An Inductive Fuzzy Classification Approach applied to Individual Marketing

Michael Kaufmann, Andreas Meier

Abstract — **A data mining methodology for an inductive fuzzy classification is introduced. The induction step is based on deriving fuzzy restrictions from data whose membership functions are inferred from normalized likelihood ratios of target class membership. A case study is presented where this data mining process is applied in predictive analytics for an individual marketing campaign in the online channel of a financial service provider.**

Index Terms — **Data Mining, Fuzzy Classification, Fuzzy Logic, Prediction, Individual Marketing, Fuzzy Target Groups**

I. INTRODUCTION AND MOTIVATION

Fuzzy Logic proposed by L. A. Zadeh is based on intuitive reasoning and takes into account human subjectivity and imprecision. Unlike classical data mining techniques such as cluster or regression analysis, fuzzy logic enables the use of classification via fuzzy propositions cominbed by fuzzy logic operators.

Fuzzy classification creates more subtle and smooth distinctions between equivalence classes than crisp classification. [1] and [2] presented the implementation of a classification query language in order to filter data using linguistic terms. They proposed the investigation of a design methodology for the application of fuzzy classification in data mining. In this paper, a data mining methodology using inductive fuzzy classification is introduced in order to address this issue.

[3] emphasizes the inductive aspect of data mining, namely learning, and suggests a motivation: Prediction. This explains how pattern models are to be discovered and for what they can be used. If applied correctly, data models not only describe the past, but also help to forecast the future. In this sense, successful data mining is a practical demonstration of inductive logic: Learning generalizations from specific examples. Hence, this paper presents a methodology for the probabilistic induction of a multivariate fuzzy classification based on likelihood ratios.

M. Kaufmann (corresponding author) is PhD-Student at the Research Center for Fuzzy Marketing Methods (www.fmsquare.org) at the University of Fribourg Switzerland (e-mail: Michael.Kaufmann@unifr.ch).

A. Meier is Professor of Information Systems Resarch. He is the head of the Research Center for Fuzzy Marketing Methods at the University of Fribourg Switzerland.

The paper is structured as follows: Section II introduces inductive fuzzy classification theoretically. Section III proposes a pragmatic methodology for an inductive fuzzy classification. Section IV presents a case study of a database marketing campaign where the target group has been selected using the proposed method. Section V summarizes the results and shows issues for further research.

II. THEORY OF INDUCTIVE FUZZY CLASSIFICATION

This section summarizes the theoretical background of classification, induction and fuzzy restrictions in order to develop a theory of Inductive Fuzzy Classification.

A. Classification

Classification is the process of grouping elements satisfying the same characteristic constraints into a set. In a class logic, a class C = {e | P(e)} is defined as a set C of elements e satisfying predicate P. In the multivariate case, P refers to multiple element attributes as a formula combined by logical operators.

B. Inductive Classification

Inductive classification, a form of supervised learning ([3], [4]), is the process of learning from examples to decide whether an element e belongs to a given class y based on its attributes. A training data set d is a $(n+1){\times}m$ matrix with n columns $X_1, \ldots, X_n$, and a column Y indicating the class membership. The columns X_i, $i \in \{1, \ldots, n\}$ are called dependent variables or attributes, and Y is called the target variable or the class label. The rows represent data elements e_j, $j \in \{1, \ldots, m\}$. The matrix content consists of attribute values x_{ij} representing the value of the i-th attribute for the j-th element, and of labels y_j representing the class membership for the j-th element. In case of a binary classification, for each row index j the label y_j is equal to 1 if and only if the element e_j is in class y.

The learning process induces a model M from the training set d based on the distribution of the target variable on the dependent variables, mapping from the Cartesian product of the dependent variable's domains into the set {0,1}. The model can be used to predict the class membership of data elements. For an element $e'_j \in d'$ in a new dataset d' with unknown class membership, a prediction $Y' = M(x_{1j}, \ldots, x_{nj})$ can be computed by applying the model to the element's attribute values. When elements are grouped into an

equivalence class y' = {e$_j$ | Y'$_j$ = 1}, this kind of classification is inductive because the classification constraints predicate has been learned from the data.

C. Fuzzy Classification

Fuzzy classification is the process of grouping elements into a fuzzy set whose membership function is defined by the truth value of a fuzzy constraint predicate. A fuzzy class C is a fuzzy set [5] defined by a fuzzy proposition [6] 'e is R', where e is an element of a universe of discourse U and R is a fuzzy restriction of U. The degree of membership

$$\mu_C(e) = v(e \text{ is } R)$$

is defined by the truth value v of the corresponding fuzzy proposition.In the multivariate case, P is a fuzzy restriction of U defined by a combination of fuzzy restrictions of attribute domains combined by fuzzy logic operators.

D. Inductive Fuzzy Classification

Consequently, *Inductive Fuzzy Classification* (IFC) is the process of grouping elements into a fuzzy set whose membership function is derived by supervised learning from data. In this approach, the predictive model is represented by a multivariate fuzzy classification. The fuzzy membership functions of each attribute are induced from data using a normalization of the likelihood ratio function. The multivariate aggregation of the univariate fuzzy attribute domain restrictions provides the prediction of the classifier, which can be defuzzified using an alpha-cut. The induction step is not learning of fuzzy rules themselves, as it is in many cases (e.g. [7], [8]) but inducing membership functions to fuzzy restrictions of multiple attribute domains. The focus is not on the combination of fuzzified inputs, but on the definition of multivariate fuzzy set memberships for induction, as presented in [9] and [10]. In fact, by using likelihood ratios for induction, this is a Bayes-like approach, whose usefulness has been shown by [11].

III. A METHODOLOGY FOR INDUCTIVE FUZZY CLASSIFICATION

As proposed by [1], a design methodology is needed in order to apply fuzzy classification to data mining. In the following section, a conceptual framework for inductive fuzzy classification is proposed. The aim is to derive a multivariate fuzzy class inductively in order to obtain a predictive classification. The resulting classifier is able to perform a fuzzy classification which is inductive by learning the fuzzy constraint predicate P from the likelihood of class membership. The resulting fuzzy class is based on a formula of fuzzy propositions on the dependent variables, whose truth values and corresponding fuzzy set membership functions are derived from normalized likelihood ratios.
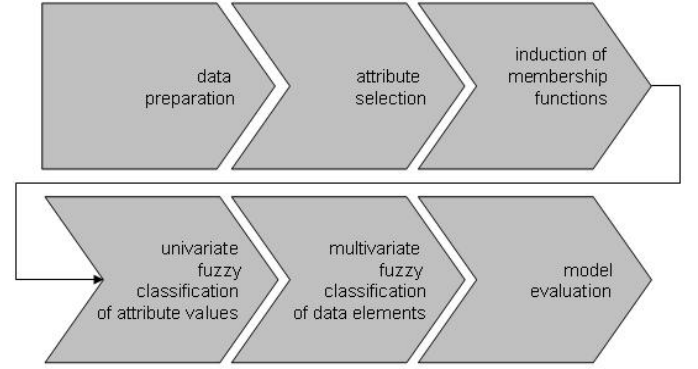

Figure 1: Inductive fuzzy classification process

As shown in Figure 1, the process for deriving a fuzzy class by induction consists of preparing the data, selecting the relevant attributes, inducing membership functions to the fuzzy restrictions on the attribute domains from the training set, transforming the attribute values into univariate fuzzy set membership values, classifying data elements by aggregating the fuzzified attributes into a multivariate fuzzy classification of data elements, and an evaluation of the predictive performance of the resulting classifier.

A. Inductive Fuzzy Classification Process Overview

The idea of the process is to develop a well performing classifier which accurately predicts the membership of data elements e in the target class y. The classifier will perform a fuzzy classification by assigning data elements to a fuzzy class y' having the following property:

The higher the degree of membership of a data element in the fuzzy class y', the greater the likelihood of class membership in the crisp target class y. In order to accomplish this, a training set is prepared form source data, and the relevant attributes are selected using an interestingness measure.

For every attribute X$_i$, a fuzzy restriction y$_i$ on its domain is defined based on the principle of maximum likelihood of target class membership. Each y$_i$ is induced from the data such that the degree of membership of an attribute value x$_{ij}$ in y$_i$ is proportional to the conditional relative frequency of Y=1. Then, in the univariate classification step, each relevant variable X$_i$ is fuzzified by transforming it into a degree of membership in its associated fuzzy restriction y$_i$. The multivariate fuzzy classification step consists of aggregating the fuzzified attributes into one multivariate fuzzy class of data elements y'. This fuzzy equivalence class then corresponds to an Inductive Fuzzy Classification which can be used for prediction. The last step of the process is model evaluation by analyzing the prediction performance of the classifier. This is done by comparing the forecasts with the real class memberships in a test set.

In the following subsections, every step of the IFC process is described in detail.

B. Data Preparation

In order to analyze the data, a training set is composed by combining data from various sources into a single coherent matrix. All possibly relevant element attributes are merged into one table. The class label for the target variable has to be defined, calculated, and added to the training set. The class label is restricted to a binary variable where 1 indicates membership in the target class.

C. Attribute Selection

Only a few variables in the training set are actually relevant for predicting the class label. In order to select the relevant attributes, a variety of measures is available, for example mutual information or chi squared independence. The resulting list is a descending ranking of dependent variables by their relevance. Only the best n variables should be taken into account, where the n-th best variable should still show a significant dependence with the target variable.

D. Induction of Membership Functions

Intuitively, the aim is to assign to every data element a membership degree in the fuzzy data element class y'. In order to accomplish this, y' is derived as an aggregation of fuzzy restrictions: For every attribute X_i, a fuzzy restriction y_i is defined on its domain which represents the likelihood of Y=1. The higher the likelihood of Y=1 is in relation to Y=0, the greater should be the degree of membership of a value x_{ij} in y_i.

The likelihood of class membership is transformed into a membership function using the following method: For each value $x \in dom(X_i)$, the fuzzy restriction y_i is defined as the (sampled) conditional probability for Y=1, divided by the sum of the conditional probabilities of x for Y=1 and Y=0 respectively. Thus, the induced membership degree of x in y_i is defined by

$$\mu_{y_i}(x) := \frac{p(X_i = x \mid Y = 1)}{p(X_i = x \mid Y = 1) + p(X_i = x \mid Y = 0)} \quad (1)$$

According to the definition of likelihood L(y|x)=p(x|y) and likelihood ratio $LR_y(x)=L(y|x)/L(not(y)|x)$, equation (1) corresponds to a normalization of the likelihood ratio (NLR):

$$\mu_{y_i}(x) := \frac{p(X_i = x \mid Y = 1)}{p(X_i = x \mid Y = 1) + p(X_i = x \mid Y = 0)}$$

$$= \frac{1}{1 + \dfrac{p(X_i = x \mid Y = 0)}{p(X_i = x \mid Y = 1)}} = \frac{1}{1 + \dfrac{L(0 \mid x)}{L(1 \mid x)}}$$

$$= \frac{1}{1 + \dfrac{1}{LR_{Y=1}(x)}} =: NLR_{Y=1}(x)$$

For categorical variables, the calculation of the membership function is straight forward: The degree of membership for each value is defined by the corresponding NLR. However, for continuous variables, the membership function could be continuous too. In that case, a solution is to calculate the NLR for quantiles of $dom(X_i)$, and to approximate a continuous function.

For examples of membership function induction for categorical or continuous attributes, see the case study in section IV.

E. Univariate Fuzzy Classification of Attribute Values

Once the membership functions have been induced as described in the previous section, the relevant attributes are fuzzified. In order to transform the attributes into degrees of membership, each variable X_i is transformed by applying the membership function induced by formula (1). This step simply transforms every value x_{ij} into the membership degree of the value in the corresponding fuzzy restriction y_i.

F. Multivariate Fuzzy Classification of Data Elements

In order to obtain a fuzzy membership degree in the fuzzified target class y' for data elements e_j, multiple attribute membership degrees are aggregated using fuzzy logic operators. For simplicity, assume the following equation:

$$\mu_{y_i}(e_j) = \mu_{y_i}(x_{ij})$$

That is, for every fuzzy restriction on an attribute domain there is a corresponding fuzzy set of data elements with the same name. Consequently, the multivariate fuzzy class y' is defined as an aggregation using a function F :

$$\mu_{y'}(e_j) := F(\mu_{y_1}(e_j), \ldots, \mu_{y_n}(e_j))$$

There are different possibilities for multivariate fuzzy aggregation. In literature, the focus is often on generating a fuzzy rule base by searching a function as a combination of fuzzy logic operators.

A simpler candidate for the aggregation function is the gamma operator [12] which is composed of a multiplication of an algebraic disjunction and an algebraic conjunction weighted by an exponent parameter $\gamma \in [0,1]$:

$$\begin{aligned} &\Gamma_\gamma(\mu_{y_1}(e_j), \ldots, \mu_{y_n}(e_j)) \\ &:= \mu_{\bigcup_i y_i}(e_j)^\gamma \cdot \mu_{\bigcap_i y_i}(e_j)^{1-\gamma} \\ &:= (1 - \prod_i 1 - \mu_{y_i}(x_{ij}))^\gamma \cdot (\prod_i \mu_{y_i}(x_{ij}))^{1-\gamma} \end{aligned} \quad (3)$$

This aggregation operator compensates between conjunction and disjunction to a degree specified by the γ-parameter.

Thus, by combining the inductively fuzzified attributes into a multivariate fuzzy class of data elements defined by the membership function $\mu_{y'}: U \rightarrow [0,1]$, a multivariate model is obtained from the training set. This model can be used for Inductive Fuzzy Classification of unlabeled data for prediction. Using an alpha cut for an $\alpha \in [0,1]$ leads to a binary classifier.
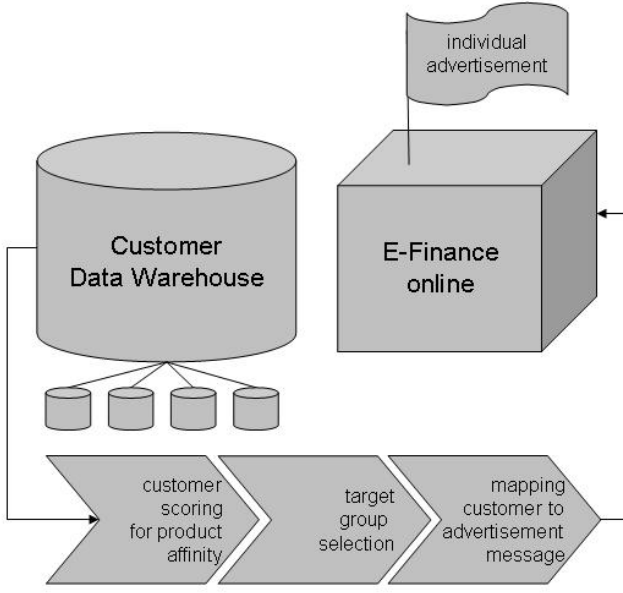
Figure 2: Analytics applied to individual marketing in the online channel

G. Model Evaluation

In order to evaluate the predictive performance, the classifier is applied to a hold-out test set, and the predictions are compared with the actual class label. There are different metrics to evaluate the predictive performance of the induced model, for example the mutual information between the predictions Y' and the class labels Y, or alternatively their correlation. Using such a measure, the model parameters $\gamma \in [0,1]$ and $\alpha \in [0,1]$ can be optimized to yield a maximal predictive performance.

IV. CASE STUDY: MEMBERSHIP FUNCTION INDUCTION FOR FUZZY TARGET GROUPS

PostFinance Inc. (www.postfinance.ch) is a profitable business unit of Swiss Post. Its activities contribute significantly to the financial services market in Switzerland. PostFinance is an analytic enterprise feeding business processes with information gained from predictive analytics. The marketing process uses
data mining scores on product affinity for individual customers. Target groups for online marketing campaigns are selected using inductive classification on customer databases, as illustrated by Figure 2. For every customer, the individual advertisement message is mapped according to the target group selected by predictive scoring.
Inductive Fuzzy Classification was applied in an online marketing campaign of PostFinance promoting investment funds. The aim of the inductive target group selection was to forecast the customers with an enhanced probability of buying investments funds. The resulting fuzzy classifier yielded a fuzzy target group for an individual marketing campaign, where every customer is part of the target group only to a certain degree.

Table 1: Frequencies, conditional probabilities and NLRs for the categorical attribute customer segment.

X_1: customer segment	Y=1	Y=0	$p(X_1 \mid Y=1)$	$p(X_1 \mid Y=0)$	$NLR_{Y=1}(X)$
Basis	11'455	308'406	0.38	0.82	0.32
SMB	249	2'917	0.01	0.01	0.52
Gold	12'666	54'173	0.43	0.14	0.75
Premium	5'432	10'843	0.18	0.03	0.86
Total	29'802	376'339			

In the following sections, the application of the data mining process steps presented in section III to the target group selection is described.

A. Data Preparation

In order to prepare a test set for model induction, a sample customer data set was selected from the customer data warehouse. This data set contained an anonymous customer number associated with customer specific attributes. As target variable, the class label was set to 1 if the customer had investment funds in his product portfolio, and to 0 else.

B. Attribute Selection

Mutual information of the dependent variables with the target variable was chosen as a useful ranking method. The following attributes were selected as relevant:

- Customer segment (0.047 bit)
- Number of products (0.036 bit)
- Overall balance (0.035 bit)
- Loyalty (0.021 bit)
- Customer group (0.016 bit)
- Balance on private account (0.014 bit)
- Age (0.013 bit)

C. Induction of Membership Functions

Using the method presented in section III, for each of the relevant attributes the fuzzy restriction corresponding to the likelihood of having investment funds was induced. In the following section, the induction processes for a categorical and a continuous attribute are described in detail.

As first example, in the domain of the categorical attribute "customer segment" (X_1), there are four values "Basis", "SMB", "Gold" and "Premium". For customers who have not yet bought investment funds, the aim is to define a degree of membership in a fuzzy restriction on the customer segment domain in order to classify them for their likelihood to buy that product in the future.

The frequencies and conditional probabilities are presented in Table 1. The first column indicates the customer segment. Column Y=1 contains the number of customers of each segment who have bought investment funds. Column Y=0 contains the number of customers of each segment who have not bought that product.
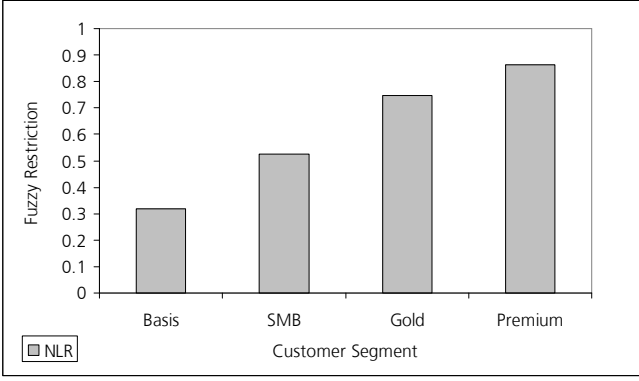
Figure 3: Membership function resulting from induction for the categorical attribute „customer segment"
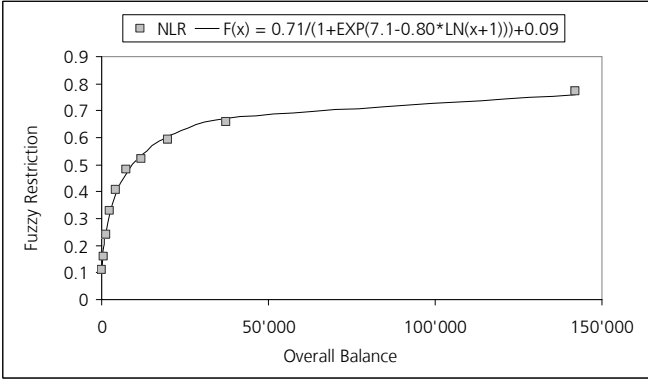

Figure 4: Membership function resulting from induction for the continuous attribute „overall balance"

According to formula (1), the membership degree of segment "Basis" in the fuzzy restriction y_1 was induced as follows:

$$\mu_{y_1}(\text{"Basis"}) = \frac{p(X_1 = \text{"Basis"} \mid Y = 1)}{p(X_1 = \text{"Basis"} \mid Y = 1) + p(X_1 = \text{"Basis"} \mid Y = 0)}$$

$$= \frac{0.38}{0.38 + 0.82} = 0.32 = NLR_{Y=1}(\text{"Basis"})$$

The other degrees of the membership were induced analogically. The resulting values are shown by Table 1 in column NLR. The corresponding membership function is illustrated by Figure 3.

As second example, for the continuous attribute "overall balance" the membership function was induced in the following way: First, the NLR for deciles of the attribute's domain was calculated analogically to the previous example, represented by grey squares in Figure 4. Then, a function A / $(1 + \exp(B - C \ln(x + 1)))+D$ was fitted by optimizing the parameters A to D using the method of [13]. The resulting membership function for the customer overall balance is shown as a line in Figure 4.

D. Univariate Fuzzy Classification of Attribute Values

Every relevant attribute of the original data set was transformed to a fuzzy membership degree using SQL (Structured Query Language) directly in the database.

Table 2: Optimization of the gamma parameter for multivariate fuzzy aggregation

	I(Y;Y')	I(Y;Y'=1)	LR(Y'=1)
Gamma = 0	0.018486	0.4357629	6.7459743
Gamma = 0.5	0.0185358	0.4359879	6.7488694
Gamma = 1	0.0209042	0.4787567	7.2359172

Categorical variables were transformed using a case differentiation statement. For example, the attribute "customer segment" was transformed using the following SQL command:

```
select case
when customer_segment = 'Basis' then 0.31
when customer_segment = 'SMB' then 0.52
when customer_segment = 'Gold' then 0.74
when customer_segment = 'Premium' then 0.86
end as fuzzy_customer_segment
from ads_funds
```

Continuous variables were transformed using a function expression. For example, the attribute "overall balance" was fuzzified using the following SQL expression:

```
select
0.71/(1+EXP(7.1-0.8*LN(overall_balance + 1)))+0.09
as fuzzy_overall_balance
from ads_funds
```

E. Multivariate Fuzzy Classification of Data Elements

The individual fuzzy attribute domain restrictions were then aggregated to a multivariate fuzzy class of customers.

This was done using an SQL statement implementing the gamma operator defined in equation (3). In order to define the gamma parameter, different performance measures were calculated. As shown in Table 2, a gamma of 1, corresponding to a algebraic disjunction or full compensation, was most successful.

Thus, the multivariate fuzzy classification was performed using the following SQL statement:

```
select
(1-fuzzy_number_of_products)
* (1-fuzzy_customer_segment)
* (1-fuzzy_overall_balance)
* (1-fuzzy_loyalty)
* (1-fuzzy_age)
* (1-fuzzy_balance_on_private_account)
* (1-fuzzy_customer_group)
) as multivariate_fuzzy_classification
from ads_funds
```

As a result, a fuzzy class of customers was calculated, whose degree of membership indicates the product affinity for investment funds and represents a product affinity score. This can be used in individual marketing for defining target groups for different products.

Table 3: Resulting product selling ratios per target group

Test group	Y=1	Y=0	Sold products
1: Fuzzy classification	31	4939	0.63%
2: Crisp classification	15	5037	0.30%
3: Random selection	10	5016	0.20%

F.   Model Evaluation

In order to test the resulting fuzzy classifier, a pilot marketing campaign was performed using the resulting fuzzy target group. A target group of 5000 customers with the highest membership degree was selected from the fuzzy class using an alpha cut (test group 1).

As a comparison, 5000 other customers were selected using a crisp classification (test group 2), using the following crisp constraints:

- Customer segment in {Gold, Premium, SMB}
- Customer_group = 50 Plus
- Loyalty > 14
- Number_of_products > 1
- Age between 35 and 75
- Balance on private account > 3000
- Overall balance > 20000

Third, 5000 customers were selected randomly (test group 3). To each of those 15000 customers, an online advertisement for investment funds was shown. After the marketing campaign, the product selling ratio within three months was measured. The results are shown in Table 3. The product selling ratio for the target group selected by inductive fuzzy classification was the most effective.

V.   CONCLUSION

A methodology for an inductive fuzzy classification was introduced. The induction process consists of generating fuzzy restrictions on attribute domains from normalized likelihood ratios.

An individual marketing campaign using fuzzy inductive target group selection showed that fuzzy classification led to a higher product selling ratio than crisp classification or random selection. In the case study, Inductive Fuzzy Classification's predictions of product affinity were more accurate than those of the crisp classification rules on exactly the same attributes. The conclusion is that in comparison to crisp classification, fuzzy classification has an advantageous feature that leads to better predictive performance.

Further research could encompass the following issues:

- Reference implementation of an IFC data mining tool supporting and automating the IFC process steps. An interpreter for an inductive fuzzy classification language (IFCL) is in development by a Master's Student Project [14]
- Evaluation of the IFC data mining methodology using different benchmark data sets in comparison to conventional prediction techniques. Evaluation of the mechanism how fuzzy classification provides more accurate prediction results than crisp classification.
- Development of a conceptual framework for business applications of fuzzy predictive analytics in the context of information systems research.

REFERENCES

[1] A. Meier, G. Schindler, N. Werro. Fuzzy classification on relational databases. In Galindo M. (Ed.): Handbook of Research on Fuzzy Information Processing in Databases. Volume II. Information Science Reference, pp. 586-614, 2008.

[2] Nicolas Werro. Fuzzy classification for online customers. PhD Thesis, University of Fribourg, 2008.

[3] Ian H. Witten and Eibe Frank. Data mining, practical machine learning tools and techniques, second edition. Morgan Kaufmann Publishers, 2005.

[4] Eyke Hüllermeier. Fuzzy methods in machine learning and data mining: Status and prospects. Fuzzy Sets and Systems 156(3), 387-406, 2005.

[5] L. A. Zadeh. Fuzzy sets. Information and Control, 8, 338–353, 1965.

[6] L. A. Zadeh. Calculus of fuzzy restrictions. In: L.A .Zadeh et al., Eds, Fuzzy sets and their applications to cognitive and decision processes. NewYork: Academic Press, 1975.

[7] W. Dianhui, T.S. Dillon, E.J. Chang. A data mining approach for fuzzy classification rule generation. IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th, pp.2960-2964 vol.5, 25-28 July 2001.

[8] Y. Hu, R. Chen and G. Tzeng. Finding fuzzy classification rules using data mining techniques. Pattern Recogn. Lett. 24, 1-3, 509-519, 2003.

[9] C.J. Kim and B.D. Russell. Automatic generation of membership function and fuzzy rule using inductive reasoning. Third International Conference on Industrial Fuzzy Control and Intelligent Systems, pp. 93-96, 1993.

[10] T. Hong and S. Wang. Determining appropriate membership functions to simplify fuzzy induction. Intell. Data Anal. 4, 1 (Jan. 2000), 51-66. 2000.

[11] S. Roychowdhury. Bayes-like Classifier with Fuzzy Likelihood. Fuzzy Systems, 2006 IEEE International Conference on, pp. 431-436, 2006.

[12] H.-J. Zimmermann. Fuzzy set theory and its applications. London: Kluwer Academic Publishers, 1992.

[13] E.G. John. Simplified curve fitting using spreadsheet add-ins. Int. J. Engng Ed. 14, 5 375-380, 1998.

[14] Philippe Mayer. Implementation of an inductive fuzzy classification language interpreter. Master's Dissertation, University of Fribourg, Switzerland (in Progress).