Semantic Context Classification by Means of Fuzzy Set Theory

Yu SUN*, Richard Khoury* Fakhri Karray* and Otman Basir*

*Dept. of Electrical and Computer Engineering
University of Waterloo, Waterloo, Ontario N2L 3G1
Email: {yusun, rkhoury, karray, obasir}@pami.uwaterloo.ca

***Abstract*— Comprehension of semantic meaning is at the heart of modern natural language processing (NLP). Currently, research in statistical NLP has focused primarily on the statistical representation of lexical combinational occurrences. Due to the limitations of current computer technology, however, representing a lexical combination is restricted to a finite length. As such, we focus attention on obtaining approximate but simpler and satisfactory solutions through soft computing techniques; in particular, the fuzzy set theory. It is difficult, however, to apply conventional fuzzy membership functions for general linguistic items. As such, we specifically propose a method of constructing membership functions for linguistic items based on the level of semantic patterns. For testing purpose, the proposed methodology is applied in text classification and the accompanying experimental results are compared with the output provided by a probabilistic based approach.**

I. Introduction

We may agree that natural language is characterized by apparent uncertainty, yet efficiency in understanding. The success of using flexible expressions by writer/reader argues that the modeling of natural language will have to provide some means of fuzzy and dynamic knowledge representation. Pertaining to language processing, fuzzy techniques are associated with imprecision and approximate reasoning. One of successes of fuzzy techniques is that they are able to provide quick, simple and satisfactory solutions in some domains. On the other hand, there has been little fundamental progress achieved in the literature of natural language understanding through fuzzy techniques. Thus, we focus our attention on discovering a methodology of natural language understanding through fuzzy techniques, in particular, on fuzzy inferencing through membership functions. Although ambiguity and vagueness can be intuitively tackled through fuzzy techniques, it is difficult to represent such uncertainties by means of a probabilistic latent variable model that is often continuous. For distance latent terms such as "hot/cold", membership functions have long been developed. The main advantage of this type of fuzzy membership function is that it pertains to probabilistic latent representation of variables in the universe of discourse. Therefore, it can express any counting of information in the universe of discourse. In contrast, it is difficult to construct anything similar for general linguistic items. For example, one of the obvious challenges is how to construct a membership function for the term "company". The issue pertains to how to represent the fuzzy concept. Specifically, our research constructs membership functions based on the complex of syntactic and semantic frameworks - semantic patterns. Accordingly, we introduce one terminology: "template". A "template" might be regarded as "an aspect of an event". A template consists of a group of patterns and each pattern covering a group of homogeneous sentences. The essence of the construction of membership functions is to discover an intermediate pattern to uniquely identify the template, a process similar to discovering the "mean" of a data distribution. The target is to find the "mean" expression for the complete sentences. In the accompanying experiments, we apply the proposed methodology to the problem of context classification and adopt two different approaches to test it: probability measures of fuzzy events and a classical fuzzy inferencing operation through membership functions. Both approaches demonstrate that the classification using constructed trapezoid membership functions is superior to one using a standard probabilistic method.

The remainder of the paper is organized as follows: In Section 2, the classical statistical and fuzzy approaches are reviewed. In Section 3, we focus on introduction of the proposed methodology of constructing membership functions. In Section 4, experimental results illustrate the performance of the proposed approach, followed by comments and discussions in Section 5. Concluding remarks and future work are given in Section 6.

II. Overview and Background

Inspired largely by the work of the current view of language processing, the focus of language processing is *isolated* linguistic items. And the according research has focused primarily on the statistical representation of linguistic occurrences, such as IDF [1] and TF/IDF [2]. The major issue regarding the type of approach pertains to the lack of capability in tackling the difficulties caused by synonymy and polysemy within the English language domain as argued in [3]. In order to decrease the ambiguity of words, one type of semantic approaches took advantage of knowledge from the predefined syntactic patterns and semantic slots [4], [5]. However, they need an extremely large and often hand-coded training corpus. For instance, the 1 million-word Tagged Brown Corpus is considered too small. One corpus that is often used to construct statistical methods is the 4.5 million-word Penn TreeBank. Not

0-7803-9361-9/05/ $ 20.00©2005 IEEE

surprisingly, building a satisfactory tagged corpus by hand is a huge project.

The term "fuzzy logic" was developed along with the theory of fuzzy sets proposed by Lotfi Zadeh [6]. In earlier research on fuzzy techniques as applied to natural language processing, researchers focused on modeling modifier terms in natural language, such as "much", "older", etc. [7], [8], [9], and [10]. In the recent progress of fuzzy techniques, researchers had tried to construct a *fuzzy thesaurus* by employing a fuzzy relation system, such as [11] and [12]. The goal is aimed at providing a global modeling semantic association between linguistic items throughout the universe of discourse. Due to the complexity of natural language, however, such a fuzzy thesaurus only covers a minor part of vocabulary and, as a consequence, does not make much of a contribution in the broader sense to context understanding.

The progress of context manipulation described in these studies has shown that modeling the knowledge of human language is the key. Knowledge then is crucial with regard to the data structure of an ideal semantic lexicon, on how to numerically represent common lexical items, on how to depict the binary and transitive relations between lexical items, and on how to infer the meaning of new utterances. Again, these studies do not paint a clear picture. Thus, these challenges become the main topics of our proposal.

III. Proposed Approach

The proposed approach describes three aspects: how to describe the semantic information of linguistic items by means of a contextual environment, how to depict distance latent relations between items, and how to construct membership functions and inference the meaning of new utterances.

A. How to Describe Linguistic Items

For the purpose of comprehension of semantic meanings, the proposed measure is constructed based on an environment integrating the syntactic and semantic information - semantic patterns. According to our assumptions, a document represents one central event (or topic), each event may be expressed by a diverse array of aspects, each aspect covers several patterns, and each pattern is uniquely identified by a verb. As the earlier introduction, an "aspect" of an event is in particular denoted as a "template", and the template covers a group of homogeneous patters. For example, considering one of aspects, the action of "Take-In", in event *Takeover* in the domain of business English, we have the following defined patterns, each of which is uniquely identified by one verb (Table I). Based on the semantic patterns, one might define a set of vocabularies for template *Take-In*. For example, template "Take-In for Takeover event" has 6 patterns. Of them 2 are listed as follows:

Template **"Take-In for Takeover event"**

Pattern.01

- Pattern: [$_{Arg0}$ Company | Person] [$_{Arg1}$ acquire] [$_{Arg2}$ Company | Person | Financial-Instrument] [$_{Arg3}$ Preposition-Money].

Pattern.02

- Pattern: [$_{Arg0}$ Company | Person] [$_{Arg1}$ bid] [$_{Arg2}$ F-I | Real-Estate | Good] [$_{Arg3}$ Preposition-Money]

Based on the set of vocabularies, we could create a group of triplet, which are achieved by combinatorial organization of terms on left, verb, and terms on right, such as $< arg_0, verb, arg_2 >$ and $< arg_0, verb, arg_3 >$. One might notice that the sequences of these triplets are highlighted, *i.e.* identified as the terms on the left of the central verb and those on the right. In a later stage, this position-sensitive characteristic shows the contribution to construction of membership functions.

B. How to Represent Distance Latent Relations Between Items

Following Rieger [13] with the modification, semantic distances between words are computed through the words' correlational coefficient and are given as follows:

$$\alpha(v_i, v_j) = \frac{\sum_{t=1}^{T}(h_{it} - e_{it})(h_{jt} - e_{jt})}{(\sum_{t=1}^{T}(h_{it} - e_{it})^2 \sum_{t=1}^{T}(h_{jt} - e_{jt})^2)^{\frac{1}{2}}} \qquad (1)$$

where, coefficient $\alpha(v_i, v_j)$ measures the pairwise relativity of $(v_i, v_j) \in V \times V$ (V denotes the universe of vocabulary and legal vocabulary pair. For example, we compute the distance between "company" and "acquire", the distance between "$<_{arg0}$ $company,$ $_{arg2}F - I >$" and "acquire", but not the distance between "$<_{arg0}$ $company,$ $_{arg0}person >$" and "acquire".); T denotes the total number of documents; h_{it} denotes the number of word i appearing in the document t; $H_i = \sum_{t=1}^{T} h_{it}$ denotes the total number of occurrence of word i in all documents; l_t denotes the length of document t which measures the total number of vocabularies appearing in document t; $L = \sum_{t=1}^{T} l_t$ denotes the total length of all documents; Mean e_{it} is defined as $e_{it} = \frac{H_i}{L} l_t$. After defining the correlational coefficient, we compute the similarity of words $\mathrm{Sim}(v_i, v_j)$ measured by Euclidian metric as:

$$\mathrm{Sim}(v_i, v_j) = (\sum_{k=1}^{N}(\alpha(v_i, v_k) - \alpha(v_j, v_k))^2)^{\frac{1}{2}} \qquad (2)$$

Obtaining the similarity of lexical co-occurrence described above, we next investigate the degree of intensity of such

TABLE I: Verbs and Associated Semantic Patterns

Verb	Semantic Pattern
Acquire	{Company | Person} *acquire* {Company | Person | Financial-Instrument} {for | at} Money
Bid	{Company | Person} *bid* {Financial-Instrument | Real-Estate| Good} {for | at} Money
Buy	{Company | Person} *buy* percentage {Company | Person | Financial-Instrument | Good} from {Company | Person} for Money-Price
Invest	{Company | Person} *invest* Money {in | for } {Financial-Instrument | Company | Person}
Purchase	{Company | Person} purchases {number Financial-Instrument | Company | Person | Real-Estate }
Take over	{Company | Person} *take over* {Company | Person} {Financial-Instrument} for Money

co-occurrence associated with each event. Throughout our research, we define the following events: *Takeover*, *Merge*, and *Earning*. In particular, we use conditional probability to describe degrees of intensity:

$$Prob(arg_i, verb|event) = \frac{freq(arg_i, verb, event)}{freq(event)} \quad (3)$$

C. How to Construct Membership Functions

Based on the preceding sections, we are now discovering an intermediate pattern to represent a template. This idea of "intermediate pattern" is similar to the concept of "Mean": a template consists of a group of patterns and each pattern covers a group of homogeneous sentences. Our target is to find the "mean" expression for the complete sentences. Thus, the construction of membership functions is underpinned by the rules finding intermediate patterns.

Let $P_i^{(k)}$, $k = 1, 2$, be the *k-th* pattern of template i (*e.g.* in *Take-In for event Takeover* template, there exist pattern 01 $P_i^{(1)}$ and pattern 02 $P_i^{(2)}$ in this example). Let $X_{i,j}^{(k)}$ be the distance of semantic category S_j from $P_i^{(k)}$ (equivalent the distance of arg_j to the corresponding verb). Without any loss of generality, let us assume that there are only two archetypal patterns $P_i^{(1)}$ and $P_i^{(2)}$ for any template i. Let pattern $P_i^{(g)}$ be the *intermediate* pattern lying between $P_i^{(1)}$ and $P_i^{(2)}$. Given the distribution of $X_{i,j}^{(k)}$ for the two archetypal patterns $P_i^{(k)}$ obtained through the triplets distribution, the challenge is to investigate the expression of $X_{i,j}^{(g)}$ for pattern $P_i^{(g)}$ in template i through $X_{i,j}^{(k)}$. Then the following rules are applied:

- Rule 1: Pattern $P_i^{(g)}$ consists of linguistic items that are involved in either $P_i^{(1)}$ and/or $P_i^{(2)}$;
- Rule 2: compute the sample *Mean* and *Std* for each category S_j involved in pattern $P_i^{(k)}$
- Rule 3: compute the weight ω of each pattern:
  Because the instances of a pattern occur in the corpus with different frequencies, we cannot compute $Mean$ and Std of the intermediate pattern $P_i^{(g)}$ by simply averaging the sum. As such, we shall compute the weight of each pattern $P_i^{(k)}$. As predefined, each verb corresponds to one and only one semantic pattern within one event. Thus, computing the weight of a semantic pattern is equivalent to computing the weight of the corresponding verb in the corpora:

$$\omega_v = \frac{freq(v|e)}{\sum_{\forall \text{ verb } i \text{ in event } e} freq(v_i|e)} \quad (4)$$

- Rule 4: If S_j involved in both $P_i^{(1)}$ and $P_i^{(2)}$ has the same direction of verb[1][14], then the $Mean$ and Std of $X_{i,j}^{(g)}$ is computed through a weighted computation of $X_{i,j}^{(1)}$ and $X_{i,j}^{(2)}$:
  1) given $Mean$ and Std for $X_{i,j}^{(k)}$:
     $Mean_1 X_{i,j}^{(1)}$, $Std_1 X_{i,j}^{(1)}$
     $Mean_2 X_{i,j}^{(2)}$, $Std_2 X_{i,j}^{(2)}$;
     and
     weight ω_1 for pattern $P_i^{(1)}$
     weight ω_2 for pattern $P_i^{(2)}$.
  2) compute $Mean$ and Std of $X_{i,j}^{g}$, we have:

$$y = \sum_{k=1}^{n} \prod_{j=1, j\neq k}^{n} \frac{\omega_g - \omega_j}{\omega_k - \omega_j} y_k, \ \omega_g = \frac{1}{n}\sum_{v=1}^{n} \omega_v \quad (5)$$

- Rule 5: If S_j involved in both $P_i^{(1)}$ and $P_i^{(2)}$ has the contradictory sign, the $Mean$ and Std of $X_{i,j}^{(g)}$ is computed as:

$$Mean_g = \omega_1 Mean_1 \bigcap \omega_2 Mean_2 \quad (6)$$

$$Std_g = \omega_1 Std_1 \bigcap \omega_2 Std_2 \quad (7)$$

- Rule 6: If S_j is involved in only one of $P_i^{(1)}$ and $P_i^{(2)}$, the $Mean$ and Std of $X_{i,j}^{(g)}$ is computed as the corresponding average, that is:

$$\begin{aligned} Mean_g &= \frac{1}{2}\omega_k Mean_k \\ Std_g &= \frac{1}{2}\omega_k Std_k \end{aligned} \quad (8)$$

Based on this rule, we could construct an intermediate pattern $P_i^{(g)}$ to represent the template i. Thus, the membership function for linguistic item S_j might be computed according to its distribution $Mean_g$ and Std_g in intermediate pattern $P_i^{(g)}$, whose form is depicted in Fig. 1, where $0 \leq \mu \leq 1$. In Fig. 1, μ is computed as follows:

$$\mu = \frac{\sum_{k=1}^{T} \Pr((\text{item } S_j, \ \text{verb}_k) \mid \text{event})}{T} \quad (9)$$

where,T denotes the total number of verbs occurring in Template i, whose similarity to the linguistic item S_j, $X_{i,j}^{(k)}$, falls in the boundary from $(M_g - Std_g)$ to $(M_g + Std_g)$. That is, $(M_g - Std_g) < X_{i,j}^{(k)} < (M_g + Std_g)$.

D. How to Infer Meanings of New Utterances

Using the rule base described above, a fuzzy class $F_{i,j}^{(g)}$ for S_j in pattern $P_i^{(g)}$ might be represented by the trapezoid membership function $\mu_{i,j}^{(g)}$ shown in Fig. 1. Let $\Delta_{i,j}^{(g)}$ be the set of classes $F_{i,j}^{(g)}$ that corresponds to pattern $P_i^{(g)}$. Given the input context, represented by a vector $\overline{U} = \{I_1, I_2, \cdots, I_n\}$, where I_k corresponds to k-*th* item in the input context, such as "Company", "Financial-Instrument", etc, it should be noted that I_j is not represented by a symbolic item. Rather, I_j is numerically represented by the similarity of co-occurrence associated with events. Through the input $\overline{U}$, the confidence

[1]NOTE: S_j cannot be treated as involved in both patterns unless it appears with the same subscript of **arg** (in other words, it is **position-sensitive**). For instance, <Arg0: Company> in pattern.01 and <Arg0: Company> in pattern.02 qualify to this rule, while <Arg2: Company> in pattern.01 and <Arg0: Company> in pattern.02 does not.

c_i that corresponds to template i is computed by the following equation:

$$c_i = \bigotimes_{F_{i,j}^{(g)} \in \Delta_{i,j}^{(g)},\ 1 \le j \le n} r_{i,j}^{(g)}$$
$$r_{i,j}^{(g)} = \mu_{i,j}^{(g)}(I_j) \qquad (10)$$

Finally the observed event can be determined by the following equation:

$$\text{Event} = \arg\max_i c_i \qquad (11)$$

where, operator $\otimes$ is $S - norm(max)$ and i indicates three events: *Takeover*, *Merge*, and *Earning* in the accompanying experiment.

IV. Experimental Results

In order to evaluate the proposed methodology, we apply it to the problem of context classification. It is necessary to compare its performance with that of an existing approach under similar conditions. We take probability measures of fuzzy events [15] as the benchmark. We use "Reuters-21578, Distribution 1.0" [16] as our experimental corpora. Reuters-21578 dataset was originally saved in 21 sets, each corresponding to one event and containing up to 1000 documents (natural context without label), with a fixed splitting between test and training data (3,299 *vs.* 9,603) where the 10 top-sized events are selected, including Earn, Takeover, Money-fx, Crude, *etc.* We select three events out of them - "Earning", "Merge", and "Takeover".

A. Experiment on Classification

Based on the similarity and probability outputs by (2) through (3), we use the rules to construct membership functions for the linguistic items. In particular, in this section, we will compare the performance of two methods: probability measures of fuzzy events and the proposed trapezoid membership function.

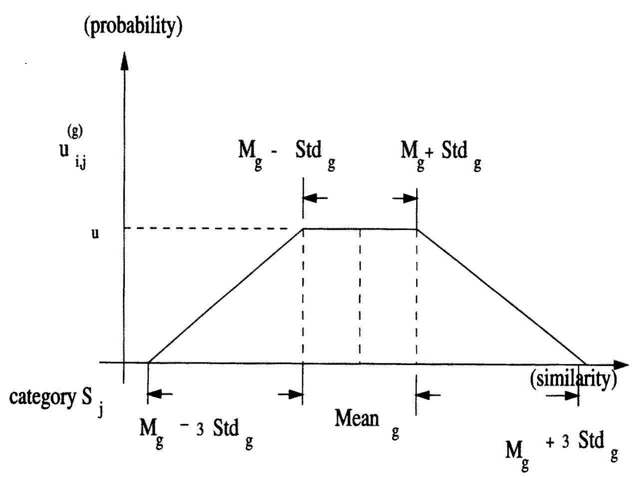

Fig. 1: The Form of Membership Function for a Linguistic Item Associated with the Particular Verb and Event

1) *Classification context by probability measures of fuzzy events*

A business event can be regarded as a fuzzy event modeled by a probability space. We assume that there are N non-repeating linguistic items from c_1 to c_N throughout all business events and the total occurrences of each term in the corpus is d_k, $k = 1, \cdots, N$. Thus, the problem might be modelled by (Ω, k, P), where Ω is the universe of co-occurrences of linguistic terms $\Omega = \{ c_1, \cdots, c_N \}$ and P is the conditional probability measure of the occurrences of terms, *i.e.*,

$$Pr\{c_k|verb\} = \frac{freq(c_k, verb)}{freq(verb)} \qquad (12)$$

The reason why we compute the conditional probability surrounding a verb is due to the structure of triplets of co-occurrence associated with events, *e.g.* <subject, verb, object>. We thus define a fuzzy set S, $S \subseteq \Omega$, for each event, such as *Takeover*. The membership degree, $\mu_S(c_k)$, describes the fuzzy degree of a co-occurrence c_k belonging to event S. In particular, we adopt $prob(co - occurrence|event)$ in (3) as $\mu_S(c_k)$. An event represented by fuzzy notation is denoted as (13)

$$\text{Event} = \sum_{c_k \in S} \mu_S(c_k)/c_k \qquad (13)$$

The ultimate goal is to classify the context according to its central event. The key for achieving this goal is the precise measurement of the closeness of co-occurrences associated with events. By introducing the probability measures of fuzzy events as described by (14), the event of a context can be "computed" accordingly:

$$P(S|verb) = \sum_{k=1}^{N} \mu_S(c_k) Pr(c_k|verb) \qquad (14)$$

where, c_k indicates the legal co-occurrences (*i.e.* the triplets) that appear in the context, and N is the number of these co-occurrences. The program determines the event S with maximum probability described by (15) as the final decision.

$$\max_{\forall \text{ event } S} (P(S|verb)) \qquad (15)$$

We obtain $\mu_S(c_k)/c_k$ and $Pr(c_k|verb)$ through a probabilistic analysis; In particular, through the outputs of triplets. Thus, we have:

$$\mu_S(c_k)/c_k = Pr((\text{arg}_i, verb, \text{arg}_j)|event) \qquad (16)$$

How then to compute $Pr(c_k|verb)$? During implementation, we transform the triplet into pairs of 2-tuplets, such as $< company, buy, money >$ transformed to $< company, buy >$ and $< buy, money >$. In order to apply this transformation, we have to assume the occurrences of $< company, buy >$ and $< buy, money >$

are independent. By this assumption, we might have the following simplification:

$$\begin{aligned} & Pr(< arg_i, arg_j > |verb) \\ & = Pr(arg_i|verb) \cdot Pr(arg_j|verb) \end{aligned} \tag{17}$$

Due to uneven training data, 1310 triplets are extracted belonging to event "Takeover", 994 triplets belonging to event "Earning" and 431 triplets belonging to event "Merger". Thus, we have more instances of "Takeover" than those of the other two events, which causes the sum of probability of "Takeover" to be consistently greater than that of the others. Thus, we use the probability of each event as the normalization value.

$$\text{EventProbability} = \frac{\text{number of triplets in event } e}{\text{number of triplets in all events}} \tag{18}$$

Then, we have a new equation as follows:

$$\begin{aligned} P(S|verb) &= \frac{\sum_{k=1}^{N} \mu_S(c_k) Pr(c_k|verb)}{\text{EventProbability}} \\ &= \frac{\sum_{k=1}^{N} Pr((arg_i, verb, arg_j)_k|e)}{\text{EventProbability}} \cdot \\ & \quad Pr((arg_i)_k|verb) \cdot \\ & \quad Pr((arg_j)_k|verb) \end{aligned} \tag{19}$$

Accordingly, we have the following results in Table II, where the figures in **bold** indicates the correct classification given by human evaluation, while the legend * indicates the wrong classification and the figures in *italics* indicates the true event if it were correctly classified. As we discover, the classification results are not satisfactory at all.

2) *Classification context by trapezoid membership functions*

As an alternative, we classify the context by the proposed fuzzy approach (10) and (11) rather than by the probability measures of fuzzy events. The experiment is based on the same environment. In our implementation, the defuzzification uses the *area* of the membership function. In particular, we test the performance based on $\otimes = S - norm$ and the accompanying results in Table III.

TABLE II: classification Results Through Equation (19)

Document No.	Takeover	Earning	Merger
1	**0.0718**	0.0009	0.0043
2	**0.0211**	0.014	0.002
3	0.0005	**0.0013**	0
4	**0.0162**	0	0.0015
5 *	0.0028	*0.0005*	0
6 *	0	*0*	0
7	0.0017	**0.0355**	0
8 *	0	0	*0*
9 *	0.0163	*0.0042*	0
10	0.002	**0.0089**	0.0017
11 *	0.0041	*0*	0

TABLE III: Classification Results Through Defuzzification If $\otimes = S - norm$

Document No.	Takeover	Earning	Merger
1	**652**	319	612
2	**701**	534	599
3	231	**385**	223
4	**452**	232	418
5	157	**362**	119
6	97	**374**	157
7	129	**557**	64
8	636	519	**3082**
9	361	**466**	326
10	285	**451**	232
11	126	**288**	177

V. Comments and Discussions

The experimental results demonstrate that our proposed approach is more capable of context classification than a standard probabilistic method. As we discover, the probabilistic method has to make an assumption of independence between terms (17). Compared to this, the proposed fuzzy inferencing method does not assume the independence as a matter of course. In contrast, the method takes advantage of a means of *area* fuzzification (*i.e.* the max operator) to compute the value of interdependence(*i.e.* $Pr(c_k|verb) = \mu^{(g)}_{(arg_i)} \otimes \mu^{(g)}_{(arg_j)}$). The experimental results demonstrate this improvement. Although we only have 11 documents for testing, we have reason to qualify the generality of the experiment: we select Reuters' news as the training and test data set. Within the same data set, news show very homogeneous written style. As such, the 11 documents represent considerable amount of similar instances.

It would have been very natural and intuitive to implement fuzzy techniques for linguistic understanding applications in similar ways as was done for other applications [17]. However, the challenges are numerous: (i) how to build continuous membership functions to depict the distribution of the corresponding linguistic items? (ii) how to numerically represent linguistic items as the input of built membership functions? To tackle the two issues, we propose to construct membership functions pertaining to continuous representation of lexical items in the universe of discourse. In order to depict the approximate distribution of linguistic items in the universe of discourse, the membership functions should provide information such as (a) syntactic information, (b) semantic information associated with other linguistic terms, and (c) arbitrary global distribution (as the input) and local event-dependent distribution (as the output) confining the represented linguistic terms to a certain applicable domain. Reviewing these steps, the proposed rules adhere to these requirements. For (a), the construction of membership functions is built upon semantic patterns containing the syntactic and semantic information. For (b), the semantic patterns provide a restrictive environment in which only the categories occurring have relationships between one another. Through identifying corresponding semantic patterns, a computer program extracts triplets of co-occurrences, which adhering more closely to the distribution of the universe of discourse than those obtained through $3 - gram$, for example. For (c), we compute the distributions of linguistic items within semantic patterns so that we could root the origin - the verb, through which we transform the relative similarity to an arbitrary similarity within the given domain. Due to the fact that the construction of membership functions is greatly dependent on the behavior of the parsing program, however, human intervention becomes indispensable. Human intervention is essentially an engineering work including the definition of lexicon, the definition of domain-specific semantic patterns, the adjustment of triplets' output adhering to syntactic regulations, etc. Meanwhile, this engineering work is also greatly domain dependent, which means that if the domain changes, it has to adjust accordingly. This is indeed a very challenging task and further work needs to be done to customize this process even further.

VI. Concluding Remarks and Future Work

The presented work could be used for further investigation and could represent the seed for future research work. The optimization and normalization of membership functions could continue to be investigated to complement the results presented here. This should lead ultimately to a substantial improvement to the field of fuzzy linguistic computing.

References

[1] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, vol. 28, no. 1, pp. 11–21, 1972.

[2] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing and Management: an International Journal*, vol. 24, no. 5, pp. 513–523, 1988.

[3] D. Hull, J. Pedersen, and H. Schuetze, "Document routing as statistical classification," *AAAI Spring Symposium on Machine Learning in Information Access*, vol. 12, pp. 49–54.

[4] W. Lehnert, "Symbolic/subsymbolic sentence analysis: Exploiting the best of two worlds," in *Advances in Connectionist and Neural Computation Theory*, J. Barnden and J. Pollack, Eds. 1990, pp. 135–164, Ablex Publishers, NJ.

[5] S. G. Soderland, "Learning text analysis rules for domain-specific natural language processing," Tech. Rep. UM-CS-1996-087, 1996.

[6] L. Zadeh, "Fuzzy sets and systems," in *Systems Theory*, J. Fox, Ed., Brooklyn, NY, 1965, pp. 29–37, Polytechnic Press.

[7] L. Zadeh, "A fuzzy-set-theoretic interpretation of linguistic hedges," *Journal of Cybernetics*, vol. 2, pp. 4–34, 1972.

[8] L. Zadeh, "Quantitative fuzzy semantics," *Inf. Sci.*, vol. 3, pp. 159–176, 1971.

[9] L. Zadeh, "Pruf and its application to inference from fuzzy propositions," in *Proc. IEEE Conf. Decision control*, 1977, pp. 1359–1360.

[10] L. Zadeh, "Pruf-a meaning representation language for natural language," *Fuzzy Reasoning and its Applications*, pp. 1–66, 1981.

[11] G. Akrivas and G. Stamou, "Fuzzy semantic association of audiovisual document descriptions," in *Proc. of Int. Workshop on Very Low Bitrate Video Coding (VLBV)*, 2001, http://www.image.ece.ntua.gr/publications.php.

[12] P. Subasic and A. Huettner, "Affect analysis of text using fuzzy semantic typing," *IEEE Transactions on Fuzzy Systems, Special Issue*, pp. 483–496, August 2001.

[13] B. B. Rieger, "Tree-like dispositional dependency structures for non-propositional semantic inferencing," in *Proceedings of the 7th Intern. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU-98), Paris (Editions EDK)*, R. Yager and B. Bouchon-Meunier, Eds., 1998, pp. 351–358.

[14] Y. Sun, F. Karray, O. Basir, and M. Kamel, "Fuzzy methodology for enhancement of context semantic understanding," in *FUZZY-IEEE 2005 at Reno, USA*, May 2005.

[15] L. Zadeh, "Probability measures of fuzzy events," *Journal of Mathematical analysis and Applications*, vol. 23, pp. 421–427, 1968.

[16] D. Lewis, "Reuters-21578 text categorization test collection, distribution 1.0," September 1997, http://kdd.ics.uci.edu/databases/reuters21578/reuters21578.html.

[17] F. Karray and C. de Silva, *Soft Computing and Intelligent Systems Design*, Addison Wesley, 2004.