguistic payoffs. In this environment, at best, one can associated with such payoff values a linear ordering indicating the decision makers preference among the various potential payoffs. Thus, in the following we shall assume that the payoff values are drawn from the following linear scale,

$$L = \{L_1, L_2, .., L_m\}$$

such that $L_i < L_j$ if $i < j$, hence $C_{ij} \in L$.

In attempting to use our general approach to finding the valuation of alternatives in this environment we are faced with a problem in attempting to implement the steps (3) and (4) of the last algorithm. As we see these operations, which are essentially weighted averages, require that we perform operations not available to us in this environment where the payoffs are drawn from an ordinal scale. For example, in trying to implement (3) even though we have numeric values for the weights, the w_{jk}, the b_k are just drawn form L and thus we can't implement this operation.

Two approaches exist for extending our decision making methodology to this less precise environment [22]. The first approach is based upon the use of the weighted median operator introduced in [21]. As we shall see this approach only requires the payoffs be drawn from a scale having a linear ordering. The second approach is based upon the work of Delgado, Verdegay and Vila [4]. This second approach requires a slightly more sophisticated scale but still doesn't requires crisp numeric values. In this second approach we use a uniform scale which is slightly stronger than a simple linear ordering.

This approach requires the use of sources vault from preference theory [4]. In the following we shall only discuss the first approach more details about the second can be found in [22].

5 Weighted Median Aggregation

Definition.The weighted median aggregation denoted by WM is defined as

$$WM[(\omega_1, x_1), (\omega_2, x_2), ..., (\omega_n, x_n)]$$

with $\sum_{i=1}^{n} w_i = 1$
as the solution of the following algorithm

1. Reindex the arguments so that (y_j) corresponds to the argument which has j^{th} largest value, y_j is the j^{th} largest of the x_i.
2. Consider the reindexed pairs (u_j, y_j) where u_j is the weighted associate with the x_i that is the j^{th} largest.
3. Let
$$T_k = \sum_{j=1}^{k} u_j$$
the sum of the weights associated with k largest arguments.
4. The weighted median is the value y_{k^*} such that
$$T_{k^*-1} < 0.5$$
$$T_{k^*} \geq 0.5$$

As shown in [21] while this aggregation approach only requires a linear ordering on the values it has many of the properties associated with weighted average; monotonicity, commutativity and idempotency.

Example: Consider the following decision problem where we assume that the payoffs are drawn from the following scale:

$$L = \{Ex.Low(EL), VeryLow(VL), Low(L), Medium(M),$$
$$High(H), VeryHigh(VH), Ex.High(EH)\}$$

and the payoffs matrix is:

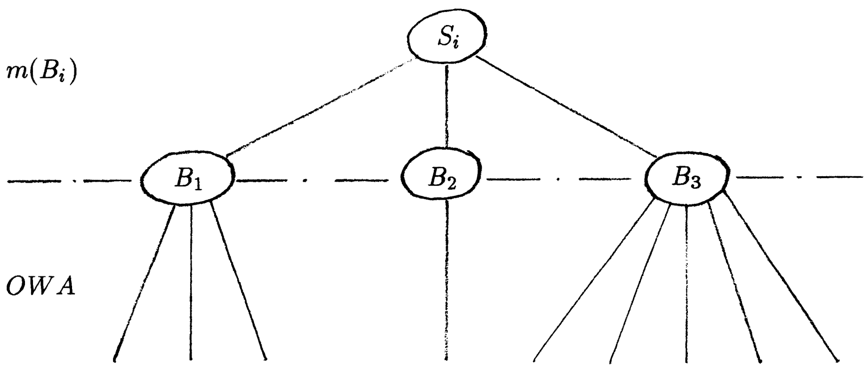
	s_1	s_2	s_3	s_4	s_5
A_1	VL	VH	EH	EL	L
A_2	M	L	VH	M	EH

Assume that our knowledge of the state of nature consists of the following belief structure m.

Focal elements	*Weights*
$B_1 = \{s_1, s_2, s_3\}$	0.4
$B_2 = \{s_5\}$	0.4
$B_3 = \{s_1, s_2, s_3, s_4, s_5\}$	0.2

We assume the degree of optimism the decision-maker is $\alpha = 0.25$

The representation of this example by means of a hierarchical structure is:



This tree has two levels the first corresponding to the objective information $m(B_i)$ about the uncertainty and the second one to the subjective information (the OWA operator) about the decision attitude.

Two ways of calculating the valuation of an alternative emerge in this situation. In one we first calculate using the appropriate OWA vector the valuation of each of the focal elements. This is done using the weighted median. We next using these valuations and the focal element weights calculate the overall evaluation of this alternative. This again requires the use of the weighted median.

The second approach takes advantage of our view of the OWA weights as a kind of probability. In this approach we assign to each payoff appearing in a focal element a probability equal to the product of the associated OWA weight and the focal element weight. We then use the weighted median to calculate the valuation the alternative.

These two approaches are the same when we work with numeric payoffs instead of linguistic and the operator that we use is the weighted average (expected value). This is due to the fact that the mean of the means is equal to the total mean. In the case of the weighted mean, as we shall see in the following example, this is not the case

5.1 First Approach

We first consider the alternative A_1.

1. For each focal element B_i we calculate the possible associated payoffs.

$$\begin{aligned} P_1 &= (EH, VH, VL) \\ P_2 &= (L) \\ P_3 &= (EH, VH, L, VL, EL) \end{aligned}$$

2. For $\alpha = 0.25$ the solution to the MP problem for $n = 1, 3, 5$ are

$$\begin{aligned} w_1 &= [1] \\ w_3 &= [0.11, 0.27, 0.62]^T \\ w_5 &= [0.05, 0.08, 0.15, 0.26, 0.46]^T \end{aligned}$$

3. To calculate the effective value of each P_i we calculate

$$\begin{aligned} V_1(P_1) &= WM_1[(0.11, EH)(0.27, VH)(0.62, VL)] = VL \\ V_2(P_2) &= WM_1[(1, L)] = L \\ V_3(P_3) &= WM_1[(0.05, EH)(0.08, VH)(0.15, L)(0.26, VL)(.46, EL)] = VL \end{aligned}$$

4. Next we calculate the overall valuation of alternative A_1

$$V_1 = WM_1[(m(B_i), V_1(P_i)] =$$

$$= WM_1[(0.4, VL)(0.4, L)(0.2, VL)] = VL$$

In the same way, we calculate the WM for A_2, and we obtain

$$V_2 = WM_1[(0.4, L)(0.4, EH)(0.2, M)] = M$$

thus we see A_2 is preferred to A_1.

5.2 Second Approach

In the second approach the steps one and two are the same as in the preceding approach but it differs in the next steps. Assuming p_{ik} is the k^{th} largest element in the payoff vector corresponding to the i^{th} focal element and w_{ik} is the associated OWA weight we calculate the probability

$$d_{ik} = w_{ik} \times m(B_i)$$

We then obtain a weighted median over all the pairs (d_{ik}, p_{ik}).

Example We consider the same data as in the previous example.
Consider the alternative A_1.
1 and 2 step are the same.
3.- for $\alpha = 0.25$

$m(B_i)$	k	W_k	P_k
0.4	3	$[0.11, 0.27, 0.62]$	$[0.044, 0.108, 0.248]$
0.4	1	$[1]$	$[0.4]$
0.2	5	$[0.05, 0.08, 0.15, 0.26, 0.46]$	$[0.01, 0.016, 0.03, 0.052, 0.092]$

4.- Next we calculate the overall valuation of alternative A_1.

$$V_1 = WM_2[(0.044, EH)(0.108, VH)(0.248, VL)(0.4, L)(0.01, EH)(0.016, VH)$$
$$(0.03, L)(0.052, VL)(0.092, EL)]$$
$$= WM_2[(0.092, EL)(0.3, VL)(0.43, L)(0.124, VH)(0.054, EH)] = L$$

We see that the result is different from the first approach, we obtain a label less extreme.

Let us now consider A_2 .
1, 2 and 3 are the same
4.-We calculate the overall valuation of alternative A_2 .

$$V_2 = WM_2[(0.044, VH)(0.108, M)(0.248, L)(0.4, EH)(0.01, EH)(0.016, VH)$$
$$(0.03, M)(0.052, M)(0.092, L)]$$
$$= WM_2[(0.41, EH)(0.06, VH)(0.19, M)(0.34, L)] = M$$

Thus we see A_2 is preferred to A_1.

5.3 Differences between Approaches

The principal difference between this two approaches is that with the second option the result will be more near that the hypothetical expected value than the first option.

To proof this assertion we need take into account the concept of entropy,

$$H(1) = -\sum_{i=1}^{r} m(B_i) \log m(B_i)$$

where r is the number of focal elements, and $m(B_i) = p_j$ is the probability associated with only one $E_j \subset B_i$.

$$H(2) = -\sum_{i=1}^{r}\sum_{k=1}^{t} m(B_i) \times \mathrm{OWA}_k^i \log(m(B_i) \times \mathrm{OWA}_k^i)$$

where OWA_k^i means a vector of dimension $k = |B_i|$ associated for each i to B_i and $\sum_{k=1}^{t} \mathrm{OWA}_k^i = 1$ by definition of OWA operator.
Then

$$H(2) = -\sum_i m(B_i) \log(m(B_i) \times \sum_k \mathrm{OWA}_k^i) - \sum_i \sum_k m(B_i) \times \mathrm{OWA}_k^i \log \mathrm{OWA}_k^i$$

$$= H(1) - \sum_i \sum_k m(B_i) \mathrm{OWA}_k^i \log \mathrm{OWA}_k^i$$

and

$$H(1) \leq H(2) \leq \log n$$

We know from information theory that if the entropy is near to the maximun value then the distribution of probability is near to the uniform distribution, for which mean and median have close values.

In the above we have introduced two approaches to solving decision problems in the payoffs are drawn from a scale which has only the structure of a linear ordering. The key to handling this type of problem is the use of the weighted median as the primary aggregation operation.

Acknowledgements

This work is partiality supported by the DGICYT under project PB95-1181.

6 Conclusion

In the preceding we have looked at the problem of decision making in environments in which there exist some uncertainty about the state of nature.

We discussed a general approach to the representation of this uncertainty using the Dempster-Shafer belief structure. A general methodology for representing the decision makers attitude was described using OWA operators. We presented a general methodology for evaluating the worth of each of the alternatives, central to this approach was the use of a weighted average operation. We next considered the case of nonnumeric payoffs. It was noted that in these nonnumeric environments a need arises for a operation to replace the weighted average. We showed that in the case of only a linear ordering on the payoffs we can use a weighted median operation to replace the weighted average.

References

[1] R.Bellman and L.A.Zadeh. "Decision making in a fuzzy environment". *Management Science*, 1970; 17: 141-164.

[2] M.J.Bolaños, M.T.Lamata and S.Moral."Decision making problems in a general environment". *Fuzzy Sets and Systems*, 1988; 18: 135-144.

[3] P.P.Bonissone. "A fuzzy sets based linguistic approach". Approximate Reasoning in *Decision Analysis,* M.M.Gupta, E.Sanchez (eds), North Holland Publishing Company, 1982.

[4] M.Delgado, J.L.Verdegay and A.Vila. "On aggregation operations of linguistic label". *International Journal of Intelligent Systems* 1993; 8: 351-370.

[5] A.P.Dempster. "A generalization of Bayesian inference", *Journal of the Royal Statistical Society,*1968; 205-247.

[6] P.C.Fishburn. *"Utility theory for decision making"*. John Wiley , 1970.

[7] M.T.Lamata. "Problemas de decisión con información general". Tesis doctoral. Universidad de Granada,1986.

[8] M.T.Lamata and S.Moral. "Decisión; Utilidades lingüísticas". III Congreso Español sobre tecnologias y lógica fuzzy, 1993: 321-328.

[9] M.T.Lamata and S.Moral. "Calculus with linguistic probabilities and belief". *Advances in the Dempster-Shafer theory of Evidence.* (M.Fedrizzi, J.Kacprzyk, R.R.Yager, eds). John Wiley , 1994.

[10] R.D.Luce and H.Raiffa. *"Games and Decisions: Introduction and Critical Survey.* John Wiley and Sons: New York, 1967.

[11] M.O'Hagan, "Aggregating template rule antecedents in real-time expert systems with fuzzy set logic". *Proceeding 22nd Annual IEEE* Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, Ca., 681-689, 1988.

[12] H.T.Nguyen and E.A.Walker. "On decision making using belief functions", *Advances in the Dempster-Shafer theory of Evidence.* (M.Fedrizzi, J.Kacprzyk, R.R.Yager, eds). John Wiley , 1994.

[13] S.B.Richmond.*Operations Research for Management Decisions*, Ronald Press:New York.

[14] M.Roubens and P.Vincke, (1989) *Preference Modeling.* Springer-Verlag:Berlin, 1968.

[15] G. Shafer. *"A Mathematical Theory of Evidence"*, Princeton University Press:Princeton, N.J, 1976.

[16] J.Q.Smith. *Decision Analysis. A Bayesian Approach*, 1988.

[17] T.M.Strat. "Decision analysis using belief functions", *Advances in the Dempster-Shafer theory of Evidence.* (M.Fedrizzi, J.Kacprzyk, R.R.Yager, eds). John Wiley , 1994.

[18] R.R.Yager. "Decision making under Dempster-Shafer uncertainties". *Int.J.of Int. Sys*, 1992; 20: 233-245.

[19] R.R.Yager. "Families of OWA operators". *Fuzzy Sets and Systems*, 1993; 59: 125-148.

[20] R.R.Yager . " An approach to ordinal decision making" *International Journal of Approximate Reasoning*, 1995; 12: 237-261.

[21] R.R.Yager. "Fusion of ordinal information using weighted median aggregation", *Technical Report #MII-1520 Machine Intelligence Institute, Iona College*, New Rochelle, NY, 1995.

[22] R.R.Yager and M.T.Lamata, " Decision Making Under Uncertainty with Nonnumeric Payoffs" *Submitted to Int. J of I. Syst*

[23] L.A.Zadeh. "The concept of a linguistic variable and its application to approximate reasoning: Part 1", *Information Sciences* , 1975; 8: 199-249.

[24] L.A.Zadeh. "Fuzzy sets and information granularity". *Advances in fuzzy sets theory and applications*, M.M.Gupta, E. Sanchez (eds), North Holland Publishing Company, New York, 1979; 3-19.

ON THE ROLE OF IMMEDIATE PROBABILITY IN VARIOUS DECISION MAKING MODELS

Kurt J. Engemann and Ronald R. Yager
Iona College
New Rochelle, NY 10801 USA

ABSTRACT

Recently, new approaches have been presented which contribute to decision analysis modeling. These methods provide a means of handling various ways of representing events, decision makers' attitudes and payoffs. We will discuss how the new approaches may be used with each other as well as in conjunction with traditional methods. The role of immediate probability in various decision making models will be highlighted.

KEYWORDS

Ordered weighted averaging operator, immediate probability, weighted median, belief function, utility.

INTRODUCTION

A decision maker frequently encounters situations requiring the selection of one of several alternatives. In such cases, decision models may be used. The model to be used to aid in the selection of the best alternative depends upon the nature of the information describing the situation. The method should incorporate the decision maker's degree of optimism regarding the state of nature and his utility for the payoffs. The payoffs may be on ordinal or numeric scales. The decision maker's knowledge of the state of nature may be that of uncertainty, risk, or in the form of a

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

belief function. Immediate probability may be effectively used in a variety of decision models. We will review recent approaches and suggest how these methodologies may be used in coordination with one another.

THE DECISION MAKER'S ATTITUDE AND IMMEDIATE PROBABILITY

A key ingredient of a decision is the attitude of the decision maker. The decision maker may be an individual or an organization. Risk perception plays an integral role in the decision and is a function of the payoffs and their likelihood [Mellers and Chang 1994]. March and Shapira [1987] suggest that the payoff magnitude is the more prominent of the two components. Some researchers report that as the level of perceived risk increases, a person is less likely to engage in risk-taking behavior [Dunegan, Duchon and Barton 1992] [Straw, Sandelands and Dutton 1981]. On the other hand, Kahneman and Tversky [1979] indicate that some decision makers who perceive high levels of risk respond with risk-seeking behavior. Risk propensity is the consistency of a decision maker to either take or avoid actions that he perceives as risky [Sitkin and Pablo 1992]. An individual with low risk-taking propensity will typically weigh negative outcomes more highly [Schneider and Loppes 1986]. An individual with high risk-taking propensity is more likely to weigh positive outcomes more highly, thereby overestimating the probability of a gain relative to the probability of a loss [Brockhaus 1980] [Vlek and Stallen 1980].

Let A_i be an alternative available to the decision maker, S_j a possible state of nature, and C_{ij} the payoff to be received by the decision maker if he selects action A_i and the state of nature is S_j. The process of decision making under risk is characterized by knowledge of the probabilities, p_j, of the states of nature. The standard procedure in this case is to calculate the expected payoff for each alternative and to select the alternative that has the optimal expectation.

Decision makers who know the probabilities of the states of nature may not always use this expected value approach exactly. The expected value is a long run concept and the decision maker may be faced with an immediate nonrecurring situation. In a hybrid model of decision making under risk [Yager, Engemann and Filev 1995] [Engemann, Filev and Yager 1996], the decision maker has the flexibility of modifying the event probabilities to reflect his degree of optimism. This is somewhat analogous to a decision maker modifying monetary payoffs through a utility curve which reflects his degree of risk aversion. However, the two approaches are not the same and may be used separately or in conjunction with each other.

The decision maker's optimism modifies the event probabilities in the following way. The decision maker's optimism is represented by an ordered weighted average

(OWA) operator with weights w_1, w_2, ..., w_n, where $\sum_{j=1}^{n} w_j = 1$, $w_j \geq 0$ for all j.
The degree of optimism is defined as Ω, ($1 \geq \Omega \geq 0$), where the OWA weights satisfy:

$$\frac{1}{n-1} \sum_{j=0}^{n} (n - j)\, w_j = \Omega.$$

A decision maker with a neutral disposition has an Ω of .5. Greater than .5 represents optimism, while less than .5 represents pessimism.

Denote by T a transformation of the second indexes of the set of outcomes c_{ij} associated with the i-th alternative to a new set b_{ij}, (j=1,2,...,n):

$$T: \quad (c_{i1}, c_{i2}, ..., c_{in}) \rightarrow (b_{i1}, b_{i2}, ..., b_{in})$$

such that $b_{i1} \geq b_{i2} \geq ... \geq b_{in}$. For each alternative i, we apply the same transformation T to the set of probabilities p_1, p_2, ..., p_n. We denote the new set of probabilities $\bar{p}_{i1}$, $\bar{p}_{i2}$, ..., $\bar{p}_{in}$. Obviously such a transformation in general doesn't imply $\bar{p}_{i1} \geq \bar{p}_{i2} \geq ... \geq \bar{p}_{in}$. The original event probabilities (*pristine probabilities*) are modified for each alternative i, by incorporating the decision maker's level of optimism, into *immediate probabilities* $\hat{p}_{i1}, \hat{p}_{i2}, ..., \hat{p}_{in}$ that are defined as follow:

$$\hat{p}_{ij} = \frac{w_j \bar{p}_{ij}}{\sum_{k=1}^{n} w_k \bar{p}_{ik}}$$

The *immediate expected value* of the i-th alternative is the expected payoff that is calculated on the basis of these immediate probabilities, $\hat{p}_{ij}$:

$$IEV_i = \sum_{j=1}^{n} \hat{p}_{ij}\, b_{ij} = \frac{\sum_{j=1}^{n} w_j \bar{p}_{ij}\, b_{ij}}{\sum_{j=1}^{n} w_j \bar{p}_{ij}}$$

Consider the following illustration in which we are given the following information in a decision making situation:

	S_1	S_2	S_3
A_1:	20	15	12
A_2:	10	12	25
P_j:	0.2	0.55	0.25

Assume $\Omega = .7$, then $w_1 = .56$; $w_2 = .29$; $w_3 = .15$ (using O'Hagan's method [1990].)

The reordered outcomes are:

$b_{11} = 20$ $\quad b_{12} = 15$ $\quad b_{13} = 12$
$b_{21} = 25$ $\quad b_{22} = 12$ $\quad b_{23} = 10$

The reordered probabilities under the same transformation of the indexes are:

$\bar{p}_{11} = 0.20$ $\quad \bar{p}_{12} = 0.55$ $\quad \bar{p}_{13} = 0.25$

$\bar{p}_{21} = 0.25$ $\quad \bar{p}_{22} = 0.55$ $\quad \bar{p}_{23} = 0.20$

The immediate probabilities are:

$\hat{p}_{11} = 0.36$ $\quad \hat{p}_{12} = 0.52$ $\quad \hat{p}_{13} = 0.12$
$\hat{p}_{21} = 0.43$ $\quad \hat{p}_{22} = 0.48$ $\quad \hat{p}_{23} = 0.09$

The value of the alternatives using these immediate probabilities are:

$IEV_1 = (.36*20 + .52*15 + .12*12) = 16.45$
$IEV_2 = (.43*25 + .48*12 + .09*10) = 17.34.$

Therefore alternative 2 should be selected. Note that the expected values of the alternatives using the original event probabilities and ignoring the decision maker's optimism would have resulted in selecting alternative 1:

$EV_1 = (.2*20 + .55*15 + .25*12) = 15.25$
$EV_2 = (.2*10 + .55*12 + .25*25) = 14.85.$

We can see that the degree of optimism modifies the event probabilities via the OWA weights and therefore the expected values.

The effect of this modification to the event probabilities is that, for each alternative, the probabilities of preferred outcomes increases and the probabilities of unpreferred outcomes decreases as the level of optimism increases. The reverse is true as the decision maker becomes more pessimistic. Also, note that the immediate probabilities may vary from alternative to alternative. The decision maker selects the alternative with the optimal IEV.

The decision maker may discount the importance of the pristine probabilities in the calculation of the immediate probabilities. Let α be the degree of discounting, where $0 \leq \alpha \leq 1$. Using Shafer's [1976] concept of discounting a belief structure and Dempster's [1967] rule to combine two belief structures, we get

$$\hat{p}_{ij} = \frac{\alpha w_j + (1-\alpha) w_j \bar{p}_{ij}}{\sum_{k=1}^{n} (\alpha w_k + (1-\alpha) w_k \bar{p}_{ik})}$$

for the case of discounting the pristine probabilities. We see that if $\alpha = 0$, no discounting, then the $\hat{p}_{ij}$ equal the original definition of the immediate probabilities. If $\alpha = 1$, complete discounting, then the $\hat{p}_{ij}$ equal the dispositional probabilities, w_j.

The results of discounting the dispositional probabilities are analogous. The decision maker may discount the importance of the dispositional probabilities in the calculation of the immediate probabilities. Let α be the degree of discounting, where $0 \leq \alpha \leq 1$. We get

$$\hat{p}_{ij} = \frac{\alpha p_j + (1-\alpha) w_j \bar{p}_{ij}}{\sum_{k=1}^{n} (\alpha p_k + (1-\alpha) w_k \bar{p}_{ik})}$$

for the case of discounting the dispositional probabilities. We see that if $\alpha = 0$, no discounting , then the $\hat{p}_{ij}$ equal the original definition of the immediate probabilities. If $\alpha = 1$, complete discounting, then the $\hat{p}_{ij}$ equal the pristine probabilities.

IMMEDIATE PROBABILITY AND UTILITY THEORY

In expected utility theory, the utilities of the outcomes are weighted by their probabilities. Actual alternative selections often violate the axioms of expected utility theory, as illustrated by Allais [1953]. Kahneman and Tversky [1979] present a variation of Allais' example as follows:

Problem 1: Choose A or B
A: (2,500 with probability .33, 2,400 with probability .66, 0 with probability .01);
B: (2,400 with certainty)

Problem 2: Choose C or D
C: (2,500 with probability .33, 0 with probability .67);
D: (2,400 with probability .34, 0 with probability .66).

Choices of A and C or alternatively, B and D, are consistent with expected utility theory, whereas choices of A and D, or B and C are not. Note that Problem 2 is obtained from Problem 1 by eliminating a .66 chance of winning 2400 from both prospects under consideration. According to expected utility theory, with $U(0) = 0$, preferences A and C imply $U(2{,}400) < .33U(2{,}500) + .66U(2{,}400)$ or $.34U(2{,}400) < .33U(2{,}500)$, while preferences B and D imply the reverse inequality. In the study, over 60% chose B and C, violating expected utility theory.

The blending of expected utility theory and immediate probability provides an explanation of these seemingly contradictory choices made by the majority of the respondents. The immediate expected values (IEV), as defined above are:

$$IEV_A = \hat{p}_{A1}U(2500) + \hat{p}_{A2}U(2400) + \hat{p}_{A3}U(0); \quad IEV_B = U(2400)$$
$$IEV_C = \hat{p}_{C1}U(2500) + \hat{p}_{C2}U(0); \quad IEV_D = \hat{p}_{D1}U(2400) + \hat{p}_{D2}U(0)$$

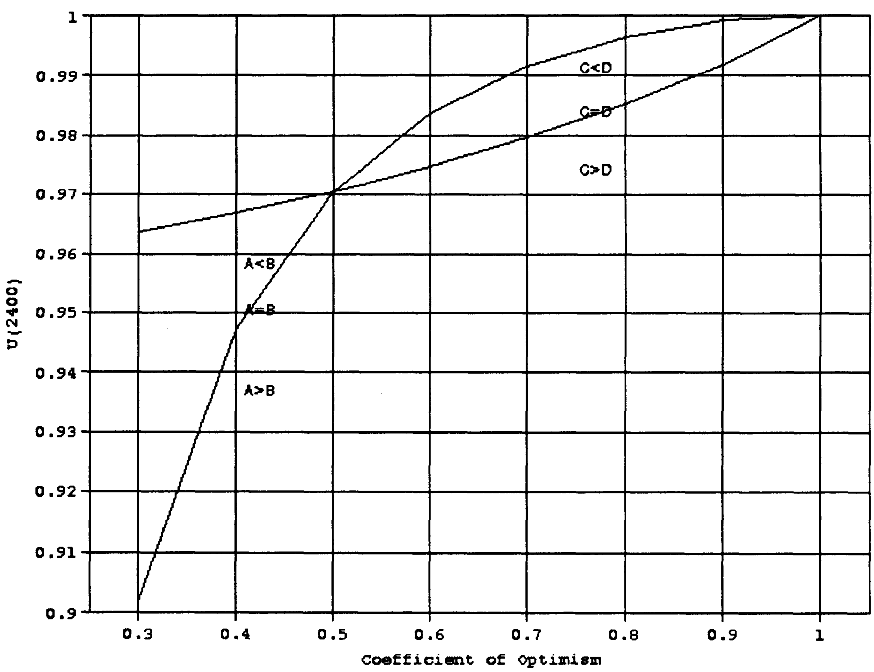
where the immediate probabilites are:

$$\hat{p}_{A1} = (.33w_1^{(3)})/(.33w_1^{(3)}+.66w_2^{(3)}+.01w_3^{(3)})$$
$$\hat{p}_{C1} = (.33w_1^{(2)})/(.33w_1^{(2)}+.67w_2^{(2)})$$
$$\hat{p}_{A2} = (.66w_1^{(3)})/(.33w_1^{(3)}+.66w_2^{(3)}+.01w_3^{(3)})$$
$$\hat{p}_{C2} = (.67w_1^{(2)})/(.33w_1^{(2)}+.67w_2^{(2)})$$
$$\hat{p}_{A3} = (.01w_1^{(3)})/(.33w_1^{(3)}+.66w_2^{(3)}+.01w_3^{(3)})$$
$$\hat{p}_{D1} = (.34w_1^{(2)})/(.34w_1^{(2)}+.66w_2^{(2)})$$
$$\hat{p}_{D2} = (.66w_1^{(2)})/(.34w_1^{(2)}+.66w_2^{(2)}).$$

$w_j^{(k)}$ are the dispositional probabilities for cardinality k. Define U(2500) = 1 and U(0) = 0. For Problem 1 we let IEV_A = IEV_B and find: U(2400) = $(.33w_1^{(3)})/(.33w_1^{(3)}+.01w_3^{(3)})$. For Problem 2 we let IEV_C = IEV_D and find: U(2400) = $(21.78 - 10.56w_1^{(2)})/(22.78 - 11.56w_1^{(2)})$. A risk neutral decision maker has U(2400) = .96 (i.e 2400/2500). A risk avoider has U(2400) > .96, and a risk seeker has U(2400) < .96. The dispositional probabilities, $w_j^{(k)}$, are a function of the decision maker's coefficient of optimism, Ω. Using this approach all four preference combinations (A and C, A and D, B and C, B and D) are explained by the decision maker's utility and optimism.

Figure 1 displays graphically the preferences for the alternatives. Over 80% preferred B over A, and likewise C over D. The least preferred combination D and A is characterized by a particular constrained blending of dispositional optimism and risk aversion. The modal combination B and C is generally characterized by dispositional pessimism and risk taking. Note that a risk taker may be viewed as undervaluing the payoffs since his convex utility curve returns values lower than that of the risk neutral person.

FIGURE 1: PREFERENCES FOR ALTERNATIVES



INFORMATION DESCRIBING EVENT LIKELIHOOD

In decision making under risk we know P_j, the probability that S_j will be the state of nature. The standard procedure in this case is to use expected values. In decision making under uncertainty, we assume no knowledge about the state of nature other than that it is an element in some set $S = \{S_1, S_2, ..., S_n\}$. The methodology used in the selection of the optimal alternative in this environment depends on the attitude held by the decision maker. Among the decision approaches available are: pessimistic, optimistic, Hurwicz, and normative. In general, in the case of decision making under uncertainty we select alternative that optimizes $V_i = F(C_{i1}, C_{i2},...,C_{in})$, where F is some aggregation function whose form depends upon the decision makers attitude.

A general formulation to the optimal alternative selection problem under uncertainty based upon the OWA aggregation operators was introduced by Yager [1988] [1992]. In particular, we can use these operators to provide the aggregated value for each alternative.

We can calculate $V_i = F(C_{i1},...,C_{in})$, where F is an OWA aggregation operator. We then select the alternative that has the highest V value. Let $b_{i1},...,b_{in}$ are the ordered set of the payoffs, C_{ij}. The OWA weights, $w_1,...,w_n$, can be interpreted as probabilities of the jth best thing happening. Then

$$V_i = \Sigma_j \, w_j b_{ij}$$

is the expected payoff in this case. Thus the OWA aggregation provides a kind of expected value similar to that used in decision making under risk. Using the OWA approach, the probabilities, or weights, are assigned not to a particular state of nature, but to the preference ordered position of the payoff. Using this interpretation we can see that the pessimistic strategy is effectively a situation in which a probability of one is assigned to the worst thing happening given any selection of alternative. In the optimistic approach we are assuming a probability of one is assigned to the probability of the best thing happening. In the normative case we are assuming equal probability for each of the preference positions. The Hurwicz strategy effectively assigns a probability α that the best thing will happen and probability $1-\alpha$ that the worst thing will happen.

The two cases, risk and uncertainty, are actually special cases of a broader formulation using the Dempster-Shafer belief structure [Dempster 1967] [Shafer 1976]. In addition to being able to capture these classic representations of our knowledge about the environment, the belief structure allows us to represent various other forms of information a decision maker may have about the state of nature.

A belief structure, m, on the set Y consists of a collection of non-empty subsets of Y, B_i, and an associated set of weights $m(B_i)$ such that:

$$m(B_i) \geq 0$$
$$\Sigma_j\, m(B_j) = 1.$$

Y is the set of possible events. i.e. the state of nature. The subsets B_i are called the focal elements of the belief structure. $m(B_i)$ can be interpreted as the probability of B_i occurring. So we have probabilities associated with the subsets of Y but not with individual events. The occurrence of a focal element induces an outcome on Y which is in the subset B_i. Our concern is the determination of which event within Y will occur.

A problem of considerable interest is that of selecting an appropriate alternative in situations in which our knowledge about the state of nature is in the form of a belief structure. Yager [1992] provides a methodology to do this using OWA operators.

The following summarizes the methodology, assuming we have obtained the payoff matrix, the belief function, m, about the state of nature and the decision makers degree of optimism, Ω:

1. Determine the w_j, for each different cardinality of focal elements, with the degree of optimism Ω.
2. Determine the payoff collection, M_{ik}, if we select alternative A_i and the focal element B_k occurs, for all i,k.
3. Determine the aggregated payoff, $V_{ik} = F(M_{ik})$, for all i, k.
4. Calculate the generalized expected value, C_i, where $C_i = \Sigma_k\, V_{ik}m(B_k)$, for each alternative.
5. Select the alternative with the largest C_i as the optimal.

The procedure for determining the best alternative combines the schemes used for both decision making under risk and uncertainty. In a manner similar to decision making under risk we obtain a generalized expected value, C_i, for each alternative A_i. To obtain this expected value we use evidential knowledge by means of the weights associated with the focal elements. Note that $m(B_k)$ is the probability that B_k will be the set that determines the state of nature.

V_{ik} is the expected payoff, using OWA weights, when we select alternative A_i and focal element B_k occurs. The determination of the value V_{ik} can be seen as equivalent to the problem of decision making under uncertainty. In particular for a given A_i and the knowledge that the state of nature lies in B_k we have a collection of possible payoffs, M_{ik}. In this case each element S_j in B_k contributes one element to M_{ik}, its payoff under S_j. In order to determine the value of V_{ik} from M_{ik} we use the procedure developed for decision making under uncertainty using OWA operators.

Engemann, Miller and Yager [1993, 1996] demonstrated how to apply this approach to risk management.

IMMEDIATE PROBABILITIES AND BELIEF FUNCTIONS

The decision maker whose knowledge of the state of nature is in the form of a belief function may want to further incorporate his level of optimism by modifying the pristine probabilities of the focal elements. Immediate probabilities may be used in conjunction with decision making using belief functions.

The following summarizes the methodology, assuming we have obtained the payoff matrix, the belief function, m, about the state of nature and the decision makers degree of optimism, Ω:

1. Determine the w_j, for each different cardinality of focal elements, with the degree of optimism Ω.
2. Determine the payoff collection, M_{ik}, if we select alternative A_i and the focal element B_k occurs, for all i,k.
3. Determine the aggregated payoff, $V_{ik} = F(M_{ik})$, for all i, k.
4. Use a transformation, T, of the second indexes of the set of aggregated payoffs V_{ik} associated with the i-th alternative to a new set b_{ik}:

   $$T: \quad (V_{i1}, V_{i2}, \ldots, V_{iq}) \rightarrow (b_{i1}, b_{i2}, \ldots, b_{iq})$$

   such that $b_{i1} \geq b_{i2} \geq \ldots \geq b_{iq}$.
5. For each alternative i, apply the same transformation T to the set of probabilities $m(B_1), m(B_2), \ldots, m(B_q)$. Denote the new set of probabilities $_{i1}$, $\bar{p}_{i2}, \ldots, \bar{p}_{iq}$.
6. Calculate the immediate probabilities $\hat{p}_{i1}, \hat{p}_{i2}, \ldots, \hat{p}_{iq}$ that are defined as follows:

$$\hat{p}_{ij} = \frac{w_j \bar{p}_{ij}}{\sum_{k=1}^{q} w_k \bar{p}_{ik}}$$

7. Calculate the immediate expected value of the i-th alternative calculated on the basis of these immediate probabilities, $\hat{p}_{ij}$:

$$IEV_i = \sum_{j=1}^{q} \hat{p}_{ij}\, b_{ij} = \frac{\sum_{j=1}^{q} w_j \bar{p}_{ij}\, b_{ij}}{\sum_{j=1}^{q} w_j \bar{p}_{ij}}$$

8. Select the alternative with the largest IEV_i as the optimal.

Consider the following example. The payoff matrix is as follows:

	S_1	S_2	S_3	S_4	S_5	S_6
A_1	20	0	20	30	40	50
A_2	20	10	0	30	20	40
A_3	40	5	10	40	30	20

The focal elements are:

Focal element		Weights
$B_1 =$	$\{S_1\}$	0.2
$B_2 =$	$\{S_2, S_3\}$	0.3
$B_3 =$	$\{S_4, S_5, S_6\}$	0.5

We recall that M_{ik} is the collection of payoffs that are possible if we select alternative A_i and the focal element B_k occurs. We next determine the payoff collections M_{ik}.

$M_{11} = \{20\}$, $M_{12} = \{0, 20\}$, $M_{13} = \{30, 40, 50\}$,
$M_{21} = \{20\}$, $M_{22} = \{0, 10\}$, $M_{23} = \{20, 30, 40\}$,
$M_{31} = \{40\}$, $M_{32} = \{5, 10\}$, $M_{33} = \{20, 30, 40\}$.

Assume that the decision maker is optimistic with a degree of optimism of 0.75. We calculate V_{ik}, using the maximum entropy OWA weights. We recall that $V_{ik} = F(M_{ik})$.

$V_{11} = 1.00(20) = 20$
$V_{12} = .75(20) + .25(0) = 15$
$V_{13} = .62(50) + .27(.40) + .11(30) = 45.1$
$V_{21} = 1.00(20) = 20$
$V_{22} = .75(10) + .25(.0) = 7.5$
$V_{23} = .62(40) + .27(30) + .11(20) = 35.1$
$V_{31} = 1.00(40) = 40$
$V_{32} = .75(10) + .25(5) = 8.75$
$V_{33} = .62(40) + .27(30) + .11(20) = 35.1$

The reordered outcomes are:

$b_{11} = 45.1$; $b_{12} = 20$; $b_{13} = 15$.
$b_{21} = 35.1$; $b_{22} = 20$; $b_{23} = 7.5$
$b_{31} = 40$; $b_{32} = 35.1$; $b_{33} = 8.75$.

The reordered probabilities under the same transformation of the indexes are:

$\bar{p}_{11} = 0.5$; $\bar{p}_{12} = 0.2$; $\bar{p}_{13} = 0.3$

$\bar{p}_{21} = 0.5$; $\bar{p}_{22} = 0.2$; $\bar{p}_{23} = 0.3$

$\bar{p}_{21} = 0.2$; $\bar{p}_{22} = 0.5$; $\bar{p}_{23} = 0.3$

The immediate probabilities are:

$\hat{p}_{11} = .5(.62)/(.5(.62)+.2(.27)+.3(.11)) = .83$
$\hat{p}_{12} = .2(.27)/(.5(.62)+.2(.27)+.3(.11)) = .14$
$\hat{p}_{13} = ..3(.11)/(.5(.62)+.2(.27)+.3(.11)) = .03$
$\hat{p}_{21} = .5(.62)/(.5(.62)+.2(.27)+.3(.11)) = .83$
$\hat{p}_{22} = .2(.27)/(.5(.62)+.2(.27)+.3(.11)) = .14$
$\hat{p}_{23} = .3(.11)/(.5(.62)+.2(.27)+.3(.11)) = .03$
$\hat{p}_{31} = .2(.62)/(.2(.62)+.5(.27)+.3(.11)) = .43$

$\hat{p}_{32} = .5(.27)/(.2(.62)+.5(.27)+.3(.11)) = .46$

$\hat{p}_{33} = .3(.11)/(.2(.62)+.5(.27)+.3(.11)) = .11$

The immediate expected values of the alternatives are:

$IEV_1 = (.83*45.1 + .14*20 + .03*15) = 40.6$
$IEV_2 = (.83*35.1 + .14*20 + .03*7.5) = 32.1$
$IEV_3 = (.43*40 + .46*35.1 + .11*8.75) = 34.2.$

Therefore alternative 1 should be selected.

INFORMATION DESCRIBING PAYOFFS

Payoff aggregation techniques are usually based upon the use of an averaging operator. The median provides an aggregation operation which only requires that the aggregated objects are ordered. Thus while the median aggregation, like the average, can work in numeric domains, it also can be used in situations in which we are not combining numbers but ordered objects. A new aggregation operator which is an extension of the median, is called the weighted median [Yager 1995] [Engemann, Miller and Yager 1995]. With this operator we can provide an aggregation of weighted objects where the objects to be aggregated need only be drawn from an ordered set while the weights can be numbers.

Assume $D = \{ (w_1, a_1), (w_2, a_2), \dots, (w_n, a_n) \}$ are a collection of pairs where a_i is a score and w_i is its associated weight. Assume that the a_i are reordered such that b_j is the jth largest of the a_i. Furthermore, let u_j be the weight that is associated with the a_i that becomes b_j. Thus if $b_j = a_5$ then $u_j = w_5$. Once having the ordered collection $D = \{ (u_1, b_1), (u_2, b_2), \dots, (u_n, b_n) \}$ to calculate the weighted median we proceed as follows. We denote

$$T_i = \sum_{j=1}^{i} u_j,$$

the sum of the first i weights. From this we define the Weighted Median (D) = b_k where k is such that $T_{k-1} < 0.5$ and $T_k \geq 0.5$. Thus the weighted median is the ordered value of the argument for which the sum of the weights first crosses the value 0.5.

The weighted median and the weighted average share the properties of idempotency, commutativity/symmetry, and monotonicity [Yager 1995].

IMMEDIATE PROBABILITY AND THE WEIGHTED MEDIAN

Immediate probabilities may be used in conjunction with the weighted median. The following example will illustrate this. Assume the scale of payoffs is: very small (vs), small (s), medium (m), large (l), very large (vl). To the decision maker larger is preferred. The payoff matrix is as follows:

	S_1	S_2	S_3	S_4	S_5
A_1	m	vl	vl	vs	l
A_2	vs	vl	m	s	l
p_j	.1	.1	.3	. 3	.2

Using the pristine probabilities we find:

A_1	b_i	u_j	T_i	A_2	b_i	u_j	T_i
	vl	.1	.1		vl	.1	.1
	vl	.3	.4		l	.2	.3
	l	.2	.6		m	.3	.6
	m	.1	.7		s	.3	.9
	vs	.3	1.0		vs	.1	1.0

Since the Weighted Median (A_1) = large and the Weighted Median (A_2) = medium, A_1 would be preferred by a decision maker with a neutral disposition.

Now assume that the decision maker is pessimistic with a degree of optimism of 0.3. The maximum entropy OWA weights [O'Hagan 1990] are w_5 = .396, w_4 = .257, w_3 = .167, w_2 = .109, w_1 = .071. Then using the immediate probabilities we find:

A_1	b_i	u_j	T_i	A_2	b_i	u_j	T_i
	vl	.03	.03		vl	.04	.04
	vl	.15	.18		l	.11	.15
	l	.15	.33		m	.26	.41
	m	.12	.45		s	.39	.80
	vs	.55	1.0		vs	.20	1.0

Therefore by including the decision makers pessimistic disposition the Weighted Median (A_1) = very small and the Weighted Median (A_2) = small. Now A2 would be preferred by the decision maker.

CONCLUSION

We have shown how several recent approaches in decision modeling may be used in conjunction with one another. A decision maker's attitude is an important component of the decision process and may be integrated into a decision model in cases of risk or uncertainty. Immediate probabilities play a key role in blending the decision maker's disposition with event probabilities.

REFERENCES

Allais, M., Le Comportement de l'Homme Rationnel devant le Risque, Critique de Postulats et Axiomes de l'Ecole Americaine, *Econometrica*, 21, 503-546, 1953.

Brockhaus, R., Risk Taking Propensity of Entrepreneurs, *Academy of Management Journal*, 23(3), 509-520, 1980.

Dempster, A. P., Upper and Lower Probabilities Induced by a Multi-valued Mapping, *Annals of Mathematical Statistics*, 325-339, 1967.

Dunegan, K., Duchon, D. and S. Barton, Affect, Risk, and Decision Criticality: Replication and Extension in a Business Setting, *Organizational Behaviour and Human Decision Processes*, 53(3), 335-351, 1992.

Engemann, K. J., Miller, H. E. and Yager, R. R., A General Methodology for Decision Making Under Uncertainty with a Risk Management Application, *Proceedings of the 2nd International Symposium on Uncertainty Modeling and Analysis*, 194-199, April 1993.

Engemann, K., Miller, H. and R. Yager, Decision Making Using the Weighted Median Applied to Risk Management, *Proceedings of the Third International Symposium on Uncertainty Modeling and Analysis and the Annual Conference of North American Fuzzy Information Processing Society*, 9-13, 1995.

Engemann, K., Filev, D. and R. Yager, Modelling Decision Making Using Immediate Probabilities, *International Journal of General Systems*, 24(3), 281-294, 1996.

Engemann, K., Miller, H., and Yager, R., Decision Making with Belief Structures: An Application in Risk Management, *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 4(1), 1-26, 1996.

Kahneman, D., and A. Tversky, Prospect Theory: An Analysis of Decisions under Risk, *Econometrica*, 47(2). 263-291, 1979.

March. J. and Z Shapira, Managerial Perspectives on Risk and Risk Taking, *Management Science*, 33(11), 1404-1418, 1987.

Mellers, B. and S. Chang, Representations of Risk Judgments, *Organizational Behaviour and Human Decision Processes*, 57(2), 167-184, 1994.

O'Hagan, M. , Using Maximum Entropy-ordered Weighted Averaging to Construct a Fuzzy Neuron, *Proceedings of the 24th Annual IEEE Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 618-623, 1990.

Schneider, S. and L. Lopes, Reflection in Preferences Under Risk: Who and When May Suggest Why, *Journal of Experimental Psychology: Human Perception and Performance*, 12(4), 535-548, 1986.

Shafer, G. A., *Mathematical Theory of Evidence*, Princeton, New Jersey: Princeton University Press, 1976.

Sitkin, S. and A. Pablo, Reconceptualizing the Determinants of Risk Behavior, *Academy of Management Review*, 17(1), 9-38, 1992.

Staw, B., Sandelands, L. and J. Dutton, Threat-Rigidity Effects in Organizational Behaviour: A Multilevel Analysis, *Administrative Science Quarterly*, 26(4), 501-524, 1981.

Vlek, C. and P. Stallen, Rational and Personal Aspects of Risk, *Acta Psychologica,* 45(1-3), 273-300, 1980.

Yager, R. R., On Ordered Weighted Averaging Aggregation Operators in Multi-criteria Decision Making, *IEEE Transactions on Systems, Man and Cybernetics,* 18, 183-190, 1988.

Yager, R. R., Fusion of Ordinal Information Using Weighted Median Aggregation, *Technical Report #MII-1520,* Iona College, New Rochelle, NY, 1995.

Yager, R. R., On Mean Type Aggregation, *IEEE Transactions on Systems, Man and Cybernetics*, 26, 209-221, 1996.

Yager, R. R., Decision Making Under Dempster-Shafer Uncertainties, *International Journal of General Systems,* 20, 233-245, 1992.

Yager, R., Engemann, K.and D. Filev, On the Concept of Immediate Probabilities, *International Journal of Intelligent System,* 10(4), 374-397, 1995.

RISK MANAGEMENT USING FUZZY LOGIC AND GENETIC ALGORITHMS

Teresa C. Rubinson
TCR, Inc.
18 Wildwood Lane
Weston, CT, USA 06883
rubinson@tcrinc.com

Georgette Geotsi
Sacred Heart University
5151 Park Avenue
Fairfield, CT, USA 06432
geotsi@shu.sacredheart.edu

ABSTRACT

Risk management is a complex and subjective task. It involves identifying risk factors, setting risk thresholds, and determining appropriate actions to reduce risk [Bernstein, 1995]. Fuzzy logic techniques are especially well suited to subjective problems of this type. Traditionally, fuzzy techniques have relied on user supplied data to define fuzzy functional parameters. However, in real life people are notoriously inaccurate and unreliable in reporting their preferences, especially as the complexity and uncertainty of the problem increases. In this paper, we describe a methodology which uses a genetic algorithm to automate and validate derivation of fuzzy functional parameters.

KEYWORDS

Risk management, fuzzy logic, genetic algorithms, Ordered Weighted Averaging (OWA) Operators, fuzzy multiple criteria decision making, fuzzy ranking

INTRODUCTION

In this paper, we discuss the problem of scoring and ranking options or actions according to subjective risk features they possess. This type of problem appears in many forms in risk management. For example, one might wish to identify "high risk" investment portfolios so that proactive efforts can be taken to manage or compensate for the investment risk. In another case, one might wish to rank loan recipients according to the potential default risk they pose.

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

Many treatments for problems of this type are described in the literature. These treatments include statistical analysis, artificial neural networks, and fuzzy logic techniques. Both statistical and artificial neural network techniques produce predictive models based on analysis of historical data. Thus, these techniques are not effective in predicting behavior that has not been observed in the past. Alternatively, a typical "fuzzy" approach to risk prediction is to build a fuzzy expert system. As described in [Cox, 1994], this type of system contains fuzzy rules which are used to evaluate each candidate and to prescribe a specific course of action or to assign a ranking score. However, this approach is practical only if the risk factors and corresponding actions are sufficiently well understood to allow definition of a fuzzy rule base.

However, there are many risk scenarios -- particularly those relating to economic conditions and market forces -- where the underlying system dynamics are in flux and are poorly understood. A prime example of this is the derivative market. Market forces exert strong influences affecting the risk involved in trading derivative instruments. Conversely, derivative instruments have profound influences on the market as a whole. Derivative instruments -- which include SWAPS, exotic options, mortgage backed securities, etc. -- are a fairly recent phenomenon. Failures in the derivative markets -- such as the default of Orange County -- tend to be spectacular, but fairly few and far between, given the variety and number of derivative instruments being traded in the market as a whole. Thus, the historical data available on the derivatives market is limited in both quantity and quality. Therefore, none of the methods described thus far are wholly adequate for assessing risk in this type of situation.

This paper outlines a methodology to assess risk in situations where there is a high degree of uncertainty and a lack of historical data. This methodology is an extension of fuzzy multi-criteria problem solving approaches. This approach is viable even when the data and/or rules underlying system behavior are unavailable or are not well known or well defined. Our approach involves collecting subjective data on what analysts perceive are relevant risk factors and their relative importance. From this input, we are able to rapidly build individual or group models for risk assessment. This capability is important, because in an organizational context both individual preferences and corporate culture shape how risk is defined [Cox, 1994]. This capability also allows us to study the effects of various points of view to help build consensus on a corporate risk model. Additionally, our fuzzy logic tools are computationally fast, and can be easily modified to reflect new conditions and information.

FUZZY MULTIPLE CRITERIA DECISION MAKING: THEORETICAL BACKGROUND

In [Bellman and Zadeh, 1970], Bellman and Zadeh re-examined the classical problem of optimizing an objective function subject to constraints. They suggested a fuzzy reformulation which combines objective function(s) and constraint(s) into a single decision function. In discussing the rationale for their approach, Bellman and Zadeh suggested that, in a fuzzy environment, a single decision function representing a confluence of goals *and* constraints is the best analogue to the classical optimization problem, which also attempts to satisfy objectives *and* constraints simultaneously [Bellman and Zadeh, 1970]. In [Bellman and Zadeh, 1970], the aggregation of the objective function and the constraint is performed using the "min" operator.

Two important features of operator behavior in light of the problem at hand are compensation and adaptability. Compensation refers to the ability of the operator to "compensate" for a low degree of membership in one set by a higher degree of membership in another. The "logical and" represented by the "min" operator does not allow for any compensation. That is, in aggregating decision criteria, the "min" operator determines the solution outcome by evaluating the minimum performance measure. Therefore, this is potentially a serious shortcoming in performing risk assessment. The adaptability of an operator refers to its ability to be modified in context dependent situations. The "min" operator, as defined by Bellman and Zadeh, can not be adapted to provide this discernment. This, too, is a significant limitation of the “min” operator.

In response to these limitations, Yager has developed a class of aggregation operators -- called Ordered Weighted Averaging Aggregation (OWA) operators -- to specifically address the need for adaptability and compensation in the decision function, while maintaining many of the desirable features of the "min" operator. Yager has formally defined an OWA operator in [Yager, 1988] as:

> A mapping F from $I^n \rightarrow I$ (where $I = [0,1]$) is called an OWA operator of dimension n if associated with F, is a weighting vector $\mathbf{W} = [W_1, W_2, \ldots, W_n]$ such that:
>
> 1) $W_i \in [0,1]$,
>
> 2) $\sum_{i=1}^{n} W_i = 1$
>
> and where $F(a_1, a_2, \ldots, a_n) = W_1 b_1 + \ldots + W_n b_n$
> and b_i is the ith largest element in the collection $a_1, \ldots a_n$.

As defined above, the components of the vector W must be greater than or equal to zero, and less than or equal to one. A unique feature of these OWA operators is that they allow "combining of criteria under the guidance of a quantifier . . . . That is, the requirement that 'most' of the criteria be satisfied corresponds to one of these OWA operators" [Yager, 1988]. For example, one may wish that all criteria be satisfied in the evaluation procedure, or alternatively, that at least one criterion be satisfied. These examples correspond, respectively to "and" and "or" fuzzy operators. To mitigate these extremes, Yager has suggested use of the OWA operators as fuzzy quantifiers to mathematically represent the linguistically expressed requirement that "most" or "some" criteria must be satisfied.

Another desirable feature of Yager's OWA operators is that they are easy to implement in a computer program. The most difficult aspect in defining an OWA operator is the determination of the weighting vector W. Once W has been determined, the decision function F is computed as the inner product of W and an ordered vector B. In the sections that follow, we describe a genetic algorithm technique to help in the determination of W.

In [Yager, 1988], Yager describes how to compute an overall evaluation score using the OWA operator. The first step in the ordered weighted aggregation process is to compute the individual criteria weighting $\mathbf{a_j}$, as shown below:

$$\mathbf{a_j} \quad = \quad (\alpha_j \cup p) * (A_j(x))^{(\alpha j \cup q)}$$

where: $\mathbf{a_j}$ = *final risk score*

α_j = *criteria weight*

$A_j(x)$ = *individual fuzzy membership score for criteria j*

$1 = p + q$

p = *degree of "andness"* $= 1 - q$

q = *degree of "orness"* $= (1/(n-1)) * [\sum_{j=1}^{n} ((n-i) * w_j)]$

n = *number of decision criteria*

w_j = *weighting vector, as defined above*

The $\mathbf{a_j}$ are used, in turn, to compute an overall risk evaluation score. The overall evaluation score is determined -- according to the definition of the OWA operator given above -- by sorting the $\mathbf{a_j}$ from high to low, thus forming a vector b. Next, the vector product of b and W is computed, yielding a final score.

The approach suggested by Yager [Yager, 1988] relies on user supplied data to set the model parameters (which include $A_j(x)$, w_j, and α_j, as defined above). In real life, however, people are frequently inaccurate and unreliable in reporting their preferences. This is especially true as the dimensionality and complexity of the problem increases. In [Rubinson and Geotsi, 1996], genetic algorithm techniques to automate derivation of the OWA functional parameters are described. This paper builds on the findings presented in [Rubinson and Geotsi, 1996]. In the discussion that follows, we demonstrate how the criteria importance weights (α_j) and the components of the OWA vector, (w_j), can be derived using a genetic algorithm.

GENETIC ALGORITHMS: THEORETICAL BACKGROUND

Genetic Algorithms (GAs) are based on the Darwinian notion of survival of the fittest. GAs are designed to breed solutions using genetic cross-over and mutation operators. Genetic cross-over is a process by which elements of prior solutions are re-combined to produce new solutions. Mutation is a process by which elements in a given solution are randomly altered to produce population variability. By virtue of the features of random mutation and cross-over, GAs are able to produce a wide range of solutions and thus effectively search a large, complex solution space. Solutions produced by GAs are evaluated and scored using a "fitness" function. The fitness score, in conjunction with pre-determined probability settings, determines whether or not the solution will be allowed to produce offspring for the next generation of solutions. The major procedure steps of a GA are summarized in Table 1 below [Michalewicz, 1994].

Table 1: Major Steps in a Genetic Algorithm

1. Produce initial population of solutions
2. Cross-over initial solutions to produce new solutions
3. Mutate some solutions
4. Evaluate solutions using fitness function
5. Discard unfit solutions
6. Repeat Steps 2-4 above, until a dominant solution emerges or until time and resource constraints are exhausted.

As can be inferred from Table 1, there are a number of significant challenges relating to the design of a Genetic Algorithm experiment. Briefly summarizing, some of these challenges include [Michalewicz, 1994]:

- Design of genetic chromosome to represent potential solutions
- Specification of fitness function to evaluate solutions

- Design of genetic operators
- Selection of GA parameters, including population sizes, probabilities (for mutation, cross-over, etc.)
- Determination of stopping criteria

The chromosome must contain enough information to fully specify a prospective solution. It must also be robust enough to withstand the effects of mutation and cross-over and yet still yield reasonable solutions.

The fitness function must be designed to correctly analyze the desired features of prospective solutions.

The cross-over and mutation operations must be designed to promote the breeding of appropriate solutions. Although genetic variability is desirable, one must also guard against random mutations and cross-over that will transform good solutions into unfit solutions that will in turn produce even worse offspring in future generations.

Additionally, there are many parameters that need to be set in a GA experiment. These parameters determine the population sizes produced during breeding, the probability of cross-over and mutation, and so on. Table 2 summarizes some of the standard parameters which must be set in a GA experiment.

Table 2: Key Genetic Algorithm Parameters

- Number of Experiments
- Total Number of Trials
- Population Size
- Structure Length of Chromosome
- Crossover Rate
- Mutation Rate
- Generation Gap
- Scaling Window
- Report Interval
- Structures Saved
- Maximum Number of Generations Without Evaluation
- Population Distribution
- Random Seed
- Convergence Settings and Threshold

Finally, one must determine when to stop the GA breeding and to terminate the experiment. Sometimes this determination is straightforward. This is the case when a dominant solution emerges, showing little variability from one generation to the next, suggesting that further iterations will yield little or no improvement in the final result. In other cases, the results of the GA experiment are not so clear cut. If the GA does not produce a solution that converges to a single value, it is possible that there is no unique solution. This, in turn, may mean that there are several more or less equally satisfying alternative solutions. When there is no convergence on a single solution, the solution producing the best fitness score calculated thus far is often selected as the final result. In this case, there is the possibility that an even better solution might be obtained if the GA were allowed to continue. Alternatively, a lack of convergence in the solution may indicate problems in the formulation or parameter settings used in the GA experiment. All these possibilities must be carefully considered when interpreting the results of a GA experiment.

In summary, GAs are very effective in searching complex solution spaces. GAs have been successfully applied to numerous optimization problems, including "wire routing, scheduling, adaptive control, game playing, cognitive modeling, transportation problems, traveling salesman problems, optimal control problems, database query optimization, etc." [Michalewicz, 1994, p. 15] Although GAs do not guarantee an optimal solution, in our experience, GAs are usually very effective in locating the correct neighborhood for the solution, thus avoiding the pitfalls of local minima and maxima. After the correct solution neighborhood has been identified, one may wish to refine the solution search using branch-and-bound or other solution techniques. We are currently researching techniques for refining the output of a GA experiment using a variety of supplemental optimization techniques. We plan to report our findings in this area in a future paper.

RISK MANAGEMENT METHODOLOGY

In this section, we describe a methodology for using the OWA decision function to score and rank individuals according to risk factors they possess. Our methodology consists of the following steps:

1. Data Preparation
2. Genetic Algorithm Experiment
3. Model Validation and Testing

Step 1: Data Preparation

In this step, we prepare and collect the data needed for analysis. First, we collect data specifying the risk factors to be included in the OWA multi-criteria decision function. This involves translating each risk factor into a corresponding fuzzy membership function. The fuzzy membership function represents, in essence, a linguistic definition for each decision variable or risk factor. Fuzzy membership functions are derived by interviewing experts and/or decision makers and by asking them to assign fuzzy scores to possible values of each risk factor. For example, to define "Low Liquidity", a decision maker would be given a series of available cash balances for which they would be asked to provide a corresponding fuzzy score. The fuzzy score represents the degree to which the decision maker agrees or not that a specific cash balance is consistent with the linguistic variable being defined (i.e., "Low Liquidity"). These (value, fuzzy score) pairs are then analyzed using automated curve fitting techniques to derive an estimation of the fuzzy membership function equation. The fuzzy membership functions are used to determine the values of A_j (x), as defined earlier.

Second, we construct a *training* data set containing individuals with *known* risk factors and associated overall fuzzy risk scores. Note that the risk factors in this data set must be transformed into fuzzy membership values using the appropriate fuzzy membership function definitions. The training data sets should consist of records of the following form:

Individual Name	Risk_Factor_1	, . . . ,	Risk_Factor_n	Overall Fuzzy Risk Score

Depending on the need for accuracy and sensitivity analysis, one may wish to construct multiple training sets. Each training set would then be used in a separate GA experiment. The result would be several independently derived solution sets for the OWA function parameters.

For example, one might wish to develop a training set containing only very high risk individuals so that the GA would tune the OWA function parameters as carefully as possible for this risk group. Alternatively, one might wish to develop a training set containing a range of low to high risk individuals. This type of training set is helpful in determining an overall risk scoring scheme. Using both types of training sets helps to elucidate whether or not different OWA weighting schemes need to be used to score different classes within a given population.

One might also wish to construct a training set consisting of records that have been scored differently by different experts or decision makers. The GA could then be used to see if a good compromise scheme can be found to satisfy most of the decision making participants.

In any case, it is important to carefully consider the make-up of the individuals to be included in the training data set(s) and the potential ramifications on the GA results. Unlike statistical or artificial neural network techniques, which favor quantity in the test samples, the GA technique favors quality in the test samples. Thus, it is better to have fewer, but more consistent records in the training set, if at all possible. With the method we propose, it is possible to train a GA with only one record in the training set, although it is advisable to have as many training records as is feasible. In our experience, we often have to deal with very sparse training data sets for the types of imprecise, complex problems we wish to study.

The GA partially compensates for a lack of training data by producing many candidate solutions which can then be examined and empirically validated [Rubinson and Geotsi, 1996].

Finally, we construct a data set containing the individuals we wish to rank and score. Note that the risk factors in this data set must also be transformed into fuzzy membership values using the appropriate fuzzy membership function definitions. This data set consists of records of the following form:

Individual Name	Risk_Factor_1	Risk_Factor_2	, . . . ,	Risk_Factor_n

Step 2: Genetic Algorithm Experiment

In this step, we perform a genetic algorithm experiment to find optimal criteria weights for each fuzzy membership function and the components of the OWA weighting vector. We recommend the GAUCSD 1.4 Genetic Algorithm tool to perform these experiments. This software is readily available on the Internet and it is well-documented.

To perform the GA experiment, one must define an appropriate chromosome structure. For the purposes of this paper, the chromosome defined for the GA experiment represents criteria weighting (α_j) values and the (w_j) components of the OWA weighting vector. We have had good success representing each α_j and w_j by a 30 bit string, encoded in binary. Other chromosome representations are possible, and are the subject of on-going investigations we are conducting.

One must also define an initial population to begin the GA breeding process. We recommend a super-uniform initial population, as this type of population is easy to generate randomly and has yielded good results in the experiments we have conducted.

The GAUCSD 1.4 tool generates standard default settings for the GA parameters (see Table 2) for a given GA experiment. These settings provide a reasonable starting point for a GA experiment. It is recommended that these settings be used as initial settings, if only for bench-marking purposes. In our experience, we have found the GA procedure to be highly robust and relatively insensitive to small changes in the parameter settings. However, each GA experiment should be evaluated on a case by case basis, since good parameter settings are largely determined through an empirical process of trial and error.

The remaining task prior to initiating the GA procedure is to define a fitness function. The GAUCSD 1.4 tool allows one to write customized evaluation functions in C. In our research, we are investigating various learning paradigms typically associated with Artificial Neural Net training techniques. One of this simplest of these paradigms is the least mean squared error goal [Hecht-Nielsen, 1970]. This paradigm is based on the idea of minimizing the deviation between a known score and a calculated score (computed using parameters w_j and α_j supplied by the GA). When the least mean squared goal is used as the fitness function, the GA must evaluate all known scores in the training data set against the score provided by a particular candidate solution for α_j and w_j. The least mean squared goal penalizes larger deviations more than smaller deviations, and is indifferent to whether or not the deviation is positive or negative. The decision to use this type of paradigm is highly dependent upon whether one wishes to model this type of situation.

After these preliminary tasks are completed, it is relatively easy, given sufficient machine resources, to run the GA experiment using the GAUCSD tool. As a matter of good experimental practice, it is highly recommended that the GA procedure be calibrated using an artificially constructed data set for which all parameters are known. If the GA returns consistent, expected results, then one can safely proceed with the actual GA experiment.

Step 3: Model Validation and Testing

This phase typically involves substituting the parameters derived by the GA into the OWA decision function and then comparing the results against a known hold-out sample constructed specifically for the purpose of model validation. After the components of the fuzzy ranking function are derived from the GA, it is a simple matter to substitute them into the OWA ranking equation, to derive an overall ranking score for each option under consideration. The OWA ranking model is refined (by performing additional GA experiments using various training data sets and parameter settings) until it performs to some pre-specified level of performance or is shown to be unacceptable.

In assessing the model performance, it is often helpful to graph the results. In Figure 1, the actual risk score is plotted against the corresponding score calculated using the OWA function and GA derived parameters. In this example, the actual and calculated scores do not match well, as can be seen by the fact that when the actual risk score equals 1, the calculated risk score assumes values any where from .25 to 1. Similar results hold for actual risk scores equal to .5 and 0, respectively. Poor or inclusive results may reflect one or more of the following:

- sub-optimal GA specification and parameter selection
- inconsistent training data
- sub-optimal fuzzy model specification

The contribution, if any, of each of these factors to the final result should be carefully analyzed and considered in the design of the GA experiment.

Actual Versus Calculated Fuzzy Risk Score

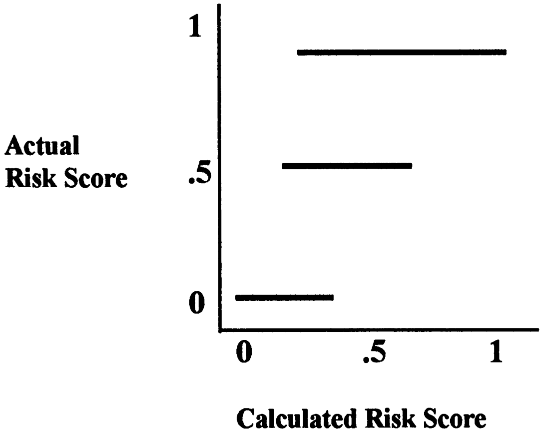


Figure 1: Example of Poor Scoring

Similarly, in Figure 2, the actual risk score is plotted against the corresponding score calculated using the OWA function and GA derived parameters. In this example, the actual and calculated scores do match well, as can be seen by the fact that when the actual risk score equals 1, the calculated risk score assumes values very close to 1. A similar result holds for actual risk scores equal to .5 and 0, respectively.

Actual Versus Calculated Fuzzy Risk Score

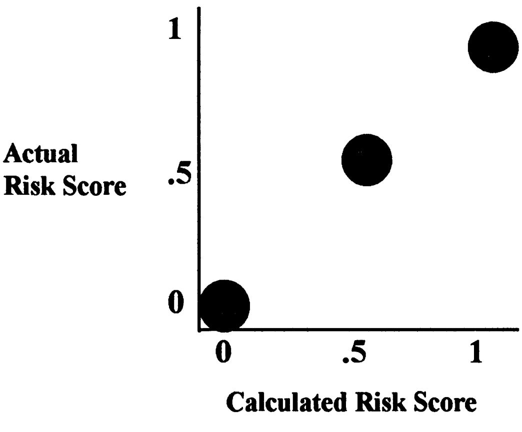


Figure 2: Example of Good Scoring

After a model has been developed and validated, it can be used to score, categorize or identify unknown individuals and to make predications about their behavior.

SUMMARY

In this paper we described a method to perform risk assessment using a fuzzy OWA operator. Additionally, we discussed how genetic algorithms can be used to automatically determine the parameters of an OWA multi-criteria decision function.

ACKNOWLEDGMENTS

We wish to thank Dr. Charles Palmer, Dr. Ronald Yager, and Dr. Aaron Kershenbaum for their support and guidance during the course of this research.

REFERENCES

Bellman, R., E., and Zadeh, L., A. "Decision Making in a Fuzzy Environment." *Management Science* 17, 1970, pp. 151-169.

Bernstein, Peter. *Portable MBA in Investment.* John Wiley & Sons, 1995, pp. 243-307.

Cox, Earl. *The Fuzzy Systems Handbook.* AP Professional, 1994, pp. 435-469.

Deboeck, Guido, editor. *Trading on the Edge, Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets.* John Wiley & Sons, 1995, pp. 204-205.

Hecht-Nielsen, Robert. *Neurocomputing.* Addison-Wesley Publishing Company, 1990.

Michalewicz, Zbigniew. *Genetic Algorithms + Data Structures = Evolution Programs. Second Edition.* Springer Verlag, 1994.

Rubinson, Teresa C. *A Fuzzy Multiple Attribute Design and Decision Procedure for Long Term Network Planning.* Ph.D. Dissertation, Polytechnic University, 1992.

Rubinson, Teresa C, and Geotsi, Georgette, M.. "Estimation of Subjective Preferences Using Fuzzy Logic and Genetic Algorithms." Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems; 1996 July 1-5; Granada, Spain, pp. 781-786.

Rubinson, T., and Yager, R., R "Fuzzy Logic and Genetic Algorithms for Financial Risk Management." Proceedings of IEEE/IAFE Conference on Computational Intelligence for Financial Engineering; 1996, pp.-90-95.

Saaty, T. "Hierarchies, priorities, and eigenvalues." University of Pennsylvania, Philadelphia, 1972.

Yager, R., R. "Fuzzy decision making including unequal objective." *Fuzzy Sets and Systems.* Volume 1, 1978, pp. 87-95.

Yager, R., R. "On Order Weighted Averaging Aggregation Operators in Multicriteria Decisionmaking." *IEEE Trans. on Systems, Man, and Cybernetics.* Volume 18, No. 1, January/February 1988, pp. 183-190.

OWA Operators for doctoral student selection problem

Christer Carlsson*, Robert Fullér** and Szvetlana Fullér**

* Institute for Advanced Management Systems Research, Åbo Akademi University, Åbo, Finland
*** Department of Computer Science, Eötvös Loránd University, Budapest, Hungary

Abstract

Yager [7] introduced a new aggregation technique based on the ordered weighted averaging (OWA) operators. In this article we illustrate the applicability of OWA operators to a doctoral student selection problem at the Graduate School of Turku Centre for Computer Science.

1 The case

The Graduate School of Turku Centre for Computer Science (TUCS) offers a programme for gaining the Doctoral (PhD) degree in Computer Science and Information Systems. It is open for students from everywhere. The teaching language of the school is English. Prerequisites are either a Master's or a Bachelor's degree in Computer Science or in a closely related field. Study time is expected to be 4 years when starting from Master's level and 6 years from Bachelor's level.

The Graduate School offers advanced level courses in Computer Science and supervision of students within existing research projects. The main areas of research are *Algorithmics, Discrete Mathematics, Embedded Systems, Information Systems, Software Engineering.* Students are expected to take courses from at least two of these areas. Each student is assigned a supervisor from one of the fields indicated above.

The Graduate School is open for applications. There are no specific application forms. Applicants to TUCS graduate school should write a letter to the

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

Director of TUCS. The letter should contain a formal application to the school, together with the following enclosures:

- Curriculum vitae
- Financing plan for studies
- Application for financial support, if requested
- Two letters of recommendation with referees' full contact addresses
- Official copy of examinations earned with official English translation
- Certificate of knowledge of English
- Short description of research interests

As certificate of knowledge of English, TOEFL test (minimum 550 points) or corresponding knowledge in English is required for applicants outside Finland.

Since the number of applicants (usually between 20 and 40) is much greater than the number of available scholarhips (around 6) we have to rank the candidates based on their performances. It can also happen that only a part of available scholarships will be awarded, because the number of *good* candidates is smaller than the number of available places.

2 Problem Formulation

The problem of selecting *young promising doctoral researchers* can be seen to consist of three components. The first component is a collection

$$X = \{x_1, \ldots, x_p\}$$

of applicants for the Ph.D. program.

The second component is a collection of 6 criteria (see Table 1) which are considered relevant in the ranking process.

Research interests	(excellent)	(average)	(weak)
- Fit in research groups	○	○	○
- On the frontier of research	○	○	○
- Contributions	○	○	○
Academic background			
- University	○	○	○
- Grade average	○	○	○
- Time for acquiring degree	○	○	○
Letters of recommendation	Y	N	
Knowledge of English	Y	N	

Table 1 Evaluation sheet.

For simplicity we suppose that all applicants are *young* and have Master's degree acquired more than one year before. In this case all the criteria are meaningful, and are of approximately the same importance.

For applicants with Bachelor's degree the first three criteria *Fit in research groups*, *Contributions* and *On the frontier of research* are meaningless, because we have an undergraduate student without any research record. An applicant with Bachelor's degree or *just acquired* Master's degree should have excellent university record from a good university to be competitive.

For *old* applicants we encounter the problem of trade-offs between the age and the research record, and in this case their ratings on the last three criteria *University*, *Grade average* and *Time for acquiring degree* do not really matter. An *old* applicant should have a very good research record and a history of scientific cooperation with a TUCS research group to be competitive.

The third component is a group of 11 experts whose opinions are solicited in ranking the alternatives. The experts are selected from the following 9 research groups

- Algorithmics Group
- Coding Theory Group
- Computational Intelligence for Business
- Information Systems Research group

- Institute for Advanced Management Systems Research
- Probabilistic Algorithms and Software Quality
- Programming Methodology Group
- Strategic Information Systems Planning
- Theory Group: Mathematical Structures in Computer Science

So we have a **Multi Expert-Multi Criteria Decision Making** (ME-MCDM) problem. The ranking system described in the following is a two stage process. In the first stage, individual experts are asked to provide an evaluation of the alternatives. This evaluation consists of a rating for each alternative on each of the criteria, where the ratings are chosen from the scale $\{1, 2, 3\}$, where 3 stands for *excellent*, 2 stands for *average* and 1 means *weak* performance. Each expert provides a 6-tuple

$$(a_1, \ldots, a_6)$$

for each applicant, where $a_i \in \{1, 2, 3\}$, $i = 1, \ldots, 6$. The next step in the process is to find the overall evaluation for an alternative by a given expert.

In the second stage we aggregate the individual experts evaluations to obtain an overall value for each applicant.

3 OWA Operators

Ronald R. Yager [7] introduced a new aggregation technique based on the ordered weighted averaging operators.

Definition 3.1 *An OWA operator of dimension n is a mapping $F: R^n \rightarrow R$, that has an associated n vector*

$$w = (w_1, w_2, \ldots, w_n)^T$$

such as $w_i \in [0, 1]$, $1 \le i \le n$, and

$$\sum_{i=1}^{n} w_i = w_1 + \cdots + w_n = 1.$$

Furthermore

$$F(a_1, \ldots, a_n) = \sum_{j=1}^{n} w_j b_j = w_1 b_1 + \cdots + w_n b_n$$

where b_j is the j-th largest element of the bag $< a_1, \ldots, a_n >$.

A fundamental aspect of this operator is the re-ordering step, in particular an aggregate a_i is not associated with a particular weight w_i but rather a weight is associated with a particular ordered position of aggregate. When we view the OWA weights as a column vector we shall find it convenient to refer to the weights with the low indices as weights at the top and those with the higher indices with weights at the bottom.

It is noted that different OWA operators are distinguished by their weighting function. We point out three important special cases of OWA aggregations:

- *Max*: In this case $w^* = (1, 0 \dots, 0)^T$ and
$$Max(a_1, \dots, a_n) = \max\{a_1, \dots, a_n\}.$$
- *Min*: In this case $w_* = (0, 0 \dots, 1)^T$ and
$$Min(a_1, \dots, a_n) = \min\{a_1, \dots, a_n\}.$$
- *Average*: In this case $w_A = (1/n, \dots, 1/n)^T$ and
$$F_A(a_1, \dots, a_n) = \frac{a_1 + \dots + a_n}{n}$$

We can see the OWA operators have the basic properties associated with an *averaging operator* (commutative, monotonic and idempotent).

Compensative connectives have the property that a higher degree of satisfaction of one of the criteria can compensate for a lower degree of satisfaction of another criterion. *Oring* the criteria means full compensation and *anding* the criteria means no compensation. In order to classify OWA operators in regard to their location between *and* and *or*, Yager [7] introduced a measure of *orness*, associated with any vector w as follows

$$orness(w) = \frac{1}{n-1}\sum_{i=1}^{n}(n-i)w_i$$

It is easy to see that for any w the $orness(w)$ is always in the unit interval. Furthermore, note that the nearer w is to an *or*, the closer its measure is to one; while the nearer it is to an *and*, the closer is to zero. Generally, an OWA operator with much of nonzero weights near the top will be an *orlike* operator,

$$orness(w) \geq 0.5$$

and when much of the weights are nonzero near the bottom, the OWA operator will be *andlike*

$$andness(w) := 1 - orness(w) \geq 0.5.$$

The standard degree of orness associated with a Regular Increasing Monotone (RIM) linguistic quantifier Q

$$orness(Q) = \int_0^1 Q(r)\, dr$$

is equal to the area under the quantifier [13]. This definition for the measure of orness of quantifier provides a simple useful method for obtaining this measure.

Consider the family of RIM quantifiers

$$Q_\alpha(r) = r^\alpha, \ \alpha \geq 0. \tag{1}$$

It is clear that

$$orness(Q_\alpha) = \int_0^1 r^\alpha \, dr = \frac{1}{\alpha + 1}$$

and $orness(Q_\alpha) < 0.5$ for $\alpha > 1$, $orness(Q_\alpha) = 0.5$ for $\alpha = 1$ and $orness(Q_\alpha) > 0.5$ for $\alpha < 1$.

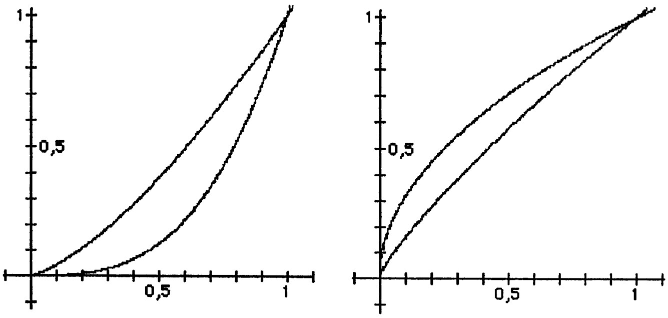


Figure 1 Risk averse and risk pro linguistic quanitfiers.

In [7] Yager suggested an approach to the aggregation of criteria satisfactions guided by a regular non-decreasing quintifier Q. If Q is RIM quantifier then we measure the overall success of the alternative $x = (a_1, \dots, a_n)$ by

$$F_Q(a_1, \dots, a_n)$$

where F_Q is an OWA operator derived from Q, i.e. the weights associated with this quantified guided aggregation are obtained as follows

$$w_i = Q(\frac{i}{n}) - Q(\frac{i-1}{n})$$

for $i = 1, \dots, n$.

Taking into consideration that we have 6 criteria (see Table 1) the weights derived from Q_α are determined as

$$w_1 = \left[\frac{1}{6}\right]^\alpha - 0, \quad w_2 = \left[\frac{2}{6}\right]^\alpha - \left[\frac{1}{6}\right]^\alpha, \quad w_3 = \left[\frac{3}{6}\right]^\alpha - \left[\frac{2}{6}\right]^\alpha$$

$$w_4 = \left[\frac{4}{6}\right]^\alpha - \left[\frac{3}{6}\right]^\alpha, \quad w_5 = \left[\frac{5}{6}\right]^\alpha - \left[\frac{4}{6}\right]^\alpha, \quad w_6 = 1 - \left[\frac{5}{6}\right]^\alpha$$

Furthermore, whatever is the linguistic quantifier, Q_α, representing the statement *most criteria are satisfied by* x, we see that

$$1 \leq F_\alpha(a_1, \dots, a_6) \leq 3$$

holds for each alternative $x = (a_1, \dots, a_6)$ since $a_i \in \{1, 2, 3\}$, $i = 1, \dots, 6$.

We search for an index $\alpha \geq 0$ such that the associated linguistic quantifier Q_α from the family (1) approximates the experts' preferences as much as possible. After interviewing the experts we found that all of them agreed on the following principles

(i) if an applicant has more than two weak performances then his overall performance should be less than two,

(ii) if an applicant has maximum two weak performances then his overall performance should be more than two,

(iii) if an applicant has all but one excellent performances then his overall performance should be about 2.75,

(iv) if an applicant has three weak performances and one of them is on the criterion *on the frontier of research* then his overall performance should not be above 1.5,

¿From (i) we get

$$F_\alpha(3, 3, 3, 1, 1, 1) = 3 \times (w_1 + w_2 + w_3) + w_4 + w_5 + w_6 < 2,$$

that is,

$$3 \times \left[\frac{3}{6}\right]^\alpha + 1 - \left[\frac{3}{6}\right]^\alpha < 2 \iff \left[\frac{1}{2}\right]^\alpha < \left[\frac{1}{2}\right] \iff \alpha > 1,$$

and from (ii) we obtain

$$F_\alpha(3, 3, 3, 2, 1, 1) = 3 \times (w_1 + w_2 + w_3) + 2 \times w_4 + w_5 + w_6 > 2$$

that is,

$$3 \times \left[\frac{3}{6}\right]^\alpha + 2 \times \left(\left[\frac{4}{6}\right]^\alpha - \left[\frac{3}{6}\right]^\alpha\right) + 1 - \left[\frac{4}{6}\right]^\alpha > 2 \iff \left[\frac{1}{2}\right]^\alpha + \left[\frac{2}{3}\right]^\alpha > 1$$

which holds if $\alpha < 1.293$. So from (i) and (ii) we get

$$1 < \alpha \leq 1.293,$$

which means that Q_α should be *andlike* (or risk averse) quantifier with a degree of compensation just below the arithmetic average.

It is easy to verify that (iii) and (iv) can not be satisfied by any quantifier Q_α, $1 < \alpha \leq 1.293$, from the family (1). In fact, (iii) requires that $\alpha \approx 0.732$ which is smaller than 1 and (iv) can be satisfied if $\alpha \geq 2$ which is bigger than 1.293. Rules (iii) and (iv) have priority whenever they are applicable.

In the second stage the technique for combining the expert's evaluation to obtain an overall evaluation for each alternative is based upon the OWA operators. Each applicant is represented by an 11-tuple

$$(b_1, \dots, b_{11})$$

where $b_i \in [1, 3]$ is the unit score derived from the i-th expert's ratings. We suppose that the b_i's are organized in descending order, i.e. b_i can be seen as the worst of the i-th top scores.

Taking into consideration that the experts are selected from 9 different research groups there exists no applicant that scores overall well on the first criterion "Fit in research group". After a series of negotiations all experts agreed that the support of at least four experts is needed for qualification of the applicant.

Since we have 11 experts, applicants are evaluated based on their top four scores

$$(b_1, \dots, b_4)$$

and if at least three experts agree that the applicant is excellent then his final score should be 2.75 which is a cut-off value for the best student. That is

$$F_\alpha(3, 3, 3, 1) = 3 \times (w_1 + w_2 + w_3) + w_4 = 2.75,$$

that is,

$$3 \times \left[\frac{3}{4}\right]^\alpha + 1 - \left[\frac{3}{4}\right]^\alpha = 2.75 \iff \left[\frac{3}{4}\right]^\alpha = 0.875 \iff \alpha \approx 0.464$$

So in the second stage we should choose an *orlike* OWA operator with $\alpha \approx 0.464$ for aggregating the top six scores of the applicant to find the final score.

If the final score is less than 2 then the applicant is disqualified and if the final score is at least 2.5 then the scholarship should be awarded to him. If the final score is between 2 and 2.5 then the scholarship can be awarded to the applicant pending on the total number of scholarships available.

Example 1 *Let us choose $\alpha = 1.2$ for the aggregation of the ratings in the first stage. Consider some applicant with the following scores*

Criteria	C_1	C_2	C_3	C_4	C_5	C_6
Expert 1	3	2	3	2	3	1
Expert 2	2	3	3	2	3	2
Expert 3	2	2	3	2	2	1
Expert 4	3	2	3	3	3	2
Expert 5	2	2	3	2	3	1
Expert 6	3	2	3	2	3	1
Expert 7	1	2	3	2	3	2
Expert 8	1	2	3	2	3	1
Expert 9	1	2	2	2	3	2
Expert 10	1	2	2	3	3	1
Expert 11	1	2	2	2	2	1

The weights associated with this linguistic quantifier are

$$(0.116, 0.151, 0.168, 0.180, 0.189, 0.196)$$

After re-ordering the scores in descending order we get the following table

							Unit score
Expert 1	3	3	3	2	2	1	2.239
Expert 2	3	3	3	2	2	2	2.435
Expert 3	3	2	2	2	2	1	1.920
Expert 4	3	3	3	3	2	2	2.615
Expert 5	3	3	2	2	2	1	2.071
Expert 6	3	3	3	2	2	1	2.239
Expert 7	3	3	2	2	2	1	2.071
Expert 8	3	3	2	2	1	1	1.882
Expert 9	3	2	2	2	2	1	1.920
Expert 10	3	3	2	2	1	1	1.882
Expert 11	2	2	2	2	1	1	1.615

In the second stage we choose $\alpha = 0.464$ and obtain the following weights

$$(0.526, 0.199, 0.150, 0.125).$$

The best four scores of the applicant are

$$(2.615, 2.435, 2.239, 2.239).$$

The final score is computed as

$$F_\alpha(2.615, 2.435, 2.239, 2.239) = 2.475.$$

So the applicant has good chances to get the scholarship.

4 Summary and Conclusions

We have presented a two stage process for doctoral student selection problem. In the first stage we have used an *andlike* OWA operator to implement some basic rules derived from certain (extremal) situations. In the second stage we have applied an *orlike* OWA operator, because the final score of applicants should be high if at least three experts find his record attractive (we do not require support from *all experts*).

It can happen (and it really happened) that some experts (a minority) forms a coalition and deliberately *overrate* some candidates in order to qualify them even though the majority of experts finds these candidates overall weak. We can resolve this problem by adding an extra criterion to the set of criteria measuring the competency of individual experts, or we issue an alarm message about the attempted cheating.

To determine the most appropriate linguistic quantifier in the first stage we can also try to *identify* interdependences between criteria [1, 2, 3].

References

[1] C. Carlsson and R. Fullér, On fuzzy screening system, in: *Proceedings of EUFIT'95 Conference*, August 28-31, 1995 Aachen, Germany, Verlag Mainz, Aachen, 1995 1261-1264.

[2] C.Carlsson and R.Fullér, On linear interdependences in MOP, in: *Proceedings of CIFT'95*, June 8-10, 1995, Trento, Italy, University of Trento, 1995 48-52.

[3] C.Carlsson and R.Fullér, Compound interdependences in MOP, in: *Proceedings of EUFIT'96 Conference*, September 2-5, 1996, Aachen, Germany, (to appear).

[4] J.C. Fodor and M. Roubens, *Fuzzy Preference Modelling and Multicriteria Decision Support*, Theory and Decision Library, Series D, system Theory, Knowledge Engineering and Problem Solving, Vol. 14, Kluwer, Dordrecht, 1994.

[5] J. Kacprzyk and R.R. Yager, Using fuzzy logic with linguistic quantifiers in multiobjective decision making and optimization: A step towards more human-consistent models, in: R.Slowinski and J.Teghem eds., *Stochastic versus Fuzzy Approaches to Multiobjective Mathematical Programming under Uncertainty*, Kluwer Academic Publishers, Dordrecht, 1990 331-350.

[6] R.R. Yager, Fuzzy decision making using unequal objectives, *Fuzzy Sets and Systems*, 1(1978) 87-95.

[7] R.R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making, *IEEE Transactions on Systems, Man and Cybernetics* 18(1988) 183-190.

[8] R.R.Yager, Fuzzy Screening Systems, in: R.Lowen and M.Roubens eds., *Fuzzy Logic: State of the Art*, Kluwer, Dordrecht, 1993 251-261.

[9] R.R.Yager, Families of OWA operators, *Fuzzy Sets and Systems*, 59(1993) 125-148.

[10] R.R.Yager, Aggregation operators and fuzzy systems modeling, *Fuzzy Sets and Systems*, 67(1994) 129-145.

[11] R.R.Yager, On weighted median aggregation, *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 1(1994) 101-113.

[12] R.R.Yager, Constrained OWA aggregation, Technical Report, #MII-1420, Machine Intelligence Institute, Iona College, New York, 1994.

[13] R.R.Yager, Quantifier guided aggregation using OWA operators, Technical Report, #MII-1504, Machine Intelligence Institute, Iona College, New York, 1994.

5

OWA OPERATORS IN MULTICRITERIA AND MULTIPERSON DECISION MAKING

BEYOND MIN AGGREGATION IN MULTICRITERIA DECISION: (ORDERED) WEIGHTED MIN, DISCRI-MIN, LEXIMIN[1]

Didier DUBOIS, Hélène FARGIER and Henri PRADE
Institut de Recherche en Informatique de Toulouse (IRIT) – CNRS, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France, Email: {dubois, fargier, prade}@irit.fr

ABSTRACT

Conjunctive aggregation based on min operation provides too crude a ranking of the possible alternatives in multiple criteria aggregation, since decisions are only compared on the basis of the worst-rated criteria, and also since the levels of importance of the different criteria or constraints are not taken into account. Various types of weighted min operations are distinguished. Two refinements of the min-based ordering (and of the Pareto ordering which corresponds to a fuzzy set inclusion) are presented and relations between them are laid bare. These refinements aim to increase the discriminating power of the min-based aggregation, yet keeping its non-compensatory nature. A relationship between the leximin ordering and ordered weighted averages (OWA) is also discussed. Lastly, ordered weighted min operations are introduced and are shown to be of interest when only most of the criteria have to be taken into account in the evaluation.

1 - INTRODUCTION – NOTATIONS

The notion of a fuzzy constraint as introduced by Bellman and Zadeh (1970) intends to represent constraints as well as criteria by fuzzy subsets C_i of the set S of possible decisions. Viewed as a constraint, C_i is called a fuzzy constraint and $\mu_{C_i}(s_u) = 1$ then means that decision $s_u \in S$ totally satisfies the constraint C_i, while $\mu_{C_i}(s_u) = 0$ means that it totally violates C_i (s_u is unacceptable). If $0 < \mu_{C_i}(s_u) < 1$, s_u satisfies C_i only partially. Hence, like an objective function, a fuzzy constraint rank-orders the feasible decisions. However, contrary to an objective function, a

[1] This chapter is a revised and restructured version of a paper entitled "Ordered weighted operations, "discri-min" and leximin in multicriteria decision", presented at the 2nd World Automation Congress (WAC'96) in the framework of the 1st International Symposium on Soft Computing for Industry (ISSCI'96) on May 27-30, 1996 in Montpellier (France).

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

fuzzy constraint also models a threshold beyond which a solution will be rejected. In fact, a fuzzy constraint can be viewed as the association of a constraint (defining the support of C_i) and a criterion which rank-orders the solutions satisfying the constraint, according to preference. However, in contrast with objectives which are usually allowed to compensate each other, fuzzy constraints are combined in a conjunctive way.

The min operation is the greatest associative and monotonic operation which can model conjunction in a logical sense. Using min, the global level of satisfaction of a set of fuzzy constraints is the level of satisfaction of the least satisfied one(s). As a result, many possible decisions may receive the same evaluation, although it would be possible to distinguish between them by considering the levels of satisfaction of other constraints. In this perspective, two successive refinements of min operation are presented in Section 3 ("Discri-Min") and Section 4 ("Leximin"). Refinements of min-based orderings have been felt necessary in different areas such as cooperative decision-making (Moulin, 1988), reasoning under inconsistency (Dubois et al., 1992), and more recently in the processing of flexible queries to a database (Ribeiro and Muntz, 1996).

It may also happens in practice that no possible decision satisfies all the constraints to a positive degree (see, e.g., Felix, 1992), i.e., the result of min-aggregation is zero for any possible decisions. In such a case a relaxation of the min-aggregation is a natural way for coping with this problem; see Section 2 for various ways of weighting constraints and relaxing them, and Section 5 for the use of ordered weighted min operations when only the satisfaction of most constraints is required.

An obvious advantage of the min-operation, which will be preserved in the various refinements and extensions presented, is that it can be defined on any totally ordered scale L, either infinite as [0,1], or made of a finite number of levels. In the following, we use L = [0,1] for simplicity (except where the contrary is explicitly stated).

Let us first introduce some notations. Let us denote by $\mathcal{C} = \{C_1, \ldots, C_n\}$ the set of fuzzy constraints involved in a decision problem. The set of constraints satisfied by a solution s_u, denoted by $\mathcal{S}_u$, is the fuzzy subset of $\mathcal{C}$ defined by its membership function $\mu_{\mathcal{S}_u}(C_i) = u_i = \mu_{C_i}(s_u) \in [0,1]$. The vector of satisfaction levels of a solution s_u will be denoted $\vec{u} = (u_1, \ldots, u_n)$. It provides the same information as $\mathcal{S}_u$.

According to Bellman and Zadeh's approach to decision making, the best decisions are those maximizing the satisfaction degree of the least satisfied constraint. They are called min-optimal solutions. As they notice, the global ordering over S is a complete preordering which corresponds to a fuzzy set C, which is nothing but the intersection $C = C_1 \cap \ldots \cap C_n$ of the fuzzy constraints, defined by

$$\mu_C(s_u) = \min_{i=1,n} \mu_{C_i}(s_u).$$

Another natural way of ranking the solutions is given by the well-known Pareto ordering. In the fuzzy context, *the Pareto-ordering corresponds to a fuzzy set inclusion*:

• $\vec{u} \geq_{\text{Pareto}} \vec{v} \Leftrightarrow \mathcal{S}_v \subseteq \mathcal{S}_u \Leftrightarrow \mu_{\mathcal{S}_v} \leq \mu_{\mathcal{S}_u} \Leftrightarrow \forall i, \mu_{C_i}(s_v) \leq \mu_{C_i}(s_u)$,

s_u is Pareto-optimal iff $\nexists\, s_v \in S$ such that $\vec{u} \neq \vec{v}$ and $u_i \leq v_i$ for $i = 1,n$. Clearly, the Pareto-ordering is a partial ordering: many decisions can be incomparable. It is worth noticing that the min-ordering defined by $\vec{u} \succcurlyeq_{\min} \vec{v}$ if and only if $\min_i u_i \geq \min_i v_i$, satisfies a weak Pareto principle (if $u_i \geq v_i$ for $i = 1, n$, then $\vec{u} \succcurlyeq_{\min} \vec{v}$), but there are Pareto-optimal decisions which are not min-optimal, and min-optimal decisions which are not Pareto-optimal.

2 - TWO TYPES OF WEIGHTED MIN

Like averaging operations, min operations (as well as the DeMorgan dual disjunctive connective 'max') can be weighted. Weighting constraints in order to acknowledge their respective levels of importance may also help breaking ties between possible decisions (especially when $\forall s_u$, $\min_i \mu_{C_i}(s_u) = 0$). Importance may be a matter of degree and is not relative to the comparison of a pair of decisions, but is an attribute attached to each constraint. Relaxing constraints by assuming that all the constraints have not the same importance can be done in different ways in the context of conjunctive aggregation.

Let $w_i \in [0,1]$ be the importance level of fuzzy constraint C_i. $w_i = 0$ means that the constraint can be forgotten. The convention $\max_{i=1,n} w_i = 1$ is assumed in order to have an appropriate scaling of the levels of importance. A natural requirement is that the result of the weighing is 0 if $w_i = 1$ and $u_i = \mu_{C_i}(s_u) = 0$, i.e., if constraint C_i is not satisfied at all but is fully important. Moreover if $w_i = 0$ (C_i has no importance at all), the result of the weighting should be 1, which is the neutral element for min operation. Lastly, if $w_i = 1$ and $u_i = 1$, the result of the weighting should be 1 again, since C_i is completely satisfied. Denoting by $w_i \perp u_i$ the result of the weighting, we should have more generally $\forall w_i$, $w_i \perp u_i = 1$ if $u_i = 1$, and $\forall u_i$, $w_i \perp u_i = 1$ if $w_i = 0$. Thus we should have

$$(\forall i, w_i > 0 \Rightarrow \mu_{C_i}(s_u) = 1) \Rightarrow \mu_C(s_u) = 1. \tag{1}$$

All these requirements for $\perp$ are satisfied by multiple-valued implications, i.e., $w_i \perp u_i = w_i \rightarrow u_i$; Yager (1984) already noticed the role of implications in the weighting process. Depending on the choice of the implication, various types of weighting are obtained; in the following, we only consider the ones, associated with min operation, which are compatible with the idea of a purely ordinal scale.

It may be found natural to require that the complete satisfaction of a set of fuzzy constraints by a possible decision s_u demands that all the constraints in $\mathcal{S}$, whatever their (non-zero) importance, be completely satisfied, i.e., that (1) holds as an equivalence:

$$\mu_C(s_u) = 1 \Leftrightarrow (\forall i,\ w_i > 0 \Rightarrow \mu_{C_i}(s_u) = 1). \qquad (2)$$

Moreover, we may require that $\mu_C(s_u) = 0$ *only if* $\exists i$ such that *both* $w_i = 1$ (the constraint has the maximum level of importance) and $\mu_{C_i}(s_u) = 0$ (the decision does not at all fulfil the constraint). Thus, $\mu_C(s_u)$ will be allowed to be strictly positive if $\mu_{C_i}(s_u) = 0$ provided that $w_i < 1$, i.e., the constraint is not completely important (since then the constraint can be forgotten to some extent). An example of aggregation operation meeting these requirements is (Dubois and Prade, 1986):

$$\mu_C(s_u) = \min_i \max(1 - w_i, \mu_{C_i}(s_u)). \qquad (3)$$

Note that now $\mu_C(s_u) \geq \min_{i=1,n} \mu_{C_i}(s_u)$, which indeed expresses a relaxation. This expression is a conjunction of the degrees of satisfaction $\mu_{C_i}(s_u)$ weighted by the levels of importance w_i (Dubois et al., 1988). Min-aggregation is recovered when $w_i = 1, \forall i$. Clearly, C_i is ignored when $w_i = 0$ as expected. Denoting by $\mathcal{W}$ the fuzzy set of important constraints (i.e., $\mu_{\mathcal{W}}(C_i)= w_i$), (2) expresses a strong type of inclusion of $\mathcal{W}$ into $\mathcal{S}_u$ (where $\mu_{\mathcal{S}_u}(C_i) = \mu_{C_i}(s_u)$), based on Dienes implication $a \rightarrow b = \max(1 - a, b)$.

This form of weighted minimum as given by (3) can be formally justified by the following axioms:

a) weighting the constraints produces a relaxing effect, i.e., we have the condition $\min_i w_i \perp u_i \geq \min_i u_i$. This condition is verified for any w_i and u_i provided that $w_i \perp u_i \geq u_i$;
b) limit cases: $1 \perp u_i = u_i$ and $0 \perp u_i = 1$;
c) $w_i \perp 0$ is a strictly decreasing function of w_i. Namely the more prioritary the constraint, the more its violation decreases the overall satisfaction level. In particular, due to b) $1 \perp 0 = 0$ and $0 \perp 0 = 1$. So $w_i \perp 0$ is of the form $n(w_i)$ where n is an order-reversing function (a negation), and $w_i \perp 0 > 0$ if $w_i < 1$;
d) $w_i \perp u_i$ is an increasing function of u_i, in the wide sense. Indeed the more satisfied the constraint, the more feasible the solution. Hence $w_i \perp u_i \geq w_i \perp 0$.

Putting together (a) and (d) leads to $w_i \perp u_i \geq \max(w_i \perp 0, u_i)$. It shows that the weighting scheme (3) is the lower bound of any form of weighted minimum satisfying natural axioms stated above, if we let $w_i \perp 0 = 1 - w_i$. In particular, any triangular conorm may be chosen for aggregating $w_i \perp 0$ and u_i (for instance $a + b - ab$ is particularly suitable for numerical ratings). So $\perp$ should be a S-implication in the sense of Trillas and Valverde (1985).

As shown by (3), even if C_i is completely violated, the global level of satisfaction is only upper bounded by $1-w_i$ because of this violation (instead of 0 in case C_i is imperative). In practice, it may be useful in this case to add a "safeguard" constraint $\hat{C}_i$ in case some decision among the ones which violates C_i are really inacceptable, thus changing (3) into

$$\mu_C(s_u) = \min_i \min(\mu_{\hat{C}_i}(s_u), \max(1 - w_i, \mu_{C_i}(s_u))). \tag{4}$$

The above weighting scheme relaxes constraints by considering that a non-prioritary constraint cannot be fully violated. This is expressed in axiom (c) whereby $w_i \perp 0 > 0$ when $w_i < 1$. There is a different way of relaxing the constraints, namely by considering that the constraint is sufficiently satisfied if the level of satisfaction for s_u reaches some threshold θ_i, i.e., $\mu_{C_i}(s_u) \geq \theta_i$. Then $\mu_{C_i}(s_u)$ will be changed into $\mu_{C'_i}(s_u) = 1$ in the min-aggregation process and C_i will not be taken into account further. If $\mu_{C_i}(s_u) < \theta_i$ we may either consider that the constraint is satisfied at the level which is reached, i.e., $\mu_{C'_i}(s_u) = \mu_{C_i}(s_u)$, or in order to avoid discontinuity, make $\mu_{C'_i}(s_u)$ equal to the *relative* level of satisfaction $\mu_{C_i}(s_u) / \theta_i$ (which requires a numerical scale like [0,1] and not a simple completely ordered scale). Then we have

$$\mu_C(s_u) = \min_{i=1,n} \mu_{C'_i}(s_u) = \min_{i=1,n} \theta_i \rightarrow \mu_{C_i}(s_u) \tag{5}$$

where $a \rightarrow b$ is Gödel implication ($a \rightarrow b = 1$ if $a \leq b$, $a \rightarrow b = b$ if $a > b$) in the first case, and Goguen's in the second one, namely $a \rightarrow b = \min(1, b/a)$ if $a \neq 0$ and $a \rightarrow b = 1$ if $a = 0$. We still have $\mu_C(s_u) \geq \min_{i=1,n} \mu_{C_i}(s_u)$. However, it contrasts with the use of Dienes implication in (3), and (2) is weakened into

$$\mu_C(s_u) = 1 \Leftrightarrow \forall i, \mu_{C_i}(s_u) \geq \theta_i \tag{6}$$

which expresses an ordinary fuzzy set inclusion of $\mathcal{W}$ into $\mathcal{S}_u$. Here C_i is "forgotten" as soon as $\mu_{C_i}(s_u) \geq \theta_i$. For $\theta_i = 1$, $C'_i = C_i$ is recovered. Weighted minimum schemes where the weights act as membership thresholds satisfy axioms (a), (b), (d), but (c) is replaced by

c') if $\theta_i \leq u_i$ then $\theta_i \perp u_i = 1$.

It is easy to verify that $\theta_i \perp u_i = \sup\{x, \theta_i * x \leq u_i\}$ (where $*$ is increasing in both places, $\theta_i * x \leq \min(\theta_i, x)$, $\theta_i * 1 = \theta_i$, and $1 * x = x$) verifies axioms (a), (b), (c'), (d). Hence $\perp$ can be chosen as a residuated implication with respect to a triangular norm (Trillas and Valverde, 1985).

We might also think of using in (5) the implication obtained by contraposition from Gödel implication, letting $\mu_{C'_i}(s_u) = 1 - \theta_i$ if the threshold is not reached

($\mu_{C_i}(s_i) < \theta_i$). But this no longer satisfies the requirement (a) stating that $\mu_C(s_u) \geq \min_{i=1,n} \mu_{C_i}(s_u)$.

3 - "DISCRI-MIN"

The Pareto ordering corresponds to the notion of fuzzy set inclusion and is thus not a refinement of the min-ordering. Moreover the discrimination power of the min-ordering is low as already said, while the Pareto-ordering leads to incomparabilities. Keeping in mind the idea of inclusion, we propose to define a refinement of both these orderings by comparing the decisions on the basis of level-cut inclusion. Hence, we consider a new ordering (a new preference relation), $>_{LSDC}$, first introduced in constraint satisfaction problems by Fargier et al. (1993), and defined in the finite case by: $\vec{u} >_{LSDC} \vec{v}$ iff $\exists\ \alpha \in L$ s.t.

$$\text{(i) } \forall\ \beta \in \text{s.a. } \beta < \alpha,\ (\mathcal{S}_u)_\beta = (\mathcal{S}_v)_\beta \quad ; \quad \text{(ii) } (\mathcal{S}_u)_\alpha \supset (\mathcal{S}_v)_\alpha,$$

where $(\mathcal{S}_w)_\gamma = \{C_i \in \mathcal{C} \mid \mu_{C_i}(s_w) \geq \gamma\}$ is a level-cut of $\mathcal{S}_w$. LSDC stands for 'least satisfied discriminating constraint'. The relation $>_{LSDC}$ is irreflexive and transitive: it is a strict partial ordering. It can be verified that it is a refinement of $>_{min}$ and of the Pareto ordering. The use of $>_{LSDC}$ consists in practice in comparing pointwise satisfaction degrees like in $>_{Pareto}$, and focusing on the lowest satisfaction degrees among the constraints satisfied *at different degrees* by the competing decisions: decisions are compared on the basis of the least satisfied discriminating constraints. $>_{LSDC}$ ordering will be referred to as "discri-min", for the sake of brevity.This seems to be in accordance with the intuition, since only the constraints making a difference between two alternatives can help make a final decision. This ordering of solutions was actually first proposed by Behringer (1977) with a totally different definition (see the discussion by Dubois and Fortemps (1996)).

Let us note $\mathcal{D}(u,v) = \{C_i \in \mathcal{C} \mid u_i \neq v_i\}$ the set of constraints which are satisfied by s_u and s_v to a different extent. An equivalent characterization of $>_{LSDC}$ has been proposed by (Fargier et al., 1993):

$$\vec{u} >_{LSDC} \vec{v} \Leftrightarrow \min_{C_i \in \mathcal{D}(u,v)} u_i > \min_{C_i \in \mathcal{D}(u,v)} v_i \qquad (7)$$

The "discri-min" refinement is based on the idea that the constraints on which two decisions receive the same evaluation, have no importance when comparing the decisions. Indeed, it can be easily checked that the same ordering between $\vec{u}$ and $\vec{v}$ is obtained by comparing $\min_{i=1,n} w_i \rightarrow \mu_{C_i}(s_u)$ and $\min_{i=1,n} w_i \rightarrow \mu_{C_i}(s_v)$ where the w_i's are such that $w_i = 0$ if $\mu_{C_i}(s_u) = \mu_{C_i}(s_v)$ and $w_i = 1$ otherwise, where Dienes, or Gödel, implication is used (since in both cases, $w_i \rightarrow u_i = u_i$ if $w_i = 1$).

It is worth noticing that this second formulation of $>_{LSDC}$ implies that this partial ordering cannot be represented by a numerical function. Indeed, suppose there is a function f such that $\vec{u} >_{LSDC} \vec{v} \Leftrightarrow f(\vec{u}) > f(\vec{v})$. Then clearly the relation $u \; R \; v \Leftrightarrow f(\vec{u}) = f(\vec{v})$ is an equivalence relation. It expresses that neither $\vec{u} >_{LSDC} \vec{v}$ nor $\vec{v} >_{LSDC} \vec{u}$. But this relation, although equivalent to $\min_{C_i \in \mathcal{D}(u,v)} u_i = \min_{C_i \in \mathcal{D}(u,v)} v_i$, is not an equivalence relation (it is not transitive) contrary to R. Obviously there exists a 2-argument function $M(\vec{u}, \vec{v})$ such that $\vec{u} >_{LSDC} \vec{v} \Leftrightarrow M(\vec{u}, \vec{v}) > M(\vec{v}, \vec{u})$. We argue that $>_{LSDC}$, which is technically a stratified inclusion (inclusion of level-cuts) actually relates to a fuzzy set inclusion index. Indeed, we have the following result:

$$\vec{u} >_{LSDC} \vec{v} \Leftrightarrow \min_{C_i \in \mathcal{C}} \mu_{\mathcal{S}_u}(C_i) \to \mu_{\mathcal{S}_v}(C_i) < \min_{C_i \in \mathcal{C}} \mu_{\mathcal{S}_v}(C_i) \to \mu_{\mathcal{S}_u}(C_i)$$

where $\to$ stands for Gödel's implication, which is defined by: $a \to b = 1$ if $a \le b$, $a \to b = b$ otherwise. The proof of this proposition is given in (Dubois et al., 1996). This Gödel-implication-based index is also proposed by De Baets (1994) to compare solutions in multi-criteria decision making problems. However, he exploits the level cuts of the fuzzy relation induced on the solution set by the inclusion index $I(\mathcal{S}_u, \mathcal{S}_v) = \min_{C_i \in \mathcal{C}} \mu_{\mathcal{S}_u}(C_i) \to \mu_{\mathcal{S}_v}(C_i)$, while the LSDC ordering coincides with the crisp relation obtained by comparing $I(\mathcal{S}_u, \mathcal{S}_v)$ and $I(\mathcal{S}_v, \mathcal{S}_u)$.

4 - LEXIMIN AS AN OWA

Moulin (1988) has proposed another refinement of the egalitarist maximin ordering: the leximin ordering. The idea is to keep all the information pertaining to $\mathcal{S}_u$, instead of summarizing it by a collective utility function as with $>_{min}$, by considering a ranked multi-set of satisfaction degrees. Let $\vec{u}$ be a vector of satisfaction degrees, and denote by $\vec{u}^*$ another such vector such that $u^*_i = u_{\gamma(i)}$ where γ is a permutation such that $u_{\gamma(1)} \le u_{\gamma(2)} \le \ldots \le u_{\gamma(n)}$ obtained by reordering the components of $\vec{u}$. Indeed, two multi-sets built on an ordered referential can always be compared using the leximin ordering $>_{Lex}$:

- $\vec{u}^* >_{Lex} \vec{v}^*$ iff $\exists\, k \le n$ such that $\forall\, i < k$, $u^*_i = v^*_i$ and $u^*_k > v^*_k$

where u_i^* (resp. v_i^*) stands for the i-th component of $\vec{u}^*$ (resp. $\vec{v}^*$). The two possible decisions are indifferent if the corresponding multi-sets are the same.

As shown by Moulin (1988), the leximin-ordering is a refinement of both the Pareto-ordering and the min-ordering:

- $\vec{u} >_{min} \vec{v} \Rightarrow \vec{u} >_{Lex} \vec{v}$
- $\vec{u} >_{Pareto} \vec{v} \Rightarrow \vec{u} >_{Lex} \vec{v}$.

But it is NOT a refinement of the utilitarist ordering based on a sum of satisfaction levels: a solution $\vec{u}$ whose least satisfied constraint is less satisfied than the least satisfied constraint of another one $\vec{v}$ will not be leximin-preferred, regardless of the sum of the other partial satisfaction degrees. The leximin ordering is still egalitarist. Finally, Moulin (1988) has shown that this ordering cannot be represented in general by a single function, despite the fact that it is a complete preordering. However we shall see at the end of this section that this can be done, if we restrict ourselves to a finite scale in a numerical setting.

Leximin optimal decisions are always LSDC-optimal decisions, and thus indeed min-optimal and Pareto-optimal: $>_{Lex}$ is the most selective among these preference relations. Reciprocal implications are not verified. The leximin ordering can discriminate among equivalent decisions according to the min-preordering ($\min u_i = \min v_i$) and among incomparable ones with respect to $>_{Pareto}$ and $>_{LSDC}$. This kind of lexicographic notion of optimum originates in approximation problems where the Tschebyscheff distance is used (Rice, 1962).

A representation theorem for the leximin ordering by means of a 2-place numerical function using fuzzy cardinality notions is established in (Dubois et al., 1996):

$$\vec{u} >_{lex} \vec{v} \Leftrightarrow \max \{\alpha \,/\, |(\mathcal{S}_u{}^c)_\alpha| < |(\mathcal{S}_v{}^c)_\alpha|\} > \max \{\alpha \,/\, |(\mathcal{S}_v{}^c)_\alpha| < |(\mathcal{S}_u{}^c)_\alpha|\}. \quad (8)$$

where $\mathcal{S}_u{}^c$, the complement of $\mathcal{S}_u$, is the fuzzy set of constraints violated by s_u. *Viewing the leximin ordering in the context of fuzzy set theory, we can thus interpret it in terms of fuzzy cardinality.* Note that saying that the leximin-ordering tends to favor solutions that violate as few fuzzy constraints as possible is not equivalent to favoring solutions that satisfy as many fuzzy constraints as possible (while these two notions would be equivalent in the crisp case). The latter option corresponds to a leximax-ordering (Dubois et al., 1996).

As we are going to show, it is possible to encode the leximin ordering in terms of a particular ordered weighted average (OWA) operation, if we only use a finite number of satisfaction levels inside [0,1]. This result has been independently obtained in a slightly different way by Yager (1996). Let us first briefly recall the definition of an OWA (Yager, 1988). Let $\vec{u} = (u_1, \ldots, u_n) \in [0,1]^n$ a n-tuple. Let $\vec{w} = (w_1, \ldots, w_n) \in [0,1]^n$ be a n-tuple of weights such that $\sum_{i=1,n} w_i = 1$. Then $OWA_{\vec{w}}(\vec{u})$ is defined by the two-step procedure:

i) rank-order the u_i's increasingly, i.e., $u_{\gamma(1)} \leq \ldots \leq u_{\gamma(n)}$ where γ is a permutation;

ii) $OWA_{\vec{w}}(\vec{u}) = \sum_{j=1,n} w_j \cdot u^*_j$ with $u^*_j = u_{\gamma(j)}$.

N.B.: In Yager (1988) the u_i's are ranked decreasingly. But this is just a matter of convention. The same class of ordered weighted averages is captured.

Let us first explain the idea of capturing the leximin ordering by an OWA with the help of an example. Let us suppose that we only use the 10 levels 0, .1, .2, .3, .4,

.5, .6, .7, .8, .9 in [0,1], for instance. It is clear that comparing two vectors like for example $\vec{u}^* = (.2,.3,.5,.6)$ and $\vec{v}^* = (.2,.3,.3,.9)$ and concluding that $\vec{u}^* >_{lex} \vec{v}^*$ (or equivalently $\vec{u} >_{lex} \vec{v}$) is the same as comparing the numbers 2356, which corresponds to the weighted sum $.2 \cdot 10^4 + .3 \cdot 10^3 + .5 \cdot 10^2 + .6 \cdot 10$, with the number 2339! In order to remain inside the interval [0,1], we divide here by 10^4, i.e., $0.2356 = 0.2 \cdot 1 + 0.3 \cdot 10^{-1} + 0.5 \cdot 10^{-2} + .6 \cdot 10^{-3}$.

The weights of this weighted sum belong now to [0,1]. In order to have a genuine OWA operation with $\sum_{i=1,n} w_i = 1$, we have just to multiply by an appropriate constant, i.e., if we deal with n-tuples, $w_i = \alpha \cdot 10^{-i+1}$, $i = 1,n$ and $\alpha = 1 / (\sum_{i=1,n} 10^{-i+1})$. Capturing the leximax ordering only requires the same weights be used in the reversed order.

This can be readily generalized to any finite scale extracted from [0,1] where ε is the interval between two successive levels (here $\varepsilon = 0.1$), by letting $w_i = \alpha' \cdot \varepsilon^{-i+1}$ where α' is the normalization constant. Note that here the length of the interval between two successive levels inside [0,1] can be always made constant since leximin is only sensitive to the orderings. By this trick it is always possible to find an OWA operation, i.e., to compute $\vec{w}$, such that $\vec{u} <_{lex} \vec{v} \Leftrightarrow \mathrm{OWA}_{\vec{w}}(\vec{u}) < \mathrm{OWA}_{\vec{w}}(\vec{v})$.

It can be seen that the ordering remains unchanged when the w_i's are computed by using ε_0 with $0 < \varepsilon_0 < \varepsilon$ in place of ε. Thus the ordering becomes stable for some ε small enough, and coincides with to the leximin ordering when $\varepsilon \to 0$.

5 - ORDERED WEIGHTED MIN

Conjunctive aggregation can be relaxed by only requiring the satisfaction of k over n constraints, or more generally, the satisfaction of most constraints. By "delocalizing" the weights w_i in (3), we can turn (3) into a (fuzzily) quantified conjunction, corresponding to the requirement that a decision s_u satisfies 'at least k', or more generally 'most' constraints in $\mathcal{C}$ (rather than 'all' the constraints or more generally all the important constraints). This can be done in the following way (see, e.g., Dubois et al., 1988):

i) rank-order the degree $\mu_{C_i}(s_u) = u_i$ *decreasingly*, where σ is a permutation of $\{1, \ldots, n\}$ in order to only consider the best satisfied constraints in the weighting process, i.e., $u_{\sigma(1)}) \geq u_{\sigma(2)} \geq \ldots \geq u_{\sigma(n)}$;
ii) let I be a fuzzy subset of the set of integers $\{0, 1, 2, \ldots, n\}$ s.t. $\mu_I(0) = 1$, $\mu_I(i) \geq \mu_I(i + 1)$. For instance, the requirement that "at least k" constraints are important will be modelled by k weights equal to 1, i.e., $w_i = \mu_I(i)$ in (3) with $\mu_I(i) = 1$ if $0 \leq i \leq k$, $\mu_I(i) = 0$ for $i \geq k + 1$ in order to express that k

constraints are important ($\mu_I(i)$ is going to be used as the weight associated with the i-th most satisfied constraint) ;

iii) the aggregation operation is then defined by

$$\mu_C(s_u) = \min_{i=1,n} \max(1 - \mu_I(i), u_{\sigma(i)}). \tag{9}$$

Thus, the satisfying of a constraint is all the less important (prioritary) as a sufficient number of other constraints are already satisfied to a high degree. When $\mu_I(i) = 1$ for $0 \leq i \leq n$, (9) reduces to $\mu_C(s_u) = u_{\sigma(n)} = \min_{i=1,n} \mu_{C_i}(s_u)$ as expected. When $\mu_I(1) = 1$ and $\mu_I(2) = \ldots = \mu_I(n) = 0$, (9) reduces to $u_{\sigma(1)} = \max_{i=1,n} \mu_{C_i}(s_u)$.

The expression (9) can be easily modified for accommodating relative quantifiers Q like 'most', by changing $1 - \mu_I(i + 1)$ into $\mu_Q\left(\frac{i}{n}\right)$ for $i = 0, n - 1$ and $\mu_Q(1) = 1$ where μ_Q is increasing (a required proportion of at least $\frac{k}{n}$ amounts to have k non-zero weights among n). What has been computed here is an *ordered weighted minimum* operation (OWmin). It has been proved that it comes down to computing the median of the set of numbers made by the $u_{\sigma(i)}$'s and the $1 - \mu_I(i)$'s. See Dubois and Prade (1986). OWmin can thus be related to the idea of fuzzy cardinality, at it is the case for leximin or OWA aggregation. However, a single OWmin operation, in contrast with OWA operations, cannot capture the leximin ordering. As a median, an OWmin operation can be viewed as a Sugeno integral (see, e.g., Dubois and Prade (1980) for Sugeno integrals as medians), just as OWA's are Choquet integrals (Grabisch, 1995).

6 - CONCLUSION

Conjunctive aggregation plays a central role in constraint satisfaction problems (CSP) such as design, planning, scheduling problems (Dubois et al., 1995), and in the expression of compound queries to databases (e.g., Dubois et al., 1988)). The various refinements and extensions of the min-operator are then very useful for handling fuzzy constraints. Best solutions of fuzzy CSP's can be found using extensions of classical CSP methods (Dubois et al., 1994; Fargier et al., 1995; Dubois and Fortemps, 1996).

REFERENCES

Behringer, F. (1977) "On optimal decisions under complete ignorance: A new criterion stronger than both Pareto and maxmin." Europ. J. of Operations Research, 1:295-306.

Bellman, R., Zadeh, L.A. (1970) "Decision-making in a fuzzy environment." Management Sciences, 17:141-164.

De Baets, B. (1994) "Ordering alternatives in MCDM problems." In *Fuzzy logic and Intelligent Technologies in Nuclear Science*, Da Ruan et al., eds., Singapore: World Scientific, 48-55.

Dubois, D., Fargier, H., Prade, H. (1994) "Propagation and Satisfaction of flexible constraints." In *Fuzzy Sets, Neural Networks and Soft Computing*, R.R. Yager, L.A. Zadeh, eds., New York: Van Nostrand Reinhold, 166-187.

Dubois, D., Fargier, H., Prade, H. (1995) "Fuzzy constraints in job-shop scheduling." J. of Intelligent Manufacturing, 6(4):215-234

Dubois, D., Fargier, H., Prade, H. (1996) "Refinements of the maximin approach to decision-making in fuzzy environment." Fuzzy Sets and Systems, 81:103-122.

Dubois, D., Fortemps, P. (1996) "Computing improved optimal solutions to maxmin flexible constraint satisfaction problems." Joint Report in preparation, IRIT, Toulouse and IMAGE, Faculté Polytechnique de Mons.

Dubois, D., Lang, J., Prade, H. (1992) "Inconsistency in possibilistic knowledge bases: To live with it or not live with it." In *Fuzzy Logic for the Management of Uncertainty*, L.A. Zadeh, J. Kacprzyk, eds., New York: Wiley, 335-351.

Dubois, D., Prade, H. (1980) *Fuzzy Sets and Systems — Theory and Applications.* New York: Academic Press.

Dubois, D., Prade, H. (1985) "Fuzzy cardinality and the modeling of imprecise quantification." Fuzzy Sets and Systems, 16:199-230.

Dubois, D., Prade, H. (1986) "Weighted minimum and maximum operations in fuzzy set theory." Information Sciences, 39:205-210.

Dubois, D., Prade, H., Testemale, C. (1988) "Weighted fuzzy pattern matching." Fuzzy Sets and Systems, 28:313-331.

Fargier, H., Dubois, D., Prade, H. (1995) "Problèmes de satisfaction de contraintes flexibles: Une approche égalitariste." Revue d'Intelligence Artificielle, 9(3):311-354.

Fargier, H., Lang, J., Schiex, T. (1993) "Selecting preferred solutions in fuzzy constraint satisfaction problems." In *Proc. of the 1st Europ. Congress on Fuzzy and Intelligent Technologies (EUFIT'93)*, ELITE, Aachen, Germany, 1128-1134.

Felix, R. (1992) "Towards a goal-oriented application of aggregation operators in fuzzy decision making." In *Proc. of the Inter. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'92)*, Mallorca, Spain, July 6-10, 585-588.

Grabisch, M. (1995) "Fuzzy integral in multicriteria decision making." Fuzzy Sets and Systems, 69(3):279-298.

Moulin, H. (1988) *Axioms of Cooperative Decision-Making*. Cambridge, UK: Cambridge University Press.

Ribeiro B.A.N., Muntz, R. (1996) "Fuzzy ranking of approximate answers." In *Proc. of the 1996 Workshop on Flexible Query-Answering Systems (FQAS'96)*, H. Christiansen, H.L. Larsen, T. Andreasen, eds., Roskilde, Denmark, May 22-24, 41-56.

Rice, J. (1962) "Tschebyscheff approximation in a compact metric space." Bull. American Math. Soc., 68:405-410.

Trillas, E., Valverde, L. (1985) "On implication and indistinguishability in the setting of fuzzy logic." Reprinted in *Readings in Fuzzy Sets for Intelligent Systems*, D. Dubois, H. Prade, R.R. Yager, eds., San Francisco, CA: Morgan & Kaufmann, 97-104.

Yager, R.R. (1984) "General multiple objective decision making and linguistically quantified statements." Int. J. of Man-Machine Studies, 21:389-400.

Yager, R.R. (1988) "On ordered weighted averaging aggregation operators in multi-criteria decision making." IEEE Trans. on Systems, Man and Cybernetics, 18:183-190.

Yager, R.R. (1996) "On the analytic representation of the leximin ordering and its application to flexible constraint propagation." Tech. Report #MII-1602, Machine Intelligence Institute, Iona College, New York.

OWA OPERATORS IN GROUP DECISION MAKING AND CONSENSUS REACHING UNDER FUZZY PREFERENCES AND FUZZY MAJORITY

Janusz Kacprzyk*, Mario Fedrizzi** and Hannu Nurmi***

* Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01–447 Warsaw, Poland

E-mail: `kacprzyk@ibspan.waw.pl`

** Department of Informatics and Managerial Sciences
University of Trento
Via Inama 5, 38100 Trento, Italy

E-mail: `fedrizzi@cs.unitn.it`

*** Department of Political Science, University of Turku
20500 Turku, Finland

E-mail: `hnurmi@sara.utu.fi`

Abstract

We discuss the use of Yager's (1988) ordered weighted averaging (OWA) operators with importance qualification (Yager, 1993, 1996) for dealing with a fuzzy majority, meant as a linguistic quantifier (most, almost all, ...), in group decision making and consensus formation under fuzzy preferences. We show how new solution concepts in group decision making, and new "soft" degrees of consensus can be defined.

Keywords: group DM, consensus, fuzzy preference, fuzzy majority, linguistic quantifier, fuzzy logic, OWA operator.

1 Introdution

This paper concerns *group decision making* the essence of which is basically as follows. Suppose that we have a set of $n \geq 2$ options (alternatives), $S =$

$\{s_1, \ldots, s_n\}$, and a set of $m \geq 2$ individuals, $I = \{1, \ldots, m\}$. For each individual k, $k \in I$, we have an individual fuzzy preference relation in $S \times S$ which assigns a value in the unit interval for the strength of preference of one option over another.

The problem is then to find an option (or a set of them) which is best acceptable by the group of individuals as a whole. In practice, this means that some majority is fixed, exemplified by "at least a half" or "at least 75%", and the acceptance is meant by such a majority of indivuduals.

The second problem considered in this paper is that of *consensus reaching*. Traditionally, consensus is meant as a full and unanimous agreement (that is, in our context, that all individuals agree as to all options) which is utopian in virtually all practical cases in which it is fully sufficient if, say, *most* of the individuals agree as to *almost all* of the options. Consensus is then, clearly, to a degree (from $[0, 1]$). As it can be seen, a fuzzy majority is also crucial here.

A strict majority usually assumed may be therefore inadequate in many practical cases. A natural line of reasoning is to somehow make that strict concept of majority closer to its real human perception by making it more vague. A good, often cited example in a biological context may be found in Loewer and Laddaga (1985):

> " ... It can correctly be said that there is a consensus among biologists that Darwinian natural selection is an important cause of evolution though there is currently no consensus concerning Gould's hypothesis of speciation. This means that there is a widespread agreement among biologists concerning the first matter but disagreement concerning the second ... "

and it is clear that a rigid majority as, e.g., more than 75% would evidently not reflect the essence of the above statement. However, it should be noted that there are naturally situations when a strict majority is necessary, for obvious reasons, as in all political elections.

A natural manifestations of such a "soft" majority are the so-called *linguistic quantifiers* as, e.g., most, almost all, much more than a half, etc. Such linguistic quantifiers can be, fortunately enough, dealt with by fuzzy-logic-based calculi of linguistically quantified statements (cf. Zadeh, 1983). These calculi have been applied by the authors to introduce a fuzzy majority (represented by a fuzzy linguistic quantifier) into group decision making and consensus formation models (Fedrizzi and Kacprzyk, 1988; Kacprzyk, 1984, 1985b, 1986, 1987a; Kacprzyk and Fedrizzi, 1986, 1988, 1989; Kacprzyk, Fedrizzi and Nurmi, 1990, 1992a; Nurmi and Kacprzyk, 1991; Nurmi, Fedrizzi and Kacprzyk, 1990), and also in an implemented decision support system for consensus reaching (Fedrizzi, Kacprzyk and Zadrożny, 1988; Fedrizzi, Kacprzyk, Owsiński and Zadrożny, 1994).

Yager's (1988) ordered weighted averaging (OWA) operators can also be used for handling fuzzy linguistic quantifiers, and in this paper we will present how fuzzy preference relations and fuzzy majorities can be employed for deriving solution of group decision making and of degrees of consensus using the OWA operators with importances.

2 Fuzzy linguistic quantifiers and the ordered weighted averaging (OWA) operators

A *linguistically quantified statement* may be exemplified by, say, "most experts are convinced" or "almost all good cars are expensive", and may be generally written as

$$Qy\text{'s are } F \tag{1}$$

where Q is a linguistic quantifier (e.g., most), $Y = \{y\}$ is a set of objects (e.g., experts), and F is a property (e.g., convinced).

We may assign to the particular y's (objects) a different importance (relevance, competence, ...), B, which may be added to (1) yielding a *linguistically quantified statement with importance qualification* generally written as

$$QBy\text{'s are } F \tag{2}$$

which may be exemplified by "most (Q) of the important (B) experts (y's) are convinced (F)".

The problem is to find the truth of such linguistically quantified statements, i.e. truth(Qy's are F) or truth(QBy's are F) knowing truth(y is F), $\forall y \in Y$.

Traditionally, Zadeh's (1983) calculus is employed for this purpose. A fuzzy linguistic quantifier Q is assumed to be a fuzzy set in $[0,1]$. For instance, $Q =$ "most" may be given as

$$\mu_Q(x) = \begin{cases} 1 & \text{for } x \geq 0.8 \\ 2x - 0.6 & \text{for } 0.3 < x < 0.8 \\ 0 & \text{for } x \leq 0.3 \end{cases} \tag{3}$$

which is an example of a *proportional* fuzzy linguistic quantifier; we will deal with such quantifiers only as they reflect the essence of a fuzzy majority.

Property F is defined as a fuzzy set in Y. For instance, if $Y = \{X, W, Z\}$ is the set of experts and F is a property "convinced", then $F =$ "convinced" $= 0.1/X + 0.6/W + 0.8/Z$ which means that expert X is convinced to degree 0.1, expert W to degree 0.6 and expert Z to degree 0.8. If $Y = \{y_1, \ldots, y_p\}$, then truth(y_i is F) $= \mu_F(y_i)$, $i = 1, \ldots, p$.

The value of truth(Qy's are F) is then determined in two steps:

$$r = \frac{1}{p} \sum_{i=1}^{p} \mu_F(y_i) \tag{4}$$

$$\text{truth}(Qy\text{'s are } F) = \mu_Q(r) \tag{5}$$

In case of importance qualification, B is defined as a fuzzy set in Y, and $\mu_B(y_i) \in [0,1]$ is a degree of importance of y_i. We rewrite first "QBy's are F" as "$Q(B$ and $F)y$'s are B" which leads to the following counterparts of (4) and (5):

$$r' = \frac{\sum_{i=1}^{p} [\mu_B(y_i) \wedge \mu_F(y_i)]}{\sum_{i=1}^{p} \mu_B(y_i)} \tag{6}$$

$$\text{truth}(QBY'\text{s are } F) = \mu_Q(r') \tag{7}$$

The above calculus can be reformulated using the OWA operators. An OWA operator of dimension p is a mapping $F : [0,1]^p \to [0,1]$ if associated with F is a weighting vector $W = [w_1, \ldots, w_p]^T$ such that: $w_i \in [0,1]$, $w_1 + \cdots + w_p = 1$, and

$$F(x_1, \ldots, x_p) = w_1 b_1 + \cdots + w_p b_p \tag{8}$$

where b_i is the i-th largest element among $\{x_1, \ldots, x_p\}$.

Then

$$F(x_1, \ldots, x_p) = W^T B \tag{9}$$

The OWA weights may be found from the (monotone nondecreasing) membership function of a fuzzy linguistic quantifier Q by (Yager, 1988)

$$w_k = \begin{cases} \mu_Q(k) - \mu_Q(k-1) & \text{for } k = 1, \ldots, p \\ \mu_Q(0) & \text{for } k = 0 \end{cases} \tag{10}$$

Thus, if OWA_Q means the aggregation guided by a fuzzy linguistic quantifier Q, i.e. with the OWA weights derived vy (10), then we can write

$$\text{truth}(Qy'\text{s are } F) = \text{OWA}_Q(\text{truth } y_i \text{ is } F) = W^T B \tag{11}$$

The OWA operators with importance coefficients associated with the particular data, denoted OWA_I, are relevant for our purposes. Suppose that we have a vector of data $A = [a_1, \ldots, a_n]$, and a vector of importances $V = [v_1, \ldots, v_n]$ such that $v_i \in [0,1]$ is the importance of a_i, $i = 1, \ldots, n$, $(v_1 + \cdots + v_n \neq 1$, in general), and the OWA weights $W = [w_1, \ldots, w_n]^T$ corresponding to Q is determined via (10). Using Yager's (1993, 1996) proposal, which is simple and intuitively appealing, the problem boils down to a of the OWA weights w_i into $\overline{w}_i$. Then, (8) becomes

$$F_I(a_1, \ldots, a_n) = \overline{W}^T \cdot B = \sum_{j=1}^{n} \overline{w}_j b_j \tag{12}$$

Basically, from the a_i's, $i = 1, \ldots, n$, we obtain first the b_j's, the j-th largest elements of $\{a_1, \ldots, a_n\}$. Next, we denote by u_j the importance of b_j, i.e. of the a_i which is the j-th largest; $i, j = 1, \ldots, n$. Finally, the new weights $\overline{W}$ are defined as

$$\overline{w}_j = \mu_Q\left(\frac{\sum_{k=1}^{i} u_k}{\sum_{k=1}^{n} u_k}\right) - \mu_Q\left(\frac{\sum_{k=1}^{i-1} u_k}{\sum_{k=1}^{n} u_k}\right) \tag{13}$$

Example 1 If $A = [a_1, a_2, a_3, a_4] = [0.7, 1, 0.5, 0.6]$, $U = [u_1, u_2, u_3, u_4] = [1, 0.6, 0.5, 0.9]$, and Q="most" be given by (3), then $B = [b_1, b_2, b_3, b_4] = [1, 0.7, 0.6, 0.5]$, and $\overline{W} = [0.04, 0.24, 0.41, 0.31]$, and

$$F_I(A) = \sum_{i=1}^{4} \overline{w}_i b_i = 0.04 \cdot 1 + 0.24 \cdot 0.7 + 0.41 \cdot 0.6 + 0.31 \cdot 0.5 = 0.609$$

□

We have now the necessary formal means to proceed to our discussion of group decision making and consensus formation models under fuzzy preferences and a fuzzy majority.

Notice that we consider here the OWA operators in their basic, numeric form. In an ordinal setting, i.e. for non-numeric data (which are only ordered), we refer the interested reader to, e.g., Delgado, Verdegay and Vila (1993), Herrera, Herrera-Viedma and Verdegay (1996), or some other papers in this volume.

3 OWA operators with importance qualification in group decision making

Let the set of options be $\mathcal{S} = \{s_1, \ldots, s_n\}$, and a set of m individuals be $\mathcal{I} = \{1, \ldots, m\}$. Suppose that each individual k is assigned some relevance (importance, competence, ...), $v_k \in [0, 1]$. Each individual $k \in \mathcal{I}$ provides his or her *individual fuzzy preference relations* $\mu_{R_k} : \mathcal{S} \times \mathcal{S} \longrightarrow [0, 1]$ such that

$$\mu_{R_k}(s_i, s_j) = \begin{cases} 1 & \text{if } \mathrm{s}_i \text{ is definitely preferred over } \mathrm{s}_j \\ \mathrm{c}\in(0.5,\ 1) & \text{if } \mathrm{s}_i \text{ is sligtly preferred over } \mathrm{s}_j \\ 0.5 & \text{if there is indifference} \\ \mathrm{d}\in(0,\ 0.5) & \text{if } \mathrm{s}_j \text{ is sligtly preferred over } \mathrm{s}_i \\ 0 & \text{if } \mathrm{s}_j \text{ is definitely preferred over } \mathrm{s}_i \end{cases} \tag{14}$$

which is represented by a matrix $[r^k_{ij}] = [\mu_{R_k}(s_i, s_j)]$; $i, j = 1, \ldots, n$; $k = 1, \ldots, m$; $r^k_{ij} + r^k_{ji} = 1$, and $r^k_{ii} = 0$, $\forall i, j, k$.

Fuzzy preference relations may be used to devise solution concepts along the two lines of reasoning (cf. Kacprzyk, 1984, 1985b, 1986):

- a direct approach:

$$\{R_1, \ldots, R_m\} \longrightarrow \text{ solution}$$

- an indirect approach

$$\{R_1, \ldots, R_m\} \longrightarrow R \longrightarrow \text{ solution}$$

that is, in the first case we determine a solution just on the basis of individual fuzzy preference relations, and in the second case we form first a *social fuzzy preference relation*, R, which is then used to find a solution.

Now we will use the OWA operators with importance qualification to redefine some solution concepts, mainly those defined in Kacprzyk (1984, 1985b, 1986, 1987a) [cf. also Kacprzyk, Fedrizzi and Nurmi (1990; 1992a, b), Kacprzyk and Nurmi (1989)].

3.1 Direct derivation of a solution – the core

For the direct approach, $\{R_1, \dots, R_m\} \longrightarrow$ solution, the *core* is intuitively appealing and often used. It is defined as a set of undominated options, i.e. those not defeated in pairwise comparisons by a required majority (strict!) $r < m$, that is $C = \{s_j \in \mathcal{S} : \exists s_i \in \mathcal{S}$ such that $r^k_{ij} > 0.5$ for at least r individuals$\}$.

Nurmi (1981) extended the core to the fuzzy α-core defined as $C_\alpha = \{s_j \in \mathcal{S} : \exists s_i \in \mathcal{S}$ such that $r \geq \alpha > 0.5$ for at least r individuals$\}$, i.e. as a set of options not sufficiently (at least to degree α) defeated by the required (crisp) majority.

For a fuzzy majority as, say, "most" defined by (3), we start with

$$h^k_{ij} = \begin{cases} 1 & \text{if } r^k_{ij} > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

which reflects if option s_j defeats (in individual k's opinion) option s_i or not.

Then

$$h^k_j = \frac{1}{n-1} \sum_{i=1, i \neq j}^{n} h^k_{ij} \tag{16}$$

is the extent to which individual k is not against option s_j .

Next

$$h_j = \frac{1}{m} \sum_{k=1}^{m} h^k_j \tag{17}$$

is to what extent all the individuals are not against s_j.

And

$$v^j_{Q,I} = \text{OWA}^I_Q(h_j) \tag{18}$$

is to what extent Q (say, most) of the important I individuals are not against s_j; OWA^I_Q means the aggregation of Q h_j's via the OWA operator corresponding to Q (10), with importance qualifications (I) using (12).

The fuzzy Q, I-core is now defined as a fuzzy set

$$C_{Q,I} = v^1_{Q,I}/s_1 + \dots + v^n_{Q,I}/s_n \tag{19}$$

i.e., as a fuzzy set of options that are not defeated by Q (say, most) of important I individuals.

By introducing a threshold of the degree of defeat in (15), we can define the fuzzy $\alpha/Q/I$-core.

First,

$$h_{ij}^k(\alpha) = \begin{cases} 1 & \text{if } r_{ij}^k \leq \alpha < 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

and then, following the line of reasoning (16)–(19), and using $h_{ij}^k(\alpha)$, $h_j^k(\alpha)$ and $v_{Q,I}^j(\alpha)$, respectively, we define the fuzzy α/Q-core as

$$C_{\alpha/Q/I} = v_{Q,I}^1(\alpha)/s_1 + \cdots + v_{Q,I}^n(\alpha)/s_n \tag{21}$$

i.e., as a fuzzy set of options that are not sufficiently (at least to degree $1-\alpha$) defeated by Q of the important (I) individuals.

We can also explicitly introduce the strength of defeat as, for example,

$$\overline{h}_{ij}^k = \begin{cases} 2(0.5 - r_{ij}^k) & \text{if } r_{ij}^k < 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

and then, following the line of reasoning (16)–(19), but using $\overline{h}_j^k$, $\overline{h}_j$ and $\overline{v}_{Q,I}^j$ instead of h_j^k, h_j and $v_{Q,I}^j$, respectively, we define the fuzzy $s/Q/I$-core, as

$$C_{s/Q/I} = \overline{v}_{Q,I}^1/s_1 + \cdots + \overline{v}_{Q,I}^n/s_n \tag{23}$$

i.e., as a fuzzy set of options that are not strongly defeated by Q of the important (I) individuals.

Example 2 Suppose that we have four individuals, $k = 1, 2, 3, 4$, whose individual fuzzy preference relations are:

$$R_1 = \begin{array}{r|cccc} & j=1 & 2 & 3 & 4 \\ \hline i=1 & 0 & 0.3 & 0.7 & 0.1 \\ 2 & 0.7 & 0 & 0.6 & 0.6 \\ 3 & 0.3 & 0.4 & 0 & 0.2 \\ 4 & 0.9 & 0.4 & 0.8 & 0 \end{array} \qquad R_2 = \begin{array}{r|cccc} & j=1 & 2 & 3 & 4 \\ \hline i=1 & 0 & 0.4 & 0.6 & 0.2 \\ 2 & 0.6 & 0 & 0.7 & 0.4 \\ 3 & 0.4 & 0.3 & 0 & 0.1 \\ 4 & 0.8 & 0.6 & 0.9 & 0 \end{array}$$

$$R_3 = \begin{array}{r|cccc} & j=1 & 2 & 3 & 4 \\ \hline i=1 & 0 & 0.5 & 0.7 & 0.1 \\ 2 & 0.5 & 0 & 0.8 & 0.4 \\ 3 & 0.3 & 0.2 & 0 & 0.2 \\ 4 & 1 & 0.6 & 0.8 & 0 \end{array} \qquad R_4 = \begin{array}{r|cccc} & j=1 & 2 & 3 & 4 \\ \hline i=1 & 0 & 0.4 & 0.7 & 0.8 \\ 2 & 0.6 & 0 & 0.4 & 0.3 \\ 3 & 0.3 & 0.6 & 0 & 0.1 \\ 4 & 0.7 & 0.7 & 0.9 & 0 \end{array}$$

Suppose now that the fuzzy linguistic quantifier is Q = "most" defined by (3), and $I = 0.5/1 + 0.7/2 + 1/3 + 0.3/4$. Then, e.g.:

$$C_{\text{"most"},I} \approx 0.5/s_2 + 1/s_4$$

$$C_{0.3/\text{"most"},I} = 0.8/s_4$$

$$C_{s/\text{"most"},I} = 0.5/s_4$$

Notice that though the results obtained for the particular cores are different, for obvious reasons, s_4 is clearly the best choice which is evident if we examine the given individual fuzzy preference relations. □

3.2 Indirect derivation of a solution – the consensus winner

For the indirect approach, $\{R_1, \ldots, R_m\} \to R \to$ solution, we will not deal here with the first step, $\{R_1, \ldots, R_m\} \to R$, and assume that the social fuzzy preference relation $R = [r_{ij}]$ is given, for simplicity, by

$$r_{ij} = \begin{cases} \frac{1}{m} \sum_{k=1}^{m} a_{ij}^k & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

where

$$a_{ij}^k = \begin{cases} 1 & \text{if } r_{ij}^k > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

We redefine a solution concept of much intuitive appeal, the *consensus winner*. We start with

$$g_{ij} = \begin{cases} 1 & \text{if } r_{ij} > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

which expresses whether s_i defeats s_j or not, and then

$$g_i = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} g_{ij} \tag{27}$$

which is the mean degree to which s_i is preferred over all other options options.

Next

$$z_Q^i = \mathrm{OWA}_Q(g_i) \tag{28}$$

is the extent to which s_i is preferred over Q other options; $\mathrm{OWA}_Q(.)$ means the aggregation of Q g_i's using the OWA operator corresponding to Q (10).

Finally, we define the fuzzy Q-consensus winner as

$$W_Q = z_Q^1/s_1 + \cdots + z_Q^n/s_n \tag{29}$$

i.e., as a fuzzy set of options that are preferred over Q other options.

And analogously as in the case of the core, we can introduce a threshold to (26), i.e.

$$g_{ij}(\alpha) = \begin{cases} 1 & \text{if } r_{ij} \geq \alpha > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

and then, following the reasoning (27) and (28), we can define the fuzzy α/Q-consensus winner as

$$W_{\alpha/Q} = z_Q^1(\alpha)/s_1 + \cdots + z_Q^n(\alpha)/s_n \tag{31}$$

i.e., as a fuzzy set of options that are sufficiently (at least to degree α) preferred over Q other options.

We can also explicitly introduce the strength of preference into (26) by

$$\overline{g}_{ij} = \begin{cases} 2(r_{ij} - 0.5) & \text{if } r_{ij} > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

and then, following the reasoning (27) and (28), we can define the fuzzy s/Q-consensus winner as

$$W_{s/Q} = \overline{z}_Q^1/s_1 + \cdots + \overline{z}_Q^n/s_n \tag{33}$$

i.e., as a fuzzy set of options that are strongly preferred over Q other options.

The inclusion of importance is here somewhat "tricky". First, as to importances of the individuals, these have to do with the aggregation of the individual preference relations into the social one (24). This is not trivial, however, and beyond the scope of our discussion. On the other hand, one may also deal with the importances of the options, and this may be done in a straightforward way applying to (29), (31) and (33) the OWA_Q^I aggregation as in Section 3.1.

We will therefore present a simplified examples, without importances.

Example 3 For the same individual fuzzy preference relations as in Example 2, and using (25) and (24), we obtain the following social fuzzy preference relation

$$R = \begin{array}{r|cccc} & j=1 & 2 & 3 & 4 \\ \hline i=1 & 0 & 0 & 1 & 0 \\ 2 & 0.75 & 0 & 0.75 & 0.25 \\ 3 & 0 & 0.25 & 0 & 0 \\ 4 & 1 & 0.75 & 1 & 0 \end{array}$$

If now $Q =$ "most" is given by (3), then we obtain

$$W_{\text{"most"}} = \frac{1}{15}/s_1 + \frac{11}{15}/s_2 + 1/s_4$$

$$W_{0.8/\text{"most"}} = \frac{1}{15}/s_1 + \frac{11}{15}/s_4$$

$$W_{s/\text{"most"}} = \frac{2}{15}/s_1 + \frac{11}{15}/s_2 + 1/s_4$$

Notice that here once again option s_4 is clearly the best choice which is obvious by examining the social fuzzy preference relation.

4 "Soft" degrees of consensus

We will briefly show how to use the OWA-based aggregation to redefine a "soft" degree of consensus proposed in Kacprzyk (1987a), and in Kacprzyk and Fedrizzi (1986, 1988, 1989), and Fedrizzi and Kacprzyk (1988). This degree is meant to overcome a "rigidness" of the conventional concept of consensus in which (full) consensus occurs only when "*all* the individuals agree as to *all* the issues".

The new degree measures the degree to which, say "most individuals agree as to almost all of the options". The use of the OWA operators in this context was proposed by Fedrizzi and Kacprzyk (1993), Kacprzyk and Fedrizzi (1995), and Fedrizzi, Kacprzyk and Nurmi (1993).

We start with the degree of strict agreement between individuals k_1 and k_2 as to their preferences between s_i and s_j

$$v_{ij}(k_1, k_2) = \begin{cases} 1 & \text{if } r_{ij}^{k_1} = r_{ij}^{k_2} \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

where $k_1 = 1, \ldots, m-1$; $k_2 = k_1 + 1, \ldots, m$; $i = 1, \ldots, n-1$; $j = i+1, \ldots, n$.

The degree of agreement between k_1 and k_2 as to their preferences between all the pairs of options is

$$v(k_1, k_2) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} v_{ij}(k_1, k_2) \tag{35}$$

The degree of agreement between k_1 and k_2 as to their preferences between Q_1 pairs of options is

$$v_{Q_1}(k_1, k_2) = \mathrm{OWA}_{Q_1}(v(k_1, k_2)) \tag{36}$$

Next, the degree of agreement of all the pairs of individuals as to their preferences between Q_1 pairs of options is

$$v_{Q_1,R} = \frac{2}{m(m-1)} \sum_{k1=1}^{m-1} \sum_{k2=k1+1}^{m} v_{Q_1}(k_1, k_2) \tag{37}$$

and, finally, the degree of agreement of Q_2 pairs of individuals as to their preferences between Q_1 pairs of options, called the degree of Q_1/Q_2-consensus, is

$$\mathrm{con}(Q_1, Q_2) = \mathrm{OWA}_{Q_2}(v_{Q_1}) \tag{38}$$

Since (34) may be viewed too rigid, we can use a sufficient agreement (at least to degree $\alpha \in [0,1]$) similarly as in (20), and even the strength of agreement as in (22), and obtain by following (34)–(38) the degree of sufficient agreement (at least to degree α) of Q_2 pairs of individuals as to their preferences between Q_1 pairs of options, called the degree of $\alpha/Q_1/Q_2$-consensus, and the degree of strong agreement of Q_2 pairs of individuals as to their preferences between Q_1 pairs of options, called the degree of $s/Q_1/Q_2$-consensus.

And analogously as in Section 3, we can add the importance of pairs of individuals, I and of pairs of options, J, and replace the OWA_{Q_1} and OWA_{Q_2} by $\mathrm{OWA}_{Q_1}^{I,J}$ and OWA_{Q_2}, respectively, in (36)–(38) to obtain the $Q_1/Q_2/I/J$-consensus, $\alpha/Q_1/Q_2/I/J$-consensus, and $s/Q_1/Q_2/I/J$-consensus.

Example 4 Suppose that $n = m = 3$, $Q_1 = Q_2 =$ "most" are given by (3), $\alpha = 0.9$, $s(x)$ is defined by (22), and the individual preference relations are:

$R_1 = [r_{ij}^1] =$

	$j=1$	2	3
$i=1$	0	0.1	0.6
2	0.9	0	0.7
3	0.4	0.3	0

$$R_2 = [r^2_{ij}] = \begin{array}{c|ccc} & j=1 & 2 & 3 \\ \hline i=1 & 0 & 0.1 & 0.7 \\ 2 & 0.9 & 0 & 0.7 \\ 3 & 0.3 & 0.3 & 0 \end{array}$$

$$R_3 = [r^3_{ij}] = \begin{array}{c|ccc} & j=1 & 2 & 3 \\ \hline i=1 & 0 & 0.2 & 0.6 \\ 2 & 0.8 & 0 & 0.7 \\ 3 & 0.4 & 0.3 & 0 \end{array}$$

If we assume the relevance of the alternatives to be $b^B_i = 1/s_1 + 0.6/s_2 + 0.2/s_3$, the importance of the individuals to be $b^I_k = 0.8/1 + 1/2 + 0.4/3$, $\alpha = 0.9$ and $Q =$ "most" given by (3), then we obtain the following degrees of consensus:

$$\text{con}(\text{"most"}, \text{"most"}, I, B) \approx 0.35$$
$$\text{con}^{0.9}(\text{"most"}, \text{"most"}, I, B) \approx 1.0$$
$$\text{con}^{s}(\text{"most"}, \text{"most"}, I, B) \approx 0.75$$

□

5 Concluding remarks

We presented the use of the ordered weighted avarage (OWA) operators with importance qualification for handling fuzzy linguistic quantifiers representing a fuzzy majority in group decision making and consensus formation. We showed how some popular solution concepts and degrees of consensus could be redefined.

Bibliography

Barrett, C.R., Pattanaik, P.K. and Salles, M. (1990). On choosing rationally when preferences are fuzzy. Fuzzy Sets and Systems, 34, 197–212.

Barrett, C.R., Pattanaik, P.K. and Salles, M. (1992). Rationality and aggregation of preferences in an ordinally fuzzy framework. Fuzzy Sets and Systems, 49, 9–13.

Bezdek, J.C., Spillman, B. Spillman, R. (1978). A fuzzy relation space for group decision theory, Fuzzy Sets and Systems, 1, 255–268.

Blin, J.M. (1974). Fuzzy relations in group decision theory, Journal of Cybernetics, 4, 17–22.

Blin, J.M. and Whinston, A.P. (1973). Fuzzy sets and social choice, Journal of Cybernetics, 4, 17–22.

Carlsson, Ch. et al. (1992). Consensus in distributed soft environments, European Journal of Operational Research, 61, 165–185.

Delgado, M., Verdegay, J.L. and Vila, M.A. (1993). On aggregation operations of linguistic labels, International Journal of Intelligent Systems, 8, 351–370.

Fedrizzi, M. and Kacprzyk, J. (1988). On measuring consensus in the setting of fuzzy preference relations. In J. Kacprzyk and M. Roubens (Eds.): Non-Conventional Preference Relations in Decision Making, Springer–Verlag, Berlin, pp. 129–141.

Fedrizzi, M. and Kacprzyk, J. (1993). Consensus degrees under fuzzy majorities and preferences using OWA (ordered weighted average) operators, Proceedings of Fifth IFSA World Congress '93 (Seoul, Korea), Vol. I, pp. 624-626.

Fedrizzi, M., Kacprzyk, J. and Nurmi, H. (1993). Consensus degrees under fuzzy majorities and fuzzy preferences using OWA (ordered weighted average) operators, Control and Cybernetics, 22, 71–80.

Fedrizzi, M., Kacprzyk, J. and Nurmi, H. (1996). How different are social choice functions: a rough sets approach, Quality and Quantity, 30, 87–99.

Fedrizzi, M., Kacprzyk, J., Owsiński, J.W. and Zadrożny, S. (1994). Consensus reaching via a GDSS with fuzzy majority and clustering of preference profiles, Annals of Operations Research, 51, 127–139.

Fedrizzi, M., Kacprzyk, J. and Zadrożny, S. (1988). An interactive multi-user decision support system for consensus reaching processes using fuzzy logic with linguistic quantifiers, Decision Support Systems, 4, 313–327.

Herrera, F., Herrera-Viedma, E. and Verdegay, J.L. (1996). A model of consensus in group decision making under linguistic assessments. Fuzzy Sets and Systems, 78, 73–88.

Kacprzyk, J. (1984). Collective decision making with a fuzzy majority rule, Proc. WOGSC Congress, AFCET, Paris, pp. 153–159.

Kacprzyk, J. (1985a). Zadeh's commonsense knowledge and its use in multicriteria, multistage and multiperson decision making. In M.M. Gupta et al. (Eds.): Approximate Reasoning in Expert Systems, North–Holland, Amsterdam, pp. 105–121.

Kacprzyk, J. (1985b). Group decision-making with a fuzzy majority via linguistic quantifiers. Part I: A consensory-like pooling; Part II: A competitive-like pooling, Cybernetics and Systems: an International Journal, 16, 119–129 (Part I), 131–144 (Part II).

Kacprzyk, J. (1986). Group decision making with a fuzzy linguistic majority, Fuzzy Sets and Systems, 18, 105–118.

Kacprzyk, J. (1987a). On some fuzzy cores and 'soft' consensus measures in group decision making. In J.C. Bezdek (Ed.): The Analysis of Fuzzy Information, Vol. 2, CRC Press, Boca Raton, pp. 119–130.

Kacprzyk, J. (1987b). Towards 'human consistent' decision support systems through commonsense-knowledge-based decision making and control models: a fuzzy logic approach, Computers and Artificial Intelligence, 6, 97–122.

Kacprzyk, J. and Fedrizzi, M. (1986). "Soft" consensus measures for monitoring real consensus reaching processes under fuzzy preferences, Control and

Cybernetics, 15, 309–323.

Kacprzyk, J. and Fedrizzi, M. (1988). A "soft" measure of consensus in the setting of partial (fuzzy) preferences, European Journal of Operational Research, 34, 315–325.

Kacprzyk, J. and Fedrizzi, M. (1989). A "human-consistent" degree of consensus based on fuzzy logic with linguistic quantifiers, Mathematical Social Sciences, 18, 275–290.

Kacprzyk, J. and Fedrizzi,M., Eds. (1990). Multiperson Decision Making Models Using Fuzzy Sets and Possibility Theory, Kluwer, Dordrecht.

Kacprzyk, J. and Fedrizzi, M. (1995). Consensus degrees under fuzziness via ordered weighted average (OWA) operators. In Z. Bien and K.C. Min (Eds.): Fuzzy Logic and its Applications in Engineering, Information Sciences and Intelligent Systems, Kluwer, Dordrecht, pp. 447–454.

Kacprzyk, J., Fedrizzi, M. and Nurmi, H. (1990). Group decision making with fuzzy majorities represented by linguistic quantifiers. In J.L. Verdegay and M. Delgado (Eds.): Approximate Reasoning Tools for Artificial Intelligence, Verlag TÜV Rheinland, Cologne, pp. 126–145.

Kacprzyk, J., Fedrizzi, M. and Nurmi, H. (1992a). Fuzzy logic with linguistic quantifiers in group decision making and consensus formation. In R.R. Yager and L.A. Zadeh (Eds.): An Introduction to Fuzzy Logic Applications in Intelligent Systems, Kluwer, Dordrecht, 263–280.

Kacprzyk, J., Fedrizzi, M. and Nurmi, H. (1992b). Group decision making and consensus under fuzzy preferences and fuzzy majority, Fuzzy Sets and Systems, 49, 21–31.

Kacprzyk, J., Nurmi, H. and Fedrizzi, M., Eds. (1997). Consensus under Fuzziness, Kluwer, Boston.

Kacprzyk, J. and Roubens, M., Eds. (1988). Non-Conventional Preference Relations in Decision Making, Springer–Verlag, Heidelberg.

Loewer, B. and Laddaga, R. (1985). Destroying the consensus, in Loewer B., Guest Ed., Special Issue on Consensus, Synthese, 62 (1), pp. 79–96.

Nurmi, H. (1981). Approaches to collective decision making with fuzzy preference relations, Fuzzy Sets and Systems, 6, 249–259.

Nurmi, H. (1987). Comparing Voting Systems, Reidel, Dordrecht.

Nurmi, H. (1988). Assumptions on individual preferences in the theory of voting procedures. In J. Kacprzyk and M. Roubens (Eds.): Non-Conventional Preference Relations in Decision Making, Springer-Verlag, Heidelberg, pp. 142–155.

Nurmi, H., Fedrizzi, M. and Kacprzyk, J. (1990). Vague notions in the theory of voting. In J. Kacprzyk and M. Fedrizzi (Eds.): Multiperson Decision Making Models Using Fuzzy Sets and Possibility Theory, Kluwer, Dordrecht, pp. 43–52.

Nurmi, H. and Kacprzyk, J. (1991). On fuzzy tournaments and their solution concepts in group decision making, European Journal of Operational Research, 51, 223–232.

Nurmi H., Kacprzyk, J. and Fedrizzi, M. (1996). Probabilistic, fuzzy and rough concepts in social choice, European Journal of Operational Research, 95, 264–277.

Tanino, T. (1990). On group decision making under fuzzy preferences. In J. Kacprzyk and M. Fedrizzi (Eds.): Multiperson Decision Making Models using Fuzzy Sets and Possibility Theory, Kluwer, Dordrecht, pp. 172–185.

Torra, V. (1997) The weighted OWA operator, International Journal of Intelligent Systems, 12, 151–166.

Yager, R.R (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking, IEEE Trans. on Systems, Man and Cybernetics, SMC-18, 183–190.

Yager, R.R. (1993). On the Issue of Importance Qualifications in Fuzzy Multi-Criteria Decision Making. Tech. Report MII-1323, Machine Intelligence Institute, Iona College, New Rochelle, NY.

Yager, R.R. (1996). Quantifier quided aggregation using OWA operators, International Journal of Intelligent Systems, 11, 49–73.

Zadeh, L.A. (1983). A computational approach to fuzzy quantifiers in natural languages. Computers and Mathematics with Applications, 9, 149–184.

APPLICATIONS OF THE LINGUISTIC OWA OPERATORS IN GROUP DECISION MAKING

F. Herrera, E. Herrera-Viedma and J.L. Verdegay

Department of Computer Science and Artificial Intelligence
University of Granada
18071–Granada, Spain
E-mail: herrera@decsai.ugr.es

Abstract

Assuming a group decision making problem where the experts express their opinions by means of linguistic preference relations, the application of the Linguistic OWA operator guided by fuzzy majority is analyzed. Two different perspectives of the use of the LOWA operator are presented: (i) in the selection process, to aggregate individual linguistic preference relations in a collective one and to calculate different linguistic choice degrees of the alternatives, and (ii) in the consensus reaching process to obtain the linguistic consensus measures. In all cases, the concept of fuzzy majority is represented by means of a fuzzy linguistic quantifier used to obtain the weights that the LOWA operator needs in its aggregation way.

1 Introduction

There are many problems in which the solutions depend on the synthesis of information supplied by different sources (e.g. the strategic planning, the diagnosis, the capital investment, etc.). When the final solution consists of making a decision they may be modeled as a particular case of a Group Decision Making (GDM) problem.

A GDM problem may be defined as a decision situation in which (i) there are two or more experts, each of them characterized by his own perceptions,

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

attitudes, motivations, and personalities, (ii) who recognize the existence of a common problem, and (iii) attempt to reach a collective decision. It is usually solved in a process of two phases:

1. consensus phase, which attempts to achieve the maximum possible consensus degree about final decision, and

2. selection phase, which attempts to make the best decision according to the experts' opinions.

Any GDM model must include (i) a comprehensive representation of experts' opinions, and (ii) a data fusion technique that combines those individual opinions and allows to reach the final decision. Futhermore, they must be able to deal with the fuzziness of human judgements. To do so, many authors have used the Fuzzy Sets Theory to model the imprecise information [12].

In a fuzzy framework, fuzzy preference relations are the classical representation used to provide the experts' preferences. These may have a numerical or linguistic nature. In the first case, it is assumed that the experts are able to express their preferences with exact numerical values (e.g., numbers in the [0,1] interval) [11]. In the second case, it is supposed that the experts cannot estimate their preferences with exact numerical values, and then, they use linguistic assessments instead of numerical values (e.g., linguistic terms of a pre-established label set) [4].

We assume a GDM model in which the experts provide their preferences by means of the linguistic preference relations and we present how the application of the Linguistic OWA (LOWA) operator guided by fuzzy majority [3, 7] contributes to solve the GDM problem. We show how the LOWA operator can be applied from two different perspectives:

1. In the selection phase:

   It is applied to aggregate the individual linguistic preference relations in a collective one and to derive different linguistic choice degrees of the alternatives from that collective linguistic preference relation.

2. In the consensus phase:

   It is used to determine some linguistic consensus measures according to the majority experts' preferences.

To do so, the paper is structured as follows: Section 2 presents the basic elements used to develop our study, i.e., the type of label set used to provide the opinions and the considered GDM problem; Section 3 shows the LOWA operator guided by fuzzy majority; Section 4 analyzes the two applications of the LOWA operator in the resolution process of considered GDM problem; and finally, some concluding remarks are pointed out.

2 Preliminaries

We use label sets with an odd cardinal, representing the middle term an assessment of "approximately 0.5", with the rest of the terms being placed symmetrically around it and the limit of granularity 11 or no more than 13. The semantic of the elements in the label set is given by fuzzy numbers defined on the [0,1] interval, which are described by membership functions. Because the linguistic assessments are just approximate ones given by the experts, we can consider that linear trapezoidal membership functions are good enough to capture the vagueness of those linguistic assessments, since it may be impossible or unnecessary to obtain more accurate values. This representation is achieved by the 4-tuple, $(a_i, b_i, \alpha_i, \beta_i)$, the first two paremeters indicate the interval in which the membership value is 1; the third and fourth parameters indicate the left and right widths. Moreover, the term set, $S = \{s_0, \ldots, s_T\}$, must have the following characteristics:

1) The set is ordered: $s_i \geq s_j$ if $i \geq$ j.
2) There is the negation operator: Neg$(s_i) = s_j$ such that $j = T$-i.
3) Maximization operator: Max$(s_i, s_j) = s_i$ if $s_i \geq s_j$.
4) Minimization operator: Min$(s_i, s_j) = s_i$ if $s_i \leq s_j$.

Example 1. For example, this is the case of the following term set:

MA	$Maximum$	$(1, 1, 0.25, 0)$
VM	$Very\_Much$	$(.75, .75, .15, .25)$
MU	$Much$	$(.6, .6, .1, .15)$
M	$Medium$	$(.5, .5, .1, .1)$
L	$Little$	$(.4, .4, .15, .1)$
VL	$Very\_Little$	$(.25, .25, .25, .15)$
MI	$Minimum$	$(0, 0, 0, .25)$

In the following we shall use this set of seven labels in all examples.

Then, a linguistic setting of the GDM problem is as follows [5]. Let $X = \{x_1, \ldots, x_n\}$ $(n \geq 2)$ be a finite set of alternatives to be analyzed by a finite set of experts, $E = \{e_1, \ldots, e_m\}$ $(m \geq 2)$. Then, assuming a label set, S, each expert, e_k, provides his opinions on X as a linguistic preference relation, $P^k \subset$ XxX, with membership function

$$\mu_{P^k} : X\text{x}X \rightarrow S,$$

where $\mu_{P^k}(x_i, x_j) = p_{ij}^k$ denotes the linguistic preference degree of the alternative x_i over x_j. Without any loss of generality, we suppose that P^k is reciprocal in the sense, $p_{ij}^k =$ Neg(p_{ji}^k), and by definition, $p_{ii}^k =$ - (Undefined).

Example 2. Suppose a set of four alternatives and a group of four experts. Then, using the above label set, assume that they express their opinions by means of the following linguistic preference relations:

$$P^1 = \begin{bmatrix} - & VL & VM & VL \\ VM & - & M & M \\ VL & L & - & VL \\ VM & L & VM & - \end{bmatrix} P^2 = \begin{bmatrix} - & L & M & VL \\ M & - & VM & L \\ L & VL & - & VL \\ VM & M & VM & - \end{bmatrix}$$

$$P^3 = \begin{bmatrix} - & M & VM & MI \\ M & - & VM & L \\ VL & VL & - & VL \\ MA & M & VM & - \end{bmatrix} P^4 = \begin{bmatrix} - & L & VM & VL \\ M & - & L & VL \\ VL & M & - & VL \\ VM & VM & VM & - \end{bmatrix}.$$

Other more complex settings with heterogeneous groups of experts and heterogeneous set of alternatives may be found in [6, 4, 9, 10].

3 The LOWA Operator Guided by Fuzzy Majority

We manage the linguistic information by means of an operator of combination of linguistic values based on direct computation, the *LOWA operator* [3, 7]. It is based on the *OWA operator* defined by Yager [14] and the *convex combination of linguistic labels* defined by Delgado et al. [1].

Definition 1. *Let $A = \{a_1, \ldots, a_m\}$ be a set of labels to be aggregated, then the LOWA operator, ϕ, is defined as*

$$\phi(a_1, \ldots, a_m) = W \cdot B^T = \mathcal{C}^m\{w_k, b_k, k = 1, \ldots, m\} =$$

$$= w_1 \odot b_1 \oplus (1 - w_1) \odot \mathcal{C}^{m-1}\{\beta_h, b_h, h = 2, \ldots, m\}$$

where $W = [w_1, \ldots, w_m]$, is a weighting vector, such that, (i) $w_i \in [0,1]$ and, (ii) $\Sigma_i w_i = 1$,. $\beta_h = w_h / \Sigma_2^m w_k, h = 2, \ldots, m$, and $B = \{b_1, \ldots, b_m\}$ is a vector associated to A, such that,

$$B = \sigma(A) = \{a_{\sigma(1)}, \ldots, a_{\sigma(n)}\}$$

in which, $a_{\sigma(j)} \leq a_{\sigma(i)}$ $\forall\, i \leq j$, with σ being a permutation over the set of labels A. $\mathcal{C}^m$ is the convex combination operator of m labels and if m=2, then it is defined as

$$\mathcal{C}^2\{w_i, b_i, i = 1, 2\} = w_1 \odot s_j \oplus (1 - w_1) \odot s_i = s_k, \quad s_j,\ s_i\ \in\ S,\ (j \geq i)$$

such that, $k = min\{T, i + round(w_1 \cdot (j - i))\}$, where "round" is the usual round operation, and $b_1 = s_j$, $b_2 = s_i$.
If $w_j = 1$ and $w_i = 0$ with $i \neq j$ $\forall i$, then the convex combination is defined as:

$$\mathcal{C}^m\{w_i, b_i, i = 1, \ldots, m\} = b_j.$$

Its properties and its extensions are studied in [7] and in [2] respectively.

As we have mentioned, the weights represent the concept of fuzzy majority in the aggregation of LOWA operator. As in [11] we propose to specific the fuzzy majority rule by means of a fuzzy linguistic quantifier [16]. Then, they can be obtained as [14, 15]:

$$w_i = Q(i/n) - Q((i-1)/n), i = 1, \ldots, n,$$

where Q is a non-decreasing proportional fuzzy linguistic quantifier represented by the membership function [16]:

$$Q(r) = \begin{cases} 0 & \text{if } r < a \\ \frac{r-a}{b-a} & \text{if } a \leq r < b \\ 1 & \text{if } r \geq b \end{cases}$$

with $a, b, r \in [0, 1]$.

Some examples of non-decreasing proportional fuzzy linguistic quantifiers with their respective parameters (a, b) are: "most" $(0.3, 0.8)$, "at least half" $(0, 0.5)$, and "as many as possible" $(0.5, 1)$.

When a fuzzy linguistic quantifier, Q, is used to compute the weights of LOWA operator, ϕ, it is symbolized by ϕ_Q.

Finally, we must quote that aggregations of the LOWA operator in the GDM models are based on two types of fuzzy majority [7, 10]:

- *Fuzzy majority of alternatives*, when information to be aggregated are values that characterize different pairs of alternatives, symbolized by Q_a. For example, when we calculate the linguistic choice degrees of alternatives.

- *Fuzzy majority of experts*, when information to be aggregated are values provided by different experts, symbolized by Q_e. For example, when we derive a collective linguistic preference relation from individual ones.

4 The Use of LOWA Operator in GDM Processes

As we said at the beginning, a GDM problem is solved in a process of two phases: (i) Consensus and (ii) Selection. These phases are combined in a sequential resolution scheme. Firstly, the consensus phase is applied. In each step, the degree of existing consensus among experts' opinions is established by means of a consensus measure. If the consensus degree is satisfactory, then the selection phase is applied in order to obtain a solution. Otherwise, the experts are persuaded to update their opinions. In this way, a GDM process may be defined as a dynamic and iterative process where the experts, via exchange of information and rational arguments, update their opinions until that they become sufficiently similar [6].

Assuming GDM problems in a linguistic context (as in Section 2) and using the LOWA operator guided by fuzzy majority, we have proposed different selection [4, 5, 7, 10] and consensus [6, 8, 9] models. In the next subsections, we shall summarize the applications of LOWA operator in these models. They shortly, are:

1. To obtain a collective linguistic preference relation from individual ones in the aggregation of selection phase.

2. To calculate linguistic choice degrees of alternatives in the exploitation of selection phase.

3. To determinate different linguistic consensus measures that guide the consensus phase.

4.1 The LOWA Operator in the Selection Phase

As is well known, basically two approaches may be considered to develop the selection phase of a GDM problem. A direct approach

$$\{P^1, \ldots, P^m\} \rightarrow \text{ solution}$$

according to which, on the basis of the individual preference relations, a solution is derived, and an indirect approach

$$\{P^1, \ldots, P^m\} \rightarrow P^c \rightarrow \text{ solution}$$

providing the solution on the basis of a collective preference relation, P^c, which is a preference relation of the group of individuals as a whole.

Both approaches are developed along three activity states [10]:

- *Aggregation State.* The goal of this state is to aggregate individual linguistic information. These information units may be the experts' opinions as well as choice degrees obtained from the experts' opinions.

- *Exploitation State.* The goal of this state is to calculate choice degrees of alternatives from linguistic preference relations (individual or collective).

- *Selection State.* The goal of this state is to find out the solution of the alternatives according to different choice degrees.

Depending on the approach considered, these states are applied in one order or another:

- Direct approach: *Exploitation + Aggregation + Selection.*

- Indirect approach: *Aggregation + Exploitation + Selection.*

Here, we present a selection process by indirect approach as in [7]. In this case, the aggregation state defines a collective liguistic preference relation which indicates the global preference between every ordered pair of alternatives according to the experts' opinions. The exploitation state transforms the global information about the alternatives into a ranking of them, supplying in the selection state a set of solution alternatives.

4.1.1 Collective Linguistic Preference Relation Based on the LOWA Operator

Assuming a set of linguistic preference relations, $\{P^1, \ldots, P^m\}$, and a fuzzy linguistic quantifiers, Q_e, the collective linguistic preference relation, P^c, is obtained from all individual linguistic preference relations using the LOWA operator guided

$$P^c = \phi_{Q_e}[P^1, \ldots, P^m].$$

Therefore, each value, $p^c_{ij} \in S$, represents the preference of alternative x_i over alternative x_j according to the fuzzy majority experts' opinions.
Example 3. Suppose the framework presented in Example 2 and the fuzzy linguistic quantifier, Q_e="at least half" (0,0.5). Then, P^c with the weighting vector, $W = [0.5, 0, 5, 0, 0]$, is:

$$P^c = \begin{bmatrix} - & M & VM & VL \\ MU & - & VM & M \\ L & M & - & VL \\ MA & MU & VM & - \end{bmatrix}.$$

4.1.2 Linguistic Choice Degrees Based on the LOWA Operator

In order to select the alternatives best acceptable to the group of experts as a whole, we may use different linguistic choices degrees of alternatives based on the LOWA operator guided by fuzzy majority of alternatives. In indirect approaches they are derived from the collective linguistic preference relation, P^c, [4, 5, 10] and in direct approaches from the individual linguistic preference relations [7, 10]. Here, we shall consider the first approach.

Specifically, in order to illustre the application of the LOWA operator, we use the *quantifier guided non-dominance degree* [10], which is based on the concept of Orlovski's non-dominated alternatives [13].

Given an alternative, x_i, its quantifier guided non-dominance degree, $QGNDD_i$ is defined as:

$$QGNDD_i = \phi_{Q_a}(Neg(p^s_{ji}), j = 1, ..., n, j \neq i),$$

where p^s_{ji} represents the degree to which the alternative x_i is strictly dominated by the alternative x_j, and it is obtained as:

$$p^s_{ji} = s_0 \text{ if } p_{ij} > p_{ji},$$

$$\text{or } p^s_{ji} = s_h \in S \text{ if } p_{ji} \geq p_{ij} \text{ with } p_{ji} = s_l, p_{ij} = s_t \text{ and } l = t + h.$$

Therefore, $QGNDD_i$ quantifies, linguistically, the degree to which one alternative, x_i, is not dominated by a fuzzy majority of the alternatives of set X.

Then, in the selection state, the application of linguistic choice degree of each alternative over X allows us to obtain the set of solution alternatives, X^{QGNDD}, as:

$$X^{QGNDD} = \{x_i | x_i \in X, QGNDD_i = sup_{x_j \in X} QGNDD_j\}.$$

Example 4. Assuming the issues obtained in Example 3, then, the strict linguistic preference relation, P^s, is:

$$P^s = \begin{bmatrix} - & MI & M & MI \\ VL & - & VL & MI \\ MI & MI & - & MI \\ VM & VL & MU & - \end{bmatrix}.$$

Considering Q_a="at least half", then $W = [0.67, 0.33, 0]$ and the linguistic choice degrees are:

$$\{QGNDD_1 = MA, QGNDD_2 = MA, QGNDD_3 = MU, QGNDD_4 = MA\},$$

and therefore, the solution is $X^{QGNDD} = \{x_1, x_2, x_4\}$.

4.2 The LOWA Operator in the Consensus Phase

In [6] we have proposed a consensus model based on two types of linguistic consensus measures: a) *linguistic consensus degrees*, to measure the agreement degree among experts' opinions, and b) *linguistic distances*, to evaluate the closeness among each expert's opinions and group opinion. These measures are applied in three acting levels: level of preference (pairs of alternatives), level of alternative, and level of relation. Therefore, the model contains six linguistic consensus measures: *the preference linguistic consensus degree, the alternative linguistic consensus degree, the relation linguistic consensus degree, the preference linguistic distance, the alternative linguistic distance, and the relation linguistic distance.*

We have proposed two computation schemata of these measures. One is based on the concept of *strict coincidence*, i.e., differentiating between 0 (no coincidence over the evaluations assigned to the preference) and 1 (total coincidence) [6, 8]. The other one is based on the concept of fuzzy coincidence, i.e., considering the large spectrum of possible coincidence degrees [9]. Both schemata compute the consensus measures by means of the LOWA operator guided by fuzzy majority, in such a way that:

- the linguistic consensus degrees express the existing agreement degree among the fuzzy majority of experts according to their preferences over the fuzzy majority of alternatives; and similarly,

- the linguistic distances express the existing agreement degree among an expert and the fuzzy majority of remaining experts according to their preferences over the fuzzy majority of alternatives.

Now, following the scheme described in [8], we shall show the application of LOWA operator in the consensus phase with the computation of some linguistic consensus measures: the alternative linguistic consensus degree and the alternative linguistic distance. Then, assuming a set of linguistic preference relations, $\{P^1, \ldots, P^m\}$, and two fuzzy linguistic quantifiers, Q_a and Q_e, these linguistic consensus measures are obtained as:

- Alternative Linguistic Consensus Degree

  1. Obtain an *Individual Consensus Relation (ICR)*, which contains the proportional number of experts who coincide when assigning the same value about each preference of a pair of alternatives, (x_i, x_j). Then:

  $$ICR_{ij} = \begin{cases} n_{ij}/m & \text{if } n_{ij} > 1 \\ 0 & \text{otherwise} \end{cases}$$

  $n_{ij} = MAX_{s_t \in S}\{\sharp(V_{ij}[s_t])\}$, such that,

  $$V_{ij}[s_t] = \{k \mid p^k_{ij} = s_t, \; k = 1..m\}, \forall s_t \in S.$$

  2. Obtain the preference linguistic consensus degree in each pair, (x_i, x_j), called PCR_{ij}:

  $$PCR_{ij} = Q^2_e(ICR_{ij}) \; i,j = 1, \ldots, n, \text{ and } i \neq j.$$

  Q^2_e is a fuzzy linguistic quantifier which makes linguistical values in a label set $L = \{l_i\}$, $i \in J = \{0, \ldots, U\}$, $Q^2 : [0,1] \rightarrow L$, and it is defined as follows [6]:

  $$Q^2(r) = \begin{cases} l_0 & \text{if } r < a \\ l_i & \text{if } a \leq r \leq b \\ l_U & \text{if } r > b \end{cases}$$

  l_0 and l_U are the minimum and maximum labels in L, respectively, and

  $$l_i = Sup_{l_q \in M}\{l_q\},$$

  $$\text{with } M = \{l_q \in L : \; \mu_{l_q}(r) = Sup_{t \in J} \{\mu_{l_t}(\frac{r-a}{b-a})\}\},$$

  with $a, b, r \in [0,1]$.

  3. Obtain the alternative linguistic consensus degree in each alternative, x_i, called PCR_i:

  $$PCR_i = \phi_{Q_a}(PCR_{ij}, \; j = 1, \ldots, n, \; i \neq j), \; i = 1, \ldots, n.$$

- Alternative Linguistic Distance

1. Obtain a *Label Consensus Relation (LCR)*, which contains the *consensus labels* about each preference of a pair of alternatives, (x_i, x_j). For example, we may assume $LCR = P^c$, i.e.: $LCR = \phi_{Q_e}(P^1, \ldots, P^m)$. Other different ways to calculate LCR may be found in [6, 8].
2. Obtain the preference linguistic distance in each pair, (x_i, x_j), for each e_k, called D_{ij}^k:

$$D_{ij}^k = \begin{cases} p_{ij}^k - LCR_{ij} & \text{if } p_{ij}^k > LCR_{ij} \\ LCR_{ij} - p_{ij}^k & \text{if } LCR_{ij} \geq p_{ij}^k \\ s_T & \text{otherwise} \end{cases} \quad i \neq j$$

with $i, j = 1, \ldots, n$, and $k = 1, \ldots, m$ and, where if $p_{ij}^k = s_t$ and $LCR_{ij} = s_v$ then $p_{ij}^k - LCR_{ij}$ is defined as s_w, such that, $w = t - v$.
3. Obtain the alternative linguistic distance in each alternative, x_i, for each e_k, called D_i^k:

$$D_i^k = \phi_{Q_a}(D_{ij}^k, j = 1, \ldots, n,\ j \neq i) \quad k = 1..m,\ i = 1..n.$$

The computation of remaining linguistic consensus measures is described in [8].

Example 5. Suppose the framework presented in Examples 2 and the issues obtained in Example 3. Then:

- Alternative Linguistic Consensus Degrees:

ICR is the following:

$$ICR = \begin{bmatrix} - & 0.5 & 3/4 & 3/4 \\ 3/4 & - & 0.5 & 0.5 \\ 3/4 & 0.5 & - & 1 \\ 3/4 & 0.5 & 1 & - \end{bmatrix}.$$

Then, the preference linguistic consensus degrees in each pair of alternatives with $L = S$ are:

$$PCR = \begin{bmatrix} - & MA & MA & MA \\ MA & - & MA & MA \\ MA & MA & - & MA \\ MA & MA & MA & - \end{bmatrix}.$$

And the alternative linguistic consensus degrees in each alternative using the LOWA operator with the weighting vector associated to $Q =$"at least half", $W = [0.67, 0.33, 0]$, are:

$$\{PCR_i = MA, i = 1...4\}.$$

Therefore the consensus state over all alternatives, according to the used concept of fuzzy majority, is total.

- Alternative Linguistic Distances:

  Assuming $LCR = P^c$, the preference linguistic distances in each pair for e_4 are:

$$D^4 = \begin{bmatrix} - & VL & MI & MI \\ VL & - & M & L \\ MI & MI & - & MI \\ VL & VL & MI & - \end{bmatrix}.$$

  And the alternative linguistic distances in each alternative for e_4 using the LOWA operator with the same above weighting vector, $W = [0.67, 0.33, 0]$, are:

$$\{D_1^4 = VL, D_2^4 = M, D_4^4 = MI, D_4^4 = VL\}.$$

  Therefore, the expert e_4 presents agreement with the opinion of group in his fuzzy majority of preferences over alternative x_3, a little disagreement in his fuzzy majority preferences over alternatives x_1 and x_4, and a disagreement in his fuzzy majority preferences over alternative x_2.

5 Concluding Remarks

In this paper, we have presented the use of LOWA operator, in GDM problems in a linguistic context where the experts express their preference by means of the linguistic preference relations. We have shown how the LOWA operator may be applied to find the set of solution alternatives (its application in selection phase) and to measure the consensus state among experts (its application in consensus phase). In both cases, using the LOWA operator guided by fuzzy majority, we have got that all the results reflect the opinions of the fuzzy majority of experts over the fuzzy majority of alternatives.

References

M. Delgado, J.L. Verdegay and M.A. Vila, On Aggregation Operations of Linguistic Labels, *Int. Journal of Intelligent Systems* **8** (1993) 351-370.

F. Herrera and E. Herrera-Viedma, Aggregation Operators for Linguistic Weighted Information, *IEEE Transactions on Systems, Man, and Cybernetics*. To appear.

F. Herrera and J.L. Verdegay, Linguistic Assessments in group decision, *Proc. First European Congress on Fuzzy and Intelligent Technologies*, Aachen, 1993, 941-948.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, A Sequential Selection Process in Group Decision Making with Linguistic Assessment, *Information Sciences* **85** (1995) 223-239.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, A Linguistic Decision Process in Group Decision Making, *Group Decision and Negotiation* **5** (1996) 165-176.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, A Model of Consensus in Group Decision Making under Linguistic Assessments, *Fuzzy Sets and Systems* **78** (1996) 73-87.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, Direct Approach Processes in Group Decision Making Using Linguistic OWA Operators, *Fuzzy Sets and Systems* **79** (1996) 175-190.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, A Rational Consensus Model in Group Decision Making under Linguistic Assessments, *Fuzzy Sets and Systems*. To appear.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, Linguistic Measures Based on Fuzzy Coincidence for Reaching Consensus in Group Decision Making, *Int. Journal of Approximate Reasoning*. To appear.

F. Herrera, E. Herrera-Viedma and J.L. Verdegay, Choice Processes for Non-Homogeneous Group Decision Making in Linguistic Setting, *Fuzzy Sets and Systems*. To appear.

J. Kacprzyk, Group Decision Making with a Fuzzy Linguistic Majority, *Fuzzy Sets and Systems* **18** (1986) 105-118.

J. Kacprzyk and M. Fedrizzi, *Multiperson Decision Making Models Using Fuzzy Sets and Possibility Theory* (Kluwer Academic Publishers, 1990).

S.A. Orlovski, Decision Making with a Fuzzy Preference Relation, *Fuzzy Sets and Systems* **1** (1978) 155-167.

R.R. Yager, On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making, *IEEE Transactions on Systems, Man, and Cybernetics* **18** (1988) 183-190.

R.R. Yager, Families of OWA Operators, *Fuzzy Sets and Systems* **59** (1993) 125-148.

L. A. Zadeh, A Computational Approach to Fuzzy Quantifiers in Natural Languages, *Computers and Mathematics with Applications* **9** (1983) 149-184.

6

OWA OPERATORS IN QUERYING AND INFORMATION RETRIEVAL

AGGREGATION RULES IN COMMITTEE PROCEDURES

Javier Montero* and Vincenzo Cutello**

*Faculty of Mathematics
Complutense University
Madrid, Spain

**Department of Mathematics
University of Catania
Catania, Italy

Abstract

Very often, decision procedures in a committee compensate potential manipulations by taking into account the ordered profile of qualifications. It is therefore rejected the standard assumption of an underlying associative binary connective allowing the evaluation of arbitrary finite sequences of items by means of a one-by-one sequential process. In this paper we develop a mathematical approach for non-associative connectives allowing a sequential definition by means of binary fuzzy connectives. It will be then stressed that a connective rule should be understood as a consistent sequence of binary connective operators. Committees should previously decide about which connective rule they will be condidering, not just about a single operator.

Keywords: Fuzzy Connectives, Fuzzy Sets, Aggregation Operators.

1 Introduction.

Many real-life problems are solved by means of some information aggregation procedures. If we observe chunks of partial information and we have to elaborate a global opinion we have an aggregation process which can be modeled in terms of aggregation rules. In principle, the only property which is required is just the ability to transform the data set into a simpler representation. For

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

example, if we expect data with a fixed and known dimension n, we do not need to define how $n+1$ or $n-1$ items of information have to be aggregated into one single index. Nevertheless, in practice we rarely know in advance the dimension of the real problem we are going to be faced to. Moreover, even if we do know the dimension of the input data we have the semantic problems of using the *same* aggregation operator everytime this is the case, on inputs of different dimensions. Therefore, any aggregation rule has to be able to operate with an arbitrary dimension of the data set, even when we assume that such a data set has nice mathematical properties like homogeneity and non-redundant information.

When a committee has to decide about the qualification of candidates, the objetive is just to aggregate the opinion of each committee member into a single index. Each committee member represents a piece of information. Quite often, the number of voting members may be not fixed (e.g., unavoidable last minute absences), or the number of voting members can not be *a priori* known (this is the case when we just have a *qualitative* definition of the crisp set of voters, i.e., the list of properties giving people the right to vote). Moreover, if the number of potential voters is too big, existence of an algorithm allowing some kind of sequential reckoning will help calculus.

A standard solution to such a basic semantical and operational problem is the assumption of Associativity. Under Associativity, we can aggregate information by pieces, item by item, by applying a unique binary connective, no matter the dimension of our data set. The aggregated value can be obtained by aggregating information by means of a sequential one-by-one process. The aggregation process is fully characterized by a single binary aggregation rule, the one applied to the first couple of aggregated elements. Then the same computation is applied iteratively until the data set has been completely read. This is the case with classical fuzzy connective operators for conjunction and disjunction, t-norms and t-conorms (see, e.g., [7]).

Associativeness is a quite frequent assumption in group decision making models. Associative decision procedures are not affected by the order in which individuals express their opinions, and every aggregated opinion is considered just like another individual opinion. However, important operators are not associative.

Some decision procedures in a committee sometimes do not take into account committee highest and lowest qualifications for each candidate, and then they evaluate the mean of the remaining *middle* values. This is a particular case of *Ordered Weighted Averaging (OWA)* operators, introduced by Yager [12] in order to fill the gap between min (which is the maximal t-norm) and max (which is the minimal t-conorm). OWA operators are not associative, and their application requires that the number of items to be aggregated has been previously fixed.

Classical Mean Rule, for example, takes into account social support of each alternative, in such way that group opinions are no longer equivalent to individual opinions (see, e.g., [10]). Notice that a key OWA connective operator like the standard mean, defined as a mapping

$$M_n : [0,1]^n \rightarrow [0,1]$$

such that

$$M_n(a_1, \ldots, a_n) = \frac{\sum_{i=1}^n a_i}{n},$$

is not associative. Each mean M_n is just the mean of n numbers (it has been defined for a fixed n). M_n is just an operator, not a *rule*. When we refer to the *Mean Rule* we refer to the rule that evaluates the above mean for every n. The *Mean Rule* is not a single mapping M_n, but the complete sequence $\{M_n\}_n$ of all those mappings.

Moreover, it should be also pointed out that an operational calculus for the *Mean Rule* would not follow the above formula, but a *left* recursive calculus

$$M_n(a_1, \ldots, a_n) = \frac{(n-1)M_{n-1}(a_1, \ldots, a_{n-1}) + a_n}{n}$$

or alternatively, a *right* recursive calculus

$$M_n(a_1, \ldots, a_n) = \frac{a_1 + (n-1)M_{n-1}(a_2, \ldots, a_n)}{n}$$

where $M_2(a,b) = (a+b)/2$ (see [5, 10] for a discussion on some ethical and computational issues).

In general, it is not so easy to talk about *OWA rules.* Although several interesting families of OWA operators have been introduced in the past, showing the great flexibility in the choice of types of OWA operators (see, e.g., [13, 14, 15]), not every family of OWA operators can be properly considered as a *rule.*

In this paper we generalize the arguments introduced by the authors in [2], where each OWA rule was represented in terms of a family of binary OWA operators (see also [3, 4, 5]). Any connective rule will be conceived as a consistent family of connectives capable of solving arbitrary dimension problems. As a consequence, it is claimed that committees should always previously fix their *coherent* connective (aggregation) rule.

2 Recursive connective rules.

A connective rule should allow an aggregated value for any possible dimension of the list of items to be aggregated. That is, a connective rule should be a sequence of connective operators

$$\{\phi_n : [0,1]^n \rightarrow [0,1]\}_{n>1}$$

to be used to aggregate any finite number of items. We shall focus our attention here on those connective rules which allow the aggregation of arbitrary lists in a recursive manner. In particular, we shall consider those families of connective operators that can be defined by means of a left or a right recursive application of binary operators, once an appropriate re-arrangement of the items to be aggregated has been previously realized. In this way, a connective rule should be understood as a family of connective operators which can be recursively evaluated.

Obviously, in order to be considered as a *rule*, some *consistency* assumption has to be imposed on the family of connectives. Not every family of connectives either defines a connective rule or can be considered to *consistent*, or allow a recursive definition.

DEFINITION 1 *An ordering rule π on a set of items $A, A \neq \emptyset$, is a family of permutations*

$$\pi_B : B \longrightarrow B$$

such that for every finite sequence of items

$$B = (a_1, a_2, \ldots, a_n), \quad a_i \in A \; \forall i,$$

a new sequence is defined on B in such a way that

$$\pi_B(a) < \pi_B(b) \Rightarrow \pi_C(a) < \pi_C(b)$$

whenever $a, b \in B \cap C$.

An ordering rule tells us the exact position each new element will placed in any previously given ordered set of items.

An immediate example of ordering rule is the natural decreasing order of real numbers

$$\sigma : \Re \longrightarrow \Re$$

which assigns to each list of n numbers $(a_1, \ldots, a_n)$ its sorting permutation

$$\sigma(a_1, \ldots, a_n) = (a_{[1]}, \ldots, a_{[n]})$$

such that $a_{[i]} \geq a_{[j]}$ for all $i \leq j$.

DEFINITION 2 *A left-recursive connective rule is a family of connective operators*

$$\{\phi_n : [0,1]^n \rightarrow [0,1]\}_{n>1}$$

such that there exists a sequence of binary operators

$$\{L_n : [0,1]^2 \rightarrow [0,1]\}_{n>1}$$

verifying

$$\begin{aligned} \phi_2(a_1, a_2) &= L_2(\pi(a_1), \pi(a_2)) \text{ and} \\ \phi_n(a_1, \ldots, a_n) &= L_n(\phi_{n-1}(\pi(a_1), \ldots, \pi(a_{n-1})), \pi(a_n)) \end{aligned}$$

for some ordering rule π.

Right-recursiveness can be analogously defined.

DEFINITION 3 *A collection of connective operators*

$$\{\phi_n : [0,1]^n \to [0,1]\}_{n>1}$$

is said to be a right-recursive connective rule whenever

$$\phi_2(a_1, a_2) = R_2(\pi(a_1), \pi(a_2))$$

and

$$\phi_n(a_1, \ldots, a_n) = R_n(\pi(a_1), \phi_{n-1}(\pi(a_2), \ldots, \pi(a_n)))$$

hold for some family of binary operators

$$\{R_n : [0,1]^2 \to [0,1]\}_{n>1}$$

and some ordering rule π.

We immediately have that an operator $\phi_n : [0,1]^n \to [0,1]$ can be right-recursively defined if and only if it can be left-recursively defined. This can be seen as follows. Let

$$\phi_n(a_1, \ldots, a_n) = R_n(\pi(a_1), \phi_{n-1}(\pi(a_2), \ldots, \pi(a_n))).$$

Define the permutation $\hat{\pi}$ in such a way that for any n

$$\hat{\pi}(a) < \hat{\pi}(b) \Leftrightarrow \pi(a) > \pi(b).$$

Moreover, put $L_k(a,b) = R_k(b,a)$ for all k. Then we have

(a) $R_2(\pi(a_1), \pi(a_2)) = L_2(\pi(a_2), \pi(a_1))$

(b)

$$\begin{aligned} \phi_2(a_1, a_2, a_3) = R_3(\pi(a_1), \phi_2(\pi(a_2), \pi(a_3))) = \\ R_3(\pi(a_1), R_2(\pi(a_2), \pi(a_3))) = L_3(L_2(\pi(a_3), \pi(a_2)), \pi(a_1)) = \\ L_3(L_2(\hat{\pi}(a_1), \hat{\pi}(a_2)), \hat{\pi}(a_3)) \end{aligned}$$

(c) In general, we can see by induction that

$$\begin{aligned} R_n(\pi(a_1), \phi_{n-1}(\pi(a_2), \ldots, \pi(a_n))) = \\ L_n(\phi_{n-1}(\hat{\pi}(a_1), \ldots, \hat{\pi}(a_{n-1})), \hat{\pi}(a_n)) \end{aligned}$$

The ordering rule $\hat{\pi}$ is to be known as the *dual* ordering rule of π.

The existence of a right (left) recursion representation of a given operator does not imply in general the existence of an equivalent left (right) recursion representation by means of the same underlying ordering rule (see next section for an example). Of course, some rules will allow no one-side recursive definition (see also next section).

In some cases, such a recursive representation of a connective rule is fixed from the underlying ordering rule, as shown in the following result.

THEOREM 1 *Let*

$$\{\phi_n : [0,1]^n \to [0,1]\}_{n>1}$$

be a left-recursive connective rule with respect to the ordering rule π, *such that* $\phi_n(0,\dots,0) = 0$ *and* $\phi_n(1,\dots,1) = 1$, *with* ϕ_n *continuous and strictly increasing in each coordenate, for all* n. *Then* $\{L_n\}_{n>1}$ *is unique in its range for each ordering rule* π *such that*

$$\phi_n(a_1,\dots,a_n) = L_n(\phi_{n-1}(\pi(a_1),\dots,\pi(a_{n-1}),\pi(a_n)).$$

(Analogous result holds for right recursiveness).

Proof: First of all, notice that $L_n(0,0) = 0$ for all n. In fact, since $L_2(0,0) = \phi_2(0,0) = 0$, it follows that

$$0 = \phi_3(0,0,0) = L_3(L_2(0,0),0) = L_3(0,0)$$

and so on,

$$0 = \phi_n(0,\dots,0) = L_n(L_{n-1}(\dots L_2(0,0)\dots,0),0) = L_n(0,0)$$

Analogously, $L_n(1,1) = 1$ for all n. Moreover, every L_n is assured to be stricIty increasing too. It is direct for $n = 2$, and if we assume it is true for $n-1$, then we have

1. if $b = \pi(a_n) \in \{a_1,\dots,a_n\}$ increases and the other values remain constant, then $\phi_n(a_1,\dots,a_n)$ increases, and therefore $L_n(a,b)$ also increases;

2. if
$$a = L_{n-1}(\dots L_2(\pi(a_1),\pi(a_2))\dots,\pi(a_{n-1})),$$
increases, due to continuity of ϕ_{n-1} and the induction hypothesis it is assured the existence of a point $(\pi(b_1)\dots,\pi(b_{n-1}))$ such that $\pi(b_j) \geq \pi(a_j)$ for all j, with some strict inequality, such that
$$L_{n-1}(\dots L_2(\pi(b_1),\pi(b_2))\dots,\pi(b_{n-1}))$$
takes the new value, in such a way that
$$\phi_n(\pi(a_1),\dots,\pi(a_n)) < \phi_n(\pi(b_1),\dots,\pi(b_{n-1}),\pi(a_n))$$
and therefore L_n has increased as well.

Hence, for each a, b there is at most one c such that $L_j(c,a) = b$ (it is unique whenever the pair (a,b) belongs to the range of L_j, and therefore $c \geq a$ holds). That is, from

$$\phi_n(a_1,\dots,a_n) = L_n(c,\pi(a_n))$$

we can evaluate

$$c = L_{n-1}(\dots L_2(\pi(a_1),\pi(a_2))\dots,\pi(a_{n-1})).$$

Succesively, from this equation we can evaluate

$$L_{n-2}(\ldots L_2(\pi(a_1), \pi(a_2)) \ldots, \pi(a_{n-2}))$$

and so on, whenever we stay in the real ranges. ∎

Hence, if a recursive connective rule contains a continuous strictly increasing operator of dimension n, then consistent operators of lower dimension can be obtained according to the above result. Obviously, consistent upper dimension operators can not be freely chosen since for all n

$$\phi_n(a_1, \ldots, a_n) = L_n(\phi_{n-1}(\pi(a_1), \ldots, \pi(a_{n-1})), \pi(a_n))$$

and, analogously,

$$\phi_n(a_1, \ldots, a_n) = R_n(\pi(a_1), \phi_{n-1}(\pi(a_2), \ldots, \pi(a_n))).$$

An interesting case to analyze is the one in which left and right recursions share the same underlying ordering rule. That is, when

$$\phi_n(a_1, \ldots, a_n) = R_n(\pi(a_1), \phi_{n-1}(\pi(a_2), \ldots, \pi(a_n))) = \\ L_n(\phi_{n-1}(\pi(a_1), \ldots, \pi(a_{n-1})), \pi(a_n))$$

holds for some ordering rule π.

DEFINITION 4 *If both left and right recursiveness hold for the same ordering rule then we have* recursive rules.

In this way, recursiveness generalizes the concept of associativity, in the sense that recursive rules are the ones that can be evaluated iteratively (both sides), after an appropriate pre-arrangement of data. This ability of being iteratively evaluated is in fact the deep reason for associativity in practice. An operational calculus algorithm usually implies an iterative reckoning. But this iterative calculus does not necessarely requieres a unique binary operator. As shown above, the *Mean Rule* allows both left and right recursive definitions, although it is not associative.

The *Mean Rule* verifies an additional property: both left and right recursive definitions do not depend on the permutation, i.e., they are the same no matter the particular sequence of permutations being chosen. Left and right recursion hold for any possible ordering rule. If such a condition holds, we can talk about *commutative* recursive rules. Commutative recursive rules will be those connective rules which do not depend on any particular ordering rule.

In some way we could say that a connective rule $\{\phi_n\}_{n>1}$ is recursive if and only if a set of general associativity equations (in the sense of Mak [9]) hold for each n, once the items have been properly ordered. In fact, recursiveness holds whenever

$$\phi_n(a_1, \ldots, a_n) = \phi_n(\pi(a_1), \ldots, \pi(a_n)) = \\ R_n(\pi(a_1), \phi_{n-1}(\pi(a_2), \ldots, \pi(a_n))) = L_n(\phi_{n-1}(\pi(a_1), \pi(a_{n-1})), \pi(a_n))$$

for all n and some ordering rule π. If each one of these binary connective L_n, R_n can be assumed to be defined in the cartesian product of two nontrivial compact intervals on the real line, being continuous strictly increasing in each coordenate, then it can be shown (see [8]) that they are commutative and basically additive, in such a way that

$$\phi_n(a_1, \ldots, a_n) = \psi_0^{-1}(\psi_1(a_1) + \ldots + \psi_n(a_n))$$

for some homomorphisms in the unit interval $\psi_0, \psi_1, \ldots \psi_n$. This result can allow a particular representation of theorem 1. If we take, for example, the natural decreasing order σ as the underlying ordering rule, then each L_j is defined on a simplex $a_{j-1} \geq a_j$. Assuming the above conditions in a proper extended cartesian product of two nontrivial compact intervals, plus continuity and strict continuity, would assure such an additive solution (see [11]).

Associativity appears when the ordering rule is taken as the identity (i.e., the ordering rule keeping positions as presented), and

$$L_n = L_2 = F = R_2 = R_n \ \ \forall n$$

(that is, the whole recursive connective rule is characterized by a unique associative binary connective F, with no pre-arrangement of data).

Many connective rules $\{\phi_2, \ldots, \phi_n \ldots\}$ we can find in the literature are defined by means of a unique commutative and associative binary operator $\phi : [0,1]^2 \to [0,1]$ such that

$$\phi_n(a_1, \ldots, a_n) = \phi(\ldots, \phi(\phi(b_1, b_2), b_3), \ldots, b_n) =$$
$$\phi(b_1, \ldots, (b_{n-3}, \phi(b_{n-2}, \phi(b_{n-1}, b_n))) \ldots)$$

for $(b_1, \ldots, b_n)$ any permutation of $(a_1, \ldots, a_n)$. When we refer to a t-norm or a t-conorm as a *connective rule* we really mean the family of *connective operators* in such a way univocally defined (only one binary connective not depending on the ordering rule). The whole family of connective operators is fully characterized by its first connective operator of dimension 2, and no pre-arrangement of data is needed.

3 OWA recursive rules.

As pointed out above, sometimes only one underlying ordering rule is allowed by the decision maker. Perhaps there is only one natural way of ranking our data, and data reach to us previously pre-arranged according to such an ordering rule. If this is the case, the above concepts should be modified in order to meet such a restriction. Either a recursive definition is consistent with such an ordering rule, or such a recursive definition can not be applied. Either both recursive definitions make use of such an ordering rule, or it can not be applied as a recursive rule. For example, it may be the case that data are assumed to be ranked in its natural decreasing ordering, as happens with OWA rules.

Let us particularize the above ideas to the OWA case. First, we remind the reader some key concepts about OWA operators.

3.1 Basics on OWA operators.

OWA operators [12] are based upon the natural (decreasing) ordering. An OWA operator of dimension n is a connective operator

$$\phi : [0,1]^n \to [0,1]$$

such that for any list $(a_1, \ldots, a_n)$ then

$$\phi(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i a_{[i]}.$$

for some associated list of weights $W = (w_1, \ldots, w_n)$ such that

1. $w_i \in [0,1]$ for all $1 \leq i \leq n$
2. $\sum_{i=1}^{n} w_i = 1$

OWA operators are therefore assuming the (decreasing) natural ordering on the real line as the underlying ordering rule σ.

OWA operators are obviously commutative, monotone and idempotent, but as pointed out above, not associative in general. In fact, a binary ($n = 2$) OWA operator is associative if and only if either it is the *min* ($w_2 = 1$) operator or the *max* ($w_1 = 1$) operator. Therefore, given an OWA operator of dimension n, it can be only applied to aggregation problems of such a dimension n. If the dimension problem is modified, such OWA operator can not be applied.

Three short comments about OWA operators before going back to our operationality problem:

- A significative measure associated with OWA operators is the *orness* which estimates how close an OWA operator is to the *max* operator. It is defined as

$$orness(\phi) = \frac{1}{n-1} \sum_{i=1}^{n} (n-i) w_i.$$

  Dual to the measure of *orness* is the measure of *andness* defined as

$$andness(\phi) = 1 - orness(\phi),$$

  which therefore measures how close an OWA operator is to the min operator.

- Another important notion is *duality*. Given an OWA operator ϕ with weights $[w_1, \ldots, w_n]$, the dual $\hat{\phi}$ of ϕ is the OWA operator whose weights are $[w_n, \ldots, w_1]$. It is not difficult to see that $orness(\hat{\phi}) = andness(\phi)$.

- A particular class of OWA operator is given by the *buoyancy measures*. They are OWA operators that verify the property $w_i \geq w_j$ if $i < j$. Any buoyancy measure ϕ is such that $orness(\phi) \geq \frac{1}{2}$.

It may be the case that the existing left and right recursive definitions do not make use of the same underlying ordering rule. For example, the following two OWA operators

$$\phi_2(a_1, a_2) = \frac{1}{2}a_{[1]} + \frac{1}{2}a_{[2]}$$

and

$$\phi_3(a_1, a_2, a_3) = \frac{1}{4}a_{[1]} + \frac{1}{4}a_{[2]} + \frac{1}{2}a_{[3]}$$

allow a left recursive rule, since we can write

$$\phi_3(a_1, a_2, a_3) = \frac{1}{2}(\frac{1}{2}a_{[1]} + \frac{1}{2}a_{[2]}) + \frac{1}{2}a_{[3]}$$

But once we have chosen such a decreasing natural ordering as our ordering rule, then there is no function $h : [0,1]^2 \to [0,1]$ such that

$$\phi_3(a_1, a_2, a_3) = h(a_{[1]}, \frac{1}{2}a_{[2]} + \frac{1}{2}a_{[3]}).$$

Hence, they can not be together in the same right-recursive rule.

Moreover, not every family of OWA operators allows some one-side recursive definition based upon the same natural ordering rule. For example, no left or right recursive rule can be defined if we take $\phi'_2 = \phi_2$ as above, but

$$\phi'_3(a_1, a_2, a_3) = \frac{1}{4}a_{[1]} + \frac{1}{2}a_{[2]} + \frac{1}{4}a_{[3]}.$$

3.2 OWA recursive rules.

Although the standard associative procedure can not be considered when dealing with OWA operators, it may be the case that a recursive analysis can be applied to the decreasing ordered list $(a_{[1]}, \ldots, a_{[n]})$. Thus, practical OWA aggregation problems where the number of values to be aggregated is not previously known, should be solved by choosing one of these consistent recursive families of OWA operators, by means only of such a natural ordering rule. Each one of these families solves every aggregation problem for any arbitrary size of the input.

DEFINITION 5 *A recursive OWA rule is a recursive connective rule of OWA operators allowing left and right recursive definitions based upon the natural decreasing ordering rule, by means of binary OWA rules.*

This recursiveness definition has the advantage that aggregation weights can be computed quickly by using a dynamic programming approach (see [1]).

Such a recursiveness should not be confused with the *ordered linkage* property, considered in [6] in order to characterize OWA operators.

Anyway, we can check that once an OWA operator of dimension n has been fixed, all OWA operators of lower dimension belonging to its right and left OWA rules are almost univocally defined. In fact, it will be shown that every OWA operator can be recursively defined, both left and right, once the values to be aggregated have been pre-arranged according to the natural order in the real line. These two recursive representations will be basically unique.

THEOREM 2 *Let us consider a fixed OWA operator ϕ of dimension n. Then there exist at least one family of $n-1$ OWA operators of dimension 2*

$$L_2, \ldots, L_n$$

and another family of $n-1$ OWA operators all of them also of dimension 2

$$R_2, \ldots, R_n$$

allowing a left recursion and a right recursion, respectively, in such a way that

$$\phi(a_1, \ldots, a_n)$$

is equivalent to

$$L_n(L_{n-1}(\ldots L_3(L_2(a_{[1]}, a_{[2]}), a_{[3]}) \ldots), a_{[n]})$$

and

$$R_n(a_{[1]}, R_{n-1}(a_{[2]}, \ldots, R_3(a_{[n-2]}, R_2(a_{[n-1]}, a_{[n]})) \ldots))$$

Moreover, each one of these binary OWA operators is either unique or it can be freely chosen.

Proof: Let us assume right recursion, for example. If

$$R_n(b_1, b_2) = (1 - f(n))b_{[1]} + f(n)b_{[2]}$$

Then,

$$\phi(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i a_{[i]} = (1 - f(n))a_{[1]} + f(n)b_{[2]}$$

where

$$b_{[2]} = R_{n-1}(a_{[2]}, \ldots, R_3(a_{[n-2]}, R_2(a_{[n-1]}, a_{[n]})) \ldots))$$

in such a way that

$$f(n) = 1 - w_1.$$

Hence, if we now assume

$$R_{n-1}(b_1, b_2) = (1 - f(n-1))b_{[1]} + f(n-1)b_{[2]},$$

it must be

$$\phi(a_1,\ldots,a_n) = (1-f(n))a_{[1]} + f(n)[(1-f(n-1))a_{[2]} + f(n-1)b_{[3]}]$$

where

$$b_{[3]} = R_{n-2}(a_{[3]},\ldots,R_3(a_{[n-2]},R_2(a_{[n-1]},a_{[n]}))\ldots))$$

in such a way that

$$f(n-1) = 1 - w_2/f(n) = 1 - w_2/(1-w_1)$$

whenever $w_1 \neq 1$. In case $w_1 = 1$, it is the *max* rule, and the remaning binary OWA operators are not relevant at all. The process continues till we reach the trivial case R_2. In particular, in each step we obtain

$$f(n-i) = 1 - \frac{w_{i+1}}{1-\sum_{j=1}^{i} w_j}$$

whenever $\sum_{j=1}^{i} w_j < 1$ (otherwise, the remanining operators will be not relevent at all). ■

Notice that right recursion is unique for the *min* rule ($w_n = 1$), and left recursion will be unique for the *max* rule ($w_1 = 1$). In case $w_i \neq 0$ for all i, our OWA operator would be strictly increasing in each coordenate and theorem 1 would apply. The above result proves that in fact every OWA operator allows both left and right recursive definitions.

OWA rules as considered in this paper will consistently allow the recursive definition of each one of its operators. In other words, our *OWA rules* will be given by a sequence of OWA operators that can be explained in terms of a sequence of binary OWA operators allowing its right or left recursive representation. It is therefore natural to characterize each *recursive OWA rule* by means of the sequence of weights associated to its right or left recursive representation (see [2]).

DEFINITION 6 *A* basis function *is any mapping f that to any integer n associates a number in the unit interval (that is, $f(n) \in [0,1]$ for all n) with* $f(1) = 1$.

Each basis function f will then allow the recursive definition of two families of OWA operators. For any $n \geq 2$, we can define L_n and R_n such that

$$L_n(b_1,b_2) = (1-f(n))b_{[1]} + f(n)b_{[2]}$$

and

$$R_n(b_1,b_2) = f(n)b_{[1]} + (1-f(n))b_{[2]}$$

Then any left recursive operation

$$L_{n+1}(L_n(\ldots(L_2(a_{[1]},a_{[2]})),\ldots,a_{[n]}),a_{[n+1]})$$

and any right recursive operation

$$R_{n+1}(a_{[1]}, (R_n(a_{[2]}, \ldots, R_2(a_{[n]}, a_{[n+1]}))))$$

will always lead to OWA operators, for every $n \geq 2$, as it will be shown below. Each one of these two families of OWA operators ($\Phi = \{\phi_2, \ldots, \phi_n, \ldots\}$ if obtained via left-recursion call and $\Phi' = \{\phi'_2, \ldots, \phi'_n, \ldots\}$ if obtained via right-recursion call), will be then associated to the basis function f. According to our last theorem, left recursive (LR) and right recursive (RR) families of OWA operators will be defined, in particular, as follows:

- n is the dimension of the OWA operators ϕ_n and ϕ'_n;
- the weights of ϕ_n are denoted by $w_{1,n}, \ldots, w_{n,n}$ and $w'_{1,n}, \ldots, w'_{n,n}$ will denote the weights of ϕ'_n;
- for every $n \geq 2$ and every $i = 1, 2, \ldots, n$ we define

$$w_{i,n} = \begin{cases} f(n) & \text{if } i = n \\ (1 - f(n))w_{i,n-1} & \text{if } i < n \end{cases}$$

- for every $n \geq 2$ and every $i = 1, 2, \ldots, n$ we define

$$w'_{i,n} = \begin{cases} f(n) & \text{if } i = 1 \\ (1 - f(n))w'_{i-1,n-1} & \text{if } i > 1 \end{cases}$$

Therefore,

- $w_{i,n} = f(i) \prod_{j=i+1}^{n}(1 - f(j))$ for every $n \geq 2$ and every $i = 1, 2, \ldots, n$.
- $w'_{i,n} = f(n - i + 1) \prod_{j=n-i+2}^{n}(1 - f(j))$ for every $n \geq 2$ and every $i = 1, 2, \ldots, n$.

In view of the above equations it is immediate to check that ϕ_n and ϕ'_n are in fact OWA operators, since for any $n \geq 2$ we have

$$\sum_{i=1}^{n} w_{n,i} = \sum_{i=1}^{n} w'_{n,i} = 1$$

It is also easy to check now that not every family $\{\phi_2, \ldots, \phi_n, \ldots\}$ of OWA operators can be recursively defined by means of binary OWA operators on the basis of the decreasing natural ordering. Recursive consistency can be easily characterized by means of the weights of the OWA operators. For example, if $w_{i,k} = 0$, in order to be able to provide a left recursive characterization it must also be $w_{i,k+1} = 0$. Analogously, $w_{i+1,k+1} = 0$ must hold for a right recursive definition, whenever $w_{i,k} = 0$.

¿From theorem 2 it is implied that ϕ_n of dimension n being fixed, then all left-recursive and right-recursive *consistent* OWA operators with lower dimension $\phi_2, \ldots, \phi_{n-1}$ are univocally defined. More in general, we have the following result, which also gives a formal characterization of recursive consistency for OWA rules.

THEOREM 3 *Let us consider a family of OWA operators* $\{\phi_2, \ldots, \phi_n, \ldots\}$. *Then it can be defined by LR (i.e., it is* LR consistent*) if and only if* $w_{i,k}w_{j,k+1} = w_{j,k}w_{i,k+1}$ *for all* $i, j = 1, 2, \ldots, k$ *and every* k. *Analogously, such a family of OWA operators can be defined by RR (i.e., it is* RR consistent*) if and only if* $w_{i,k}w_{j+1,k+1} = w_{j,k}w_{i+1,k+1}$ *for all* $i, j = 1, 2, \ldots, k$ *and every* k.

Proof: Direct since every weight of the OWA operator of dimension k is multiplied by the same weight of the next binary OWA operators in order to allow the OWA operator of dimension $k + 1$. In case of right recursion, for example,

$$\sum_{i=1}^{k+1} w_{i,k+1}a_{[i]} = f(k+1)a_{[1]} + (1 - f(k+1)) \sum_{i=1}^{k} w_{i,k}a_{[i+1]}$$

Hence,

$$\frac{w_{i+1,k+1}}{w_{i,k}} = 1 - f(k+1)$$

for all $i = 1, 2, \ldots, k$. ■

Therefore, once any ϕ_n has been chosen, clear restrictions are implied in order to obtain families of OWA operators which are consistent with ϕ_n, both with respect to left recursion and right recursion. But if a left (right) recursion exists, the associated LR (RR) basis function is basically unique. Thus, each basis function is characterizing a LR (RR) consistent family of OWA operators. Moreover, it has been already pointed out above that right (left) recursive consistency for a given family $\Phi = \{\phi_2, \ldots, \phi_n, \ldots\}$ of OWA operators does not imply left (right) recursive consistency.

Being the underlying ordering rule fixed as the decreasing natural order, it may be the case that both left and right recursions exist for a given family of OWA operators $\{\phi_2, \ldots, \phi_n, \ldots\}$. The associated left and right basis functions f and g will in this case define the same family of OWA operators $\{\phi_2, \ldots, \phi_n, \ldots\}$. Many standard families of OWA opeartors do belong to this class, as shown in the next section.

The following theorem refers to duality, and the next one can be quite useful in estimating the orness of recursive OWA operators.

THEOREM 4 *Let* f *be a basis function with associated LR family of OWA operators* Φ, *and* Φ' *its associated RR family. Then* $\phi'_n \equiv \hat{\phi}_n$ *for all* n, *i.e.* ϕ'_n *is the dual of* ϕ_n *for all* n.

Proof: Immediate. ■

THEOREM 5 *Let* f *be basis function such that* $(1-f(n))f(n-1) \geq f(n)$ *for all* n, *and let us denote by* Φ *and* Φ' *the corresponding LR and RR families of OWA operators, respectively. Then every* ϕ_n *is a buoyancy measure. Therefore,* $orness(\phi_n) \geq \frac{1}{2}$ *and* $andness(\phi'_n) \geq \frac{1}{2}$.

Proof: Also immediate from the expression of weights in terms of the basis function. ■

4 Some examples.

We will now provide some interesting examples of recursive families of OWA operators.

4.1 Mean Rule.

The following result characterizes the Mean Rule as a commutative OWA rule.

THEOREM 6 *Let f be a basis function with associated LR and RR families of OWA operators being identical ($\Phi \equiv \Phi'$, that is, $\phi_n \equiv \hat{\phi}_n$ for all n). Then it must be $f(n) = 1/n$ for all n, and in turn the weights of each ϕ_n are $w_{i,n} = 1/n$ for all $i = 1, 2, \ldots, n$.*

Proof: We shall prove it by induction, just for the RR case. The result is obvious for $n = 2$. Let us assume $w_{i,n} = 1/n$ for all $i = 1, 2, \ldots, n$. Then it must be

$$w_{1,n+1} = f(n+1) = (1 - f(n+1))w_{n,n} = (1 - f(n+1))/n$$

in such a way that $f(n+1) = 1/(n+1)$. Therefore, since

$$f(n+1) = 1 - \frac{w_{i+1,n+1}}{w_{i,n}}$$

we obtain

$$w_{i,n+1} = (1 - f(n+1))(w_{i,n}) = \frac{1}{n+1}$$

■

4.2 Constant basis function.

Together with the *Mean Rule*, another case *a priori* deserving our attention are those rules characterized by a constant basis function (i.e., when there exists a value $a \in [0, 1]$ such that $f(n) = a$ for all $n \geq 2$).

In case $f(n) = a$ for all $n \geq 2$, each LR OWA operator ϕ_n will have weights

$$w_{1,n} = (1-a)^{n-1},$$

and

$$w_{i,n} = (1-a)^{n-i}a$$

for all $i = 2, \ldots, n$. Analogously, weights for each RR OWA operator ϕ'_n will be

$$w_{i,n} = (1-a)^{i-1}a$$

for all $i = 1, 2, \ldots, n-1$ and

$$w_{n,n} = (1-a)^{n-1}.$$

When $a = 1$ ($a = 0$) in left (right) recursion we obtain the *min* rule, and in right (left) recursion we obtain the *max* rule.

4.3 Harmonic OWA operators.

We recall that the n-th harmonic number is

$$H_n = \sum_{i=1}^{n} \frac{1}{i}.$$

Harmonic OWA operators are obtained by taking

$$f(n) = \frac{1/n}{H_n}.$$

Thus,

$$1 - f(n) = \frac{H_{n-1}}{H_n}.$$

By using theorem 5 it is immediate to see that the family of $LR-Harmonic$ OWA operators is a class of buoyancy measures. For example, its first OWA operators will have the following weights:

$$\begin{array}{rrrr} & & w_{1,2} = \frac{2}{3} & w_{2,2} = \frac{1}{3} \\ & w_{1,3} = \frac{6}{11}, & w_{2,3} = \frac{3}{11}, & w_{3,3} = \frac{2}{11} \\ w_{1,4} = \frac{12}{25}, & w_{2,4} = \frac{6}{25}, & w_{3,4} = \frac{4}{25}, & w_{4,4} = \frac{3}{25} \end{array}$$

4.4 A monotone fuzzy quantifier.

In [12, 13] it is shown how to obtain the evaluation of monotone fuzzy quantifiers by means of OWA operators. In particular, given a monotone non decreasing fuzzy quantifier Q such that $Q(0) = 0$ and $Q(1) = 1$, the weights $w_{i,n}$ for $i = 1, 2, \ldots, n$ of an OWA operator of dimension n to evaluate Q are defined as

$$w_{i,n} = Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right)$$

In case $Q(a) = a^r$ for some $r > 0$ we obtain that taking

$$f(n) = 1 - Q\left(\frac{n-1}{n}\right)$$

for all n, the associated left-recursive family of OWA operators does verify such a property. Hence, such a monotone fuzzy quantifier allows a left recursive definition. But it can not be right-recursively defined.

5 Final Comments.

OWA rules do play a main role in group decision making, since many aggregation procedures in practice are just particular cases. This paper generalizes previous results obtained just for OWA operators in [2]. A general approach to non-associative connective rules allowing an *operational* definition has been proposed. By *operational* we understand the ability of a recursive one-by-one evaluation, on the basis of a previous re-arrangement of the data set.

As a consequence, it has been stressed the fact that a connective rule, in order to be considered a rule, should be able to deal with any arbitrary number of items. An OWA operator is just an operator as the mean of n numbers is. None of them are *connective rules*, but single connectives. Considerably many real life decision processes require at different times to aggregate (possibly very large) lists of inputs of different dimensions. Connective rules have to be defined before knowing such a list. A connective rule is in general a rule allowing aggregation of any list, no matter its dimension.

Connective rules have been conceived here as consistent families of connective operators, allowing a representation in terms of right or left recursion of binary connective operators. Associativity is just an easy way of assuring such an operational representation.

Obviously, there are families of OWA operators that represent *rules* in the sense that they allow the evaluation of any arbitrary number of items, not allowing the recursive approach as developed in this paper, but being *consistent* in some other alternative sense. This is the case, for example, of the *Binomial* OWA rule $\{\phi_n(a_1, \ldots, a_n)\}_{n>1}$ where each ϕ_n is an OWA operator of dimension n with weights

$$w_{i,n} = \binom{n-1}{i-1} a^{i-1}(1-a)^{n-i} \quad \forall i = 1, \ldots, n$$

for some fixed $a \in (0, 1)$. Each one of these operators can be recursively defined, but the family itself does not verify the recursive OWA rule condition given in definition 5, neither the more general recursiveness definition 4. An operative description of this family of OWA operators, still by means of a sequence of binary OWA operators and the natural decreasing ordering, can be based upon the ordered linkage property of OWA operators (see [6]).

Acknowledgment: This research has been partially supported by DGICYT of Spain (grant PB95-0407) and Complutense University (grant PR295/95-6145).

References

[1] T.H. Cormen, C.E. Leiserson and R.R. Rivest *Introduction to Algorithms.* MIT Press, Cambridge, MA (1990).

[2] V. Cutello and J. Montero. Recursive families of OWA operators. In: P.P. Bonissone, Ed., *Proceedings of the Third IEEE Conference on Fuzzy Systems.* IEEE Press, Piscataway (1994); pp. 1137–1141.

[3] V. Cutello and J. Montero. Hierarchical aggregation of OWA operators: basic measures and related computational problems. *Uncertainty, Fuzziness and Knowledge-Based Systems* 3:17–26 (1995).

[4] V. Cutello and J. Montero. The computational problems of using OWA operators. In: B. Bouchon-Meunier, R.R. Yager and L.A. Zadeh, Eds., *Fuzzy Logic and Soft Computing.* World Scientific, Singapore (1995); pp. 166–172.

[5] V. Cutello and J. Montero. Information and aggregation: ethical and computational issues. In: D. Ruan, Ed., *Fuzzy Sets Theory and Advanced Mathematical Applications.* Kluwer, Boston (1995); pp. 175–198.

[6] J.C. Fodor, J.L. Marichal and M. Roubens. Characterization of the ordered weighted averaging operators. *Institut de Mathemátique, Université de Liège*, Prépublication 93.011.

[7] G.J. Klir and T.A. Folger. *Fuzzy sets, Uncertainty and Information.* Prentice Hall, Englewood Cliffs, NJ (1988).

[8] T.C. Koopmans. Representation of preference ordering with independent components of consumption. In: C.B. McGuire and R. Radner, Eds., *Decision and Organization.* North-Holland, Amsterdam (1972), 57–78 (2nd edition by the University of Minnesota Press, 1986).

[9] K.T. Mak. Coherent continuous systems and the generalized functional equation of associativity. *Mathematics of Operations Research*, 12:597–625 (1987).

[10] J. Montero. Aggregation of fuzzy opinions in a non-homogeneous group. *Fuzzy Sets and Systems*, 25:15–20 (1988).

[11] J. Montero, J. Tejada and V. Cutello. A general model for deriving preference structures from data. *European Journal Of Operational Research*, to appear.

[12] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18:183–190 (1988).

[13] R.R. Yager. Families of OWA operators. *Fuzzy Sets and Systems*, 59:125–148 (1993).

[14] R.R. Yager. MAM and MOM operators for aggregation. *Information Sciences*, 69:259–273 (1993).

[15] R.R. Yager. Aggregation operators and fuzzy systems modeling. *Fuzzy Sets and Systems*, 67:129–145 (1994).

QUANTIFIED STATEMENTS AND SOME INTERPRETATIONS FOR THE OWA OPERATOR

Patrick Bosc & Ludovic Liétard
IRISA/ENSSAT
Technopole ANTICIPA BP 447
22305 LANNION Cédex
FRANCE

Abstract. The interpretation of fuzzy quantified statements of the type "Q X are A" (where Q is a fuzzy quantifier and A is a fuzzy predicate) thanks to the OWA operator is the main topic of this paper. A meaning in terms of α-cuts of fuzzy sets is proposed and the relationships between this approach and fuzzy integrals on the one hand and Dempster-Shafer theory on the other hand, is investigated. The use of fuzzy quantified statements for database querying purposes is also illustrated.

Key-words. linguistic quantifiers, fuzzy quantified statements, Choquet fuzzy integral, upper and lower mean value, relational databases, flexible querying.

INTRODUCTION

The goal of this paper is to investigate the role of the OWA operator in the interpretation of expressions involving linguistic quantifiers, also called quantified statements. Linguistic quantifiers are defined as fuzzy sets and they are able to express a range of attitudes between conjunction (universal quantifier) and disjunction (existential quantifier). They are involved in linguistic expressions such as "most" or "about 3" so as to introduce some graduality or tolerance in some combination mechanism (for instance, "x fulfills most of the conditions $\{C_1, \ldots, C_n\}$" or "for at least half x's, P(x)"). Their potential applications are those of the usual quantifiers and they are involved in particular in decision making [16, 18, 19, 27, 29], expert systems [14, 23, 24, 36] and database flexible querying [4, 5, 17].

It is possible to distinguish between quantified statements of type "Q X are A" and "Q B X are A" where X denotes a usual set of elements (referential), A and B two vague predicates (young, well -paid) and Q a linguistic quantifier. In the first case,

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

the statement expresses that "Q elements of set X satisfy A" (e.g., "almost all the employees are young"), whereas several fairly different meanings may be assigned to the second expression, for instance "among the elements of X which satisfy B, Q satisfy A" or "among the elements of X, Q are such that if they are B, they are also A". Consequently, the remainder of this paper focuses on the first category of quantified statements. More specifically, the proposal made by R. Yager who suggests to use an OWA operator [29, 30], is investigated in order to justify this approach and to point out its semantics, some of its properties and the connections it has with Choquet fuzzy integral and Dempster-Shafer theory.

The rest of the paper is organized as follows. The second section is devoted to a brief overview of basic notions such as fuzzy measures, fuzzy integrals, possibility and necessity measures, fuzzy numbers. Linguistic quantifiers are presented in the third section where two types of representation are suggested for them. In the next section, Yager's interpretation of quantified statements by means of an OWA operator is recalled and the meaning of this approach in terms of α-cuts of a fuzzy set is proposed. In the fifth section it is shown that this interpretation can also be reached by a Choquet fuzzy integral. Another type of connection is established with Dempster-Shafer theory in the next section where the OWA-based interpretation is seen in terms of an expectation of an imprecise truth value. The last section illustrates the use of quantified statements in the framework of databases where they are components of a flexible query language.

BASIC NOTIONS: A SHORT REVIEW

This section provides the theoretical background necessary for the rest of the paper. Fuzzy sets and fuzzy predicates [32] are assumed to be familiar for the reader and we successively review the concepts of: fuzzy measure and integral, possibility and necessity measure, fuzzy number and cardinality of a fuzzy set.

Fuzzy Measures and Fuzzy Integrals

Very often, measures (e.g. probabilities) are supposed to satisfy an additivity property. Thanks to Lebesgue integral, it is then possible to extend the notion of measure to a function (which universe can be measured). For example, it is possible to calculate the area bounded by a stepwise function using Lebesgue integral. The value obtained stands for the extension of the notion of distance to the set of points represented by the considered function. The additivity property can be called into question, which leads to the fuzzy (non-additive) measures and fuzzy integrals.

A fuzzy measure [26] on a finite discrete universe X is a function g from $\mathcal{P}(X)$ (the powerset of X) to [0,1] satisfying the following properties:

- $g(X) = 1$,
- $g(\emptyset) = 0$,
- $\forall$ A, B such that $A \subset B$: $g(A) \leq g(B)$ (monotonicity).

Let f be a function from $X = \{x_1, x_2, ..., x_n\}$ into $\mathcal{R}^*$ (the subset of non negative reals) such that: $f(x_1) \le f(x_2) \le ... \le f(x_n)$ and g be a fuzzy measure. The Choquet fuzzy integral [21] of function f with respect to the measure g is given by :

$$C_g(f) = \sum_{i=1}^{n} f(x_i) * (g(A_i) - g(A_{i+1})) \quad (1)$$

where $A_i = \{x_i, x_{i+1}, ..., x_n\}$ and $A_{n+1} = \varnothing$. If we assume that $f(x_0) = 0$, this expression can be rewritten :

$$C_g(f) = \sum_{i=1}^{n} g(A_i) * (f(x_i) - f(x_{i-1})) \quad (2)$$

Example. Let us consider a set X made of 5 elements and a fuzzy measure g such that $g(F) = \|F\| / 5$ ($\|F\|$ is the usual cardinality of the crisp set F). If $f(x_1) = .1$, $f(x_2) = .4$, $f(x_3) = .8$, $f(x_4) = .9$, $f(x_5) = 1$, according to (1), we get:

$$C_g(h) = .1 * (1 - .8) + .4 * (.8 - .6) + .8 * (.6 - .4) + .9 * (.4 - .2) + 1 * (.2 - 0)$$
$$= .1 * .2 + .4 * .2 + .8 * .2 + .9 * .2 + 1 * .2 = .64$$

and with (2) we have:

$$= 1 * (.1 - 0) + 4/5 * (.4 - .1) + 3/5 * (.8 - .4) + 2/5 *(.9 - .8) + 1/5 * (1 - .9)$$
$$= 1 * .1 * 4/5 * .3 + 3/5 * .4 + 2/5 * .1 + 1/5 * .1 = .64$$

which illustrates the equivalence between formulae (1) and (2)♦

Possibility theory

Similarly to a probability, a possibility is a function intended for the estimation of event occurrences. On a finite universe [23], both possibility and probability measures are fuzzy measures, but they convey different semantics. A probability gives a frequency for the occurrence of an uncertain event and it is possible to state that all elementary events have the same eventuality. However, it is impossible to express ignorance concerning the occurrences of events. On the contrary, possibility theory aims at modeling this type of ignorance, but does not allow for expressing that all events have the same eventuality. Basically, possibilities describe preferences over situations.

A possibility measure [35] is a function Π from $\mathcal{P}(X)$ into [0, 1] such that:

- $\Pi(X) = 1$,
- $\Pi(\varnothing) = 0$,
- $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$.

$\Pi(A)$ stands for the possibility degree of the event A which may be a set of elementary events. $\Pi(A) = 1$ means that A is a completely possible event and more generally, the closer to 1 $\Pi(A)$, the more possible the event A. When the property $A \cup \overline{A} = X$ holds (non fuzzy events), the property:

$$\max(\Pi(A), \Pi(\overline{A})) = 1$$

also holds. From a practical point of view, the possibility measure is drawn from a possibility distribution describing the possibility degree of every elementary event. A possibility distribution over a universe X is a function π from X to [0, 1] such that there exists at least one element x_0 with $\pi(x_0) = 1$. Such a distribution may be seen as a normalized fuzzy set and:

$$\forall\ A \in \mathcal{P}(X),\ \Pi(A) = \sup_{x \in A} \pi(x) \qquad (3).$$

However, the knowledge of $\Pi(A)$ does not convey all the information related to the occurrence of A (in particular, if both $\Pi(A)$ and $\Pi(\overline{A})$ equal 1). The possibility measure can be completed by the necessity measure N(A) [8, 9]:

$$N(A) = 1 - \Pi(\overline{A}).$$

The closer to 1 N(A), the more certain the occurrence of event A.

Fuzzy Numbers

A fuzzy number $\mathcal{N}$ is a number which value is ill-known (or imprecise). It is represented by a possibility distribution π over $\mathcal{R}$ (the real numbers) such that [9]:

- $\forall\ u, v \in \mathcal{R},\ \forall\ w \in [u, v],\ \pi(w) \geq \min(\pi(u), \pi(v))$ convexity
- $\exists n!$ such that $\pi(n) = 1$ uniqueness of the modal value.

$\pi(u)$ expresses the degree of possibility that the actual value of $\mathcal{N}$ is u.

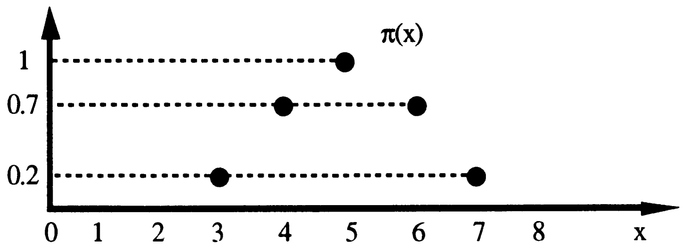


Figure 1. The fuzzy number "about 5".

Example. Consider the fuzzy number "about 5" depicted in Figure 1. This description indicates that it is completely possible that $\mathcal{N}$ equals 5 ($\pi(5)=1$) and that the values 3, 4, 6 and 7 are somewhat possible but 5 is preferred. The necessity that N is in {4, 5, 6} is .8 since: $N(\{4, 5, 6\}) = 1 - \Pi(\{0, 1, 2, 3, 7, 8\}) = 1 - 0.2$♦

Dubois and Prade [10, 11] have proposed to represent the mean value of a fuzzy number $\mathcal{N}$ (represented by a possibility distribution over the successive reals $\{x_0, x_1, x_2, ..., x_n, ..., x_m, x_{m+1}\}$ whose modal value is x_n) by the upper mean value $U(\mathcal{N})$ and the lower mean value $L(\mathcal{N})$ according to the following definitions:

- $U(\mathcal{N}) = \sum_{i=n}^{m} x_i * (\pi(x_i) - \pi(x_{i+1}))$
- $L(\mathcal{N}) = \sum_{i=1}^{n} x_i * (\pi(x_i) - \pi(x_{i-1}))$

with $\pi(x_0) = \pi(x_{m+1}) = 0$.

Example. Let us come back to the fuzzy number "about 5" introduced previously:

$$\pi = 0.2/3 + 0.7/4 + 1/5 + 0.7/6 + 0.2/7.$$

Its upper and lower mean values are respectively:

- $U(\mathcal{N}) = (0.3 \times 5) + (0.5 \times 6) + (0.2 \times 7) = 5.9$
- $L(\mathcal{N}) = (0.2 \times 3) + (0.5 \times 4) + (0.3 \times 5) = 4.1.$

The identical variation with respect to 5 (.9) is obviously due to the symmetry of the distribution used here ♦

In Dempster-Shafer theory [7, 25], $U(\mathcal{N})$ (resp. $L(\mathcal{N})$) represents the largest (resp. smallest) expectation that can be obtained when $\mathcal{N}$ is represented by a probability distribution bounded by the considered possibility distribution. Another meaning of these values is suggested in [2]: $U(\mathcal{N})$ (resp. $L(\mathcal{N})$) is the expectation of $\mathcal{N}$ when it is supposed to be greater (resp. smaller) than or equal to its modal value.

Cardinality of a Fuzzy Set

The cardinality of a fuzzy set has been defined as an imprecise value in [10, 22] and this approach has the advantage of preserving most of the properties of the cardinality of a regular set. The cardinality of F is described by a fuzzy number (a possibility distribution πc) defined as follows:

- let k be the number of values of F whose degree is 1: $\pi c(k) = 1$,
- $\forall i < k, \pi c(i) = 0$,
- $\forall j > k$ $\pi c(j)$ is the jth largest value $\mu_F(x)$.

This view is worthy of two comments:

- a) if F is a usual set, πc is exactly the description of a precise value ($\pi c(k) = 1$ and $\pi c(i) = 0$ $\forall$ $i \neq k$) which is the cardinality of this set,

- ii) $\forall$ i, j such that $i > j \geq k$, $\pi c(i) \leq \pi c(j)$; this states that the preference over the values of cardinalities decreases as elements with decreasing grades appear.

Example. Let $F = \{1/x_1 + 1/x_2 + 1/x_3 + 1/x_4 + 1/x_5 + 0.8/x_6 + 0.3/x_7\}$. Its imprecise cardinality is the fuzzy number:

$$\pi c = \{1/5, 0.8/6, 0.3/7\}.$$

It is certain that the cardinality is in the set {5, 6, 7} and completely possible that it is 5 since there are 5 elements in F with a membership degree equal to 1. The certainty that the cardinality is exactly 5 is only .2 ♦

LINGUISTIC QUANTIFIERS

The semantics of the universal and existential quantifiers ($\forall$ and $\exists$) of first order logic are limited to the conjunction (all) and the disjunction (at least one). Some intermediary attitudes between these two extremes are sometimes desirable and that is the reason why linguistic quantifiers have been introduced by L. Zadeh [36]. These quantifiers represent linguistic expressions (nearly all, about 2) and are defined by fuzzy sets.

Moreover, a linguistic quantifier is either absolute or relative. The absolute quantifiers refer to a number (about 3, nearly 2) whereas the relative quantifiers refer to a proportion (about the quarter, nearly the half). The linguistic quantifiers can be increasing (resp. decreasing) [29] which means that if the satisfaction of the predicate A by the elements of X increases then the degree of truth of "Q X are A" cannot decrease (resp. increase). "Almost all", "at least 3" are examples of increasing quantifiers whereas "at most the half" is an example of a decreasing quantifier.

This section defines two different representations for increasing fuzzy quantifiers. The first one represents a quantifier by a function of the real line. To interpret "Q X are A" is then to determine the value of a function into an imprecise argument. The second representation is more general and defines a fuzzy quantifier by a fuzzy measure and in this context the interpretation of the statement "Q X are A" relies on the application of a fuzzy measure on a fuzzy set.

Representation Using the Real Line

An absolute quantifier is represented by a function Q from an integer range to [0,1] whereas a relative quantifier is represented by a function Q from [0,1] to [0,1] [36]. In both cases, the value Q(j) is defined as the truth value of the statement "Q X are A" when exactly j elements from X fully satisfy A (whereas it is assumed that A is totally unsatisfied for the other elements). It means that function Q gives the truth value of "Q X are A" when A is a Boolean predicate (i.e. A is completly satisfied or completly unsatisfied). In the general case, A is fuzzy and to interprate "Q X are A" is to determine the value of function Q into the cardinality of a fuzzy set (if Q is absolute). If Q is relative, arguments of function Q are proportions (of elements satisfying the regular predicate A) and the general case needs to determine the proportion of elements satisfying a gradual property (i.e. satisfying the fuzzy predicate A).

Monotonous fuzzy quantifiers are such that if Q is increasing, then: i) $Q(0) = 0$, ii) $\exists k$ such as $Q(k) = 1$ and iii) $\forall$ a,b if $a > b$ then $Q(a) \geq Q(b)$. When Q is decreasing: i) $Q(0) = 1$, ii) $\exists k$ such as $Q(k) = 0$ and iii) $\forall$ a,b if $a > b$ then $Q(a) \leq Q(b)$. Figure 2 gives an example of a monotonous increasing quantifier.

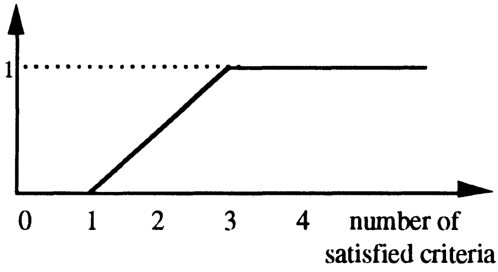


Figure 2. The fuzzy quantifier "at least about 3".

Example. Let Q be the fuzzy quantifier of Figure 2 and X be the crisp set made of 3 employees e1, e2, e3 earning respectively \$48,000, \$45,000 and \$80,000. Let A be the fuzzy predicate "well-paid" described in Figure 3.

The satisfaction of the predicate "well-paid" by these employees is given by the fuzzy set: $\{0.9/e1 + 0.75/e2 + 1/e3\}$. The evaluation of the statement "at least about 3 employees are well-paid" relies on the determination of the number of "well-paid" employees discussed in the next section♦

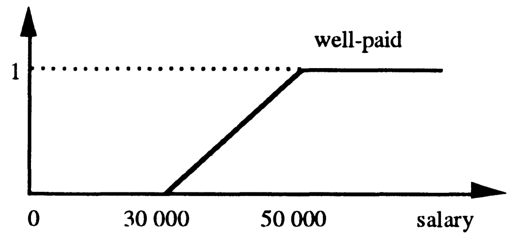


Figure 3. The fuzzy predicate "well-paid".

Representation Using a Fuzzy Measure

If we consider a quantified statement of the type "Q X are A", where A is a Boolean predicate, it is possible to define its truth value as the measure of the subset F (g(F)) comprising the elements of X which satisfy A [3, 6]. In this context, the quantifier is associated with the fuzzy measure g. One limitation concerns the nature of the quantifier due to the fact that a measure is monotonic. In the following, we consider only increasing quantifiers.

From the definition of a quantifier Q on the real line, the corresponding fuzzy measure g is obtained according to the following transformations:

- if Q is relative, $g(E) = Q(\|E\|/n)$ (4),
- if Q is absolute, $g(E) = Q(\|E\|)$ (5)

where $\|E\|$ denotes the cardinality of the set E and n the cardinality of the referential. It should be noticed that when Q is absolute, the referential X must fulfill $Q(\|X\|) = 1$ which is not always guaranteed (cf. next example). In that case, it is necessary to introduce dummy elements (which do not fulfill A at all) to comply with all the properties of fuzzy measures [3, 6].

Example. Let X be the crisp set made of 2 elements x_1, x_2 and the quantified statement "at least about 3 X are A". The quantifier "at least about 3" is represented using the measure g such that:

$$g: \{x_1, x_2, x_3\} \rightarrow [0,1] \text{ with } g(E) = Q(\|E\|),$$

where $Q(\|E\|)$ is given by Figure 2. In this case, with $X = \{x_1, x_2\}$, $Q(\|X\|)$ does not equal 1 and the referential used is $X' = \{x_1, x_2, x_3\}$ where $\mu_A(x_3) = 0$, for which $Q(\|X'\|) = 1$ holds ♦

It is worth noticing that this representation of quantifiers in terms of fuzzy measures is more powerful than the former since it is able to deal with situations where the model based on the real line does not work.

Example. Let us consider the quantifier "almost all" defined over the set $X = \{x_1, x_2\}$ by: $g(\{x_1, x_2\}) = 1$, $g(\{x_1\}) = .8$, $g(\{x_2\}) = .1$. Such a quantifier privileges x_1 with respect to x_2 and this preference cannot be handled with the real line modeling where all elements are implicitly "equivalent" ♦

Up to now, the predicate A has been assumed to be a crisp one and the case where A is fuzzy, leads to apply the measure to a fuzzy set. This topic will be dealt with in a further section.

QUANTIFIED PROPOSITIONS AND OWA

An OWA (Ordered Weighted Averaging) of n values x_i's [29] is defined by:

$$\mathrm{OWA}(x_1, \dots, x_n) = \sum_{i=1}^{n} (w_i * x_{k_i}) \quad (6)$$

where x_{k_i} is the i^{th} largest values among the x_k's and the sum of the weights w_i's equals 1. Let n be the crisp cardinality of a set X and Q be an increasing (fuzzy) quantifier defined by a function of the real line. According to Yager [29], the truth value of a quantified statement of the type "Q X are A" is given by an OWA applying to the values $\mu_A(x_i)$'s. More precisely, the weights are given by:

- $w_i = Q(i) - Q(i-1)$ if Q is absolute $\quad (Q(0) = 0)$
- $w_i = Q(i/n) - Q(i-1/n)$ if Q is relative $\quad (Q(0) = 0)$.

Each weight w_i represents the increase of satisfaction if one compares a situation where exactly (i - 1) elements are entirely A with a situation where i elements are entirely A (assuming that the other elements are entirely not satisfying A). This approach can be extended to a decreasing quantifier Q [1, 31], knowing the following equivalence:

$$\text{"Q X are A"} \Leftrightarrow \text{"Q' X are } \overline{A}\text{"},$$

where $\overline{A}$ is the negation of A and Q' (which is increasing) is the antonym of Q (i.e., $\mu_{Q'}(u) = \mu_Q(1 - u)$ if Q is relative; the reader interested in the case of an absolute quantifier can refer to [1]). In so doing, the evaluation of "Q X are A" with Q decreasing comes down to the evaluation of an equivalent statement where the quantifier involved is increasing and the weights for the calculation using an OWA can be derived as mentioned previously.

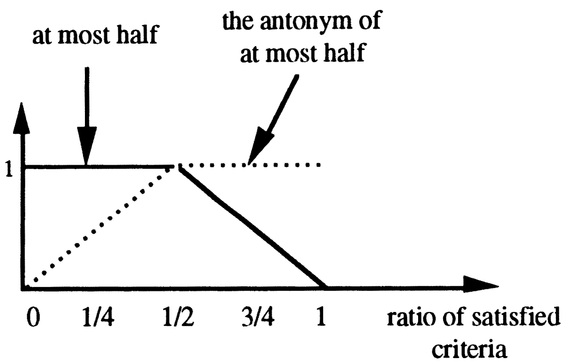


Figure 4. The quantifier "at most half" and its antonym.

Example. Let X be the crisp set made of 4 employees e1, e2, e3, e4 earning respectively $38,000, $40,000, $45,000 and $80,000. The satisfaction of the predicate "well-paid" (cf. Figure 3) is described by the fuzzy set: {.4/e1 + .5/e2 + .75/e3 + 1/e4}. If the quantifier "at most half" is the one of Figure 4, the evaluation of the statement "at most half employees are well-paid" by an OWA operator requires to tranform it into the statement: "at least half employees are not well-paid" if we call the antonym of "at most half": "at least half" (see Figure 4). The truth value of the initial expression is:

t("at most half employees are well-paid") =

t("at least half employees are not well-paid") =
.6 * (Q(1/4) - Q(0)) + .5 * (Q(1/2) - Q(1/4)) + .25 * (Q(3/4) - Q(1/2)) +
0 * (Q(1) - Q(3/4)) = .6 * .5 + .5 * .5 + .25 * 0 + 0 * 0 = .55 ♦

An important interesting question is that of the "intuitive meaning" of this kind of aggregation. In other words, how can we describe the spirit of this caculus? Let us come back to the application of formula (6) to the statement "Q X are A", we get the truth value t:

$$t = \sum_{i=1}^{n} (Q(i) - Q(i - 1)) * \mu_A(x_{k_i}).$$

First, let us assume that all the $\mu_A(x_{k_i})$'s are distinct. We can then consider that each is the level value of an α-cut of A and that the term (Q(i) - Q(i - 1)) is the intrinsic contribution (to the overall satisfaction) of this α-cut. Now, let us generalize this to the case where several $\mu_A(x_{k_i})$'s are identical, i.e., $\mu_A(x_{i_1}) = ... = \mu_A(x_{i_p}) = \beta$. The expression:

$$\sum_{j=i}^{i+p} (Q(j) - Q(j - 1)) * \mu_A(x_{k_j}) = (Q(i + p) - Q(i)) * \beta$$

represents the product of the level value of this α-cut (β) by its proper contribution to the satisfaction of the quantifier Q (since Q(i + p) - Q(i) represents the increase of satisfaction brought by the p elements of this α-cut). From this, we can state that this interpretation of a quantified statement is "α-cut based". More precisely, the truth value is a weighted mean: the sum of the level values of the α-cuts of A weighted by their proper contribution.

Example. Let X be the crisp set made of 4 employees e1, e2, e3, e4 earning respectively \$48,000, \$45,000, \$70,000 and \$80,000. The satisfaction of the predicate "well-paid" (cf. Figure 3) is described by the fuzzy set: {.4/e1 + .4/e2 + 1/e3 + 1/e4}. If the quantifier "at least about 3" is the one of Figure 2, the evaluation of the statement "at least about 3 employees are well-paid" leads to:

1 * 0 + 1 * .5 + .4 * .5 + .4 * 0 = .7	if we apply directly (6)
1 * .5 + .4 * .5 = .7	if we reason in terms of α-cuts ♦

One may notice that the case where A is a usual set is captured since in that case there is only one α-cut whose level value is 1 and if it comprises k elements, its proper contribution is: Q(k) - Q(0) = Q(k).

FUZZY INTEGRALS AND OWA

Let Q be a linguistic quantifier whose associated fuzzy measure is g. The interpretation of the quantified statement "Q X are A" consists in determining the value of the measure g over the fuzzy subset A of X. It has been shown that the

Choquet fuzzy integral is a means to make such a calculus [15] and it is then interesting to look at the connections between this view and the previous one.

Let g be a fuzzy measure over the powerset of $U = \{x_1, x_2, ..., x_c\}$ and let us consider the statement "Q X are A" where X is a subset of U whose cardinality is n $(n \leq c)$. It is necessary to consider a superset of X to deal with the case of an absolute quantifier such that $g(\|X\|) < 1$. In such a situation, the (c - n) elements of (U - X) are assumed to fulfill: $\mu_A(x_i) = 0$.

Let us consider the subset EA of U of elements satisfying (more or less) A:

$$\forall i,\ \mu_{EA}(x_i) = \mu_A(x_i).$$

The Choquet integral of function μ_{EA} with respect to g is:

$$C_g(EA) = \sum_{i=1}^{c} \mu_{EA}(x_i) * (g(F_i) - g(F_{i+1})) = \sum_{i=1}^{c} g(F_i) * (\mu_{EA}(x_i) - \mu_{EA}(x_{i-1}))$$

where $F_i = \{x_i, x_{i+1}, ..., x_c\}$, $F_{c+1} = \varnothing$, $\mu_{EA}(x_1) \leq ... \leq \mu_{EA}(x_c)$, $\mu_{EA}(x_0) = 0$.

It has been proved in [3] that $C_g(EA)$ is also given by an OWA provided that Q can be defined as a function of the real line complying with formulae (4) (resp. (5)) if Q is absolute (resp. relative).

Example. Let X be the crisp set made of 3 employees e1, e2, e3 earning respectively \$48,000, \$45,000 and \$80,000. Let Q be the increasing fuzzy quantifier "at least about 3" of Figure 2 given by the following fuzzy measure g:

$$g: \{e_1, e_2, e_3\} \to [0,1] \text{ such that: } g(E) = Q(\|E\|).$$

The satisfaction of the predicate "well-paid" (cf. Figure 3) is described by the fuzzy set: {.9/e1 + .75/e2 + 1/e3}. The interpretation of the statement "at least about 3 employees are well-paid" by the Choquet fuzzy integral leads to:

$$C_g(EA) = .75 * (g(\{e_2, e_1, e_3\}) - g(\{e_1, e_3\})) + .9 * (g(\{e_1, e_3\}) - g(\{e_3\})) + 1 * (g(e_3\}) - g(\varnothing\})) = .75 * (1 - .5) + .9 * (.5 - 0) + 1 * (0 - 0) = 0.825.$$

The interpretation with an OWA is:

$$OWA = 1 * (Q(1) - Q(0)) + .9 * (Q(2) - Q(1)) + .75 * (Q(3) - Q(2))$$
$$= 1 * 0 + .9 * .5 + .75 * .5 = .825 ♦$$

MEAN VALUES AND OWA

In [28], R.R. Yager proposes to define the truth value of the statement "Q X are A" as the value of the fuzzy quantifier Q (defined as a function of the real line) for the imprecise cardinality of the fuzzy set A of X. To do that, he suggests to use Zadeh's extension principle [34] which enables to establish (in terms of a fuzzy set) the value of a function in an imprecise datum (represented by a possibility distribution). In this particular case, the idea is to compute a fuzzy number describing the value of function Q into an imprecise cardinality.

According to this approach, the truth value of "Q X are A" is the fuzzy number $\mathcal{N}$ represented by the possibility distribution $\pi_{\mathcal{N}}$ defined by:

$$\pi_{\mathcal{N}}(i) = \max_{i = Q(j)} \pi c(j)$$

where πc describes the imprecise cardinality of the considered set. It has been proved in [2] that the upper mean value of this imprecise truth value ($\mathcal{N}$) is the one obtained with the OWA calculus.

It has also been proved [2] that the lower mean value necessarily equals the modal value (due to the specific shape of the underlying possibility distribution where all values smaller than the modal value have a null possibility degree) and then $L(\mathcal{N})$ conveys no interesting information.

The previous result gives a probabilistic flavor of the OWA since mean values are expectations. More precisely, the expression of the OWA:

$$OWA = \sum_{i=1}^{n} \mu_A(x_i) * (Q(i) - Q(i-1))$$

can be rewritten:

$$OWA = \sum_{i=0}^{n} Q(i) * (\mu_A(x_i) - \mu_A(x_{i+1}))$$

if we assume that: $\mu_A(x_0) = 1$ and $\mu_A(x_{n+1}) = 0$.

Each value $(\mu_A(x_i) - \mu_A(x_{i+1}))$ may be interpreted [12, 13] as the probability that the set $\{x_1, x_2, ..., x_i\}$ is a representation of the elements of A $((\mu_A(x_0) - \mu_A(x_1))$ being the probability associated with the empty set). Each value Q(i)stands for the satisfaction of the set $\{x_1, x_2, ..., x_i\}$ with respect to the quantifier Q (Q(0) is the satisfaction of the empty set). Knowing that the probability of a fuzzy event is the expectation of its membership function [33], the OWA operator expresses the probability of the fuzzy event: "the cardinality of the fuzzy set of elements which more or less comply with A is in agreement with Q".

Example. Let Q be the increasing fuzzy quantifier of Figure 2. Let X be the crisp set made of 3 employees e1, e2, e3 earning respectively \$48,000, \$45,000 and \$80,000. These employees satisfy "well-paid" according to the fuzzy set: {0.9/e1 + 0.75/e2 + 1/e3}. The imprecise cardinality of the set made of the "well-paid" employees is given by the possibility distribution $\pi c(1) = 1$, $\pi c(2) = .9$ and $\pi c(3) =$

.75. Thus, the imprecise truth value for the statement "at least 3 employees are well-paid" is the fuzzy number $\mathcal{N}$ described by the possibility distribution $\pi_{\mathcal{N}}$ below:

$$\pi_{\mathcal{N}} = \pi c(1)/Q(1) + \pi c(2)/Q(2) + \pi c(3)/Q(3) = 1/0 + .9/.5 + .75/1.$$

It is completely possible that the statement is false, possible at the degree .9 that it is half satisfied and possible at the degree .75 that it is true. The upper mean value of $\mathcal{N}$ is:

$$U(\mathcal{N}) = 0 * (1 - .9) + .5 * (.9 - .75) + 1 * (.75 - 0) = 0.825.$$

This result is exactly the one obtained in the preceding example ♦

QUANTIFIED STATEMENTS AND FLEXIBLE QUERYING

This section aims at an illustration of the use of quantified statements in the database area. We place ourselves in the context of ordinary relational DBMS's, where it is assumed that the stored data are precisely known. The interest for flexible queries is to introduce preferences inside criteria and/or in their aggregation so as to obtain discriminated answers (the closer to the condition an element, the better its grade). Fuzzy set theory offers a particularly well suited framework for this purpose since preference is a basic notion. Moreover, the use of fuzzy quantifiers enables to prevent the limitations of crisp ones. Let us consider the query: "retrieve the department stores where all the employees are well-paid". A store where only one employee e is such that $\mu_{\text{well-paid}}(e) = 0$ is rejected and this behavior may be too sharp. In our proposal, it is possible to soften the previous query into "retrieve the department stores where most of the employees are well-paid".

In this section, we consider an SQL-like query language (namely SQLf [4]) which supports flexible querying using fuzzy predicates. We focus on introducing fuzzy quantified statements inside queries and three constructs of the language are reviewed. For each of them, the underlying fuzzy quantified statement of the type "Q X are A" is made explicit.

Horizontal Quantification

First, we consider that the fuzzy quantifier aggregates different satisfaction degrees related to a same tuple. This type of calculus was initially proposed in [17] and it gave birth to the term "horizontal quantification". An example of such a query could be: "retrieve the employees satisfying most of the following predicates: "well-paid", "young", ..., "important position". In SQLf, this type of query is:

select * **from** R **where** Q $(fc_1, fc_2, ..., fc_n)$

where R is a regular relation, Q a fuzzy quantifier and $(fc_1, fc_2, ..., fc_n)$ is a list of n fuzzy conditions. The degree of satisfaction for a tuple t of R (with respect to this query) is the truth value of the statement "Q X are A" where $X = \{fc_1, fc_2, \ldots, fc_n\}$ and A = "to be satisfied by t" ($\forall\ fc_i, A(fc_i) = fc_i(t)$).

Example. Considering a relation EMPLOYEE(#emp, #dep, salary, position, age, city) describing employees of a plant, the query aiming at the retrieval of the employees living in Paris, fulfilling at least 2 conditions among "well-paid", "interesting position", "young" is expressed by the following SQLf query:

select #emp **from** EMPLOYEE **where** city = 'Paris' **and**
at least 2 (well-paid(salary), interesting(position), young(age)) ♦

Vertical Quantification and Partitioning

Now, the fuzzy quantifier applies to a set of tuples and the aggregation works on this set which originates the term "vertical quantification". Similarly to SQL, the partitioning is expressed in SQLf using the key-word "group by" and its logical effect is to split a relation into groups of tuples sharing the same value on one or several attributes.

The general SQLf expression for this kind of query is:

select att **from** R **group by** att **having** Q **are** A.

This query splits relation R into subsets sharing the same value for the attribute "att". The degree to which a value v of attribute "att" is an answer is the truth value of the statement "Q X are A" where X is the subset made of the tuples sharing the value v for attribute "att".

Example. The query "retrieve the departments where at least 2 employees are well-paid" may be expressed:

select #dep **from** EMPLOYEE **group by** #dep **having** at-least 2 **are** well-paid

In this example, each subset gathers the employees working in a same department ♦

Vertical Quantification and Query Nesting

An interesting feature of SQL is that it allows for the nesting of query blocks to an arbitrary depth. In particular, two select blocks can be connected by means of the existential or the universal quantifier using the key words "any" or "all". In this case, the internal block (subquery) is seen as a set of values which are compared with the value of the current tuple. For instance, the query "find the employees whose salary is greater than that of "all" the engineers" can be expressed:

select * **from** EMPLOYEE **where** salary ≥ **all**
(**select** salary **from** EMPLOYEE **where** position = "engineer").

Let us now assume that the user wants to retrieve the employees whose salary is much greater than the salary of most engineers. In SQLf, this query is:

select * **from** EMPLOYEE **where** salary >> **most of**
(**select** salary **from** EMPLOYEE **where** position = "engineer").

This query can be interpreted in the following fashion. For each employee e, the fuzzy quantified statement "Q X are A" is evaluated, where Q = "most of", X = the set of the engineers' salaries, A = e's salary much greater than an engineer's salary.

More generally, if θ is a fuzzy comparison operator, att1 (resp. att2) is an attribute of R (resp. S), Q is a fuzzy quantifier and fc is a Boolean condition, the SQLf query:

select * **from** R **where** att1 θ Q (**select** att2 **from** S **where** bc)

retrieves any tuple r of R which satisfies (more or less) the condition "Q X are A" where X is a set of att2 values issued from S (with respect to the Boolean condition bc) and A is the fuzzy predicate: r.att1 θ x (where x denotes an element of X).

CONCLUSION

The role of the OWA operator in the interpretation of expressions involving linguistic quantifiers, also called fuzzy quantified statements has been investigated. More specifically, statements of the type "Q X are A" have been considered where Q is a monotonic linguistic quantifier and A a fuzzy predicate. The focus has been put on the approach suggested by R.R. Yager based on the use of an OWA operator to evaluate this type of statement.

We have provided an interpretation of this view in terms of α-cuts of a fuzzy set. Moreover, it has been shown that this calculus is equivalent to the one obtained with a Choquet fuzzy integral by choosing an appropriate representation for the quantifier (in terms of a fuzzy measure). A probabilistic interpretation in terms of an expectation of an imprecise truth value has also been given. The interest and use of such complex expressions have been illustrated in the context of database querying where they serve to build selection conditions either on individual tuples, or on sets of tuples.

Two kinds of issues have been left out: the comparison of the OWA-based interpretation of quantified statements with other approaches (e.g., a more ordinal view provided by Sugeno fuzzy integral) one the one hand and aspects related to the implementation of flexible queries on the other hand. These two important points will be taken into account in future research works.

REFERENCES

[1] P. Bosc, L. Liétard (1993), On the extension of the OWA operator to evaluate some quantifications, *Proc. of the First European Congress on Fuzzy and Intelligent Technologies (EUFIT'93),* Aachen (Germany), 332-338.

[2] P. Bosc, L Liétard (1994), Upper mean value and OWA-based computation of quantified statements, *Proc. of the Second European Congress on Fuzzy and Intelligent Technologies (EUFIT'94)*, Aachen (Germany), 371-375.

[3] P. Bosc, L. Liétard (1994), Monotonic quantified statements and fuzzy integrals, *Proc. of NAFIPS/IFIS/NASA'94 Joint Conference,* San Antonio (Texas), 8-12.

[4] P. Bosc, O. Pivert (1995), SQLf: A relational database language for fuzzy querying, *IEEE Transactions on Fuzzy Systems,* 3, 1-17.

[5] P. Bosc, L. Liétard and O. Pivert (1995), Quantified statements in a flexible relational query language, *Proc. of the 1995 ACM Symposium on Applied Computing,* Nashville (U.S.A.), 488-492.

[6] P. Bosc, L. Liétard and O. Pivert (1995), Quantified statements and database fuzzy querying, in *Fuzziness in Database Management Systems,* (P. Bosc, J. Kacprzyk eds), Physica-Verlag, 275-308.

[7] A.P. Dempster (1967), Upper and lower probabilities induced by a multi-valued mapping, *Ann. Math. Stat.,* 38, 325-339.

[8] D. Dubois, H. Prade (1980), Fuzzy set and systems: theory and applications, *Academic Press.*

[9] D. Dubois, H. Prade (1983), Ranking fuzzy numbers in the setting of possibility theory, *Information Sciences,* 30, 183-224.

[10] D. Dubois, H. Prade (1985), Fuzzy cardinality and the modeling of imprecise quantification, *Fuzzy Sets and Systems,* 16, 3, 199-230.

[11] D. Dubois, H. Prade (1987), The mean value of a fuzzy number, *Fuzzy Sets and Systems,* 24, 279-300.

[12] D. Dubois, M.C. Jaulent (1987), A general approach to parameter evaluation in fuzzy digital pictures, *Pattern Recognition Letters,* 6, 251-259.

[13] D. Dubois, H. Prade (1990), Measuring properties of fuzzy sets: a general technique and its use in fuzzy query evaluation, *Fuzzy Sets and Systems,* 38, 137-152.

[14] D. Dubois, L. Godo, R.L. De Mantaras and H. Prade (1993), Qualitative reasoning with imprecise probabilities, *Journal of Intelligent Information Systems,* 2, 319-363.

[15] M. Grabisch, T. Murofushi and M. Sugeno (1992), Fuzzy measure of fuzzy events defined by fuzzy integrals, *Fuzzy Sets and Systems,* 50, 293-313.

[16] J. Kacprzyk, R.R. Yager (1984), Linguistic quantifiers and belief qualification in fuzzy multicriteria and multistage decision making, *Control and Cybernetics,* 13, 3, 155-172.

[17] J. Kacprzyk and A. Ziolkowski (1986), Database queries with fuzzy linguistic quantifiers, *IEEE Trans.actions on Syst.ems, Man and Cybernetics,* 16, 474-478.

[18] J. Kacprzyk, M. Fedrizzi (1989), A human consistent degree of consensus based on fuzzy logic with linguistic quantifiers, *Mathematical Social Sciences,* 18, 275-290.

[19] J. Kacprzyk (1991), Fuzzy linguistic quantifiers in decision making and control, *Proc. of International Fuzzy Engineering Symposium (IFES'91),* Yokohama (Japan), 800-811.

[20] M. Mizumoto, S. Fukami and K. Tanaka (1979), Fuzzy conditional inferences and fuzzy inferences with fuzzy quantifiers, *Proc. of the 6th International joint Conference on Artificial Intelligence,* Tokyo (Japan), 589-591.

[21] T. Murofushi, M. Sugeno (1989), An interpretation of fuzzy measure and the Choquet integral as an integral with respect to fuzzy measure, *Fuzzy Sets and Systems*, 29, 201-227.

[22] H. Prade (1990), A two-layer fuzzy pattern matching procedure for the evaluation of conditions involving vague quantifiers, *Journal of Intelligent and Robotic Systems*, 3, 93-101.

[23] M. L. Puri, D. Ralescu (1982), A possibility measure is not a fuzzy measure, *Fuzzy sets and Systems*, 7, 311-313.

[24] D.G. Schwartz (1994), A connection between fuzzy quantifiers and the classical modalities, *Proc. of NAFIPS/IFIS/NASA'94 Joint Conference*, San Antonio (Texas), 310-314.

[25] G. Shafer (1976), A mathematical theory of evidence, *Princeton Univ. Press*, Princeton (New Jersey).

[26] M. Sugeno (1974), Theory of fuzzy integrals and its applications, *Ph. D Thesis, Tokyo Institute of Technology*, Tokyo (Japan).

[27] R.R. Yager (1983) Quantifiers in the formulation of multiple objective decision functions, *Information Sciences*, 31, 107-139.

[28] R.R. Yager (1983), Quantified propositions in a linguistic logic, *International Journal of Man-Machine studies*, 19, 195-227.

[29] R.R. Yager (1988), On ordered weighted averaging aggregation operators in multicriteria decisionmaking, *IEEETransactions on Systems, Man, and Cybernetics*, 18, 183-190.

[30] R.R. Yager (1991), Fuzzy quotient operators for fuzzy relational databases, *Proc. of the International Fuzzy Engineering Symposium (IFES'91)*, Yokohama (Japan), 289-296.

[31] R.R. Yager (1993), Families of OWA operators, *Fuzzy Sets and Systems*, 59, 125-148.

[32] L.A. Zadeh (1965), Fuzzy sets, *Information and Control*, 8, 338-353.

[33] L.A. Zadeh (1968), Probability measures of fuzzy events, *J. Math. Anal. Appl.*, 23, 421-427.

[34] L.A. Zadeh (1975), The concept of a linguistic variable and its application to approximate reasoning - Part I, *Information Sciences*, 8, 199-249.

[35] L.A Zadeh (1978), Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems*, 1, 3-28.

[36] L.A. Zadeh (1983), A computational approach to fuzzy quantifiers in natural languages, *Computer Mathematics with Applications*, 9, 149-183.

USING OWA OPERATOR IN FLEXIBLE QUERY PROCESSING

Maria-Amparo Vila, Juan-Carlos Cubero, Juan-Miguel Medina and Olga Pons
Department of Computer Science and A.I. University of Granada
E.T.S. de Informatica, Avd. de Andalucia n.38 18071, Granada, Spain
email: vila@decsai.ugr.es

Abstract

The use of OWA operators has been widely used in database contexts to process flexible queries where linguistic quantifiers are involved. This paper is devoted to comparing these operators' use with other ways of addressing the same problems such as the Zadeh's approach. Furthermore a new proposal is presented in order to avoid some inconveniences which the former approaches have in a wider class of problems. Finally, the paper analyzes the different types of queries with linguistic quantifiers which may appear in connection with relational databases and proposes the most appropiate way to solving them,in each case.

KEYWORDS: Fuzzy queries, fuzzy relational databases, linguistic quantifiers, OWA operators.

1 Introduction

The importance of the OWA operators in the flexible querying context lies in their strong connection with the concept of linguistic quantifier. Though these operators were initially introduced to aggregate the importance levels of the alternatives in a decision making problem (see Yager 1988 [8]), in this paper the author had already presented the OWA relation with linguisitc quantifiers, by explaining how those weights which appear in the OWA expression could be obtained by using the membership function of any quantifier.

On the basis of these ideas, OWA operators were presented in Yager 1992 [9] as a way to compute the accomplishement degree of those sentences where linguistic valuations of proportions are included together with imprecise properties, such as:

Most of the students are young

A few of the young students are efficient

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

Obviously this approach appeared as an alternative to the Zadeh's method as presented in Zadeh 1983 [11].

Since then, different authors have used either OWA operators or Zadeh's methods to solve queries in databases which involve both linguisitc quantifiers and imprecise properties. However no comparative studies have been developed and until now there are no reasons to use one or the other method.

This comparative analysis is the objective of the present paper. Therefore here we shall here concerned with studying the advantages and drawbacks of both methods in order to solve queries. Additionally, an alternative computing method will be presented in order to avoid some of the inconveniences of the former approaches in dealing with some kinds of sentences.

The paper is organized as follows:

The first part is devoted to the comparative analysis itself, it starts off with a formal presentation of the concept of linguistic quantifier which includes the definition of coherent family of quantifiers. This concept will be the key to further developments. The problem of computing the accomplishment degree of sentences which include linguistic quantifiers and properties is studied by considering two general classes of "such as" sentences established in [11]. We compare the two aforementioned approaches for both classes and, in the case of type II sentences, a new approach is presented.

The second part of the paper deals with the use of linguistic quantifiers in flexible querying in relational databases. More specifically, we study the three major classes of "such as" queries which have appeared in the current literature:

- Queries which involve a generalization of the aggregation function COUNT
- Queries which are a generalization of the classical relational quotient
- Queries with composite conditions and importance levels.

The conclusions obtained in the first part of the paper are then applied to each case. The paper finishes off with some remarks and the references.

2 General issues about linguistic quantifiers

2.1 Initial definitions and notations

The concept of linguistic quantifier was introduced by L.A. Zadeh in 1983 ([11]) in order to model those linguistic terms which represent imprecise quantities, in a computational way. There are two basic types of linguistic quantifiers:

2.1.1 Absolute Quantifiers

They are smoothed versions of integer quantities such as "some of", "around 7", "not much more than 5". Each absolute quantifier can be associated with a fuzzy set defined on an adequate interval of non-negative integers. This fuzzy set is called the semantic representation of such a quantifier and it allows us to deal with the linguistic term in the fuzzy logic and fuzzy set context.

The semantic representation can have different shapes depeding on the meaning of the corresponding quantifier. It will be a convex fuzzy set, in the case of the quantifier expressing an approximate number (e.g. "around 7"), whereas S or Z type membership functions shall be more adequate to represent comparison quantifiers (e.g. "much more than 5").

2.1.2 Relative Quantifiers

They involve those terms with a meaning of proportion such as "most", "exists", "almost all" etc. They are the most commonly used in flexible querying processes, since the cardinal of the whole set involved in a database query is usually unknown before the query is solved. Only in the case of flexible satisfaction of conditions, where their total number is fixed, could the use of absolute quantifiers be relevant.

The semantic representations of relative quantifiers are fuzzy sets defined on the $[0, 1]$ interval. Also in this case, the shape of the membership functions depends on the quantifier meaning but now there are two ways to interpret such a meaning:

1. The exclusive way which implies representing only what the quantifier means, in such a case a fuzzy number defined on the $[0, 1]$ interval is used as semantic representation .

2. The broader way which assumes that any quantifier includes itself and those which could be considered "greater than" in some sense. The idea is to express by means of the quantifier the "minimal" proportion of objects which verifies a concrete property. For example when we say:

   "*Most* of X are A"

   we are implicitly including assertions such as "*almost all* of X are A" or "*All* of X are A" in the sense that if the latter are true the former will be true as well.

The broader interpretation of relative quantifiers is the one commonly used in flexible querying, and therefore the following hypotheses are assumed:

1. The membership function $Q(.)$ of any relative quantifier Q will be non-decreasing.

2. There must be some kind of order between the quantifiers, which, in some way represents the amount of proportion . For example, "a few" should go before "most" and "exists" and "for all" will be respectively the first and the last for all of them.

Different authors such as Yager (1992) [9] and Vila et al (1995) [7] propesed the same hypotheses, and by formalizing these ideas in [7] the following definition was proposed

Definition 2.1 *Let $\mathcal{Q} = \{Q_1, \ldots, Q_l\}$ be a linguistic quantifier set, we shall say it is coherent if it verifies that:*

i *The membership functions of $\mathcal{Q}$ elements are non-decreasing functions.*

ii *A partial order relation $\succeq$ is defined in $\mathcal{Q}$. It has as its maximal element $Q_1 = \exists$ and as its minimal one $Q_l = \forall$. Furthermore $\forall Q_i, Q_j \in \mathcal{Q}$ $Q_i \subseteq Q_j \Rightarrow Q_j \succeq Q_i$*

iii *The membership function of the quantifier $\exists$ is given by $Q_1(x) = 1$ if $x \neq 0$ and $Q_1(0) = 0$, whereas the membership functions of $\forall$ will be $Q_l(x) = 0$ if $x \neq 1$ and $Q_l(1) = 1$*

Below we shall consider relative quantifiers belonging to some coherent family

2.2 Computing the accomplishment degree of sentences including linguistic quantifiers

According to [11, 9], there are two types of vague sentences including linguistics quantifiers, whose general structures are given by:

Sentences of type I: Q of X are A

Sentences of type II: Q of D are A

where Q is a linguistic quantifier, X is a general finite referential set, and A and D are vague properties defined on X's elements and represented by fuzzy sets with membership functions $A(.)$ and $D(.)$ respectively. Some examples of such sentences are:

type I : Most of the students are competent

type II: Most of the competent students are young

Now, the question is how to compute the accomplishment degree of "such as" sentences by using the memberships functions $Q(.)$, $A(.)$ and $D(.)$ if need be. Different approaches to solving the problem have been proposed (see [11, 9, 7]), and all of them have been used in flexible querying problems. One of the most used approaches is the one based on the OWA operators. Therefore , in this paragraph, we will be more concerned with a comparative analysis than with simple descriptions.

2.2.1 Accomplishment degree of type I sentences

Zadeh's approach This author's method is based on the cardinal of the fuzzy set A, considered in and absolute or relative sense depending on whether the quantifier is in turn absolute or relative. If we consider:

$$\alpha_A = \begin{cases} ||\sum_{x \in X} A(x)|| & \text{if Q is absolute} \\ \sum_{x \in X} A(x)/|X| & \text{if Q is relative} \end{cases}$$

where $||a||$ stands for " the largest integer less than a" and $|X|$ is the X's cardinal number, then the accomplishment degree of "Q of X are A" in, Zadeh's sense, is given by:

$$Z_Q(A) = Q(\alpha_A)$$

Yager's approach. OWA operators Yager's proposal covers relative quantifiers and it is based on the relation between OWA operators and linguistics terms. Let us assume $|X| = n$ and let $a_i = A(x_i); i = 1 \ldots n$ and let $\{b_i\}$ be a permutation of the sequence $\{a_i\}$, where $\forall i \in \{1 \ldots n-1\}$ $b_i \leq b_{i+1}$. Under these conditions, if we define:

$$w_i = Q(i/n) - Q((i-1)/n) \text{ by assuming } Q(0) = 0$$

the accomplishment degree is given by:

$$Y_q(A) = \sum_{i=1}^{n} w_i b_i$$

A brief comparative analysis The difference between both approaches may be derived from the definitions. Whereas the Z_Q computation is made by summarizing $A(.)$ in α_A and next by computing how Q and α_A match, in the Y_Q, case the summarization is made by using Q, and the accomplishment degree is this summarization. However we can find some relationships between both approaches by analyzing their results in several particular cases.

Extreme quantifiers

According to the definition 2.1 the extreme quantifiers of any coherent family are $\exists$ and $\forall$ and by using the expressions given in this definition it is easy to check:

$$\forall A \; Z_\forall(A) = \begin{cases} 0 & \text{if } \alpha_A < 1 \\ 1 & \text{if } \alpha_A = 1 \quad \text{equivalently } A = X \end{cases}$$

$$\forall A \; Z_\exists(A) = \begin{cases} 1 & \text{if } \alpha_A > 0 \\ 0 & \text{if } \alpha_A = 0 \quad \text{equivalently } A = \emptyset \end{cases}$$

From the OWA operator properties we obtain:

$$\forall A \; Y_\forall(A) = min_x(A(x)) \; Y_\exists(A) = \max_x(A(x))$$

Obviously,

$$\forall A\; Z_\forall(A) \leq Y_\forall(A) \text{ and } Z_\exists(A) \geq Y_\exists(A)$$

and in these cases, Yager's approach seems, in our opinion, semantically more correct than the other which appears to be too strict.

The case of crisp sets

Let us consider that A is a crisp set such that $|A| = m$, in this case $Z_Q(A) = Q(m/n)$ and it is easy to check that: $Y_Q(A) = 1 - Q(1 - m/n)$ that is, in this case, Z_Q is the measure of m/n with respect to the possibility measure defined by Q, and Y_Q is the corresponding necessity measure.

The case of linear quantifiers

Definition 2.2 *Let us consider the set of linguistic quantifiers, $\mathcal{Q}$, represented by means of linear functions in the following way:*

$$\forall Q \in \mathcal{Q}\; Q(x) = \begin{cases} 0 & \textit{if } 0 \leq x < a_Q \\ (x - a_Q)/(b_Q - a_Q) & \textit{if } a_Q \leq x < b_Q \\ 1 & \textit{if } b_Q \leq x \leq 1 \end{cases}$$

where a_Q, b_Q are [0,1] values such that $a_Q \leq b_Q$. In such a situation we shall say that $\mathcal{Q}$ is a family of linear quantifiers.

With this definition, the $\mathcal{Q}$ elements are characterized by a pair of real values. Concretely, $\exists$ corresponds to (0,0) and $\forall$ to (1,1). It can be easily seen that:

$$Q \succeq Q' \Leftrightarrow a_Q \leq a_{Q'} \text{ and } b_Q \leq b_{Q'}$$

which gives us an adequate partial order relation on $\mathcal{Q}$, and characterizes it as a coherent family.

Now, if we assume Q is a linear quantifier, it is easy to obtain:

$$\forall A Z_Q(A) = \begin{cases} 0 & \text{if } 0 \leq \sum_1^n (b_i/n) < a_Q \\ (\sum_1^n (b_i/n) - a_Q)/(b_Q - a_Q) & \text{if } a_Q \leq \sum_1^n (b_i/n) < b_Q \\ 1 & \text{otherwise} \end{cases}$$

and if we take values $l, m \in \{1 \ldots n\}$ such that, $(l-1)/n < a_Q \leq l/n$ and $m/n \leq b_Q < (m+1)/n$ then we have:

$$w_i = Q(i/n) - Q((i-1)/n) = \begin{cases} 0 & \text{if } i < l \\ (l/n - a_Q)/(b_Q - a_Q) & \text{if } i = l \\ (1/n)/(b_Q - a_Q) & \text{if } l < i \leq m \\ (b_Q - (m/n))/(b_Q - A_Q) & \text{if } i = m+1 \\ 0 & \text{otherwise} \end{cases}$$

and the $Y_Q(A)$ expression is given by:

$$Y_Q(A) = \sum_{i=l+1}^{m} (b_i/n)/(b_Q - a_Q) + b_l(l/n - a_Q)/(b_Q - a_Q) + b_{m+1}(b_Q - m/n)/(b_Q - a_Q)$$

from the l and m definitions, we have $(l/n - a_Q) < 1/n$ $(b_Q - m/n) < 1/n$ therefore, if n is large enough we can assure:

$$Y_Q(A) \approx \sum_{i=l}^{m+1} (b_i/n)/(b_Q - a_Q)$$

By comparing the Z_Q and Y_Q expressions, one can see that the former is less accurate than the latter, as the conditions under which each one achieves the external values 1 or 0 show. Furthermore, the $\{b_i\}$ and Q values which are actually significant in the Y_Q's computation correspond to central positions.

2.2.2 Accomplishment degree of type II sentences

Zadeh's approach Now we use the cardinal of A relative to D, which is given by:

$$\alpha_{A/D} = \frac{\sum_{x \in X}(A(x) \bigcap D(x))}{\sum_{x \in X} D(x)}$$

and, as in the case above we obtain the accomplishment degree by computing how Q and $\alpha_{A/D}$ match. That is:

$$Z_Q(A/D) = Q(\alpha_{A/D})$$

Yager's approach This author's proposal is also based on the OWA operators, but by introducing D in the way of computing both the elements to be aggregated as well as the weights.

Whereas in the above case we aggregated a permutation of the $A(x_i)$ values, now we consider the sequence $\{c_i = A(x_i) \vee (1 - D_(x_i))\}$ as basic values to be aggregated and the $\{b_i\}$ values which will appear in the OWA expression are an ordered permutation of such a sequence.

Furthermore the weights are now computed by considering points in the $[0, 1]$ interval which depend on the D values, as follows:

If we call $\{d_i = D(x_i)\}$, $\{e_i\}$ an ordered permutation of such a sequence, and $d = \sum_1^n d_i$, the following sequence can be defined: $s_0 = 0, \ldots s_i = e_i/d + s_{i-1}$. And by using this sequence we compute the weights $w_i = Q(s_i) - Q(si - 1)$.

Finally the accomplishment degree is given by:

$$Y_Q(A/D) = \sum_{i=1}^{n} w_i b_i$$

Some reflections on both approaches As in the case above we shall analyze both proposals by considering the extreme cases of $\exists$ and $\forall$. Regarding Zadeh's method, we have:

$$\forall A \; Z_\forall(A/D) = \begin{cases} 0 & \text{if } \alpha_{A/D} < 1 \\ 1 & \text{if } \alpha_{A/D} = 1 \quad \text{equivalently } A = D \end{cases}$$

$$\forall A\ Z_{\exists}(A/D) = \begin{cases} 1 & \text{if } \alpha_{A/D} > 0 \\ 0 & \text{if } \alpha_{A/D} = 0 \quad \text{equivalently } A \bigcap D = \emptyset \end{cases}$$

From these expressions we can deduce that Z_Q is again a too strict approach. It only takes the values 0 or 1, and, in the case of the $\forall$ quantifier, only value 1 is achieved when $A \subseteq D$. On the contrary, in the case of $\exists$, it is necessary that $A \bigcap D = \emptyset$ to get a value different from 1.

With respect to Yager's approach it is easy to check that:

$$Y_{\forall}(A/D) = \min_x(A(x) \vee (1 - D(x)) \quad Y_{\exists}(A/D) = max_x(A(x) \vee (1 - D(x))$$

As we shall see later the first result is reasonable and consistent with fuzzy logic but the second is not, moreover the following example shows how it is even contradictory.

Example 1 *Let us consider* $X = \{1,2,3,4,5,6,7,8\}$ $A = \{1/1, 2/1, 3/1\}$ *and* $D = \{6/1, 7/1, 8/1\}$ *by definition* $Y_{\exists} = 1$ *but* A *and* D *are crisp sets with empty intersection.*

From these considerations we can deduce that the approaches presented are not suitable for dealing with all of the possible quantifiers which might appear in a query, and for this reason a new approach was presented in [7], it was developed in order to solve fuzzy queries which extend the relational quotient, but it can easily be generalized to every type II sentence.

An alternative approach (Vila et. al 1995) The basis of our proposal is the concept of coherent set of quantifiers, whose extremes are the $\exists$ and the $\forall$ quantifiers.

The idea is that the accomplishment degree of the sentence: "Q of D are A" where Q is any quantifier between $\forall$ and $\exists$ shall lie between the accomplishment degree of the sentences "All of D are A" and "There exists one D which is A" and these must be computing by using fuzzy logic-based criteria.

More specifically, the logical representation of the sentence " There exists one D which is A" is given by

$$\exists x \in X\ (D(x) \text{ and } A(x))$$

and if we take $D(x) \wedge A(x)$ as truth value of $D(x)$ *and* $A(x)$ then the accomplishment degree of the complete sentence is given by:

$$\max_x(D(x) \wedge A(x))$$

"All of D are A" has as its logical representation:

$$\forall x \in X\ (D(x) \longrightarrow A(x))$$

and if we compute the truth of $(D(x) \longrightarrow A(x))$ by means of the material implication, this value will be $A(x) \vee (1 - D(x)$ and the accomplishment degree of the sentence is given by:

$$\min_x (A(x) \vee (1 - D(x))$$

It should be pointed out that this last result coincides with Yager's approach, i.e., the degree associated with the quantifier $\forall$ is the necessity measure of A with respect to D, however in the case of the $\exists$, we use the corresponding possibility measure and this is coherent with a fuzzy logic based approach.

Now, if we consider the quantifier Q is situated between $\exists$ and $\forall$, a good way to give its associated accomplishment degree is to consider it as a mixture of the possibility and necessity which correspond to both limit quantifiers. This idea leads us to the following definition:

Definition 2.3 *Let D and A be two fuzzy set with the same domain X. Also let $\mathcal{Q}$ be a coherent set of quantifiers, and let us assume a set of [0,1] values associated to $\mathcal{Q}$ elements is established, such that:*

i $\gamma_\exists = 1 \; \gamma_\forall = 0$ **ii** $\forall Q, Q' \in \mathcal{Q} \; Q \succeq Q' \Rightarrow \gamma_Q \geq \gamma_{Q'}$

Under these conditions, $\forall Q \in \mathcal{Q}$, we define the accomplishment degree of the sentence "Q of D are A" as follows:

$$V_Q(A/D) = \gamma_Q \max_x (D(x) \wedge A(x)) + (1 - \gamma_Q) \min_x (A(x) \vee (1 - D(x))$$

The following properties of V_Q have a direct proof.

Property 2.1

V_Q is actually a fuzzy measure, that is: $\forall A_1, A_2, D$ fuzzy sets of X $A_1 \subseteq A_2 \Rightarrow V_Q(A_1/D) \leq V_Q(A_2/D)$

Property 2.2

V_Q maintains the order relation existing in $\mathcal{Q}$. That is: $\forall Q, Q' \in \mathcal{Q} \;\; Q \succeq Q' \Rightarrow V_Q(A/D) \geq V_{Q'}(A/D)$

Now the problem is to find and adequate way of to obtain the values γ_Q associated with any coherent set. Several alternatives can be chosen, our proposal is to use an approximation of the "orness" measure given by Yager in [8], according to this author, the "orness" measure of any OWA operator with weights $W = \{w_1 \dots w_n\}$ is given by:

$$o_W = \sum_{j=1}^{n} ((n-j)/(n-1)) w_j$$

if the weights are obtained by means of any quantifier, Q, the aforesaid expression becomes:

$$o_W = \sum_{j=1}^{n} ((n-j)/(n-1))(Q(j/n) - Q(j-1/n)) = \sum_{j=1}^{n} ((1/(n-1))Q(j/n)$$

By considering n to be large enough we can obtain an approximation of the "orness" measure which only depends on the quantifier itself and is given by:

$$o_Q = \int_0^1 Q(x)dx$$

it is easy to check that o_Q verifies the conditions involved in the definition 2.3

Finally, the accomplishment degree expression of the sentence: "Q of D are A" will be:

$$V_Q(A/D) = o_Q \max_x(D(x) \wedge A(x)) + (1 - o_Q) \min_x(A(x) \vee (1 - D(x))$$

The advantages of a such proposal are the following:

- It may be used through all of the ranking of a family of quantifiers from the $\exists$ to the $\forall$.
- It is similar to Yager's approach in the nearness of the $\forall$ quantifier and avoids the drawbacks of such a method for the $\exists$.
- It is based on fuzzy logic considerations instead of the fuzzy set properties and therefore its results are less strict than those in Zadeh's approach.
- Finally it should be pointed out that it can be obtained as easily as Zadeh's degree. One may think that the o_Q computation might be complicated, but since it only depends on the quantifier this computation shall only be carried out once. Moreover, if we take a parametric family of quantifiers, o_Q could have a simple expression. For example, in the case of a linear family of quantifiers such as the one in definition 2.2, which is characterized by parameters (a_Q, b_Q), the o_Q value is given by: $o_Q = (a_Q + b_Q)/2$.

  This computational simplicity is an important advantage regarding the flexible querying process when dealing with very large sets.

3 Linguistic quantifiers in flexible querying

3.1 Initial definitions and notation

From a general point of view, the use of linguistic quantifiers in information retrieval is important two broad areas:

1. Relational databases, which may include fuzzy data or not.
2. Document information systems

In this paper we deal with the first area, i.e. that of relational databases which have been the basis for a wide diversity of "such as" queries. These queries have been formulated both in crisp databases (see Kacprzyk et al. 1986,1995 [3, 4] and Bosc et al. [1, 2]) and fuzzy relational databases, i.e. relational databases which allow fuzzy values for their attributes, (see Medina 1994 [5], Vila et. al 1995 [7] and Yager 1992,1995 [9, 10]).

In this paper we shall work in a general context by considering fuzzy relational databases with a possibility-based approach (Prade and Testemale 1984 [6], Zemankova-Leech and Kandel 1984 [12]). Therefore we will assume that a Fuzzy Relational Database (FRDB) is a set of relations $R, S, \ldots$, where each relation, R, is defined in the following way:

Definition 3.1 *Let us consider a sequence of attributes $A_1, \ldots, A_n$ the basic domain of A_i will be denoted by D_i, $\forall i \in \{1, \ldots, n\}$. We shall assume that some of A_is attributes are allowed to take imprecise values which will be modeled by means of possibility distribution on the basic domains. The set $\mathcal{I}$ will be the index set of such attributes and the set of possibility distributions defined on $D_i; i \in \mathcal{I}$ will be denoted by $P(D_i)$. Under these conditions we shall define the relation R as a subset of the Cartesian product:*

$$\prod_{i \in \mathcal{I}} \mathcal{P}(D_i) \times \prod_{i \in \{1 \ldots n\} - \mathcal{I}} D_i$$

The set of attributes $\{A_1, \ldots, A_n\}$ will be called schema of R, as in the classical case. Below, the attribute sets, schemes or subeschemes of some relation, will be denoted by $\mathcal{A}$, $\mathcal{B}$, ...

With this definition crisp and fuzzy values may appear in the same relation, which shall denote the former by $x, y \ldots$ whereas the later by $X, Y, \ldots$. We shall also denote by $\bar{t}[A]$ the value of the attribute A in the tuple $\bar{t}$. It should be pointed out that crisp relations and databases appear as a particular case of this definition.

On the basis of this model different kinds of flexible queries involving linguistic quantifiers have been discussed by different authors, these will be analyzed in the following paragraphs.

3.2 Queries concerning linguistic summaries

(Medina 1994 [5], Yager 1995 [10])

This kind of queries concerns with the accomplishment degree of sentences which involve linguistic quantifiers in connection with aggregation functions. More specifically, these sentences will use a smoothed version of the aggregation function COUNT. Examples of "such as" sentences could be the following:

Example 2 *Let us consider a students database where marks corresponding to scientific subjects (Mathematics, Physics, Chemistry etc.) are included. Some queries including linguistic summaries could be to ask for the accomplishment degree of the following sentences:*

1. *More than 40 students have passed Physics*

2. Not much more than 15 students have passed Physics with high marks.

3. Most of the students have failed in Mathematics.

4. Most of the young students have passed Mathematics with high marks.

From this example, we can deduce that several kinds of such queries can appear and different general schemes must be considered, as follows:

- Queries that are similar to 1 or 2 in the example above correspond to type I sentences "Q of X are A" where Q is an absolute quantifier and A is a set of items which is the result of a simple selection. In the case 1 this set is crisp since the property involved in the selection is also crisp, in case 2, A is a fuzzy set since its corresponding property is imprecise.
- Query 3 in the example shows a type I sentence where the quantifier is relative and the set A is crisp
- Finally the fourth query is an example of a type II sentence with a relative quantifier.

All of theses sentences have mainly been studied in [5, 10], and, in both cases, they have been solved with the same approach which is Zadeh's one. However, according to the conclusions presented in 2.2.1, 2.2.2, the OWA approach should be the most appropiate in cases such as the third (i.e. type I sentences with relative quantifiers) and for sentences like the fourth (i.e. type II sentences), the operator $V_Q(.)$ ought to be used.

3.3 Generalizations of relational quotient

The classical relational quotient were defined to model queries which agree with following general structure:

> "To find those objects which connect through a relation to *all* of the elements in some set described by means of other relation"

and this structure may be made more flexible in two ways:

1. consider the involved set is the representation of some imprecise property and consequently a fuzzy set.
2. make the quantifier *all* soft by imposing the connection with *almost all, most* etc. of elements in some set.

the use of either way or both generates an important class of imprecise queries which have to be solved in the fuzzy database context.

Many authors have dealt with the fuzzy quotient problem from different approaches ([9, 1, 7]), the most general ones among these papers are those of Yager and Vila et al., which include linguistic quantifiers in the problem formulation.

This general problem of the fuzzy relational quotient can be formulated in the following way:

Definition 3.2 *Let R be a relation of an FRDB, with schema $\mathcal{A}$, let $A \in \mathcal{A}$ be a fuzzy attribute of R, and let P be a possibility distribution on the A domain which represents some imprecise property by means of a fuzzy set of objects. Let us assume that $\mathcal{A}$ includes the primary key, $\mathcal{B}$, of another relation of the FRDB which describes another object set, and $A \notin \mathcal{B}$. Also let Q be a quantifier belonging to a coherent family. Under these conditions fuzzy quotient problem can be described as:*

> *Obtaining the set of objects, described by $\mathcal{B}$, which connect through the relation R with Q of the objects belonging to the fuzzy set P*

The basic ideas appearing in [9] [7] are very similar and these lead to the following procedure:

1. A first "compression" operation of the relation R with respect to the attributes A and $\mathcal{B}$ is needed. With this operation the relation R is transformed into a new relation, T, with the following characteristisc:

   - It has $\mathcal{B}$ as a primary key
   - $\forall \bar{t} \in T$ the value of $\bar{t}[A]$ is a fuzzy set, which represents the set of objects in the domain of A which connect to $\bar{t}[\mathcal{B}]$ through the relation R

2. The fuzzy relational quotient will be a relation C with schema $\mathcal{B} \bigcup [0,1]$ where for each tuple $\bar{t} \in T$, we include the value of $\bar{t}[B]$ together with the accomplishment degree of the sentence:

   Q of P are $\bar{t}[A]$

The following example make all this process clearer:

Example 3 *Let us consider the students database and the following query:*

> *Find those students which have got a high mark in most of the hard subjects"*

To solve this we first use a relation R with schema [name_of_student,subject_h], where one tuple appears for each related pair of student and subject. The first attribute meaning is clear, and the attribute "subject_h" is a fuzzy attribute which includes the name of the subject and the matching level of the mark which the corresponding student has obtained in this subject with the property "high". Obviously, R can be obtained from the database by means of suitable selection.

By using R we can obtain a new relation, T, with the same schema where "name_of_student" is now the primary key and "subject_h" takes fuzzy sets of subjects as its values. Let us now consider a new fuzzy set of subjects, P, which represents the imprecise property "to be hard", the solution of the query presented in the example 3 is obtained by computing for each tuple $\bar{t} \in T$ the accomplishment degree of the sentence:

Most of P are in $\bar{t}[(subject\_h)]$

The way this degree is obtained depends on the authors, since in [9] $Y_Q(.)$ is used whereas in [7] the $V_Q(.)$ approach is presented. We think this last way is the most suitable in accordance with conclusions in the paragraph 2.2.2.

3.4 The use of linguistic quantifiers in composite conditions

Another way to include linguistic quantifiers in flexible querying is to make the accomplishment of a composite query less strict. This approach has been dealt with by Kacprzyk et al. (see [3, 4]) and it is given as follows:

Let us consider a composite query, organized in the following general way:

```
select <list_of_field> where
cond11 and cond12 and .... and cond1k1
or
..........
or
condn1 and condn2 and .... and condnkn
```

where each condition condij is an atomic condition which may be crisp or fuzzy and which has the general expression:

```
attribute <fuzzy_relation> fuzzy_value
```

With this formulation,the query is not different from another imprecise query, the originality of this approach arises when we generalize the query as follows:

```
select <list_of_field> where
{most, almost all, etc..}
cond11 and cond12 and .... and cond1k1
or
..........
or
{most, almost all, etc..}
condn1 and condn2 and .... and condnkn
```

i.e., the groups of "and" conditions are quantified in the sense of "not all of the conditions" must be verified.

To solve this type of query, let us denote by μ_i the accomplishment degree of the ith group of "and" conditions:

```
Q condi1 and ...and coniki
```

and by ν_{ij} $j = \{1 \ldots ki\}$ the acomplishement degree of each atomic condition condij $j = \{1 \ldots ki\}$. The ν_{ij} sequence can be viewed as a fuzzy set, A_i, defined on the referential $\{condi1, \ldots condiki\}$, and the the μ_i value is that of the sentence:

Q of $\{condi1, \ldots condiki\}$ are A_i

The authors propose two alternative ways to compute this former value, $Z_Q(A_i)$ and $Y_Q(A_i)$ and they use one t-conorm to combine all of μ_i $i = \{1 \ldots n\}$, in order to obtain the accomplishment of the whole query.

The problem is generalized by introducing importance levels into the atomic conditions. In this case, the ith group of "and" conditions in a composite query might be:

```
Q  of important condi1 and ...and coniki
```

where "important" is represented as a fuzzy set, B_i, defined on the set of conditions. If A_i has the same meaning as that of the case above the μ_i value is now the accomplshement degree of the sentence:

Q of B_i are A_i

and the way of computing this degree could be one of those proposed in 2.2.2 for these kinds of assessments. To be more specific, the authors consider $Z_Q(A_i/B_i)$ and $Y_Q(A_i/B_i)$. However, according to results in paragraph 2.2.2, $V_Q(A_i/B_i)$ should be the most situable one.

As in the case above, all of $\{\mu_i\}$ are aggregated by means of any t-conorm to acheive the final accomplishment degree.

4 Concluding Remarks

We have analyzed the different ways for computing the accomplishement degree of sentences including linguistic quantifiers, obtaining the following conclusions:

- In the case of type I sentences, the OWA operator-based method is the most suitable since it leads to less strict results than those in Zadeh's approach.

- In the case of type II sentences a new approach has been presented since Zadeh's approach is again too strict and OWA method could be inconsistent for linguistic quantifiers which are similar to the $\exists$ one. This approach also presents many computational advantages.

Next, we have applied our conclusions to the different types of queries involving linguistic quantifiers, which can be formulated for a relational database, with the following results:

- In the case of linguistic summary queries, the OWA method or our new approach are recommended, depending on whether the query leads to a type I or of type II sentence.

- Queries which are relational quotient generalizations always lead to type II sentences, therefore our new approach is here the most adequate one.

- In the case of composite condition quantification without importance, here the OWA method is the most adequate, when an importance level is introduced a type II sentence appears and the new approach should be used.

As future works, we have plan the following ones:

1. Implementing all of the aforementioned types of queries on the basis of coherent families of quantifers, described by means of parameters like the linear family.

2. Applying the the results obtained in the first part of the paper to information retrieval problems.

References

[1] Bosc P. Dubois D., Pivert O., Prade H. (1995) *Fuzzy division for regular relational databases)* Proc. of 4th. International IEEE Conference on Fuzzy Systems (FUZZ-IEEE/FIFES'95) Yokohama Japan pp. 729-734

[2] Bosc.P. Lietard L. Pivert O. (1995) *Quantified statements and databases fuzzy querying* In P. Bosc and J. Kacprzyk eds. Fuzziness in Database Management Systems pp.275-309 Physica-Verlag 1995.

[3] Kacprzyk J. Ziolkowski S. (1986) *Database queries with fuzzy linguistic quantifiers* IEEE Transc. on Systems, Man and Cybernetics SMC - 16, 474-479

[4] Kacprzyk J. Zadrozny S. (1995) *Fuzzy Queries in Microsoft Access V.2* Proceedings of the Workshop on Fuzzy Database Systems and Information Retrieval.pp. 61-66 FUZZ-IEEE/IFES'95 March 1995 Yokohama.

[5] Medina J.M. (1994) *Bases de datos relacionales difusas. Modelos teóricos y aspectos de su implementación.* Ph. Doctoral Thesis Univeristy of Granada May 1994.

[6] Prade H. and Testemale C. (1984) *Generalizing database relational algebra for the treatment of incomplete and uncertain information and vague queries* Information Science Vol. 34 pp. 115-143

[7] Vila M.A., Cubero J.C., Medina J.M., Pons O. (1995) *The generalized selection : an alternative way for the quotient operations in fuzzy relational databases* In B. Bouchon-Meunier, R. Yager, L. A. Zadeh eds. "Fuzzy Logic and Soft Computing". World Scientific Press.

[8] Yager R.R. (1988) *On ordered weighted averaging aggregation operators in multicriteria decisionmaking* IEEE Transc. on System, Man and Cybernetics Vol.18 (1) pp. 183-190

[9] Yager R.R. (1992) *Fuzzy Quotient Operators* Proc. of IPMU'92 pp. 317-322

[10] Yager R.R. (1995) *Linguistic Summaries as as tool for database discovery* Proceedings of the Workshop on Fuzzy Database Systems and Information Retrieval. pp. 79-85 FUZZ-IEEE/IFES'95 March 1995 Yokohama.

[11] Zadeh L.A. (1983) *A Computational approach to fuzzy quantifiers in natural languages* Computing and Mathematics with Applications Vol. 9 pp. 149-184

[12] Zemankova-Leech M., Kandel A. (1984) Fuzzy Relational Databases: A key for Expert Systems Verlag TUV Rheiland

APPLICATION OF OWA OPERATORS TO SOFTEN INFORMATION RETRIEVAL SYSTEMS

Gloria Bordogna and Gabriella Pasi
Istituto per le Tecnologie Informatiche Multimediali
Consiglio Nazionale delle Ricerche
Via Ampère 56 - 20131 Milano, Italy
e-mail: fuzzir@itim.mi.cnr.it

Abstract
In this contribution an overview of the use of OWA operators to model a soft retrieval activity in Information Retrieval Systems (IRSs) is presented. In particular some extensions of the Boolean document representation and of the Boolean query language are introduced, and a possible evolution of these approaches is outlined.

1. *INTRODUCTION*

The increasing development and use of wide area networks makes urgent the need for flexible systems to manage narrative information in documents such as papers, reports, books. The main aspect related to flexibility in information systems is the ability to account for the imprecision and vagueness typical of human communication. Commercial Information Retrieval Systems (IRSs) ignore these aspects: they oversimplify both the representation of the documents' content and the user-system interaction.

In recent years a branch of research in Information Retrieval has faced the problem of modeling the vagueness and imprecision which characterize the management of information [1,2,9,11]. A great deal of research in this area has faced the problem of extending the Boolean information retrieval model, which is the most widely adopted by commercial systems, with a twofold aim: to incorporate in IRSs more specific and accurate representations of documents' content and to make the query language more expressive and natural than the Boolean language [18].

Fuzzy set theory [26] has been successfully employed to define a superstructure of the Boolean model which makes it possible to extend existing Boolean IRSs without redesigning them completely [4,5,9,10,13,15]. Through these extensions the vague nature of relevance of documents to user queries is modelled.

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

The basic idea is that an IRS performs a Multi Criteria Decision Making activity at two distinct levels:

- in the documents' indexing phase, the significance of an index term in identifying the content of a document is computed based on the global satisfaction of one or more predefined criteria; in the case of multiple criteria an aggregation condition has to be applied. In the simplest fuzzy IRSs the significance degree is computed on the basis of a single criterium which evaluates a normalized sum of the index term's occurrences in the document [18];
- in query evaluation the set of archived documents is seen as the set of possible solutions, and the query specifies both the criteria to be satisfied by documents and their aggregation condition. In the first fuzzy IR models, the criteria are weighted terms and their evaluation corresponds to verify their degree of "importance" in each document, as specified by the weight. The aggregation conditions are generally specified by a Boolean operator (AND/OR). As a result, the retrieved documents are ranked in decreasing order of their global satisfaction of set of criteria.

In this paper some fuzzy IR models are described, which introduce soft aggregation conditions both in the indexing phase and in the query language. These conditions are expressed through linguistic quantifiers formalized by means of OWA operators. In this way, the IR activity becomes more flexible and adaptable to different applications (information subject areas) and user's needs.

In section 2. a fuzzy representation of structured documents is described and the evaluation of a system based on it is reported; in section 3 an evolution of this fuzzy representation is proposed; in section 4. a generalized Boolean query language is described in which weighted terms are aggregated by linguistic quantifiers formalized as OWA operators; in section 5. a new semantics for query term weights is introduced and their evaluation is modeled by OWA operators on criteria with unequal importances [22,23,24,25].

2. *OWA OPERATORS IN DOCUMENTS' INDEXING*

The aim of an Information Retrieval System is to evaluate user queries for information based on a content-analysis of the archived documents. An automatic analysis of documents requires a formal representation of their information content: documents' representations are automatically generated by the indexing procedure. Full-text indexing is a well known and common indexing procedure based on the assumption that terms in a document (called index terms) constitute an indication of the document's topics [18].

In the Boolean IR model documents are formally represented as sets of index terms: $R(d) = \{t\}$ in which $t \in T$, and the membership function correlating terms and documents is defined as: $F: D \times T \rightarrow \{0,1\}$, in which D is the set of documents, and T is the set of index terms. A value $F(d,t)=1$ ($F(d,t)=0$) indicates the presence (absence) of term t in document d. In Boolean systems there is no means to differentiate the role of terms in representing the information content of documents.

An improvement of this representation is achieved by associating with each index term a numeric weight (index term weight) which expresses the variable degree of significance that the term has in synthesizing the information content of the document [17,18].

Within fuzzy set theory, a document is then formally represented by a fuzzy set of terms with index term weights as membership values: $R(d) = \sum_{t \in T} \mu_d(t)/t$ in which $\mu_d(t) = F(d,t)$. The F correlation function takes values in the unit interval, $F: D \times T \rightarrow [0,1]$, and the index term weight $F(d,t) \in [0,1]$ represents the degree of significance of t in d; this value can be specified between no significance, $F(d,t)=0$, and full significance $F(d,t)=1$. Based on these significance degrees the retrieval mechanism can be extended with the ability to rank documents in decreasing order of their relevance to the user query, the relevance being expressed by a numeric score, called Retrieval Status Value (RSV), which denotes how well a document satisfies the query.

The quality of the retrieval results strongly depends on the criteria used to automatically compute the index term weights. Generally, to associate a numeric weight with an index term the criterium adopted is to sum the occurrences of the index term in the document and normalize the result in such a way that short documents are not penalized with respect to long ones [18]. A common definition for function F is the following:

$$F(d,t) = tf_{dt} * g(IDF_t)$$

in which :

- tf_{dt} is a normalized term frequency which can be defined as: $tf_{dt} = \frac{OCC_{dt}}{MAXOCC_d}$; OCC_{dt} is the number of occurrences of t in d, and $MAXOCC_d$ is the number of occurrences of the most frequent term in d;
- IDF_t is an inverse document frequency which can be defined as: $IDF_t = Log \frac{N}{NDOC_t}$; where N is the total number of documents in the archive and $NDOC_t$ is the number of documents indexed by t. g is a normalization function. The computation of IDF_t is particularly costly in the case of large collections which are updated online.

This representation of documents however does not take into account that the information in a document is often structured; for example a scientific paper is structured in sections such as title, authors, keywords, abstract, references etc.; in such a document, a single occurrence of a term in the title suggests that the paper is fully concerned with the concept expressed by the term, while a single occurrence in the references may suggest that the paper refers to other publications dealing with that concept. In this case the information role of each term occurrence depends on the semantics of the section where it is located. Besides this semantic dependency, the sections of a document may assume a different importance on the basis of users' needs. For example, when looking for papers written by a given person, the most important subpart is the authors section, while when looking for papers on a given topic, the title, keywords, abstract and introduction should be first analyzed. These considerations outline that the occurrences of a term in the different documents' sections have to be taken into account separately.

By integrating the fuzzy representation of documents based on index term weights with a structured view of documents, in [6] we have proposed a new fuzzy representation of documents which can be biased by user's interpretation. In this approach an archive is seen as a collection of documents sharing a common structure. A set of N index term weights or significance degrees, $F_1(d,t),\ldots,F_N(d,t)$, is associated with each document-term pair, with N the number of the documents' sections; each value $F_i(d,t)$ denotes the significance degree of term t in section i of document d. Since the global significance degree F(d,t) of a term in a document is also dependent on the subjective interests of the user, it is computed by an aggregation function specified by the user during the phase of query fomulation. The single $F_i(d,t)$s are considered as degrees of satisfaction of the criteria to be aggregated by the decision function according to a twofold specification of the user:

- users can express preferences on documents' sections by associating a numeric score $I_i \in [0,1]$ to each section i, so that the most important ones have an importance weight close to 1, the less important ones a weight close to 0. The importance degrees are used in the computation of the global significance degree F(d,t) to emphasize the role of the $F_i(t,d)$s of important sections with respect to those of less important ones;
- users can specify, through a linguistic quantifier Q, in how many documents' sections a term must be fully significant in order to be considered globally significant in the whole document; the quantifier can be chosen among *all* (the most restrictive one, indicating that the term must be fully significant in all the documents' sections), *at least one* (the weakest one indicating that the full significance of the term in one section is a sufficient condition), or *about k* which does not force the user to a useless precision in the case he/she has only a vague idea of the number of sections in which a term must be significant (it requires for a fuzzy threshold: the terms should be in *about k* or more sections).

The numeric OWA operators on criteria of different importances have been adopted to model the quantifier-guided aggregation function [22].
The criteria to be aggregated are the N significance degrees $F_1(d,t),\ldots,F_N(d,t)$ of term t in the N sections of a document d. When setting a retrieval session, the user can specify an importance weight $I_i \in [0,1]$ for each section i, and a linguistic quantifier Q which identifies the aggregation function. When processing a query the first step accomplished by the system for evaluating F(d,t) is the selection of the OWA operator associated with the linguistic quantifier Q, OWA_Q.
In [27] Zadeh introduced the definition of linguistic quantifiers as fuzzy sets, making a distinction between absolute and relative quantifiers. An absolute linguistic quantifier such as *about 10* is defined as a fuzzy set with a membership function AQ: $R^+ \rightarrow [0,1]$; when considering N criteria to be aggregated the corresponding OWA_{AQ} operator is derived by computing the elements of its weighting vector W as follows:
$w_i = AQ(i) - AQ(i-1)$ for $i=1,\ldots,N$.
A relative linguistic quantifier such as *few*, *most of*, *at least 80% of*, etc., is defined as a fuzzy set in the unit interval with membership function PQ: $[0,1] \rightarrow [0,1]$. When considering N criteria to be aggregated, the correspondent OWA_{PQ} operator of

dimension N is defined by computing the elements of its weighting vector W as follows: $w_i = PQ(i/N) - PQ((i-1)/N)$ for i=1,2,...N.
In the application considered here, as the number of the documents sections is limited (for example 10) it is more natural to use non-decreasing absolute linguistic quantifiers such as *about 6, more than 4,* etc. Moreover, in order to define the semantics of an absolute linguistic quantifier, the system manager could be supplied with a diagram XY visualized on the screen with the X axis corresponding with the number of the sections and the Y axis corresponding with the range [0,1]; he/she can choose whether to directly define a continuous function, or a discrete function by marking N distinct discrete points, as in Figure 1, in which the absolute fuzzy quantifier *about 3* is defined. In this case function AQ is defined as a discrete function on the domain $\{1,...,N\}$ in which N is the number of the documents' sections.

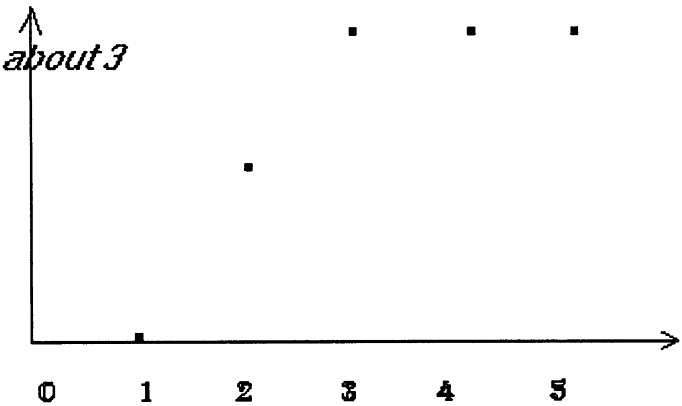


Figure 1: absolute definition of the linguistic quantifier *about 3*
(in which *3* is considered as a fuzzy threshold)

When the user does not specify any preferences on the documents' sections, the overall significance degree F(d,t) is obtained by applying directly the OWA_Q operator to the values $F_1(d,t),...,F_N(d,t)$:

$$F(d,t) = OWA_Q(F_1(d,t),...,F_N(d,t)) = \sum_{i=1}^{N} w_i \mathrm{Sup}_i(F_1(d,t),...,F_N(d,t)) \tag{1}$$

in which $\mathrm{Sup}_i(F_1(d,t),...,F_N(d,t))$ denotes the i-th highest value among the $F_1(d,t),...,F_N(d,t)$ and w_i are the weights of the weighting vector $W = [w_1,..., w_N]$ associated with OWA_Q.
When different importances $I_1,..., I_N$ are associated with the sections, it is first necessary to modify the values $F_1(d,t),...,F_N(d,t)$ in order to increase the "contrast" between the contributions due to important sections with respect to those of less important ones. The modified degrees $a_1, ... ,a_N$ of significance of the terms in sections are obtained as follows [22]:

$$a_i = (I_i \vee (1 - \text{orness}(W))F_i(d,t)^{I_i \vee \text{orness}(W)} \qquad (2)$$

in which W is the OWA_Q weighting vector. The second step is the evaluation of the overall significance degree F(d,t) by applying the operator OWA_Q to the modified degrees $a_1,..,a_N$:

$$F(d,t) = OWA_Q\,(a_1,..,a_N)$$

The fuzzy representation of structured documents has been implemented and evaluated, showing that it improves the effectiveness of a system with respect to the use of the traditional fuzzy representation [6].

$$\text{Recall} = \frac{N_{\text{retr-class}} \cap N_{\text{rel}}}{N_{\text{rel}}} \qquad \text{Precision} = \frac{N_{\text{retr-class}} \cap N_{\text{rel}}}{N_{\text{retr-class}}}$$

N_{retr_class} = number of documents retrieved in the specified class

N_{rel} = total number of documents relevant to the user query

Figure 2: Definitions of Recall and Precision

Table 1. summarizes the retrieval results of eleven queries processed by applying the traditional fuzzy representation of documents and the fuzzy representation of structured documents respectively. Salton's increasing output methodology [18] has been applied to analyse the distribution of documents judged worth retrieving by the user, N_{rel}, in various relevance classes by applying the definitions of recall and precision in Figure 2.

CLASSES	N_{retr}	VR+R+FR		VR+R		VR		
Average values	***Prec.***	***Rec.***	***Prec.***	***Rec.***	***Prec.***	***Rec.***	***Prec.***	***Agreement***
Traditional fuzzy representation	87%	58%	98%	31%	100%	20%	100%	36%
Fuzzy representation of structured documents	87%	87%	97%	55%	100%	40%	100%	61%

Table 1: Average values of Recall and Precision obtained for a sample set of queries by applying the two representations of documents.

In Table 1 each column lists the average values of Recall (Rec.) and Precision (Prec.) on the eleven test queries for documents classified by the system in different groups of relevance classes, except the second column which reports only the precision values for all the documents retrieved (N_{retr}=VR+R+FR+NR). The third column refers to the documents classified by the system in one of the most relevant classes, i.e., in one of the classes Very Relevant (VR), Relevant (R), Fairly Relevant (FR). The fourth column refers to the documents classified as Very Relevant (VR) or Relevant (R). The fifth column refers to the documents classified by the system as Very Relevant (VR). It can be observed that while the values of precision remain unchanged in the two versions of the system, the values of recall are higher by using the new representation than those obtained by using the traditional fuzzy representation. The last column reports the degrees of agreement between the system and the user classifications; also for this parameter the system with the fuzzy representation of structured documents classifies documents more closely to the user than the same system with the traditional fuzzy document representation.

3. *GENERALIZATION OF THE USE OF OWA OPERATORS IN DOCUMENT INDEXING*

The fuzzy representation of structured documents presented in the previous section suffers for some drawbacks. First of all often users find it difficult to specify the importance of the section by a number in [0,1]; it would be much more natural to specify it either in a linguistic form with the use of qualifiers such as *important, very important, fairly important, etc.* or by the aid of a graphic representation such as by moving a cursor on a bar representing the possible range of the sections' importance. Second, it has been observed that by modifying the importance of the sections even marginally, the retrieval results can change abruptly; this behaviour is not desirable, since what it is expected is that similar importance values yield almost unvaried results. A reduction of the granularity of the importance degrees would avoid this inconvenient and at the same time would supply users with a few but meaningful values [6].

Another objective of the generalization is related to the domain of the significance degrees of the sections. There are applications in which the domain [0,1] of the significance degrees of documents' sections is overdimensioned to the real needs. In fact, in some cases, the computation of the significance degree of a term in a document's section cannot be performed automatically, but it is specified by an expert on the basis of heuristic rules, not always clearly defined. In these situations, the expert finds it easier to specify the significance degree of a term by selecting a label on a finite ordinal scale such as {*none, very low, low, medium, high, very high, perfect*}. Another case takes place when the significance degree of a term in a section is derived by applying a knowledge based approach. In this situation, it can be easier to define fuzzy rules and determine a linguistic significance degree.

On the basis of these considerations, the fuzzy representation of structured documents presented in the previous section can be generalized to deal with both situations in which the significance degrees of a term in documents' sections and their importance values are numeric values in [0,1] and situations in which significances and importances are expressed by labels on a finite ordinal scale such as S = {*none, very low, low, medium, high, very high, perfect*}.

The problem is then to define a formalization of the aggregation function specified by a linguistic quantifier Q when both the criteria to be aggregated and their importances are not necessarily numeric values in [0,1], but they can be also labels. Ordinal OWA operators have been defined to aggregate values expressed as linguistic labels on an ordered ordinal scale S [23]. However, these operators are not general enough to deal also with numeric values. In the application considered here the aim is to apply a unique formalization of OWA operators which can be applied either to numeric or linguistic values. In the next subsection an approach is proposed to translate linguistic labels into numeric values and then to apply a numeric definition of OWA operators.

DEFINITION OF LINGUISTIC QUANTIFIERS ON LABELS THROUGH NUMERIC OWA.

Let us assume that the significance degrees and the importance values are expressed either by real values in [0,1] or by labels uniformely distributed on an absolute ordinal scale S such as:

$S = \{s_0, s_1, \ldots, s_{max}\} = \{none,\ very\ low,\ low,\ medium,\ high,\ very\ high,\ perfect\}$

in which $s_a, s_b \in S$, $a < b \Leftrightarrow s_a < s_b$ and $max = |S| - 1$.
Let us denote by $|S|$ the cardinality of S.
The following operations are defined:

Index of a label:

$$index : S \rightarrow [0, max] \qquad index(s_i) = i \quad \forall i\ [0, max] \tag{3}$$

Mapping a numeric value in [0,1] into a label:

$$Label: [0,1] \rightarrow S \qquad Label(x) = s_i$$
$$\text{for } \frac{i}{|S|} \le x < \frac{i+1}{|S|},\ i=0,\ldots,max \text{ and } Label(1) = s_{max} \tag{4}$$

Mapping a Label into a numeric value in [0,1]:

$$Label^{-1}: S \rightarrow [0,1] \qquad Label^{-1}(s_i) = \frac{\text{index}(s_i)}{|S|-1} \qquad \text{with } i=0,\ldots,max \tag{5}$$

In the following the procedure to aggregate through OWA operators the values representing the sections' significance degrees is described.

When no importance weights are associated with the sections and the sections' significance degrees are labels $F_i(d,t) \in S$, first the numeric values $a_1, \ldots, a_N$ are obtained through the application of the $Label^{-1}$ function defined in (5) to the linguistic values:

$$a_i = Label^{-1}(F_i(d,t)) \quad \text{for } i=1,\ldots,N$$

Then the overall significance degree F(d,t) is computed by applying the OWA_Q operator defined in (1) associated with the quantifier Q to the values $a_1,\ldots, a_N$.

$$\mathrm{F(d,t)} = \mathrm{OWA}_Q\,(a_1,\ldots, a_N) = \sum_{i=1}^{N} w_i \mathrm{Sup}_i\,(a_1,\ldots,a_N) \tag{6}$$

in which $\mathrm{Sup}_i\,(a_1,\ldots, a_N)$ is the i-th highest of the $a_1,\ldots, a_N$.
When different importance weights are specified, a more recent and better formalization than the one in (2) can be adopted of numeric OWA based aggregation of criteria with different importances [25]. In this formalization, numeric values of importance $I_i \in [0,1]$, are used to determine the weighting vector W associated with an OWA_Q operator corresponding with a relative increasing monotone linguistic quantifier Q. The underlying idea is to consider the interval [0,1] representing the range of the percentage number of the criteria and partition it into N unequal parts, depending on their importances, and not in N equal parts as it happens in the case of equal importances. Let us introduce this definition: let e_i be the numeric importance value associated with $\mathrm{Sup}_i(a_1, \ldots, a_N)$ which is the i-th largest of the numeric satisfaction degrees of the criteria a_j, i.e. $e_i = I_j$, and let:

$$e = \sum_{i=1}^{N} \mathrm{I_i} = \sum_{i=1}^{N} e_i \tag{7}$$

The weights of the vector W are obtained as follows:

$$w_i = PQ\left(\frac{1}{e}\sum_{j=1}^{i} e_j\right) - PQ\left(\frac{1}{e}\sum_{j=1}^{i-1} e_j\right). \tag{8}$$

It is important to observe that, with this formalization, the OWA weights are different for different satisfactions of the criteria. Further, each criterion having no importance plays no role. In [8] it has been shown that this formalization yields better results than the one based on the ordinal OWA operator.

In order to apply this procedure to our application, two preliminary operations are needed. First, when the importance values are expressed by labels $i_i \in S$ they have to be converted into numbers in [0,1] by applying the mapping function $Label^{-1}$ defined in (5): $I_i = Label^{-1}\,(i_i)$. Second, as the linguistic quantifiers used in this application are absolute quantifiers, we need to transform their definitions given by functions AQ into the corresponding relative functions defined on [0,1] so as to be able to determine the weighting vector of the OWA operator by taking into account the different importance degrees through formulae (7) and (8).
The absolute definition AQ of an absolute linguistic quantifier can be simply transformed into the definition PQ of a proportional quantifier by considering that $PQ(\frac{i}{N}) = AQ(\mathrm{i})$, in which N is the number of values to be aggregated. Moreover, if the absolute quantifer AQ is defined by means of a discrete function, a continuous definition of PQ is obtained by applying the following linear interpolation:

$$PQ(\frac{i}{N}) = (PQ(\frac{i}{N}) - PQ(\frac{i-1}{N}))*(N*x-(i-1)) + PQ(\frac{i-1}{N})$$
$$\text{for } x \in [\frac{i-1}{N}, \frac{i}{N}] \text{ and } i=1..N \qquad (9)$$

An example of such a transformation is shown in Figure 3.

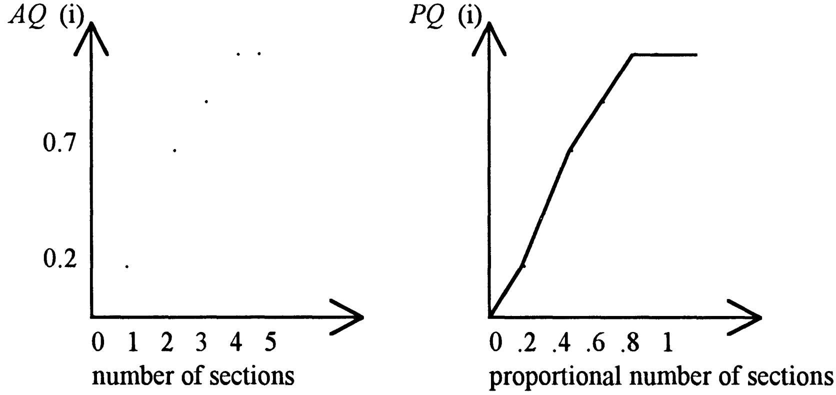


Figure 3: semantics of a quantifier Q: absolute discrete definition AQ and its corresponding proportional definition PQ

At this point the weighting vector W of the OWA_Q operator is obtained by applying (7) and (8) and then the overall significance degree F(d,t) is computed as previously by applying (6).

EXAMPLE

Let us show the application of this method through an example.
Let us assume that both the importance values and the sections' significance degrees are defined on the following ordinal scale:

S={*none, very low, low, medium, high, very high, perfect*} with | S | =7.

Let us consider five sections, N=5 with the following importances:

I = {I_1=*high*, I_2=*high*, I_3=*high*, I_4=*none*, I_5=*none*}

and for term t and document d let us consider the following sections' significance degrees:

F_1={*medium*, F_2=*high*, F_3=*very high*, F_4=*none*, F_5=*none*}

Let us compute the overall significance degree F(d,t) when the aggregation function is expressed by the absolute quantifier *about 3* defined by the discrete membership function *(A-about 3)* in Figure 1. The correspondent proportional definition (Figure 4) defined by the membership function *(P-about 3 = at least 60%)* is obtained by applying formula (11):

$$P\text{-}about\ 3\ (x) = \begin{cases} 1 & x \geq 0.6 \\ 2.5x - 0.5 & 0.2 < x < 0.6 \\ 0 & x \leq 0.2 \end{cases}$$

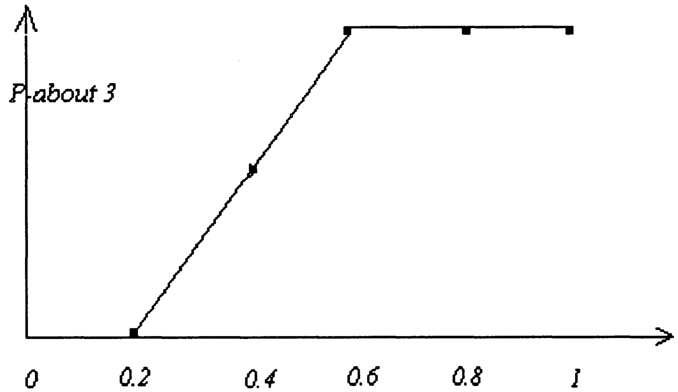


Figure 4: proportional definition of the linguistic quantifier *about 3*

The only sections of document d in which t has no significance are: section 4 with F_4=*none,* and section 5 with F_5=*none,* which have no importance at all, I_4=*none,* I_5=*none,* while in the sections with *high* importance (sections 1, 2 and 3) t has a significance: F_1=*medium,* F_2=*high,* F_3=*very high.*
First of all both the linguistic labels representing the sections' significance degrees of t in d and the linguistic importance labels are mapped into numbers in [0,1] by applying function $Label^{-1}$:

A = {$Label^{-1}$(*medium*), $Label^{-1}$(*high*), $Label^{-1}$(*very high*), $Label^{-1}$(*none*), $Label^{-1}$(*none*)} = {*0.5, 0.66, 0.83, 0, 0*}

I = {$Label^{-1}$(*high*), $Label^{-1}$(*high*), $Label^{-1}$(*high*), $Label^{-1}$(*none*),$Label^{-1}$(*none*)} =
= {I_1=*0.66*, I_2=*0.66*, I_3=*0.66*, I_4=*0*, I_5=*0*}

The importance values of the sections are ordered according to the significance of t in them:

$E = [e_1 = I_3,\ e_2 = I_2,\ e_3 = I_1,\ e_4 = I_4,\ e_5 = I_5] = [0.66,\ 0.66,\ 0.66,\ 0,\ 0]$

The weights of the vector $W_{P\text{-}about\ 3}$ are then obtained through (7) and (8) by computing:

$$e = \sum_{i=1}^{5} e_i = 1.98 \text{ and } w_i = about\ 3\left(\frac{1}{e}\sum_{j=1}^{i} e_j\right) - about\ 3\left(\frac{1}{e}\sum_{j=1}^{i-1} e_j\right)$$

$W_{P\text{-}about\ 3} = [0.3,\ 0.7,\ 0,\ 0,\ 0]$.

Finally the $OWA_{P\text{-}about\ 3}$ operator is applied to vector A and the overall significance degree of t in d is obtained: $F(d,t) = OWA_{P\text{-}about\ 3}(0.83,\ 0.66,\ 0.5,\ 0,\ 0) = 0.71$.

4. *OWA OPERATORS TO AGGREGATE SELECTION CONDITION IN QUERIES*

The first extensions of the Boolean query language within fuzzy set theory concerned the introduction of numeric weights, in order to allow the specification of the different importance of the search terms in the desired documents [3,4,9,10,13].

In [5] linguistic query weights have been introduced to overcome the main limitation of numeric weight to force the user to quantify the qualitative and rather vague concept of importance. In this extension linguistic descriptors such as *important, very important, fairly important,* are defined to specify the importance of query weights.

Another important extension of the Boolean query language has concerned the aggregation of selection criteria. The AND and OR connectives allow only for crisp aggregations which do not capture any vagueness. For example, the AND used for aggregating M selection criteria does not allow to tolerate the unsatisfactoriness of a single criterion; this may cause the rejection of useful items.

To introduce "soft" aggregation conditions, new query languages have been defined, which propose soft definitions of the operators for aggregating the selection criteria [14,19]; for example, Salton, Fox and Wu proposed a model based on a p-norm operator [16]. However, in these approaches different soft interpretations of the Boolean connectives in the same query are not supported.

Drawing on some previous works in Data Bases [12], within the framework of fuzzy set theory, a generalization of the Boolean query language has been defined [7], based on the concept of linguistic quantifiers: they are employed to specify both crisp and vague relationships among selection criteria. New aggregation operators, with a self-expressive meaning such as *at least k* and *most of*, are defined with a behavior between the two extremes corresponding to the AND and the OR connectives, which allow, respectively, requests for *all* and *at least one of* the selection criteria. Ordered Weighted Averaging (OWA) operators have been used to define the linguistic quantifiers.

With a query language adopting linguistic quantifiers the requirements of a complex Boolean query are more easily and intuitively formulated. For example when desiring that *at least 2* out of the three criteria "politics", "economy", "inflation" be satisfied, one should formulate the following Boolean query:

(politics AND economy) OR (politics AND inflation) OR(economy AND inflation)

which can be replaced by the simpler one:

at least 2(politics, economy, inflation)

The expression of any Boolean query is supported by the new language via the nesting of linguistic quantifiers. For example a query such as:

<image> AND (<processing> OR <analysis>) AND <digital>

can be translated into the following new formulation:

all (<image>, *at least 1 of*(<processing>, <analysis>), <digital>)

A quantified aggregation function can thus be applied not only to single selection criteria, but also to other quantified expressions. By this novel formulation, the ambiguities generated by the use of the AND and OR operators in a linguistic context are eliminated.The formal definition of the query language with linguistic quantifiers can be found in [7].
The following quantifiers have been defined:

- *all*: it replaces the AND;
- *at least k of*: it acts as the specification of a crisp threshold of value k on the number of selection criteria. It is defined by a weighting vector $W_{at\ least\ k\ of}$ in which w_k=1, and w_j=0, for jk. Notice that *at least 1 of* selects the maximum of the satisfaction degrees and it has thus the same semantics of the OR.
- *about k of*: this is a softer interpretation of the quantifier *at least k of* in which the k value is not interpreted as a crisp threshold, but as a fuzzy one. This means that the user is fully satisfied if k or more criteria are satisfied, but she/he gets a certain degree of satisfaction even if $k-1, k-2, \ldots, 1$ criteria are satisfied.
  This quantifier is defined by a weighting vector $W_{at\ least\ about\ k\ of}$ in which $w_i = \frac{i}{\sum_{j=1}^{k} j}$ for $i \leq k$, and $w_i = 0$ for $i > k$.
- *at least half of* : it is defined as a synonym of *at least* $\frac{n}{2}$ *of* in which n is the total number of selection criteria.
- *most of*: it is defined as a synonym of *at least* $\frac{2}{3}$ n *of* in which n is the total number of selection criteria. *Most* of the criteria as been interpreted with a semantic closer to *half of* the criteria than to *all* the criteria.

- *at least a few of:* it is defined as a synonym of *about* $\frac{n}{3}$ *of* in which n is the total number of selection criteria. Symmetrically to the previous consideration *at least a few of* the criteria has been interpreted closer to *half of* the criteria than to *at least 1 of* the criteria. This choice has been done to reduce the effect of ORing all the criteria, which may cause the retrieval of nonrelevant items.
- *more than k:* it specifies that the overall satisfaction is greater than 0 if more than k criteria are satisfied and the higher is the number of the criteria which are satisfied, the higher is the overall satisfaction degree. The weighting vector associated with this quantifier is $W_{more\ than\ k}$ in which $w_i=0$, for $i \leq k$, and $w_i=\frac{n-i+1}{\sum_{j=1}^{n-k} j}$, for *ik.*

5. *OWA OPERATORS IN AGGREGATING SEARCH TERMS WITH DIFFERENT RELATIVE IMPORTANCES*

In this section a generalization of the query language based on linguistic quantifiers previously described is proposed. The selection criteria to be aggregated are considered here as weigthed terms. The approach analysed in this section is different from the one usually adopted, in which a numeric weight is seen as a constraint on the significance degree of the term in the documents; in the traditional approach the evaluation of a weighted query with respect to a given document is performed by first evaluating the degree of satisfaction of the constraints (specified by the weighted selection criteria) in the considered document and then by aggregating the obtained numeric values. In this section another approach is defined to interpret query term weights as the indication of the "relative importance" of the search terms.
A query language is considered in which the aggregation operators are specified by means of linguistic quantifiers; the weights associated with terms are interpreted as the importances associated with the selection criteria (terms) to be aggregated. The difference with respect to the traditional approach consists in considering as selection criteria the single terms, not the weighted terms.
In the following a query evaluation procedure is defined, which satisfies the separability property of the wish list [20].
In the first subsection the traditional semantics of relative importance for weighted terms is described and it is shown why it is dependent on the applied aggregation operator.
In the second subsection the OWA operators working on criteria with different importances are applied, as formalized in section 3., formulae (7) and (8), to model the aggregation of search terms having different relative importance weights.

RELATIVE IMPORTANCE SEMANTICS IN BOOLEAN QUERIES

This semantics defines query weights as measures of the "relative importance" of each term with respect to the others in the query [15,19,21]. By assigning relative importance weights to terms in the query, the user requests that the computation of the RSV be dominated by the most heavily weighted terms. A term of 0 weight must be considered completely irrelevant, with no effect on the RSV. With this kind of semantics, a Boolean query (with no weights associated with the terms) is interpreted as the formulation of a request in which all the terms are equally "relevant" to the user with an implicit weight of value 1. A selection criterion is specified by a pair <t,w>, and it is interpreted as a constraint to be evaluated for a given document. Considering a fuzzy retrieval model in which documents are represented as fuzzy sets of terms, this semantics has been formalized within fuzzy set theory by defining the evaluation function E of a pair <t,w> for a document d.

To reflect the semantics of relative importance function E has been defined depending if the aggregation operator is either OR or AND [3,15,19,21].

For example in the definition by Bookstein [3] function E is defind as a product when the terms are linked by the OR operator:

$$E(d,< t, w>) = w * F(d,t)$$

and a ratio when the terms are linked by the AND operator:

$$E(d,< t,w >) = \begin{cases} F(d,t)/w & \text{for } F(d,t) \le w \\ 1 & \text{otherwise} \end{cases}$$

The OR operator evaluated as a maximum is conditioned by the highest RSV produced by the evaluation of its operands. Since the effect of the product with a weight $w \in [0,1]$ is to reduce the significance degree F(d,t) of t in d, the most heavily weighted terms will dominate.

On the contrary, the AND operator evaluated as a minimum is dominated by lowest RSVs resulting from the evaluation of its operands. In this case the normalized ratio F(d,t)/w is higher as w becomes lower. So, also in this case the highest weighted term will dominate the computation.

These definitions of relative importance weights force to define a query evaluation mechanism which does not satisfy the so called separability property of an IRS; this property states that the evaluation of a single query component (a weighted search term) must not influence, nor be influenced, by the evaluation of the other components, for example the specified aggregation operator. When the separability property is satisfied the query evaluation mechanism can be defined as a two step process: in the first step, the single pairs <search term, weight> in the query are evaluated for a document $d \in D$. A score in [0,1] is computed as a result of a single pair evaluation and it expresses the degree to which document d matches the pair <search term, weight>. In the second step, which can be a recursive process in the case of complex queries with nested levels, the evaluation of the aggregation operator

(AND, OR) linking couples of weighted search terms takes place. This second step produces the final RSV which is interpreted as the relevance of the document to the whole query.

The definition of a query weight semantics which discards this property is evidently seen as a drawback, as it induces a loss of generality, and then a complication of the query evaluation procedure. To simplify the query evaluation many of the approaches based on the relative semantics suggest to allow only queries in Disjunctive Normal Form; this choice implies a loss of the full potentialities of the Boolean query language.

RELATIVE IMPORTANCE SEMANTICS IN QUERIES WITH LINGUISTIC QUANTIFIERS AS AGGREGATION OPERATORS

In this subsection we propose a new evaluation of queries in which term weights are interpreted as the indications of the relative importance of terms: query term weights are interpreted as the importance weights of selection criteria constituted by terms aggregated by linguistic quantifiers defined as OWA operators. For sake of simplicity we consider here only queries constituted by M search terms and one linguistic quantifier Q aggregating the M selection criteria: $Q(<t_1,w_1>,<t_2,w_2>, ..., <t_M,w_M>)$.

The linguistic quantifiers that one can use in a query can vary with a behaviour between the two extreme cases of the AND *(all)* and the OR *(at least 1)*.

When query term weights are interpreted as the importances associated with the selection criteria (single terms) specified in the query, the mechanism whih evaluates the relevance of each document d can be defined as a two step process, which preserves the separability property :

- first the satisfaction of the selection criteria specified by the search terms t_1,...., t_M is evaluated with respect to document d. This produces M numeric values (denoted by RSV_i), one for each search term. The implicit requirement given by a search term t_i is that t_i must be the most significant as possible in the desired document; then for term t_i in d, $RSV_i=F(d,t_i)$;
- second, to aggregate the RSVi obtained at the previous step, the OWA operator weighted corresponding to quantifier Q is applied, by taking into account the different importances of the RSV_i, expressed by the weights w_i specified in the query.

The formalization of OWA operators with different importances is adopted as defined in section 3. This procedure works even when the importance weights w_i are expressed by linguistic labels on an ordinal scale. In this case they have to be converted into numbers in [0,1] by applying function $Label^{-1}$ defined in (5): $I_i = Label^{-1}(w_i)$.

Once the values RSVi have been obtained, the weighting vector W of the OWA_Q operator is obtained by applying formulae (7) and (8), by working on the query term weights w_i; the final RSV is then obtained by applying the OWA_Q operator.

6. *CONCLUSIONS*

Soft retrieval models are nowadays a research topics of great interest due to the development of wide area networks. By regarding the IR activity as a MCDM activity we have defined soft approaches to design flexible IRSs. In this paper the use of OWA operators in document indexing and querying is described. A possible evolution of preexisting models of IR can be designed based on OWA.

7. *REFERENCES*

1. Biswas G., Bezdek J. C., Marques M., and Subramanian V. (1987). Knowledge-assisted document retrieval. I. The natural-language interface. II. The retrieval process.Journal of the American Society for Information Science, 38(2), March.
2. Bolc L., Kowalski A., Kozlowska M., Strzalkowski T. (1985). A natural language information retrieval system with extentions towards fuzzy reasoning (medical diagnostic computing). Int. J. of Man-Machine Studies, 23(4), October.
3. Bookstein A (1980). Fuzzy requests: an approach to weighted Boolean searches. J. of the American Society for Information Science, 31, 240-247.
4. Bordogna G., Carrara P., Pasi G. (1991). Query term weights as constraints in fuzzy information retrieval. Information Processing & Management 27(1), 15-26.
5. Bordogna G., and Pasi G. (1993). A fuzzy linguistic approach generalizing Boolean IR: a model and its evaluation. J. of the American Society for Information Science, 44(2), 70-82.
6. Bordogna G., and Pasi G. (1995) Controlling retrieval through a user adaptive representation of documents, Int. J. of approximate reasoning, 12, 317-339.
7. Bordogna G., and Pasi G.(1995). Linguistic aggregation operators in fuzzy information retrieval. Int. J. of Intelligent systems, 10, 233-248.
8. Bordogna G., Fedrizzi M., Pasi G. (1997). A Linguistic Modeling of Consensus for a Fuzzy Majority in Group Decision Making. IEEE Transactions on Systems, Man and Cybernetics, 1.
9. Buell D. A. and Kraft D. H. (1981). Threshold values and Boolean retrieval systems. Information Processing and Management, 17(3), 127-136
10. Cater S. C., and Kraft D. H. (1989). A generalizaton and clarification of the Waller-Kraft wish-list," Information Processing & Management, v. 25, 1989, pp. 15-25.
11. Doszkocs T. (1986). Natural Language Processing in Inf. Retrieval. J. of the American Society for Information Science, 37(4), 191-196.
12. Kacprzyk J., Zadrozny S., Ziolkowski A. (1989). FQUERY III+: A 'human-consistent' database querying system based on fuzzy logic with linguistic quantifiers. Information Systems, 14(6).
13. Miyamoto S. (1990). Fuzzy sets in Information Retrieval and Cluster Analysis. Kluwer Academic Publishers.
14. Paice C. D. (1984). Soft evaluation of Boolean search queries in information retrieval systems. Information Technology: Research Development Applications, 3(1), January, 33-41.
15. Radecki T. (1979). Fuzzy set theoretical approach to document Retrieval. Information Processing & Management 15, 247-260.
16. Salton G., Fox E., and Wu H. (1983). Extended Boolean information retrieval. Communications of the ACM, 26(12), 10221036.

17. Salton G., Buckley C. (1988). Term weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513-523.
18. Salton G. (1989). Automatic text processing: The transformation, analysis and retrieval of information by computer, Addison Wesley.
19. Sanchez E. (1989). Importance in knowledge Systems. Information Systems 14(6), 455-464.
20. Waller W.G., and Kraft D.H. (1979). A mathematical model of a weighted Boolean Retrieval System. Information Processing & Management 15, 235-245.
21. Yager R. R. (1987). A note on weighted queries in information retrieval systems, J. of the American Society for Information Science 38(1), 23-24.
22. Yager R. R. (1988). On Ordered Weighted Averaging aggregation Operators in Multi Criteria Decision Making, IEEE Trans. on Systems, Man and Cybernetics 18(1), 183-190.
23. Yager, R.R., (1992). Applications and Extensions of OWA Aggregations, International Journal of Man-Machine Studies, 37, 103-132.
24. Yager R.R. (1994). Interpreting Linguistically Quantified Propositions, International Journal of Intelligent Systems, 9, 541-569.
25. Yager R.R. (1994). Quantifier Guided Aggregation Using OWA Operators, Technical Report #MII-1504, Iona College, Machine Intelligence Institute, New Rochelle, NY 10801, November, 1994.
26. Zadeh L.A. (1965). Fuzzy sets. Information and control, 8, 338-353.
27. Zadeh L.A. (1983). A computational Approach to Fuzzy Quantifiers in Natural Languages, Computing and Mathematics with Applications. 9, 149-184.

7

OWA OPERATORS IN LEARNING AND CLASSIFICATION

IMPLEMENTATION OF OWA OPERATORS IN FUZZY QUERYING FOR MICROSOFT ACCESS[1]

Janusz Kacprzyk and Sławomir Zadrożny

Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
E-mail: {kacprzyk, zadrozny}@ibspan.waw.pl

Abstract. We present an implementation of the OWA operators in FQUERY for Access (Kacprzyk and Zadrożny, 1994a, b; 1995a, b; 1997), an add-on for Microsoft Access supporting fuzzy queries of the type "find all records such that the values of, e.g. *most* of the *important* fields are as specified (possibly fuzzily)". The OWA operators are used as general aggregation operators for fuzzy linguistic quantifier quided aggregation of atomic conditions in the query.

Keywords: database querying, fuzzy querying, fuzzy logic, fuzzy linguistic quantifier, graphical user interface, Microsoft Access.

1. INTRODUCTION

The development of information technology (IT) changes everyday life of virtually all people. This way or another, everybody has to deal, more or less interactively, with some computer-based systems. Presumably the most popular software behind virtually all IT applications is a database management system (DBMS). DBMSs make it possible to organize and retrieve data efficiently. Usually, there is somebody responsible for the maintenance of a database, i.e. adding some new records, deleting or updating existing ones, etc. At the same time, there exists by far a much larger group of users of a database. From their point of view, a database may be identified with its querying interface. It may take different forms: a map in case of a city guide, a set of commands, or a highly sophisticated graphical user interface (GUI). Functionally, they are all the same as they allow for the specification of the user's requirements as to the objects he or she wants to retrieve from a database. The ease

[1] All the names of companies and their products used in this paper are trademarks or registered trademarks, and - to save space - this will not be explicitly repeated.

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

with each it can be done decides of the usefulness of a given interface, and the whole DBMS.

The human users tend to employ in querying some highly aggregated and vaguely defined concepts, relations, etc. For instance, in a real-estate agency context, the user may be interested in retrieving all houses meeting his or her criteria which refer to such terms as "*large* living room", "located *close* to a shopping mall" or "land area *not much greater* than 500 sq. meters". Moreover, the logical dependence between the required fulfillment of particular conditions may be much more subtle than this provided by the classical logical connectives AND and OR. Here is the place where the OWA operators fit perfectly by making it possible to represent an aggregation operator lying somewhere in between the classical logical connectives (AND and OR). They give the user much more control over the aggregation process of particular conditions (criteria) and, as we will show in this paper, may be very useful in the context of compound, vague queries to a database.

We will propose an implementation of the definition and manipulation of the OWA operators in the context of a fuzzy querying interface. Namely, we will describe some new extensions to previously presented FQUERY for Access (Kacprzyk and Zadrożny, 1994a, b; 1995a, b; 1997; Zadrożny and Kacprzyk, 1996), an add-on to the popular Microsoft Access DBMS (see also Kacprzyk and Ziółkowski, 1986a, b, 1987, and Kacprzyk, Zadrożny and Ziółkowski, 1988).

First, we briefly discuss the use of fuzzy elements in database querying as they are employed in FQUERY for Access. Second, we show how the OWA operators may be used in fuzzy querying. Third, we present these elements of the user interface which make it possible to introduce and manipulate the OWA operators as elements of a query. We conclude with some examples of the definition and execution of queries involving these non-standard aggregation operators.

2. SYNTAX OF A FUZZY QUERY AND REPRESENTATION OF FUZZY ELEMENTS

Basically, querying of a database consists in retrieving these *records* which *matc* specified *conditions*. The conditions of a query impose certain requirements as to th values of selected *fields* of a record. Thus, such conditions are built of field name relational operators and scalar values as, e.g., in the following example:

PRICE < 100

where 'PRICE' is a name of a field in the records, '<' is a relational operator '100' is a scalar value. Usually, a condition consists of a number of *si conditions*, like the one in (1), connected by logical connectives AND and OR.

If a relation specified in a simple condition holds for the values of given fie the record, then the record matches this part of the condition. Then, the match particular simple conditions is *aggregated* to obtain the matching of the query. The way this aggregation is done is governed by a combination of connectives applied.

Thus, in the framework of classical logic, the concept of matching is of the binary nature. For the users of a database it is often more natural to use less precisely specified queries. In case of compound queries, consisting of many simple conditions, this imprecision may have twofold roots. First, simple conditions may directly refer to some vague concepts like "high", "young" etc. or some fuzzy relations like „much less than" which are not easy to grasp in the two-valued logic framework. Second, these simple conditions may be assigned importance coefficients and be subjected to an aggregation using non-standard logical connectives. For example, instead of requiring the fulfillment of *all* simple conditions, as is the case when these conditions are connected with AND, one may be satisfied when, say, '*most*' of them are satisfied. Thus, a *linguistic quantifier* may be employed as an aggregation operator. The representation and processing of the latter type of imprecision using the OWA operators are dealt with in this paper.

If some imprecise terms and non-standard logical connectives are allowed in a query, one has to consider a *degree of matching* of particular simple conditions and the query because the matching is not a binary concept any more.

Classical querying languages do not provide any means to deal with such types of imprecision in search criteria. Thus, first we propose some extensions to the syntax of SQL - an industry standard for querying in relational DBMSs, implemented also in Microsoft Access.

Here, we consider only a subset of Microsoft Access's SQL. Our point of departure may be the following query:

SELECT <list of fields>
WHERE
$cond_{11}$ AND $cond_{12}$ AND ... AND $cond_{1k}$
OR
... (2)
OR
$cond_{n1}$ AND $cond_{n2}$ AND ... AND $cond_{ml}$

where each of $cond_{ij}$, is later referred to as *an atomic condition*, and the ANDed sequence of atomic conditions is referred to as a *subcondition*.

FQUERY for Access (cf. Kacprzyk and Zadrożny, 1994a, b; 1995a, b; 1997) extends the above scheme allowing for the use of various fuzzy terms. Namely, fuzzy values and fuzzy relations may be used inside atomic conditions as, e.g., in "PRICE is *low*". On the other hand, fuzzy quantifiers may be used instead of the classical logical connectives AND and OR. Hence, we obtain the following general form of the query accepted by FQUERY for Access:

SELECT <list of fields>
WHERE <global fuzzy quantifier>
<fuzzy quantifier> ($cond_{11}$, $cond_{12}$, ... ,$cond_{1k}$)
... (3)
<fuzzy quantifier> ($cond_{n1}$, $cond_{n2}$, ... ,$cond_{nm}$)

The part of the query following the keyword WHERE may be expressed in a more formal way as follows:

```
<condition>                     ::= <global linguistic quantifier>
                                    <set of subconditions> ;

<set of subconditions>          ::= <subcondition> |
                                    <subcondition> OR <set of subconditions>

<subcondition>                  ::= <importance coefficient> <linguistic
                                                          quantifier>
                                    <set of atomic conditions>

<set of atomic conditions>::=       <atomic condition> |
                                    <atomic condition> AND
                                    <set of atomic conditions>

<global linguistic quantifier>::= <linguistic quantifier>

<linguistic quantifier>         ::= <OWA-tag> <quantifier name>

<OWA-tag>                       ::= OWA |
```

Thus, fuzzy linguistic quantifiers may be used in a query on two levels. A global linguistic quantifier [cf. (3)] may be used instead of the ORs joining in (2) all subconditions of a given query, and other linguistic quantifiers may replace the ANDs joining atomic conditions inside the particular subconditions. Moreover, different importance may be assigned to the particular subconditions. The more important coefficient assigned to given subcondition, the more its fulfillment influences the fulfillment of the whole query.

The OWA operators may be employed to represent fuzzy quantifiers in terms of their related aggregation behavior. In comparison to the original Zadeh's (1983), the OWA operators better correspond to the logical nature of the AND and OR connectives which we are going to replace in order to make the aggregation of the matching degrees. On the other hand, in Zadeh's (1983) approach the definition of a fuzzy quantifier is much simpler, in particular as it does not depend on the number of elements being aggregated. This is very convenient in the context of querying, as the number of atomic conditions or subconditions may change several times during the formation of the query by the user. As an OWA operator requires a vector of n weights, where n is the number of elements to be aggregated, each time the user adds/removes an atomic condition in the scope of an OWA operator, this operator has to be redefined.

Thus, in FQUERY for Access, the primary way to define an OWA operator is through the definition of a linguistic quantifier in the sense of Zadeh. More precisely, the process of the definition of a linguistic quantifier boils down to the definition of its piecewise linear membership function shown in the Fig. 1.

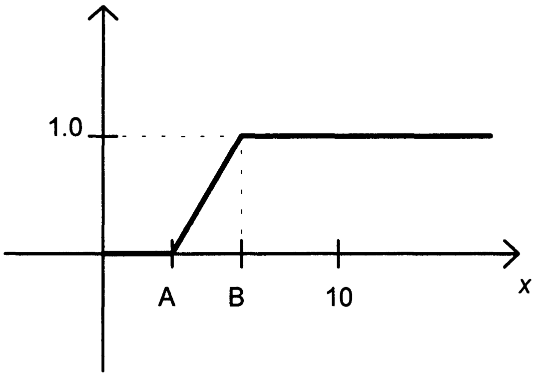


Figure 1. Example of the membership function of a linguistic quantifier

As in original Zadeh's (1983) approach, a linguistic quantifier is represented as a fuzzy set in [0.0, 10.0]. This interval is assumed here instead of [0, 1] only for technical, not conceptual reasons, hence the essence of Zadeh's approach is not changed. Thus, to define a fuzzy linguistic quantifier it is needed to provide two numbers corresponding to A and B in Fig. 1. Their interpretation is as follows. Suppose that a query (subcondition) is composed of 10 subconditions (atomic conditions). Then, for the evaluation of the matching degree for the whole query (subcondition), given the matching degrees for all subconditions (atomic conditions) - jointly referred to later on as *query parts* - the following scheme is used:

- if the number of query parts which are matched is less than A then the query (subcondition) is not matched - by a particular record - at all or, in other words, the matching degree is equal 0.0.
- if the number of query parts which are matched is greater than or equal A and less than B, then the query (subcondition) is satisfied to a degree between 0.0 and 1.0, the closer to B the higher.
- if the number of query parts which are matched is greater than or equal B, the query (subcondition) is fully satisfied, i.e. the matching degree is equal 1.0.

Notice that the above should be properly understood, because the particular query parts may be matched to a degree which is between 0.0 and 1.0; in the above explanations we have assumed that they are either matched or not, just for simplicity.

Evidently, the number of query parts can be arbitrary, say N, but, as the quantifier is defined through a membership function, we can map $[0, N]$ and the number of actually satisfied clauses onto [0, 10] preserving the meaning of the quantifier.

Therefore, Zadeh's (1983) approach to linguistic quantifiers allows to handle a varying number of aggregated arguments relatively easily. It possesses quite a clear interpretation and the user may quickly define a linguistic quantifier roughly corresponding to his or her requirements as to how partial criteria of the query should be aggregated. On the other hand, the OWA operators allow for a more precise definition of the aggregation operator and its easy fine tuning afterwards, but are rather inconvenient in case of a varying number of arguments. Thus, in our approach we are taking the best of the two methods.

First, the user defines or employs one of the predefined linguistic quantifiers meant in the sense of Zadeh (1983). Then, when the query is completely specified, but before the actual record retrieval, the user has a chance to modify the OWA operator determined from the linguistic quantifier employed.

The weights of an OWA operator of dimension N determined from a regular monotone nondecreasing linguistic quantifier Q, are calculated due to Yager (1988):

$$w_i = \begin{cases} Q(i/N) - Q((i-1)/N) & \text{for } i = 1,\ldots,N \\ 0 & \text{for } i = 0 \end{cases} \tag{4}$$

for i=1,...,N.

The user may modify the particular weights. Certainly, it is not easy to decide which weights best correspond to the needed aggregation operator. In order to assist and guide the user during the fine tuning of the OWA operator weights, the measures of ORness and dispersion are used.

These measures were introduced by Yager (1988) and are calculated for a given OWA operator F with the weight vector $[w_i]_{i=1,\ldots,N}$ as follows:

$$\text{ORness}(F) = (\sum_{i=1}^{N} (N-i) * w_i) / (N-1) \tag{5}$$

$$\text{disp}(F) = -\sum_{i=1}^{N} w_j * \ln(w_j) \tag{6}$$

The measure of ORness(F) says how similar is the OWA operator F to the logical connective OR (in terms of its aggregation behavior). Namely

$$\text{ORness}(F) = \begin{cases} 1 & \text{iff } F \text{ is OR} \quad (F \equiv \text{OWA}^{\text{OR}}) \\ a \in (0,1) & \text{iff } F \text{ is "between" OR and AND} \\ 0 & \text{iff } F \text{ is AND} \quad (F \equiv \text{OWA}^{\text{AND}}) \end{cases} \tag{7}$$

By employing the user interface described in the next section, a fine tuning of the OWA operator may be done in the following way. Instead of dealing with the particular weights separately, the user may request to increase or decrease the ORness of the currently defined operator. It is done using the following simple algorithm (the algorithm shown applies when the increase of ORness is required; in the opposite case, i.e. when the decrease of the measure is required, an analogous algorithm is used).

Let:

- the OWA operator F be defined by the vector of weights $[w_1,\ldots,w_N]$, and

- z^0 be required, increased value of „orness" measure for F; $z^0 \in (\text{ORness}(F),1)$.

Then:

Step 1. $\Delta z := z^0 - \text{ORness}(F)$

Step 2. $k := \arg\max_i \{w_i : w_i > 0\}$

$$x := 2(N-1)\Delta z / (4N - 3k) \tag{8}$$

Step 3. <u>If</u> $x > w_k$ <u>then</u> $x := w_k$

Step 4. $w_k := w_k - x$

$w_i := w_i + x/(k-1) \qquad \forall i \in [1, k-1]$

Step 5. <u>If</u> ORness$(F) < z^0$ <u>then</u> Go to Step 1.

STOP

During the whole process of weights modification the user is informed about the values of both the measures, i.e. ORness and dispersion.

3. USER INTERFACE FOR HANDLING THE OWA OPERATORS

FQUERY for Access is an add-on to Microsoft Access extending its querying capabilities with the possibility to use various fuzzy elements. From the point of view of the user interface, the presence of the package manifests itself only by one additional toolbar available in Access's native query formation environment. This means, that the user does not have to learn and use the whole new toolkit. He or she may form conventional as well as "fuzzy" queries in basically the same way.

In order to use various fuzzy elements in a query the user has to execute two actions: define fuzzy and auxiliary elements (e.g., a linguistic quantifier), and put them into the query. All actions are performed using a special toolbar available when the FQUERY for Access is installed (see Exhibit 1). Details can be found in, e.g., Kacprzyk and Zadrożny (1995a, b). Here we will concentrate on the handling of fuzzy linguistic quantifiers and OWA operators.

As already mentioned, the OWA operator may be introduced into the query through the selection of a linguistic quantifier. The definition of a linguistic quantifier consists of a name, and two numbers corresponding to A and B in Fig. 1. The name is used to identify the quantifier during the construction of a query. The definition of a linguistic quantifier is stored in a table maintained by FQUERY for Access. During the quantifier definition the user can check the shape of its membership function on a graph corresponding to Fig. 1 - see Exhibit 1.

In Exhibit 2 it is shown how a linguistic quantifier may be introduced into a query. Namely, the list of linguistic quantifiers known to FQUERY for Access (either defined earlier by the user or predefined) is displayed and the user should pick up one of them. Additionally, the user has to choose if the selected linguistic quantifier should be used as is, i.e. employing the original Zadeh's (1983) approach or the OWA operator determined from the fuzzy linguistic quantifier using (4). Then, after pushing a button, an appropriate addition is made to the current definition of the query.

In order to make the interface more flexible in respect to the manipulation of OWA operators, we introduce still another possibility to handle them inside a query. Namely, if there is no linguistic quantifier specified by the user in a query, a *default* OWA operator is placed there. In particular, if a global linguistic quantifier is omitted, the OWA^{OR} operator is put into the query by default. On the other hand, if a linguistic quantifier is omitted in a subcondition, the default operator is assumed to be the OWA^{AND} operator [cf. (3)]. These default OWA operators are not visible during the query construction. They are available for the user's modifications only at the stage of fine-tuning of the OWA operators to be described later.

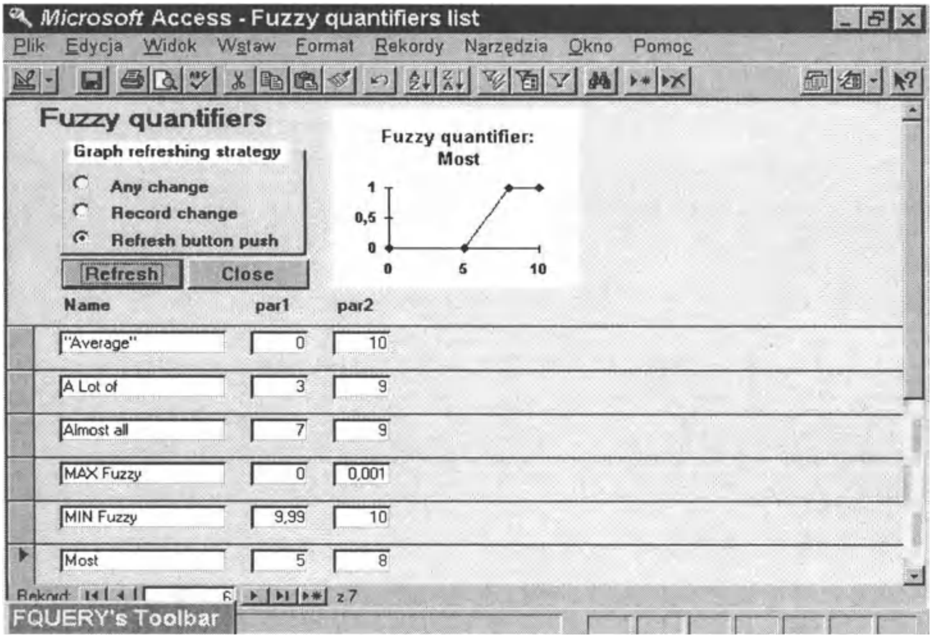


Exhibit 1. A list of linguistic quantifiers in Zadeh's (1983) sense

When the construction of the query is completed, the user can initiate the execution of querying by pressing the GO button. Then, FQUERY for Access processes the query replacing all fuzzy elements with properly chosen regular elements of the query understandable for Microsoft Access's native querying engine. This is a quite sophisticated process - its description one can find in Kacprzyk and Zadrożny (1995a, b). Here we can only mention, that each reference to a linguistic quantifier in a query is simply annotated in a FQUERY for Access's internal structure, and removed from the query.

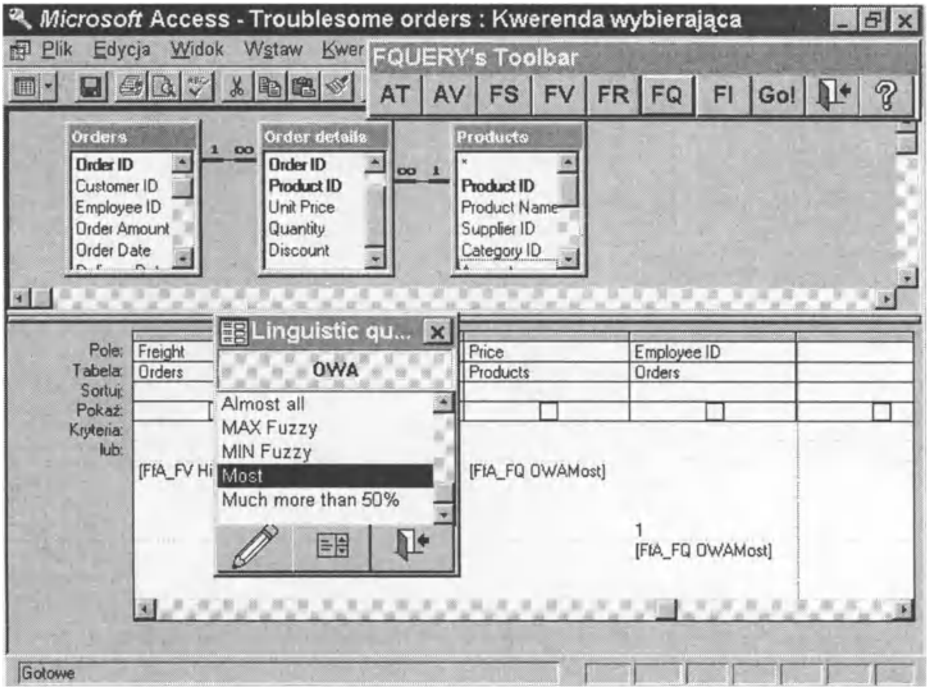


Exhibit 2. The query construction screen

Then, at the user's request, an additional step may be performed during the processing of the query. Namely, the screen, shown in Exhibit 3, is displayed allowing the user to specify more precisely the particular weights of the OWA operator. At the top of this screen all OWA operators appearing in the query, including the default ones, are listed. The user has to select one of them and, then, the weights of this operator are displayed below. Initially, these weights are calculated according to (4) for the explicitly introduced linguistic quantifiers (in terms of their related OWA operators), or correspond to particular default OWA operators.

Then, the user can modify them in several ways:

- "manually", by setting each weight separately,
- by pressing the AND button which sets the OWA weights to [0,...,1],
- by pressing the OR button which sets the OWA weights to [1,...,0],
- by automatically increasing/decreasing the ORness of the operator by a specified amount

The last modification is performed by the system automatically due to (8).

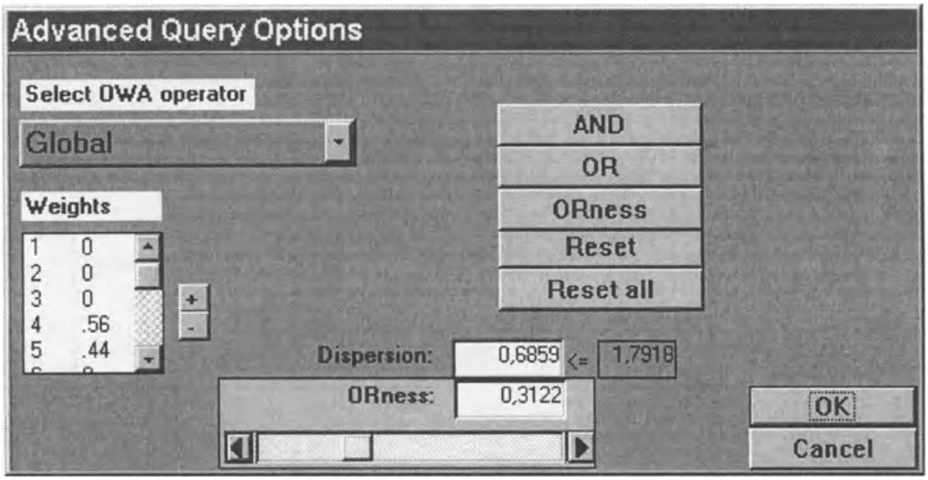


Exhibit 3. Fine tuning of the OWA operators

4. EXAMPLE OF A QUERY WITH THE OWA OPERATORS: DEFINITION AND PROCESSING

Usually, while composing a query in Microsoft Access the user interacts with an interface of a query-by-example (QBE) type. Thus, he or she does not to have construct SQL expressions directly, but can employ the tools usually provided by a graphical user interface. The same applies to handling the OWA operators. An OWA operator can be inserted into the query by pushing a button in the toolbar of FQUERY for Access. There are some rules stating when such an insertion is meaningful but we will omit them here; for detailed presentation see Kacprzyk and Zadrożny (1995a, b).

We will consider here an example based on the `Northwind.mdb` database, that is included in Microsoft Access. Let us assume that we have a database of a small trading company and we wish to retrieve a list of *troublesome orders* (orders requiring special attention).

The first problem we encounter trying to get such an information from the database is vagueness of the term *troublesome order*. Suppose that in our case there are the following factors (conditions) that may make the oders potentially troublesome:

- because our company is based in the USA, orders coming from outside of the country may require more attention (transportation formalities, customs duties, etc.),
- a short delivery time,
- a low profitability of the order which can be implied by:
  - a *low* value of the whole order,
  - *high* freight charges, (9)
  - *high* discount rates

(notice that for an order to be deemed as low profitable not necessarily all of this atomic conditions have to be completely fulfilled),

- the ordered amount (global, resulting from all orders) of the product is *much greater than* that now available in stock,
- the order is placed by a certain customer (e.g., Welli) with whom we have experienced some problems in the past,
- an employee responsible for the given order (e.g., labeled 1) has had recently a few unsuccessful transactions.

Obviously the list of factors that should be taken into account depends on a given case and can be much longer. A requirement that all of the above difficulties (conditions) occur in full scale is often unreasonable. One can claim that an order fulfilling, for example, *most* of the conditions listed above, may surely be treated as potentially troublesome.

One can easily recognize in the above formulation many fuzzy elements, which cannot be directly dealt with using, e.g., SQL. We will focus on linguistic quantifiers and possibility to exploit the OWA operators in the context of this example, but we should mention that FQUERY for Access supports all the fuzzy elements appearing in this example query, i.e. fuzzy values (e.g., *short*, *low*, *high*), fuzzy relations (e.g., *much greater than*) and fuzzy linguistic quantifiers (e.g.,*most*).

Let us assume, that with the subcondition (9) in our example is associated the linguistic quantifier 'most' in sense of Zadeh (1983) and we will use the same quantifier as the global linguistic quantifier for the whole query. This quantifier will be defined by the membership function shown in Exhibit 1.

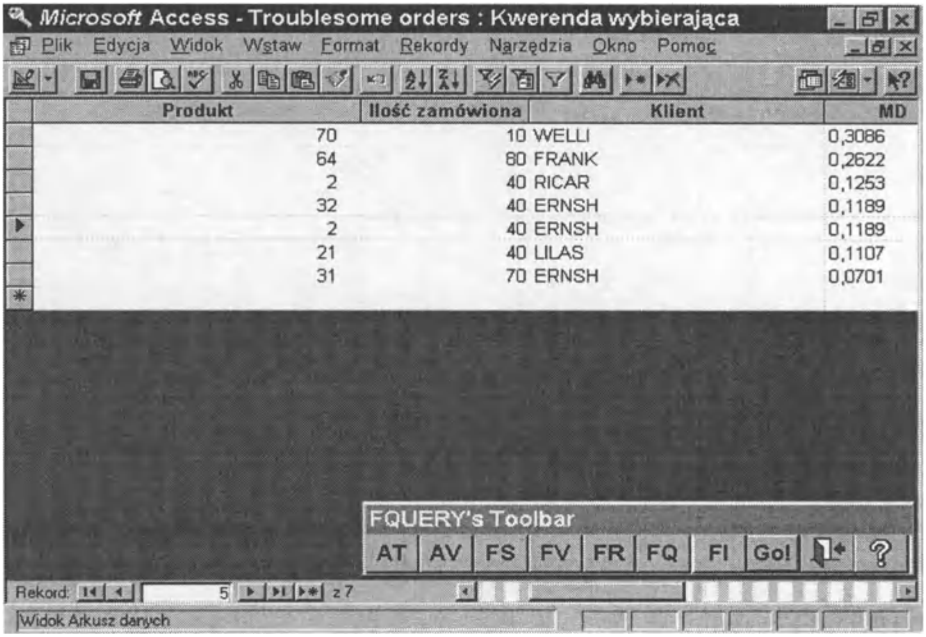


Exhibit 4 Results of querying

We will omit here the details of construction of the corresponding query - a part of it is illustrated in Exhibit 3. This exhibit shows different elements of the query put by the user using the toolbar shown at the top of the screen. After pressing the GO button, another screen is displayed where the user can set some parameters of the query. He or she may go, for instance, to the screen of fine tuning of the OWA shown in Exhibit 3. Then, the query may be executed and, assuming no weights of the OWA operators are changed, the results will be displayed as shown in Exhibit 4.

Now, let us assume, that the user decides to modify the global quantifier and increase its ORness to 0.392. It may be done easily using the scrollbar at the bottom of the screen shown in Exhibits 3 and 5. After such a modification more records will match the query as the criteria have been made weaker. Exhibit 5 shows the new weights generated automatically by the system. The dispersion has grown along with ORness.

5. CONCLUDING REMARKS

We have presented an implementation of the OWA operators in FQUERY for Access. The use of OWA operators makes it possible to essentially enhance the querying capabilities of the package allowing for the application of non-standard logical connectives. It is especially useful in the case of compound queries comprising a number of simple conditions. Combined with the use of fuzzy (imprecise) descriptions they do enhance the human frindliness and human consistence of database querying.

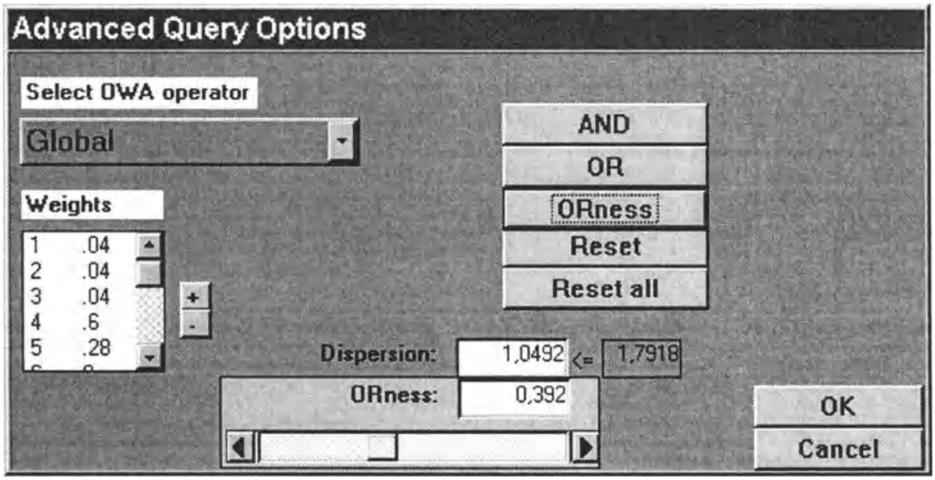


Exhibit 5. Modified weights of the OWA operator

The presented implementation suggests some possible solutions for the representation and manipulation of OWA operators in the querying. The OWA operators may be useful even when none of other fuzzy elements supported by the package are used in a query. As they may effectively replace AND and OR connectives, they may be used for an easy reconstruction of parts of the classical query formed by a query-by-example-like interface.

LITERATURE

Bordogna G., P. Carrara and G. Pasi (1995) Fuzzy approaches to extend Boolean information retrieval. In: P. Bosc and J. Kacprzyk (Eds.) Fuzziness in Database Management Systems, Physica-Verlag, Heidelberg, pp. 231 - 274.

Bosc P. and O. Pivert (1992) Fuzzy querying in conventional databases. In L.A. Zadeh and J. Kacprzyk (Eds.): Fuzzy Logic for the Management of Uncertainty. Wiley, New York, pp. 645-671.

Kacprzyk J. (1995) Fuzzy logic in DBMSs and querying, in N.K. Kasabov and G. Coghill (Eds.): Proceedings of Second New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems (Dunedin, New Zealand), IEEE Computer Society Press, Los Alamitos, CA, USA, pp. 106 - 109.

Kacprzyk J. (1996a) Database management systems: next popular fuzzy-logic-based products?, Proceedings of SIC'96 - International Panel Conference on Soft and Intelligent Computing (Budapest, Hungary), pp. 151 - 156.

Kacprzyk J. (1996b) Fuzzy logic for human-consistent database management systems, Proceedings of FUZZY'96 - Fuzzy Logic in Engineering and Natural Sciences (Zittau, Niemcy), str. 67-73, 1996.

Kacprzyk J. and S. Zadrożny (1994a) Fuzzy querying for Microsoft Access. Proceedings of the Third IEEE Conference on Fuzzy Systems (Orlando, USA), Vol. 1, pp. 167-171.

Kacprzyk J. and S. Zadrożny (1994b) Fuzzy queries in Microsoft Access: toward a 'more intelligent' use of Microsoft Windows based DBMSs, Proceedings of the 1994 Second Australian and New Zealand Conference on Intelligent Information Systems - ANZIIS'94 (Brisbane, Australia), pp. 492-496.

Kacprzyk J. and S. Zadrożny (1995a) FQUERY for Access: fuzzy querying for a Windows-based DBMS. In: P. Bosc and J. Kacprzyk (Eds.) Fuzziness in Database Management Systems, Physica-Verlag, Heidelberg, pp. 415 - 433.

Kacprzyk J. and S. Zadrożny (1995b) Fuzzy queries in Microsoft Access v. 2, Proceedings of 6th International Fuzzy Systems Association World Congress (Sao Paolo, Brazil), Vol. II, pp. 341 - 344.

Kacprzyk J. and S. Zadrożny (1996) A fuzzy querying interface for a WWW-server-based relational DBMS, Proceedings of IPMU'96 - 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (Granada, Spain), 1996, vol. 1, pp. 19 - 24.

Kacprzyk J. and S. Zadrożny (1997) Fuzzy queries in Microsoft Access v. 2, in D. Dubois, H. Prade and R.R. Yager (Eds.): Fuzzy Information Engineering - A Guided Tour of Applications, Wiley, New York, 1997, pp. 223 - 232.

Kacprzyk J., Zadrożny S. and Ziółkowski A. (1989) FQUERY III+: a 'human consistent` database querying system based on fuzzy logic with linguistic quantifiers. Information Systems 6, 443 - 453.

Kacprzyk J. and Ziółkowski A. (1986a) Retrieval from databases using queries with fuzzy linguistic quantifiers. In H. Prade and C.V. Negoita (Eds.) Fuzzy Logics in Knowledge Engineering. Verlag TÜV Rheinland, Cologne, pp. 46-57.

Kacprzyk J. and Ziółkowski A. (1986b) Database queries with fuzzy linguistic quantifiers. IEEE Transactions on Systems, Man and Cybernetics SMC - 16, 474 - 479.

Kacprzyk J. and Ziółkowski A. (1986c) Database querying schemes with fuzzy linguistic quantifiers. Proceedings NAFIPS `86, New Orleans, LA, USA.

Yager R.R. (1988) On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Transactions on Systems, Man and Cybernetics, 18, 183-190.

Yager R.R. (1994) Interpreting linguistically quantified propositions. International Journal of Intelligent Systems 9, 541-569.

Yazici A., R. George, B.P. Buckles and F.E. Petry (1992) A survey of conceptual and logical data models for uncertainty management. In L.A. Zadeh and J. Kacprzyk (Eds.): Fuzzy Logic for the Management of Uncertainty. Wiley, New York, pp. 607-643.

Zadeh L.A. (1983) A computational approach to fuzzy quantifiers in natural languages. Computers and Maths. with Appls. 9, 149 - 184.

Zadrożny S. and J. Kacprzyk (1995) Fuzzy querying using the 'query-by-example' option in a Windows-based DBMS", Proceedings of Third European Congress on Intelligent Techniques and Soft Computing - EUFIT'95 (Aachen, Germany), vol. 2, pp. 733-736.

Zadrożny S. and J. Kacprzyk (1996) Multi-valued fields and values in fuzzy querying via FQUERY for Access, Proceedings of FUZZ–IEEE'96 - Fifth International Conference on Fuzzy Systems (New Orleans, USA), vol. 2, pp. 1351 - 1357.

Zemankova M. and J. Kacprzyk (1993) The roles of fuzzy logic and management of uncertainty in building intelligent information systems, Journal of Intelligent Information Systems 2, 311-317.

Zemankova-Leech M. and Kandel A. (1984) Fuzzy Relational Databases - a Key to Expert Systems. Verlag TÜV Rheinland, Cologne.

OWA - BASED COMPUTING: LEARNING ALGORITHMS

WITOLD PEDRYCZ
University of Manitoba, Department of Electrical and Computer Engineering
Winnipeg, Canada R3T 2N2
pedrycz@ee.umanitoba.ca

1. Introductory comments

The paradigm of knowledge-based neurocomputing imposes an imperative requirement on the functional elements used in such computational architectures. What has been lacking in standard neurocomputing is an ability of the networks exploited therein to encapsulate all pieces of domain knowledge that are usually available in advance. Any successful symbiosis calls for the satisfaction of several fundamental functional postulates [2]:

- emerging topologies should easily encapsulate any prior and sometimes qualitative or imprecise domain knowledge
- an interpretation of the emerging network needs to be straightforward.

Interestingly enough, the OWA aggregation operation as originally introduced by Yager meets the above postulates. While being somewhat similar to the standard perceptron [3] [6], the OWA operation exhibits a strong logic character and allows us to model a broad spectrum of fuzzy set-operations.

The main thrust of this study is to discuss the OWA operation as a processing unit, analyze its generalizations and study several learning algorithms. We start with a brief introduction to the OWA operation and its generalizations (Section 2). Section 3 is concerned with the learning aspects. Concluding remarks are covered in Section 4.

2. The OWA operation and their generalization

Following the generic definition of the OWA operation, let us recall that it can be constructed as a weighted sum (more precisely convex combination) of some input

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

signals (inputs) $x_1, x_2, \ldots x_n$

$$y = \sum_{i=1}^{n} w_i x_i \tag{1}$$

Both **w** and **x** are elements (vector) distributed in the n-dimensional unit hypercube, $\mathbf{w} = [w_1\ w_2\ \ldots\ w_n]$, $\mathbf{x} = [x_1\ x_2\ \ldots\ x_n]$. The weight vector (**w**) exhibits an auxiliary normalization condition meaning that its coordinates sum up to 1

$$\sum_{i=1}^{n} w_i = 1 \tag{2}$$

One should stress that the arrangement of the inputs of the OWA operator plays an important role in its functioning. The coordinates of **x** are organized in a nonincreasing order,

$$x_1 \geq x_2 \geq \ldots \geq x_n. \tag{3}$$

The selection of a specific form of **w** implies a particular character of the resulting aggregation scheme. Based on that one derives

- $\mathbf{w} = [1\ 0 \ldots 0]$:	maximum
- $\mathbf{w} = [\ 0\ 0 \ldots 0\ 1]$:	minimum
- $\mathbf{w} = [1/n\ 1/n \ldots 1/n]$:	average

(4)

From a functional point of view, the OWA operator forms an n-input one output processing unit, refer to Fig. 1, $y = \mathrm{OWA}(\mathbf{x}, \mathbf{w})$. The aggregation operation (weighted sum) used there makes it very similar to the standard perceptron commonly used in neural networks[6].

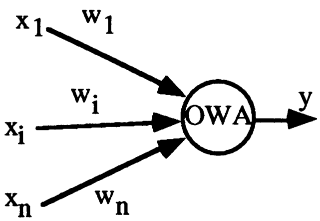


Fig. 1. OWA operator as a many input - single output processing unit

There are two interesting extensions/modifications of the basic OWA operator:

(i) Logical. The logical mutation of the OWA arises through a replacement of the product and summation by some t- and s-norms, respectively. Then we get

$$y = \underset{i=1}{\overset{n}{S}} \left(w_i \, t \, x_i\right)$$

We still keep the arrangement condition on **x** but the other one can be dropped. Now if **w** = [1 0 0 0 .. 0] the result is the minimum. The maximum is obtained for **w** = [0 ... 0 1]. An interesting link can be revealed between this particular version of the OWA operator and a fuzzy integral. Treat now x_i as the values of a fuzzy measure (λ - fuzzy measure) taken over set A_i (we assume that A_is are organized in a decreasing order

$$A_i \supset A_{i+1}$$

so that (3) holds. Thus $x_i = g_\lambda(A_i)$. Now let the values of w_i stand for the values of the function in [0, 1] to be integrated. The sequence **w** is ordered in an increasing order. Specify triangular norms: t- minimum and s-maximum. Then the OWA operator computes the fuzzy integral of **w** over **x**,

$$\mathrm{OWA}(\mathbf{x}, \mathbf{w}) = \int \mathbf{w} \circ \mathbf{x}$$

(ii) Functional. Here the original inputs (**x**) are transformed via a monotonically increasing function (f), say f(**x**). This leads to the expression

$$y = \sum_{i=1}^{n} w_i \, f(x_i) \tag{6}$$

and makes the OWA unit nonlinear. The functional transformation contributes to an enhanced processing flexibility of the original OWA model. In particular, one can concentrate on the power transformation of x, $f(x) = x^p$, $p > 0$. For $p < 1$, f(x) implies a sort of dilution effect. The concentration effect is achieved for $p > 1$.

3. Learning procedures for the OWA processing units

The number of interesting learning scenarios worth developing for the OWA neuron depends vastly upon the form (generality) of available information being used for the training purposes. The form of supervision exercised in learning is distributed between fully supervised and completely unsupervised learning with a series of situations in between. The algorithms of full supervision are usually well-developed and lead to relatively straightforward learning expressions. The learning under limited and partial supervision or without any supervision at all requires much more learning effort. Moreover the final results could be weaker. As before, we consider all inputs being the elements of the training set to be ordered vectors as required in the underlying construct.

3. 1. Supervised Learning

The formulation of the problem hinges on a collection of input-output pairs

$$\{ (\mathbf{x}(1), t(1)), (\mathbf{x}(2), t(2)) \dots (\mathbf{x}(N), t(N)) \tag{7}$$

where $\mathbf{x}(k)$ denotes an input vector while $t(k)$ is an associated output (target). The objective is to minimize a given performance index (objective function) by adjusting the weight vector of the OWA unit ($\mathbf{w}$). Thus we get

$$\min_{\mathbf{w}} Q$$

subject to

$$\sum_{i=1}^{n} w_i = 1$$

and

$$0 \leq w_i \leq 1$$

where Q is often regarded as a sum of squared errors,

$$Q = \sum_{k=1}^{N} \| OWA(\mathbf{x}, \mathbf{w}) - t(k) \|^2 = \sum_{k=1}^{N} (OWA(\mathbf{x}, \mathbf{w}) - t(k))^2 \tag{8}$$

The detailed procedure is very similar to that used in a standard perceptron. The only significant exception is that one has to maintain the constraints on the weights and assure that they are distributed within the unit interval. One among possible solutions to the problem was discussed in Yager and Filev [5]. They considered a nonlinear exponential transformation of the original variables, namely

$$w_i = \frac{\exp(\lambda_i)}{\sum_{i=1}^{n} \exp(\lambda_i)} = 1 \tag{9}$$

where λ_i is any real number. For any value of λ_i, w_i are always in [0, 1]. In virtue of (9) the sum is equal to 1. In this sense the nonlinear transformation leads to an unconstrained optimization problem. Subsequently, the ensuing optimization task requires a series of updates of λ_is so that Q becomes minimized

$$\min_{\lambda_i} Q$$

In this study we follow a different path by admitting the unconstrained gradient-based update scheme

$$w_i\,(\text{new}) = \left[\left[w_i - \alpha \frac{\partial Q}{\partial w_i} \right]\right] \tag{10}$$

i = 1, 2 ... n. The truncation operation is indicated by a double square bracket placed around the expression. This means that after each iteration the values of **w**(new) are reduced to the unit interval. The normalization of the sum of the weights is carried out afterwards.

As an example let us consider a small data set comprising 4 input-output pairs of training data

x(1) = [0.80 0.40 0.20 0.10 0.00] t(1) = 0.50
x(2) = [0.90 0.60 0.50 0.45 0.30] t(2) = 0.70
x(3) = [1.00 0.20 0.20 0.10 0.10] t(3) = 0.52
x(4) = [1.00 0.90 0.50 0.40 0.30] t(4) = 0.75

The learning scheme is given by (10) and the pace of learning is determined by the learning rate $\alpha = 0.15$. The performance index Q decreases rapidly after a few first learning epochs. Fig. 2 (i).

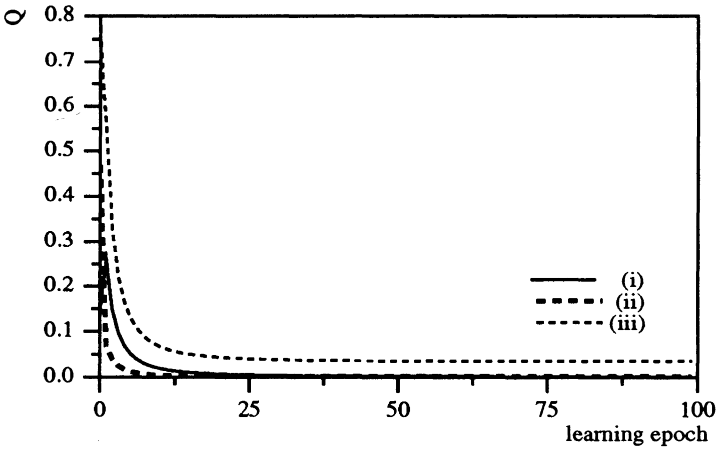
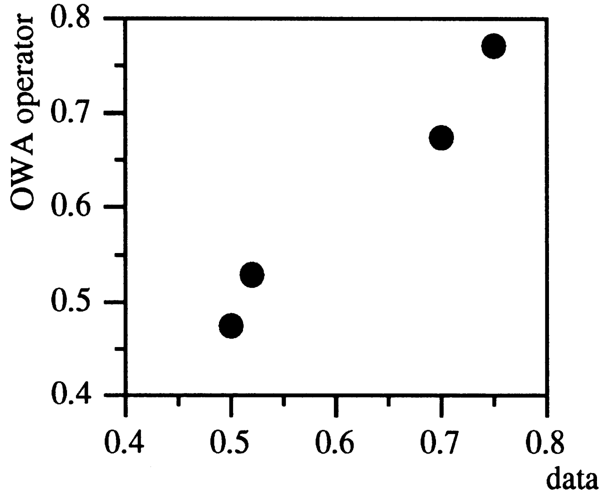
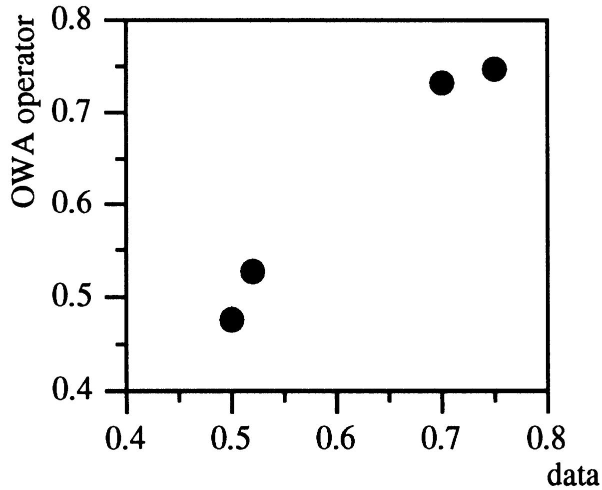


Fig.2. Performance index (Q) in successive learning epochs
(i) OWA without functional enhancement
(ii) functional enhancement, p=1/2
(iii) functional enhancement, p=2.0

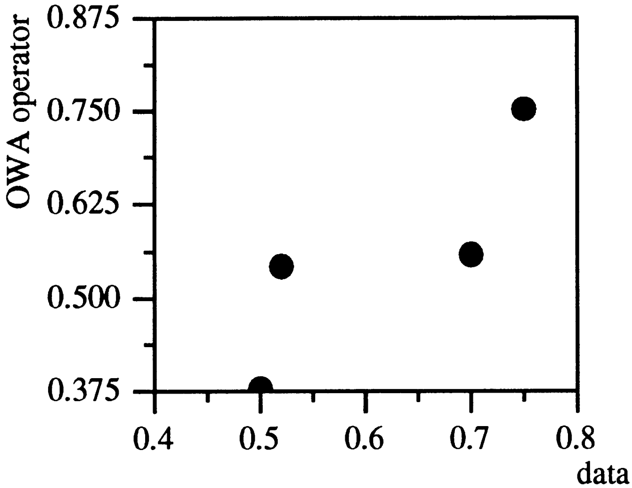
The results are summarized both in terms of the dependencies between the target and results of the OWA unit as well as the weights of the operator itself. The experiments were also conducted for the functional enhancements of the inputs. In the first case we assume a nonlinear dilution transformation of the inputs governed by $x^{0.5}$. The concentration-like transformation comes as x^2. The results of learning viewed through the original target (data) values plotted against the corresponding outputs produced by the OWA unit are summarized in Fig. 3.



(i)



(ii)



(iii)

Fig.3. Target values vis a vis outputs of the OWA operator
(i) OWA without functional enhancement
(ii) functional enhancement, p=1/2
(iii) functional enhancement, p=2.0

The weight vectors derived in each of these situations are specified below no functional modification:

w= [0.434613 0.200477 0.174082 0.121104 0.069724]

functional modification with p=1/2:

w = [0.174157 0.038273 0.662206 0.000255 0.125109]

functional modification with p=2:

w = [0.527576 0.216134 0.123728 0.092783 0.039779]

3. 2. Unsupervised learning-entropy minimization

The element of unsupervised learning emerges when several inputs are not associated with any specific output - this output needs to be determined through learning. One can formulate the problem in many possible ways depending on an assumed form of the performance index.

The one we are interested in is based upon an entropy of the outputs generated by the

OWA element. Let us consider the training data available in the form

$$\{\mathbf{x}(k), t(k)\}\ k \in K, \quad \{\mathbf{x}(l)\}\ l \in L \tag{11}$$

where the first subset (K) is a collection of standard input-output data used for fully supervised learning while the second one is used for unsupervised learning (no outputs available). Then the performance index splits into two parts

- entropy criterion

$$Q_1 = \sum_{l \in L} h\left(\mathrm{OWA}\left(\mathbf{x}(l), \mathbf{w}\right)\right) \tag{12}$$

- standard sum of squared errors taken over elements of K

$$Q_2 = \sum_{k \in K} \|\mathrm{OWA}(\mathbf{x}(k), \mathbf{w}) - t(k)\|^2 \tag{13}$$

The intent is to minimize the sum of these two

$$Q = Q_1 + \beta Q_2 \tag{14}$$

with Q_2 pertaining to the fully supervised learning mechanism and Q_1 (entropy) being responsible for the unsupervised learning. It is worth recalling that the entropy function (h) is defined as a mapping from [0, 1] to [0, 1] such that

(i) $h(0) = h(1) = 0$ (boundary conditions)

(ii) h increases monotonically over [0, 1/2] and decreases monotonically over [1/2, 1] (hence it attains a maximal value at 1/2).

The coefficient (β) used in the above combined performance index takes into account a balance to be reached by the two criteria - mapping and entropy requirement.

As a continuation of the previous example, we now consider that all data are not labelled while L includes only one element; $\mathbf{x}(l) = [0.7\ \ 0.4\ \ 0.2]$ linked with $t(l) = 0.32$.

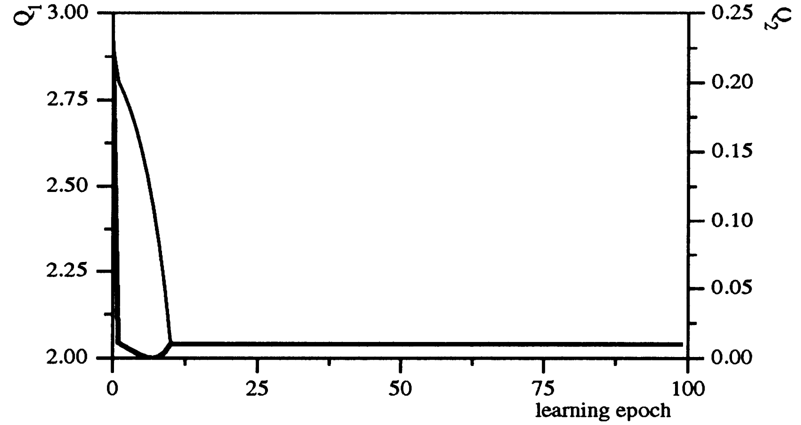
The entropy function (h) is realized as a second order polynomial over [0, 1],

$$h(u) = 4u\,(1-u). \tag{15}$$

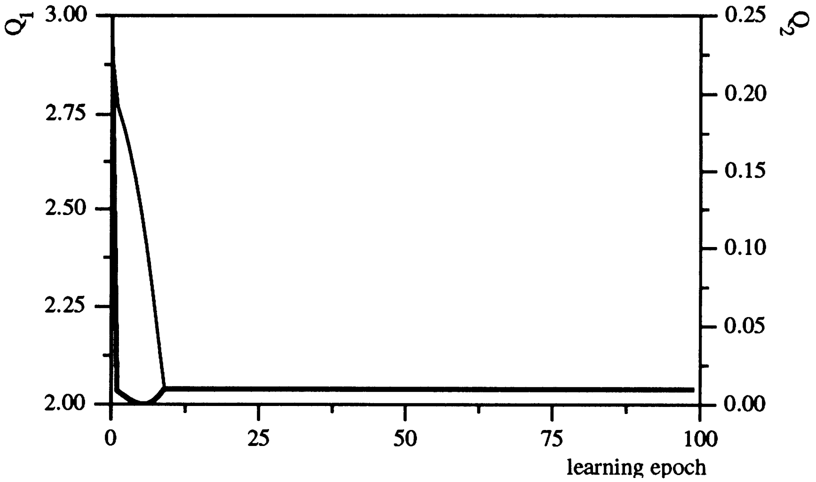
The learning scheme has the same provisions for maintaining the weights in the unit interval as well as making their sum equal 1.

The experiment is completed for $\beta = 1, 2, 5$, and 25.

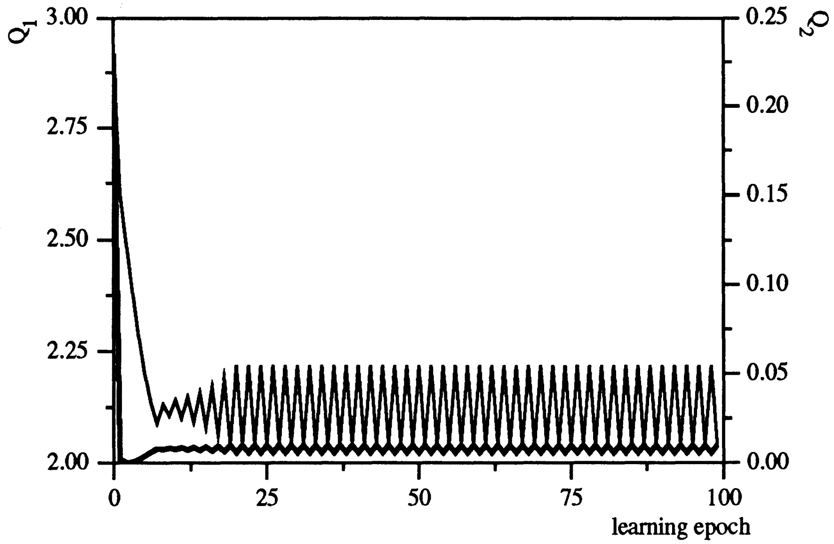
The performance of the algorithm is evaluated in sense of Q yet both Q_1 and Q_2 are reported as well to see to which extent the balance factor (β) affects the behavior of learning. The series of graphs in Fig. 4 reveals an obvious tendency. When the mapping criterion tends to become more important (β increases) then it starts to interfere with the entropy criterion (as these two seem to be in conflict). This results in significant oscillation in the values of Q_1 and Q_2 (even though the overall values of Q eventually go down). Similarly, the weights change their values as summarized in Fig. 5.



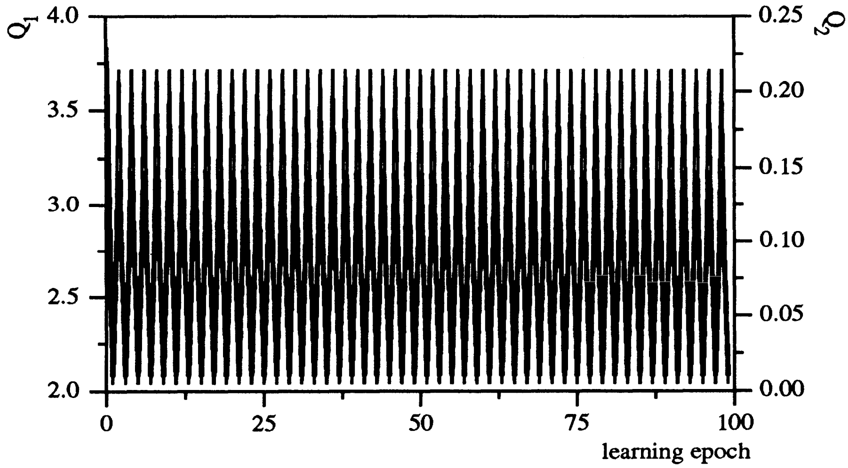
(i)



(ii)



(iii)



(iv)

Fig.4. Performance indices Q1 and Q2 for different values of β
(i) 1.0 (ii) 2.0 (iii) 5.0 (iv) 25.0

The obtained weights of the OWA operator fully reflect the requirement imposed by the assumed trade-off between mapping and entropy criteria. We get the following results

β=1.0:

w= [0.000000 0.000000 0.000000 0.000000 1.000000]

β=2.0:

w=[0.000000 0.000000 0.000000 0.000000 1.000000]

β = 5.0:

w = [0.014505 0.013531 0.007253 0.008195 0.956516]

β = 25.0:

w= [0.340510 0.223557 0.140264 0.107746 0.187922]

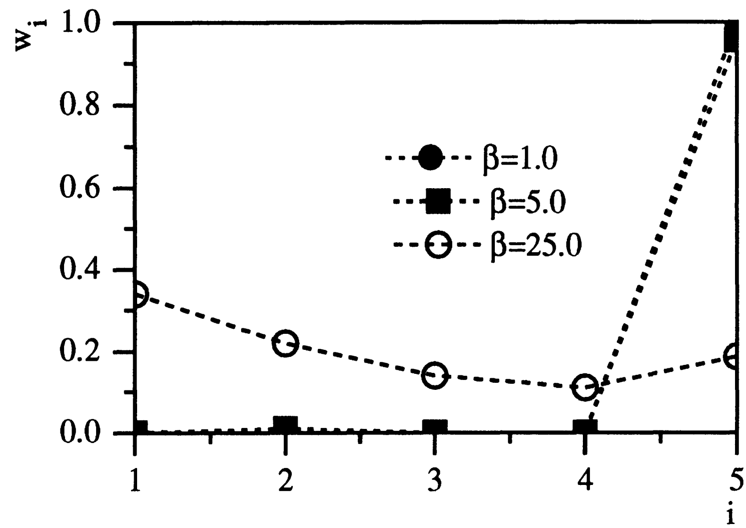


Fig. 5. Weights of the OWA unit for several values of β

One can also discuss the entropy criterion being used to optimize the weights of the OWA operator. This leads to the determination of some logic characteristics of the operator. The pertinent optimization methods were reported in [2].

4. Conclusions

This study was devoted to the learning problems encountered in OWA operators regarded as many - input perceptron - like processing units. We discussed some generalizations of the basic OWA construct and came up with supervised and unsupervised models of learning. The main point one should make is about the arrangement of the inputs of the OWA unit. The assumption in the OWA construct requires that x's are ordered and this is definitely a challenging requirement, especially when it comes to the learning phase. For instance, consider that the vectors are not ordered. We get two pairs of data of this type x(1) = [0.2 0.3 0.7] with t(1) =α and x(2) = [0.7 0.2 0.3] with t(2) = β, '$\alpha \neq \beta$. The arrangement of the inputs results in the same vector

$$\mathbf{x}(1) = [0.7\ \ 0.3\ \ 0.2] \qquad t(1) = \alpha$$

$$\mathbf{x}(2) = [0.7\ \ 0.3\ \ 0.2] \qquad t(2) = \beta.$$

yet the target values (α and β) are different. No training can alleviate this shortcoming as the problem arises at the structural not parametric level that the OWA cannot cope with. This narrows down the range of applications where the OWA operation can be used. In particular, one can envision their full utilization in processing fuzzy measures where the strict ordering comes as their inherent feature. In this sense we can witness some interesting links between fuzzy measures and OWAs at the level of the emerging computational fabric supporting these important ideas.

Acknowledgments

Support from the Natural Sciences and Engineering Research Council of Canada and MICRONET is gratefully acknowledged.

5. References

1. D.P. Filev, R. R. Yager, Learning OWA operator weights from data, *TR MII-1319C* Machine Intelligence Institute, Iona College, 1993.
2. M.O'Hagan, Using maximum entropy - ordered weighted averaging to construct a fuzzy neuron, *Proc. 24th Annual IEEE Asilomar Conference on Signals, Systems and Computers,* Paciific Grove, CA 1990, pp. 618-623.
3. W. Pedrycz (1995), *Fuzzy Sets Engineering,* CRC Press, Boca Raton, FL.
4. B. Widrow, M. E. Hoff, Adaptive switching circuits, *IRE WESCON Convention Record,* 1960, pp. 96 - 104.
5. R. R.Yager, D. P. Filev, *Essentials of Fuzzy Modeling and Control,* J.Wiley, N. York, 1994.
6. J. M. Zurada, *Introduction to Artificial Neural Systems,* West Publishing, St.Paul, MN, 1992.

OWA OPERATORS IN MACHINE LEARNING FROM IMPERFECT EXAMPLES

Janusz Kacprzyk

Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01–447 Warsaw, Poland

E-mail: kacprzyk@ibspan.waw.pl

Abstract

We show how Yager's (1988) ordered weighted averaging (OWA) operators can be employed in (inductive) learning from examples which are assumed to be imperfect in the sense of errors, misclassifications, classifications to a degree, etc. We formulate the problem as to find a concept description covering, say, *almost all* of the positive examples and *almost none* of the negative examples. Thus, by neglecting some examples, those errors are somehow "masked".

Keywords: inductive learning, learning from examples, fuzzy logic, fuzzy lingustic quantifier, OWA operator.

1 Introduction

We consider (inductive) *learning from examples*, i.e. the inferring of a (*concept*) *description* (classification rule, rule, ...) of a class (concept) from *positive* and *negative* examples.

On an inductive learning procedure the two main conditions are susally imposed:

- *completeness*, i.e. that a concept description must correctly describe *all* the positive examples,

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

- *consistency*, i.e. that a concept description must describe *none* of the negative examples,

so that a description is to be found that describes "*all* the positive examples and *none* of the negative ones".

In practice, however, the examples are imperfect in the sense that they contain errors in the values of attributes, misclassifications in the sense of negative examples classified as positive ones and vice versa, non-crisp classifications (to a degree), etc. Clearly, their location (in which examples and attribute values) is usually unknown, and corrections are impossible. This makes the above general problem formulation, i.e. to find a description that describes all the positive examples and none of the negative ones, irrelevant.

A new approach, "softer" was proposed by Kacprzyk and Iwański (1990–1992c) in which a description was sought which described, say:

- *most* of the positive examples and *at most a few* of the negative examples,
- *almost all* of the positive examples and *almost none* of the negative examples, etc.

Notice that such a new formulation somehow neglects the imperfections, if there are not too many of them, because some number of examples (possibly including the imperfect ones) are not accounted for. Basically, if we suspected more errors, we would use "milder" quantifiers as, e.g., *most* and *at most a few*, otherwise we would rather use a more "rigid" quantifiers as, e.g., *almost all* and *almost none.*

Kacprzyk and Iwański (1990–1992c) employed Zadeh's (1983) fuzzy-logic-based calculus of linguistically quantified propositions, and the results obtained were very encouraging.

In this paper we present the use of Yager's (1988) ordered weighted averaging (OWA) operators meant to provide the aggregation corresponding to a wide array of linguistic quantifiers. This was proposed by Kacprzyk (1996), and is further dealt with in this paper.

2 Linguistically quantified propositions and the OWA operators

A *linguistically quantified proposition*, exemplified by, e.g., "*most experts are convinced*", is usually written as

$$Qy\text{'s are } F$$

where Q is a *linguistic quantifier* (e.g., most), $Y = \{y\}$ is a *set of objects* (e.g., experts), and F is a *property* (e.g., convinced).

For our purposes, the problem is to find truth(Qy's are F).

Traditionally, truth(Qy's are F) can be found by using, e.g., the classic Zadeh's (1983) approach. It is simple and efficient, but may lead to unacceptable results for, e.g., fuzzy quantifiers which are not "fuzzy enough".

The OWA (ordered weighted average) operators (Yager, 1988; cf. also Kacprzyk and Yager, 1990) seems to provide an attractive means for handling a wide class of fuzzy linguistic quantifiers.

An *OWA (ordered weighted average) operator* of dimension p is a mapping

$$F : [0,1]^p \rightarrow [0,1] \tag{1}$$

if associated with F is a weighting vector $W = [w_n]^T$ such that: $w_1 \in [0,1]$, $w_1 + \cdots + w_n = 1$, and

$$F(x_1, \ldots, x_n) = w_1 b_1 + \cdots + w_n b_n \tag{2}$$

where b_i is the i-th largest element among $\{x_1, \ldots, x_n\}$.

Then

$$F(x_1, \ldots, x_n) = W^T B \tag{3}$$

If Q is a fuzzy linguistic quantifier (defined as a fuzzy set in [0,1], with monoone nondecreasing membership function), then its corresponding OWA weights are proposed to be (cf. Yager, 1988; Kacprzyk and Yager, 1990):

$$w_k = \begin{cases} \mu_Q(k) - \mu_Q(k-1) & \text{for } k = 1, \ldots, p \\ \mu_Q(0) = 0 & \text{for } k = 0 \end{cases} \tag{4}$$

Other relevant properties of the OWA operators and their relation to fuzzy lingusitic quantifier based aggregation, we refer the reader to other papers in this volume.

For notational convenience, let us generally denote by OWA_Q the aggregation (3) by the OWA operator corresponding to the fuzzy linguistic quantifier Q using (4).

3 Learning from imperfect examples using fuzzy linguistic quantifiers and the OWA operators

Since the conventional general inductive learning problem formulation: find a concept description R which covers *all* the positive examples and *none* of the negative ones is inappropriate in case of imperfect examples, a *softer* problem formulations was proposed by Kacprzyk and Iwański (1990–1992c).

Namely, we seek a concept description R that covers Q^+ (e.g., almost all) of the positive examples and Q^- (e.g., almost none) of the negative examples which may be written as

$$Q^+ \tilde{P}x\text{'s are } R \ \& \ Q^- \tilde{N}x\text{'s are } R \tag{5}$$

where $\tilde{P}$ denotes a *soft* positiveness and $\tilde{N}$ a *soft* negativeness, both to a degree between 0 and 1.

In the new formulation (5) the description is to satisfy the following milder requirement:

- Q^+-*completeness*, i.e. it has to correctly describe (in fact, as well as possible) Q^+ (e.g., *almost all*) of the positive examples,

- Q^--*consistency*, i.e. it has to describe no more than Q^- (e.g., *almost none*) of the negative examples.

To derive a formal formulation of problem (5), we assume Michalski's (1983) general methodology (cf. also Kacprzyk and Szkatuła, 1995, 1996).

A single attribute–value pair in an example is called a *selector*, $[A\ r\ a]$, where: A is an *attribute*, r is a *relation* (e.g., $=, \geq, \ldots$), and a is a *value*. For instance, an example x may be described as

$$x = [\text{height} = 190\text{ cm}][\text{color} = \textit{reddish}][\text{temperature} \gg 100^\circ\text{ C}]$$

i.e. is a conjunction of the selectors.

The value of A_i in a selector and in an example need not be the same as, e.g., s_i = [height = *high*] and x = [height = 190 cm]). We allow therefore for a *degree of identity* of these two values, $\mu_{s_i}(x) \in [0,1]$, from 1 for definitely identical to 0 for definitely different, though all intermediate values.

The fuzzification of a (possibly already fuzzy) value of an attribute which occurs in a selector has much to do with which Q^+ linguistic quantifier is used (for simplicity, Q^- is assumed related to Q^+ in that if, say, Q^+ is *almost all* then Q^- is *almost none*, etc. so that only Q^+ can be dealt with). If we have a "strict" Q^+ (not many errors are suspected), the value should be less fuzzified, otherwise, in the case of a "mild" Q^+ (more errors suspected), the value should be more fuzzified.

Suppose that we have a selector s_i = [height = v'], where v in X is a value (possibly fuzzy) of an attribute in an existing example, and a $Q^+ = (a, b, 1, 1)$, i.e. is given as a trapezoid fuzzy number.

A *fuzzified value* of v is then a trapezoid fuzzy number, too, $\overline{v} = (a', b', 1, 1)$, in which $a' \leq a$ and/or $b' \leq b$.

It is easy to see that if $a = b$, i.e. if Q^+ is like "at least a%", then a single valued v is transformed into an interval, while if $a = b = 1$, i.e if Q^+ is like "all" (no errors are allowed), we obtain that $\overline{v} = v$. In the case of a fuzzy v, we obtain by the above fuzzification a "fuzzier" $\overline{v}$.

By a *complex*, C_j, we mean the conjunction of some selectors, say $s_{j_1}, \ldots, s_{j_k}$, i.e.

$$C_j = s_{j_1} \cap \ldots \cap s_{j_k} \tag{6}$$

The *degree of covering* of x by complex $C_j = s_{j_1} \cap \ldots \cap s_{j_k}$ is defined as, $\forall x \in X$,

$$\mu_{C_j}(x) = \min[\mu_{s_1}(x), \ldots, \mu_{s_k}(x)] \tag{7}$$

The *concept description* R is assumed to be the alternative of the complexes

$$R = C_1 \cup \ldots \cup C_m \tag{8}$$

The *degree of covering* x by $R = C_1 \cup \ldots \cup C_m$ is defined as, $\forall x \in X$,

$$\mu_R(x) = \max[\mu_{C_1}(x), \ldots, \mu_{C_m}(x)] \tag{9}$$

An imprecise classification into the positive and negative examples is represented by:

- a *degree of positiveness* of x, $\mu_{\tilde{P}}(x) \in [0, 1]$, from 1 for definitely positive to 0 for definitely not positive (definitely negative), and
- a *degree of negativeness*, $\mu_{\tilde{N}}(x) \in [0, 1]$, meant analogously as above,

and we assume, for technical reasons, that $\mu_{\tilde{P}}(x) = 1 - \mu_{\tilde{N}}(x)$.

We seek now an (sub)optimal concept description R^* such that

$$\text{truth}([Q^+ \tilde{P}x\text{'s are } R] \& [Q^- \tilde{N}x\text{'s are } R]) \to \max_R \tag{10}$$

i.e. which best satisfies Q^+ of the (fuzzily) positive and Q^- of the (fuzzily) negative examples.

In a more extended form, (10) is to find an R^* such that

$$\max_R[\text{truth}(Q^+ \tilde{P}x\text{'s are } R) \wedge \text{truth}(Q^- \tilde{N}x\text{'s are } R)] =$$
$$= \max_R[\overline{\mu}_{Q^+}(R) \wedge \overline{\mu}_{Q^-}(R)] \tag{11}$$

where:

$$\overline{\mu}_{Q^+}(R) = \text{OWA}_{Q^+}[\mu_{\tilde{P}}(x) \wedge \mu_R(x)] \tag{12}$$

$$\overline{\mu}_{Q^-}(R) = \text{OWA}_{Q^-}[\mu_{\tilde{N}}(x) \wedge \mu_{\tilde{N}}(x))] \tag{13}$$

The concept description sought, R^*, is built up iteratively, adding in each iteration a new complex to the current R, i.e. the number of examples covered by R is not decreasing.

Moreover, by adding the complexes in a special way, $\overline{\mu}_{Q^+}(R)$ is increasing as quickly as possible, while $\overline{\mu}_{Q^-}(R)$ is decreasing as slowly as possible. This makes the algorithm more efficient.

We introduce first the concept of a typoid (cf. Kacprzyk and Iwański, 1990–1992c). Suppose that $x = s_1 \ldots s_n = [A_1 = a_1] \ldots [A_n = a_n]$. Assume that A_i takes on its values in a set $\{a_{i_1}, \ldots, a_{i_q}\}$. A *typoid* is then defined as an artificial example

$$\tau = s_1^*, \ldots, s_n^* = [A_1 = a_1^*] \ldots [A_n = a_n^*] \tag{14}$$

such that each $s_i^* = [A_i = a_i^*]$ is determined by

$$\sum_{x \in X} [(\mu_{\tilde{P}}(x) \wedge (1 - \mu_R(x)) \wedge \\ \wedge \mu_{s_i=[A_i=a_i]}(x)] \longrightarrow \max_{a_i \in \{a_{i_1}, \ldots, a_{i_k}\}} \tag{15}$$

i.e. into τ we put such consecutive selectors which are *most typical* for the examples that are *not covered* by R and are *most positive*, and $\mu_{s_i}(x) \in [0, 1]$ is a *degree of identity* of A_i's value in s_i and in x.

It is important to notice that the selectors taken into consideration in (and then eventually put into) the typoid are also subjected to the fuzzification as formerly described.

The algorithm for determining the R sought is now:

Step 1. To initialize, set:

- **a)** $\mu_{\tilde{P}}(x) \in [0, 1]$, for each example x; evidently, $\mu_{\tilde{N}}(x) = 1 - \mu_{\tilde{P}}(x)$.
- **b)** $R =$ "$\emptyset$" and $C =$ "$\emptyset$" to be meant that the (initial) R contains no complex, and C contains no selectors.

Step 2. $R := R \cup C$, i.e. "add" to the current R a currently formed C, and assume this as the new R.

Step 3. Form a typoid τ as formerly described.

Step 4. Find an example $x^* \in X$ which is both most positive and most similar to τ formed in Step 3, i.e.

$$\mu_{\tilde{P}}(x) \wedge sim(x, \tau) \longrightarrow \max_{x \in X} \tag{16}$$

where $sim : X \times X \longrightarrow [0, 1]$ is some function expressing the *similarity* between x and τ, from 0 for *full dissimilarity* to 1 for *full similarity* as, e.g., $sim(x, \tau) = \frac{1}{n} \sum_{i=1}^{n} \mu_{s_{i^*}}(x)$.

Step 5. Form C as follows:

Substep 5a. To initialize, set $C =$ "$\emptyset$", and $h_{max} = 0$.

Substep 5b. For each $s_i^*, i \in I'$, where I' is the set of indices of the attributes not occuring in C, and s_i^* is the i-th selector of x^* found in **Step 4**, calculate

$$h_i^* = h[\overline{\mu}_{Q^+}(R \cup (C \cap s_i^*)), \overline{\mu}_{Q^-}(R \cup (C \cap s_i^*))] \tag{17}$$

where $h : [0, 1] \times [0, 1] \longrightarrow [0, 1]$ is an averaging operator, e.g., $h(u, w) = (u + w)/2$; and $\overline{\mu}_{Q^-}(.)$ and $\overline{\mu}_{Q^+}(.)$ are given by (12) and (13), respectively, i.e.

$$\overline{\mu}_{Q^+}(R) = \mathrm{OWA}_{Q^+}[\mu_{\tilde{P}}(x) \wedge \mu_R(x)]$$

$$\overline{\mu}_{Q^-}(R) = \mathrm{OWA}_{Q^-}[\mu_{\tilde{N}}(x) \wedge \mu_{\tilde{N}}(x))]$$

Substep 5c. Find $h^* = \max_{i \in I'} h_i^*$, and i^* such that $h^* = h_{i^*}^*$;

Substep 5d. If $h^* > h_{max}$, then:

(1) $h_{max} := h^*$

(2) $C := C \cap s_{i_*}^*$

(3) $I' := I \cup \{i^*\}$

(4) go to **Substep 5b**;

else go to **Step 6**.

Step 6. If

$$\min[\overline{\mu}_{Q^+}(R \cup C), \overline{\mu}_{Q^+}(R \cup C)] > \\ > \min[\overline{\mu}_{Q^+}(R), \overline{\mu}_{Q^-}(R))] \qquad (18)$$

then go to **Step 2**.

Step 7. Output the final R, and STOP.

Example 1 Suppose that we have two attributes, A and B, with values in $\{a_1, a_2\}$ and $\{b_1, b_2\}$, respectively, and four examples $[\mu_{\tilde{N}}(.) = 1 - \mu_{\tilde{P}}(.)]$:

$$\begin{array}{ll} x_1 = [A = a_1][B = b_1] & \mu_{\tilde{P}}(x_1) = 1 \\ x_2 = [A = a_1][B = b_2] & \mu_{\tilde{P}}(x_2) = 0.8 \\ x_3 = [A = a_2][B = b_1] & \mu_{\tilde{P}}(x_3) = 0.5 \\ x_4 = [A = a_2][B = b_2] & \mu_{\tilde{P}}(x_4) = 1 \end{array}$$

For simplicity, assume that the fuzzy linguistic quantifiers are:

- $\mu_{Q^+}(u) = u$, for each $u \in [0, 1]$, which may be viewed to stand for, say, "most", and

- $\mu_{Q^-}(.) = 1 - u$, for each $u \in [0, 1]$, which may be viewed to stand for, say, "(at least) a few",

and for such quantifiers their respective OWA weights given by (4) are obvious.

First, the typoid is

$$\tau = [A = a_1][B = b_1]$$

and by following the consecutive steps of the algorithm, we finally obtain the concept description

$$R^* = [A = a_1]$$

and $\overline{\mu}_{Q^+}(R^*) \simeq 0.78$ and $\overline{\mu}_{Q^-}(R^*) \simeq 0.9$, which are very close to those obtain by using Zadeh's (1983) calculus of linguistically quantified statements (Kacprzyk and Iwański, 1990–1992c). □

4 Concluding remarks

We outlined how the OWA operators, as a means for dealing with the linguistic-quantifier-based aggregation, can be employed in learning from imperfect examples (with errors, missclassifications and non-crisp classifications). The OWA operators provide a wide array of possible aggregation types, and their use in the context considered is simple and efficient.

References

Dietterich T.G. *et al.* (1981) Learning and inductive inference. In P.R. Cohen and E.A. Feigenbaum (Eds.): *The Handbook of Artificial Intelligence*, Kaufmann, Los Altos, 323–525.

Kacprzyk J. (1996) Fuzzy quantifiers and OWA operators in inductive learning from erroneous examples. *Proceedings of FUZZ-IEEE96 – Fifth International Conference on Fuzzy Systems* (New Orleans, USA), vol. 2, pp. 1382 - 1386.

Kacprzyk J. and C. Iwański (1990) Machine learning from misclassified and erroneous examples. *Proc. of IPMU '90 Conference* (Paris, France), pp. 213–215.

Kacprzyk J. and C. Iwański (1991a) Machine learning from incomplete and imprecise examples. In B. Bouchon-Meunier, R.R. Yager and L.A. Zadeh (Eds.): *Uncertainty in Knowledge Bases*, Springer-Verlag, Berlin, pp. 424–430.

Kacprzyk J. and C. Iwański (1991b) Inductive learning under imprecision and errors. In R. Lowen and M. Roubens (Eds.): *Proc. of 4th IFSA '91 Congress* (Brussels, Belgium), Part AI, pp. 78–81.

Kacprzyk J. and C. Iwański (1992a) Fuzzy logic with linguistic quantifiers in inductive learning. In V. Novak, J. Ramik, J. Mares, M. Cerny and J. Nekola (Eds.): *Fuzzy Approach to Reasoning and Decision Making*, Kluwer, Dordrecht and Academia, Praga, pp. 29–38.

Kacprzyk J. and C. Iwański (1992b) Fuzzy logic with linguistic quantifiers in inductive learning. In L.A. Zadeh and J. Kacprzyk (Eds.): *Fuzzy Logic for the Management of Uncertainty*, Wiley, New York, pp. 465–478.

Kacprzyk J. and C. Iwański (1992c) Fuzzy quantifiers in inductive learning with perception of errors in data. *Proceedings of FUZZ-IEEE'92 – First International Conference on Fuzzy Systems* (San Diego, USA), pp. 477–484.

Kacprzyk J. and G. Szkatuła (1995) Machine learning from examples under errors in data, in B. Bouchon-Meunier, R.R. Yager and L.A. Zadeh (Eds.): *Fuzzy Logic and Soft Computing*, World Scientific, Singapore, pp. 31–36.

Kacprzyk J. and G. Szkatuła (1996) An algorithm for learning from erroneous and incorrigible examples. *International Journal of Intelligent Systems* **11**, 565–582.

Kacprzyk J. and R.R. Yager (1985) *Emergency-oriented expert systems: a fuzzy approach. Information Sciences* **37** 147–156.

Kacprzyk J. and R.R. Yager (1990) Using fuzzy logic with linguistic quantifiers in multiobjective decision making and optimization: a step towards more human-consistent models. In R. Słowiński and J. Teghem (Eds.): *Stochastic versus Fuzzy Approaches to Multiobjective Mathematical Programming under Uncertainty*, Kluwer, Dordrecht, pp. 331–350.

Michalski R.S. (1983) A theory and methodology of inductive learning. In R.S. Michalski, J. Carbonell and T. Mitchell (Eds.): *Machine Learning*. Tioga Press, Palo Alto, pp. 83–133.

Yager R.R. (1988) On ordered weighted avaraging operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics* **SMC-18** 183–190.

Zadeh L.A. (1983) A computational approach to fuzzy quantifiers in natural languages. *Computers and Mathematics with Applications* **9** 149–184.

AN APPLICATION OF OWA OPERATORS TO THE AGGREGATION OF MULTIPLE CLASSIFICATION DECISIONS

Ludmila I. Kuncheva[1]
Department of Electrical and Electronic Engineering
Imperial College, Exhibition Road, London SW7 2BT, UK
e-mail: L.Kuncheva@ic.ac.uk

Abstract

The paper considers a classification scheme made up by pooling together multiple classifiers and aggregating their decisions. The individual decisions are treated as degrees of membership assigned by the classifier to the object to be classified. We are interested in how the OWA operators compare to simple voting, linear and logarithmic techniques. In general, all the aggregation schemes appear to be of the same quality, superior to the single classifiers. It was found that OWA operators tend to generalize better than their competitors when the individual classifiers are overtrained. The idea is illustrated on a real and on an artificial data set.

Keywords: Pattern recognition, aggregation of multiple classifiers, committees of networks.

1 Introduction

Aggregating of multiple classification decisions borrows from human decision making the idea of working up a decision by collecting groups, teams, committees of individual experts. We consider a pool of classifiers each of which is supposed to make a classification decision, i.e., to assign the input object to a class from a predefined finite set of mutually exclusive classes. Provided all misclassifications are equally costly, the final goal is to minimize the number of misclassifications over the space containing all possible objects.

We will refer to the classifiers as "experts" and consider a simple scheme of expertise: the individual decisions are produced in one session without any

[1] On leave from CLBME, Bulgarian Academy of Sciences

R. R. Yager et al. (eds.), *The Ordered Weighted Averaging Operators*

communication between the experts. (For an excellent description of this correspondence the reader may refer to the paper by Ng and Abramson, 1992)[1].

Let $x \in \Re^n$ be the $n-$dimensional feature vector describing the object to be classified, and let $\Omega = \{\omega_1, \ldots, \omega_M\}$ be the set of classes. Three types of classifiers are detailed in [2]:

- Type 1 classifiers, the output of a classifier is a single class label. Therefore classifiers implement the mapping:

$$\psi_1 : \Re^n \to \Omega.$$

- Type 2 classifiers, the classifiers' outputs are some rank orderings on Ω, i.e.,

$$\psi_2 : \Re^n \to \mathcal{P}(\Omega).$$

  where $\mathcal{P}(.)$ stands for the class of all permutations of its argument set.

- Type 3 classifiers, each classifier yields some quantitative degrees of membership to each of the classes. Without losing generality we can confine the range of these degrees to the unit interval, so that each classifier implements the mapping:

$$\psi_3 : \Re^n \to [0, 1]^M. \tag{1}$$

Note that instead of $\Re^n$ we can consider any feature space: continuous, discrete, or mixed, because the matter of interest here is the output of the classifier and not the classification mapping itself.

Depending on what output is adopted, a respective aggregation operator can be picked up. The simple voting scheme is applicable to any of the above type of classifiers ([3, 4, 2]). The main result is that given a set of *independent* classifiers whose classification accuracy exceeds 50 %, the more voters we have, the higher the classification accuracy is, the limit being the optimal (Bayesian) accuracy. Voting has been found to perform reasonably well in broad range of settings which, along with its simplicity, makes it an appealing choice.

Comparatively fewer studies address the case of rank orderings (type 2) [5, 6]. The vast majority of studies in aggregation are devoted to the third type of classifiers listed above. The most popular choice has been the **linear** aggregation pool both from heuristic and rigorous statistical point of view [7, 1, 8, 2, 9, 10].

There are many ways to implement the aggregation if we have classifiers of type 3. A consequence of the flexibility of fuzzy set theory is the abundance of aggregation operators that resemble some human decision-making rationale [11, 12, 13]. Choosing one of them for a particular practical problem might appear difficult. Since their introduction by Yager in 1988 [14], the Ordered Weighted Averaging (OWA) operators have been studied thoroughly with respect to their connections with the large family of fuzzy aggregation operators[15, 13]. Those operators incorporate most naturally the simplest aggregation perspectives: pessimistic, optimistic, indifferent, competition jury, etc. Their application however has not been a focus of attention so far.

It is worth mentioning that many catchy names have been used to label the combination of multiple classification decisions: committees of networks [16], adaptive mixture of experts [17], pandemonium system of reflective agents [18], divide-and-conquer methodology [19], change-glasses approach [20], etc. An interesting direction of research stems from the idea that the parameters of the aggregation operator may not be constant over the feature space but vary with the input, expressing some input-dependent characteristics of the experts like "competence", "variance", "confidence" etc. In fact, this leads to a hard or soft partitioning of the feature space, thereby allowing for applying different aggregation schemes on different parts of the feature space [17, 21, 18, 22, 23, 20].

In this paper we are interested in the application of the OWA operators to the aggregation of multiple classification decisions. We will assume that the parameters of the operators are constant over the feature space. Admittedly, an architecture based on variable parameters would outperform the constant-parameters one. Therefore we compare the results with those from the simple voting and the linear aggregation rule. The formal description of the problem is presented in Section 2. Section 3 contains the experimental results on the two-spirals data set. In section 5 some conclusions are drawn.

2 OWA operators for combining multiple classification decisions

We consider a scheme consisting of L individual classifiers. Let $y_i(x) \in [0,1]^M$ denote the output of the ith classifier for the input vector x, $i = 1, \ldots, L$. The aggregated value for the jth class is

$$y^{(j)}(x) = \mathcal{F}\left(y_1^{(j)}(x), \ldots, y_L^{(j)}(x); \theta\right),$$

where the superscript (j) denotes the jth component of the vector, and θ is the set of parameters. We assume that both $\mathcal{F}$ and θ do not vary over the feature space.

Definition. *An L-place OWA operator $\mathcal{F}(a_1, a_2, \ldots, a_L; \theta)$ is defined by the equation:*

$$\mathcal{F}(a_1, a_2, \ldots, a_L; \theta) = \theta_1\, b_1 + \theta_2\, b_2 + \ldots + \theta_L\, b_L$$

where b_i is the ith largest element in the collection $a_1, \ldots, a_L$, and $\theta = [\theta_1 \ldots \theta_L]^T$ is a parameter vector associated with $\mathcal{F}$, such that

1. *θ_i are nonnegative;*
2. *$\sum_{i=1}^{L} \theta_i = 1$.* □

It has been pointed out that OWA operators easily represent some of the widely used aggregation operators by the following parameter vectors

- minimum

$$\theta_{\min} = [\,0\,0\,\ldots\,0\,1\,]^T$$

- maximum

$$\theta_{\max} = [\,1\,0\,\ldots\,0\,0\,]^T$$

- average

$$\theta_{average} = \left[\frac{1}{L}\,\frac{1}{L}\,\cdots\,\frac{1}{L}\right]^T$$

- competition jury

$$\theta_{jury} = \left[0\;\frac{1}{(L-2)}\;\cdots\;\frac{1}{(L-2)}\;0\right]^T \tag{2}$$

We will also use an OWA vector that can be viewed as a model of the linguistic quantifier "most". The interpretation is intended for the case where a linguistic expression is assigned to the aggregation operation, e.g. "Most experts agree on class ω_2".

- "most"

$$\theta_{\text{"}most\text{"}}(k) = \frac{k^2}{\sum_{i=1}^{L} i^2}, \quad k = 1, \ldots, L. \tag{3}$$

OWA operators possess the necessary features to be considered as an appropriate option for aggregation of classification decisions: they are monotonic on their arguments and idempotent. Note that we can assign an individual OWA vector for each class decision. By choosing the desirable linguistic expression we can either favor or neglect the class to a certain degree, and therefore move the classification boundary in the respective direction. This can be used as an alternative way to express the cost of the decision than by using a lost matrix.

If the classifiers' outputs are interpreted as posterior probabilities, aggregation by operators with monotonically decreasing coefficients $\theta_1 > \theta_2 > \ldots > \theta_L$ (e.g.,"most", maximum, etc.) favors more "confident" classifiers and punishes "skeptics". A scheme like "jury" (2) prevents rating high "confident ignorants" among the members of the committee. It is not clear in advance which of the OWA schemes will fit best the problem. This leads to the idea of estimating the OWA coefficients from the training data [24] instead of fixing them in advance. For this purpose we can use the nonnegative least square method on the sorted outputs and further rescaling so that the coefficients sum up to one. Indeed, this estimating procedure involves some imprecision (due to rescaling) but it is interesting to see whether some interpretable OWA shape can be picked up.

The competing aggregation techniques considered here are the simple voting, the plain logarithmic pool, and the weighted linear combination of classifiers' outputs with the least squares estimate of the coefficients on the training data.

3 Experiments

3.1 The two-spirals data set

The two-spirals data set is a challenging benchmark artificial problem for testing out classification techniques. It consists of two classes of 2-D vectors disposed as intertwined spirals (Fig. 1). The training and the test set, as provided initially are nearly the same which does not provide any space for testing the generalization performance of the classifier. Since we are interested in the generalization of the OWA aggregation of the individual decisions, we added random Gaussian noise to the training data with mean 0 and standard deviation 0.5 keeping the test set as originally designed.

The first-level classifiers ("experts", members of the committee) used in this study were radial-basis function (RBF) networks build on random subsets of the training set (contaminated with noise). For all experiments the **neural network toolbox** of `matlab` was used. In fact, the concrete implementation of the first-level classifiers is of no interest, provided they perform the mapping (1) and are trained independently, in the same experimental setting.

A series of experiments has been carried out showing similar results. An example is picked up to illustrate the better generalization performance of OWA classifier over its competitors. Admittedly, the most expressive example has been chosen. The tendency however was clearly presented in the rest of experiments.

Seven classifiers have been used, each one based on 29 nodes. The average of the classification accuracy on the training set of those classifiers was 62.87 %, and on the test set 58.04 %. We have deliberately selected the parameters of

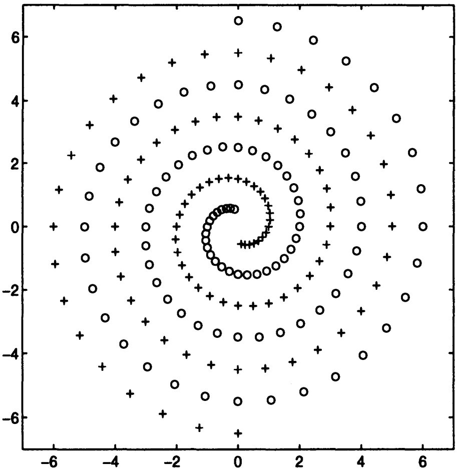


Figure 1:

the first-level classifiers so that the training would not lead to good results on the training set, and besides, we let for some overtraining. This has been done in order to study both the ability of the committee of classifiers to get better results than the single classifiers, and to check its generalization.

Figures 2 and 3 show the results from the example. The output of the combined classifier is plotted on each figure with the intensity of the gray level corresponding to the value of the function, 0 being presented by white, and 1 by black. The points from the original training set (before adding noise) which ideally should be assigned value 1 by the classifier is also depicted for reference. The classification accuracy on the test set using the respective aggregation technique shown in the brackets.

By "OWA" we denote the OWA aggregation with coefficients estimated by the nonnegative least squares algorithm (function `nnls` from `matlab`) on the *sorted* classifiers' outputs, and by "Linear" – weighted aggregation with coefficients estimated by the same function on the outputs.

For comparison, the same classification technique (RBF network) was run until the error goal was reached. This caused huge overtraining with training accuracy 100 %, and test one 49.48 %. The final result from this experiment is shown in Fig. 3. Clearly, the best choice in this example is the OWA aggregation scheme.

3.2 The "heart" data set

The second data set is taken from the database PROBEN1, ftp address:

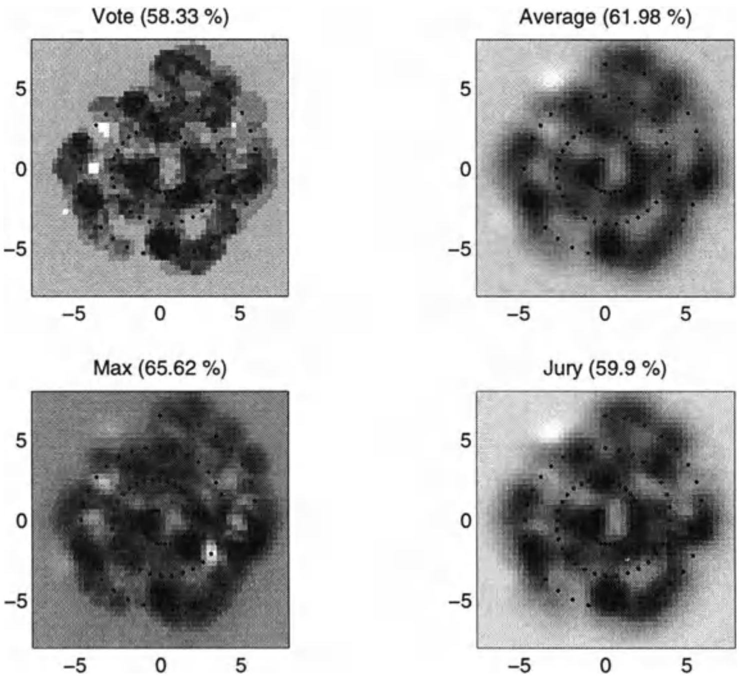


Figure 2:

ftp://ftp.ira.uka.de/pub/neuron/proben1.tar.gz.

A detailed description can be found in the Technical Report by Prechelt[25]. The data set has been supplied by Dr. Robert Detrano, V.A. Medical Center, Long Beach and Cleveland Clinic Foundation.

The data set consists of 303 patient records containing mixed variables (both continuous-valued and binary) taken from patient's history, clinical examinations, laboratory tests, etc. Two classes are considered depending on whether or not at least one of four major coronary blood vessels of a patient is reduced in diameter by more than 50 %.

The data set is called here **heart**, and the three partitions (heart1, heart2, and heart3) of the set into training and test parts are the same suggested by Prechelt[25]. Each partition comprises 228 training samples and 75 test ones.

The main task in this study has not been to achieve the best possible accuracy on the data set but to investigate the generalization abilities of some classification paradigms. Therefore no effort has been made to select optimal parameters with respect to the classification performance and the results are not supposed to be competing with those reported elsewhere.

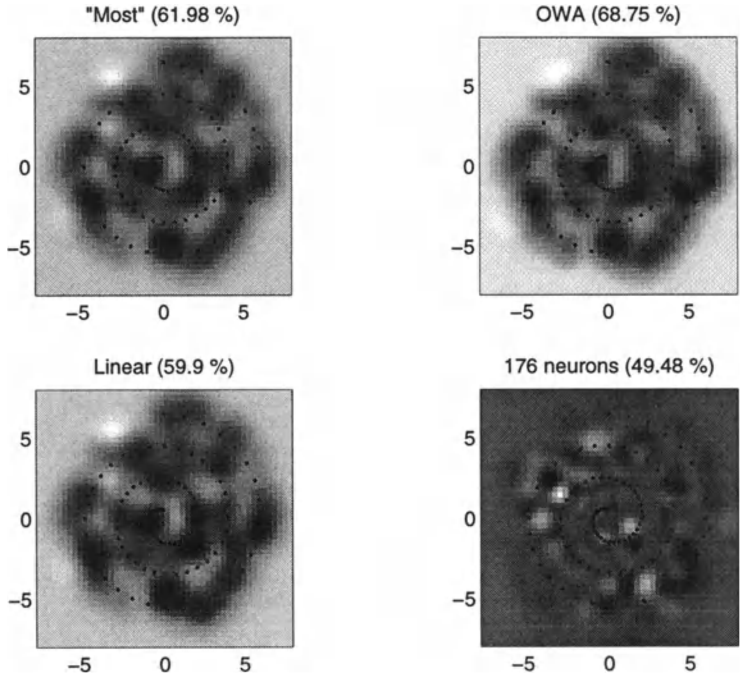


Figure 3:

After discarding some features with zero variance over the data set, the resultant feature set was reduced to 18 features. The fast back-propagation training algorithm from **`neural network toolbox`** was employed for different configurations of a multi-layer perceptron with one hidden layer. In each experiment, the first level classifiers used the same number of hidden nodes.

Tables 1 to 4 show the test results for the classification paradigms under consideration.

Five multi-layer perceptron networks (one hidden layer) have played the role of the "experts" in the committee. Four series of experiments have been carried out:

\# 1 Every classifier has 2 hidden nodes. Randomly chosen 50 % of the *training* data set is used for training the classifier;

\# 2 Classifiers have 2 hidden nodes each, the whole training data set is used;

\# 3 Classifiers have 20 hidden nodes each, 50 % of the training set is used;

\# 4 Classifiers have 20 hidden nodes each, the whole training set is used.

Classification paradigm	heart 1	heart 2	heart 3
Individual (average)	0.7013 (0.0604)	0.8285 (0.0488)	0.7483 (0.0274)
Individual (maximum)	0.7373 (0.0267)	0.8573 (0.0548)	0.7693 (0.0388)
Voting	0.7573 (0.0392)	0.8893 (0.0367)	0.8053 (0.0143)
Average	0.7733 (0.0178)	0.8853 (0.0328)	0.8080 (0.0275)
Linear	0.7653 (0.0180)	0.8880 (0.0275)	0.8067 (0.0261)
Logarithmic	0.7747 (0.0183)	0.8853 (0.0309)	0.8053 (0.0322)
Maximum	0.7813 (0.0220)	0.8920 (0.0231)	0.7987 (0.0317)
Jury	0.7707 (0.0207)	0.8853 (0.0394)	0.8053 (0.0303)
"Most"	0.7747 (0.0160)	0.8853 (0.0328)	0.8040 (0.0345)
OWA	0.7773 (0.0126)	0.8947 (0.0231)	0.8160 (0.0325)

Table 1: Test results from experiment #1: classification accuracy and standard deviations

Classification paradigm	heart 1	heart 2	heart 3
Individual (average)	0.7477 (0.0310)	0.8824 (0.0518)	0.8077 (0.0255)
Individual (maximum)	0.7547 (0.0220)	0.9053 (0.0193)	0.7947 (0.0351)
Voting	0.7707 (0.0084)	0.9093 (0.0105)	0.8267 (0.0166)
Average	0.7667 (0.0094)	0.9093 (0.0138)	0.8307 (0.0227)
Linear	0.7680 (0.0143)	0.9093 (0.0151)	0.8040 (0.0295)
Logarithmic	0.7667 (0.0094)	0.9067 (0.0126)	0.8320 (0.0220)
Maximum	0.7680 (0.0093)	0.9120 (0.0093)	0.8267 (0.0208)
Jury	0.7667 (0.0094)	0.9120 (0.0129)	0.8280 (0.0147)
"Most"	0.7667 (0.0094)	0.9067 (0.0126)	0.8320 (0.0220)
OWA	0.7693 (0.0090)	0.9093 (0.0123)	0.8227 (0.0141)

Table 2: Test results from experiment #2: classification accuracy and standard deviations

Classification paradigm	heart 1	heart 2	heart 3
Individual (average)	0.7104 (0.0483)	0.7509 (0.0636)	0.7288 (0.0399)
Individual (maximum)	0.7560 (0.0295)	0.8440 (0.0412)	0.7480 (0.0414)
Voting	0.7520 (0.0322)	0.8293 (0.1189)	0.7947 (0.0357)
Average	0.7613 (0.0231)	0.8773 (0.0439)	0.7920 (0.0383)
Linear	0.7600 (0.0259)	0.8760 (0.0333)	0.7787 (0.0261)
Logarithmic	0.7587 (0.0231)	0.8693 (0.0401)	0.7840 (0.0360)
Maximum	0.7573 (0.0287)	0.8773 (0.0406)	0.7813 (0.0351)
Jury	0.7493 (0.0325)	0.8813 (0.0317)	0.7947 (0.0328)
"Most"	0.7587 (0.0231)	0.8760 (0.0427)	0.7933 (0.0384)
OWA	0.7560 (0.0244)	0.8827 (0.0349)	0.7960 (0.0367)

Table 3: Test results from experiment #3: classification accuracy and standard deviations

Classification paradigm	heart 1	heart 2	heart 3
Individual (average)	0.7595 (0.0272)	0.8859 (0.0100)	0.7763 (0.0103)
Individual (maximum)	0.7733 (0.0208)	0.8747 (0.0157)	0.7773 (0.0227)
Voting	0.7787 (0.0129)	0.9053 (0.0284)	0.7880 (0.0247)
Average	0.7787 (0.0143)	0.9093 (0.0197)	0.7813 (0.0180)
Linear	0.7680 (0.0143)	0.8960 (0.0250)	0.7747 (0.0160)
Logarithmic	0.7800 (0.0130)	0.9093 (0.0197)	0.7800 (0.0181)
Maximum	0.7733 (0.0154)	0.9027 (0.0178)	0.7773 (0.0178)
Jury	0.7787 (0.0129)	0.9053 (0.0255)	0.7893 (0.0197)
"Most"	0.7800 (0.0130)	0.9107 (0.0178)	0.7773 (0.0178)
OWA	0.7733 (0.0126)	0.9040 (0.0186)	0.7760 (0.0138)

Table 4: Test results from experiment #4: classification accuracy and standard deviations

All training sessions have been stopped (without reaching the error goal) after the 200th epoch. Ten experiments have been carried out with each setting and the results presented in the tables are the average and the standard deviation.

The results are divided into three groups:

- One level classification results:
  - average of the constituents of the combined scheme;
  - the test accuracy of the classifier with the highest training accuracy.
- Standard aggregation techniques:
  - voting;
  - average;
  - linear – this corresponds to weighted aggregation where the coefficients are estimated via nonnegative least square function and rescaled afterwards;
  - logarithmic – the aggregation is performed by multiplying the classifiers' outputs.
- OWA group of aggregation techniques:
  - maximum;
  - jury;
  - "most"
  - OWA

  All of the notations are as described in the previous section.

It is interesting to see whether the OWA coefficients estimated from data can be matched to some interpretable profile. Figure 4 shows the averaged OWA profiles with partitions heart1, heart2, and heart3, for experiments # 1 to 4.

The OWA profiles are not expected to be identical for the two classes. It can be seen that there is no clear resemblance between the coefficients' profiles. It should be mentioned that the standard deviations of the coefficients were very large. This shows that for this particular task, the "optimal" OWA coefficients with respect to the least squares have been specific for each experiment.

4 Analysis and conclusions

The results show that OWA operators are of comparable quality to the most widely used aggregation operators of the same group (nonchanging parameters

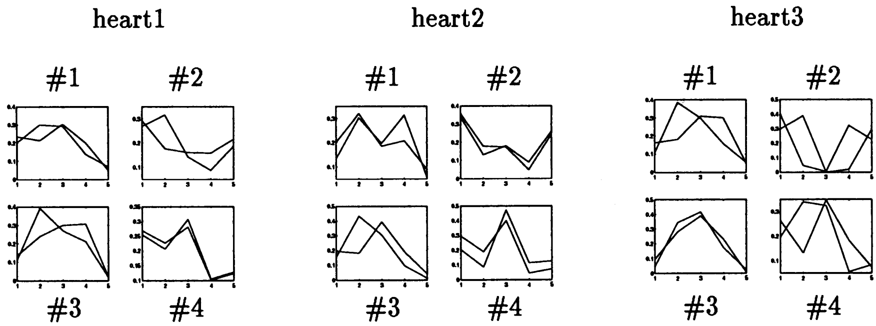


Figure 4: OWA coefficients

and aggregation type over the feature space). A better generalization capability can be expected due to the fact that they treat the classifiers as indistinguishable experts, maintaining only a profile with coefficients on the ordered decisions, rather than attaching coefficients to every expert. The risk in choosing the latter is that an "expert" whose decision fits the training set very well, and in fact overfits it, will be given most credit throughout the whole feature space. On the contrary, using OWA aggregation we do not put our "trust" in one classifier only but let the pool share it.

Indeed, OWA showed better generalization with the two-spiral data set. With the second data set, the difference in favor of OWA group of aggregation techniques is negligible. We may expect that OWA operators will perform better in case of overtrained first-level classifiers.

It is unlikely that the vast majority of practical classification problems will need sophisticated aggregation schemes. Therefore, following Occam's razor principle, the simplest aggregation techniques should be tried first [1]. OWA operators can be labeled as such, having at the same time some intellectually pleasing correspondence with human decision making.

5 Acknowledgements

This work has been supported by a fellowship kindly provided by the Royal Society and the Foreign & Commonwealth Office.

References

[1] Ng, K.-C., Abramson, B.: Consensus diagnosis: a simulation study, *IEEE Transactions on Systems, Man, and Cybernetics* **22** (1992) 916-928.

[2] Xu, L., Krzyżak, A., Suen, C.Y.: Methods of combining multiple classifiers and their applications to handwriting recognition, *IEEE Transactions on Systems, Man, and Cybernetics* **22** (1992) 418-435.

[3] Battiti, R., Colla, A.M.: Democracy in neural nets: voting schemes for classification, *Neural Networks* **7** (1994) 691-707.

[4] Lam, L., Suen, C.Y.: A theoretical analysis of the application of majority voting to pattern recognition, *Proc. 12th International Conference on Pattern Recognition, Jerusalem, Israel* (1994) 418-420.

[5] Ho, T.K., Hull, J.J., Srihari, S.N.: Decision combination in multiple classifier systems, *IEEE Transactions on Systems, Man, and Cybernetics* **16** 66-75.

[6] Tubbs, J.D., Alltop, W.O. Measures of confidence associated with combining classification results, *IEEE Transactions on Systems, Man, and Cybernetics* **21** (1991) 690-692.

[7] Jacobs, R.A.: Methods for combining experts' probability assessments, *Neural Computation* **7** (1995) 867-888.

[8] Benediktsson, J.A., Swain, P.H.: Consensus theoretic classification methods, *IEEE Transactions on Systems, Man, and Cybernetics* **22** (1992) 688-704.

[9] Hashem, S., Schmeiser, B., Yih, Y.: Optimal linear combinations of neural networks: An overview, *Proc. of the IEEE International Conference on Neural Networks, Orlando, Florida* (1994) 1507-1512.

[10] Trsep, V., Tanaguchi, M.: Combining estimators using nonconstant weighting functions, *in: Tesauro G., Touretzky, D.S., Leen, T.K., eds.: "Advances in Neural Information Processing Systems 7", MIT Press, Cambridge MA* (1995).

[11] Dubois, D., H. Prade. A review of fuzzy aggregation connectives, *Information Sciences*, **36**, 1985, 85-121.

[12] Bloch, I. Information combination operators for data fusion: A comparative review with classification, *IEEE Transactions on Systems, Man, and Cybernetics*, **26**, 1996, 52-67.

[13] Grabisch, M. On equivalence classes of fuzzy connectives – the case of fuzzy integral, *IEEE Transactions on Fuzzy Systems*, **3**, 1995, 96-109.

[14] Yager, R.R. On ordered weighted averaging aggregation operators in multicriteria decisionmaking, *IEEE Transactions on Systems, Man, and Cybernetics*, **18**, 1988, 183-190.

[15] Fodor, J. J.-L. Marichal, M. Roubens. Characterization of the ordered weighted averaging operators, *IEEE Transactions on Fuzzy Systems*, **3**, 1995, 236-240.

[16] Bishop, C. Neural Networks for Pattern Recognition, *Clarendon Press, Oxford*, 1995.

[17] Jacobs, R.A., Jordan, M.I. Adaptive mixture of local experts, *Neural Computation* **3** (1991) 79-87.

[18] Śmieja, F. The pandemonium system of reflective agents, *IEEE Transactions on Neural Networks*, **7**, 1996, 97-106.

[19] Chiang C.-C., Fu, H.-C.: A divide-and-conquer methodology for modular supervised neural network, *Proc. IEEE International Conference on Neural Networks, Orlando, Florida* (1994) 119-124.

[20] Kuncheva L.I., Change-glasses approach in pattern recognition, *Pattern Recognition Letters*, **14**, 1993, 619-623.

[21] Alpaydin, E. Combining global vs local perceptrons for classification, *Proc. International Conference on Soft Computing, SOCO'96, Reading, UK*, 1996, B291-297.

[22] Rastrigin, L.A., Erenshtein, R.H.: *Method of Gourd Recognition, Moscow, "Energoizdat"* (1981).(In Russian)

[23] Dasarathy, B.V., Sheela, B.V.: A composite classifier system design:concepts and methodology, *Proceedings of the IEEE* **67** (1979) 708-713.

[24] Filev, D., R.R. Yager. Learning OWA operator weights from data, *Proc. IIId IEEE Conference on Fuzzy Systems, Orlando, FL*, 1994, 468-473.

[25] L. Prechelt, PROBEN1 - A set of neural network benchmark problems and benchmarking rules, *Technical Report* # 21/94 (1994).

Index

S

T

U