Fuzzy Logic-Based Natural Language Processing and Its Application to Speech Recognition

JIPING SUN, FAKHRI KARRAY, OTMAN BASIR & MOHAMED KAMEL
Department of Electrical and Computer Engineering
University of Waterloo, Waterloo
Ontario, N2L 3G1, Canada

Abstract: In this paper we describe a fuzzy logic-based language processing method, which is applied to speech recognition. Our purpose is to create a system that can learn from a linguistic corpus the fuzzy semantic relations between the concepts represented by words and use such relations to process the word sequences generated by speech recognition systems. In particular, the system will be able to predict the words failed to be recognized by a speech recognition system. This will help to increase the accuracy of a speech recognition system. This will also serve as the first stage of deep semantic processing of speech recognition results by providing "semantic relatedness" between the recognized words. We report the fuzzy inference rule learning system, which we have developed and also report the experimental results based on the system.

Key-words: Fuzzy logic, Natural Language Analysis, Speech Recognition, Corpus Linguistics.

1 Introduction

The complexity of natural language has made people apply various kinds of "soft" computing techniques for its analysis. Besides statistical, connectionist and other approaches, the fuzzy logic-based approach provides another alternative for effective natural language analysis. It is commonly recognized that many phenomena in natural language lend themselves to descriptions by fuzzy mathematics, including fuzzy sets, fuzzy relations and fuzzy logic. By defining a fuzzy logic system and acquiring proper rules, we hope that difficulties in analysis of speech can be alleviated.

As far as reference is concerned, words and their meanings (the referred objects or their measurements in the world) are often in a fuzzy relationship. This is important for grounded systems such as controlling robots. On the other hand, for the mainstream of NLP research, words themselves are the objects of description. It is natural to think that the language external fuzziness could be interpolated into language and fuzzy mathematical approaches are appropriate tools in solving problems.

Fuzzy logic has been successfully applied to the description of words' meanings as related to language external phenomena. Fuzzy linguistic descriptors have been used in control systems, in which mappings can be established between fuzzy linguistic terms and physical quantities. "Hot", "cold", for example, can serve as labels for fuzzy sets to which temperature readings can be mapped into membership degrees. Fuzzy logic rules for control systems can accept fuzzy descriptors in both the premises and the consequents to simulate human-like inferencing. Another case of fuzzy application is natural language-driven database search. Here the semantics of words can be expressed as fuzzy membership functions for certain database search keys [Medina, Vila]. A language internal fuzzy treatment is found in [Subasic], in which affect types of certain words in documents are dealt with as fuzzy sets. Words representing emotions are mapped to these fuzzy sets. The difference between this case and the previous two is that the latter dealt with language internal fuzzy phenomena.

Instead of mapping physical phenomena to words, which serve as fuzzy set labels, words themselves are the set members of affect types. How can this language internal phenomena be dealt with in a more general way, for example, in a parsing and language processing system, is the main topic of this paper.

In linguistic research there has always been tendency to treat linguistic categories and structures as fuzzy entities. This is strongly reflected in the cognitive grammar tradition. In that tradition, prototypes of natural categories and lexical semantics are considered as fuzzy and gradient in membership assignment. In the 1960s, Bolinger did research on fuzzy grammar [Bolinger]. Corrigan studies natural categories and the related issues reflected in fuzzy grammar. Continuums in language and "fuzzy" degrees of subjecthood, nouniness, adjectiveness, etc. are studied by Comrie. Matthew views grammatical categories as continuous or non-discrete. Many functionalists think of linguistic categories as gradient and fuzzy [Comrie, Givon, Langacker, Lakoff].

Section 2 of the paper reviews problems presented by speech recognition systems and defines the fuzzy logic system and explains how it is applied to speech recognizer results. Section 3 describes knowledge acquisition procedures for collecting information for the logic system. Section 4 describes some experimental results. We then conclude with suggestions for future work.

2 A Fuzzy Logic for Robust Processing of Speech Recognition Data

2.1 Problems presented by speech recognition systems

Our main goal is the efficient processing of speech recognition output. We consider the major problems presented by such data. We apply the speech recognition system on restricted domains. This means the vocabulary size and senses and syntactic constructs are restricted. Here are some often-encountered phenomena in a domain-constrained speech system:

- Out-of-vocabulary words. A user may speak words that are not contained in the system lexicon. For example, in the air travel domain, a system may have only American cities while a user may ask for flight information about other cities in the world.

- Speech recognizer errors. This may match a word into a wrong word, insert or delete a word, etc. Here is an example from recognition tests. Reference: "I NEED TO GET FROM PHILADELPHIA PENNSYLVANIA TO BOSTON MASSACHUSETTS"; Recognition: "I NEED TO GET FROM PHILADELPHIA AND STOPPING AT TO BOSTON MASSACHUSETTS". We will describe fuzzy inference strategies for dealing with such a problem.

- Flexible structures. The user may use expressions that the system's grammar does not cover. For instance, while the system may have the expression "a flight from Atlanta to Boston", a user's expression "a flight originating in Atlanta, destination Boston" may be novel to the system.

- Disfluency. False start, re-phrasing, repeated words, mis-pronounced words, half-pronounced words, filled pauses, etc. These could make the system confused about word semantic relations.

2.2 The fuzzy logic system

The fuzzy logic system has two major tasks. First, it evaluates whether a recognized word is semantically appropriate with respect to the whole recognized sentence. Second, it applies a set of fuzzy inferencing rules to predict possible missing words given a correctly recognized word evaluated in the first task. The major information used is word co-occurrence data, which is transferred to fuzzy logic knowledge.

A fuzzy semantic logic is a three-tuple $L(V,F,R)$, where

V is a set of linguistic vocabulary,
F is a set of vocabulary features,

R is a set of inference rules.

A member in *V* is called a word. A word is of the form *word*($f_1=v_1$ ... $f_n=v_n$). That is to say, a word is represented by a feature-vector with features from *F* and in symbolic values.

A member in *F* is a symbol representing a feature. In our system, we use the feature set {SYNTACTIC-TYPE, SEMANTIC-TYPE, PHRASE-TYPE and CONTEXT-WORDS}.

The set *R* consists of two types of fuzzy inferencing rules: evaluation rule and prediction rule. The evaluation rule matches a word to its context and decides whether the word is correctly recognized; a prediction rule predicts possible missing words based on a successfully evaluated word. Both of these rules apply the information in word feature vectors. For each word X, if its feature vector contains a context word Y, we denote this by a co-occurrence X * Y, meaning that when word X is used, word Y is likely to be used. As every word has also syntactic and semantic types, relations can be generalized from words to their semantic types and further to their syntactic types. The type features of a word is denoted as X:T. Given the above denotations, the evaluation and prediction rules are expressed as follows.

E-rule: $X * Y \wedge X \wedge Y \rightarrow inc^E(X)$
E-rule: $X * Y \wedge X \wedge not(Y) \rightarrow dec^E(X)$

The above two rules mean that when word X is found in the recognition, and if Y co-occurs with X in the training data, then increase or decrease the evaluation of X depending on if Y is also found in the recognition. The eventual evaluation of a word sums up the individual increases and decreases for all its context words.

P-rule: $eval(X) > \gamma \wedge X * Y \rightarrow inc^P(Y)$
P-rule: $eval(X) < \gamma \wedge X * Y \rightarrow dec^P(Y)$

In the above prediction rules, if the evaluation of X is over a threshold γ, then increase the prediction strength of Y, if Y is in its co-occurrence vector.

3 Knowledge Acquisition for a Fuzzy Logic

Developing a powerful logic system requires large amounts of knowledge to be acquired either from knowledge experts or from automatic knowledge acquisition processes. We have developed a number of procedures to automatically extract information for a fuzzy semantic logic.

3.1 Automatic acquisition of semantic types

Given a general semantic lexicon (e.g. WorldNet) [Fellbaum], it is possible to classify the words in a given corpus into semantic groups specific to the domain of the corpus. We have used a simple method utilizing a given corpus and WorldNet to discover interesting semantic types. The basic idea consists of the following points:

- The WorldNet is a huge hierarchical structure of words, dividing them into various levels of abstraction for semantic grouping. Words in WorldNet are highly polysemous.
- In a restricted domain (e.g. air travel), we prefer a flat semantic categorization and most of the words should have one or two different meanings.
- We decided to use the third level abstraction from the word up for semantic grouping. This level generally corresponds to the "basic" cognitive categories. Some of the examples are "time-period", "city", "meal", "motion", "cost", etc.
- We assume that for a semantic category interesting to a domain, it should have a high frequency of occurring in the corpus. So we find all the semantic types of all the word tokens in the corpus and then select a number of high frequency ones.
- The above procedure is automatically carried out. Eventually we have got a lexicon with words carrying interesting semantic types. We experimented with the

ATIS corpus [Rudnicky]. This corpus has 152k word tokens (1571 word types). By the above procedure, 1073 word types are assigned to 66 semantic types.

3.2 Automatic acquisition of syntactic and phrase types

Besides semantic typing, a word is also defined by its syntactic typing and phrasal typing. For syntactic typing, we used a tree-based tagger trained by Penn Treebank. After automatic tagging, a set of one-level chunking rules are applied to the tag sequence and parsed it to three types of phrases: verb phrase, noun phrase and prepositional phrase. This parser is highly efficient and reliable. The ATIS corpus contains 13k sentences and all of them were parsed. Random checking of 100 sentences showed above 97% of correctness in the chunking.

After the sentences were chunked into phrases, each phrase was further typed by the semantic type of its “core” word. The core word is decided by the following criteria:

- For verb phrase, the main verb is the core word.
- For prepositional phrase and noun phrase, the last noun is the core word.
- For a ***Be***+adjective phrase, the last adjective is the core word.

Following the acquisition tasks, we obtained 66 semantic types, 12 syntactic types: noun, verb, *be*-verb, adjective, adverb, conjunction, interjection, auxiliary, numeral, pronoun, preposition, proper-noun and 3 syntactic phrase types: noun phrase, verb phrase and prepositional phrase. Based on this, we derived the semantic phrase types of the ATIS corpus. From the training section of the ATIS corpus we derived 98 semantic phrase types. A semantic phrase type is the semantic type of the core word differentiated by its syntactic type. The following list shows some of the most frequent phrase types and their frequencies in the corpus.

- City, 12108
- Flight, 5774
- Time, 3574
- Person, 3371
- Number, 3344
- Communicate, 1948
- Desire, 1947
- Leave, 1665
- Commercial-document, 1421
- Airline, 1230
- Arrival, 839
- Taxonomic-group, 805
- Motion, 788

Finally we constructed word definitions for the ATIS corpus, based on the syntactic-type, semantic-type and phrase-type information. The context-word vector is constructed for each word with respect to the phrase type it is used in. The context word is divided into 4 groups: (1) the words on its left side in the phrase, (2) the words on its right side in the phrase, (3) the words in the left neighbor phrase and (4) the words in the right neighbor phrase. A neighbor phrase is differentiated by syntactic phrase type. The word definition structure is as follows:

Word Vector:
Semantic-Type
Syntactic-Type
Phrase-Type
Left-words-in-phrase
Right-words-in-phrase
Words-in-left-neighbor-phrase
Words-in-right-neighbor-phrase

Based on the word vector, the evaluation increment of e-rules for a phrase type p is defined by

$$\frac{\Delta e = w_p(X_p - N_p / b)}{\sum_{p=1}^{n} N_p}$$

X_p is the number of context words recognized; N_p is the total number of context words. b is control parameter between 1 and N_p. w_p is weighting parameter for the phrase type p of the evaluated word, which is a function of relative importance of the word, for example, its frequency and context complexity.

4 Experimental Results

As described in section 3 and 4, we implemented an unsupervised learning system for the acquisition of fuzzy rules. The basic data we have used consists of the WorldNet semantic lexicon and the ATIS corpus. The tasks we chose in our experiment are the evaluation of recognized words and of possible missing words. The input came from speech recognition output containing errors. In our experiment, we used results from a recognition system on the 1967 sentences in the ATIS test section. The word recognition accuracy for these sentences is 81.90%. The sentence recognition accuracy is 35.13%.

The fuzzy inferencing rules used the word definition vectors trained from the ATIS training section. Depending on different levels of evaluation and prediction threshold, for each recognized sentence, the system can make more or less predictions of missing words. We then calculated how well these predicted words can cover the words in the reference sentences. The coverage then is contrasted with the ratio between the quantity of the predicted words and the original language model size. One use of the prediction is to create a dynamic language model and use it to do a re-scoring of the sentences, either through a second time recognition or through a semantic interpretation. Therefore, the indicator of performance of our system is to achieve maximal coverage and minimal model ratio. Table 1 shows the performance achieved with low evaluation threshold ($\gamma < 0.2$).

Table 1. Word prediction results tested with low evaluation threshold. Test corpus = 22.26k words. Recognition accuracy = 81.90%. Original language model size = 1282 words.

Fuzzy Logic-Based Word Prediction Results 1				
Uncovered words	Coverage percent	Average model size	Model reduction	Stop words
196 w.	99.11 %	840 w.	34.5 %	0 w.
286 w.	98.72 %	615 w.	52.1 %	43 w.
300 w.	98.65 %	599 w.	53.3 %	291 w.

In the above tests, we tested different stoplist size. By using more stop words, we could achieve larger model reduction without affecting coverage much. In Table 2, we increased the evaluation threshold to various degrees. By doing so, we could further reduce the size of the predicted words with a moderate loss of coverage.

Table 2. Word prediction results tested with increasingly high evaluation thresholds.

Fuzzy Logic-Based Word Prediction Results 2				
Uncovered words	Coverage percent	Average model size	Model reduction	Stop words
321 w.	98.60 %	580 w.	54.7 %	291 w.
343 w.	98.45 %	569 w.	55.5 %	291 w.
376 w.	98.31 %	555 w.	56.6 %	291 w.
433 w.	98.05 %	538 w.	58.0 %	291 w.
681 w.	96.94 %	482 w.	62.3 %	291 w.
881 w.	96.04 %	401 w.	68.6 %	291 w.

From the above test results we found that using this fuzzy-logic based language processing system, we can predict up to 96 percent of the words truly said by the user in a dynamic language model with the model size reduced to less than 1/3 of the original model.

5 Conclusions and Future Work

In this paper we outlined the principles and constructs of a fuzzy logic-based NLP system and have illustrated it with experimental results. We have shown that for robust natural language processing to succeed, soft computing strategies are needed at various stages of the processing. In particular, we have shown that fuzzy logic is one promising direction in this respect. In future research several directions are worth exploring. First, better knowledge acquisition methods need to be developed so that a system can collect more information in efficient ways. Second, discourse information is desirable to be incorporated into the sentence understanding process. This is important because in many situations it is the usage of the sentence in the discourse context that decides precise interpretations of its semantics. In this perspective we can explore the direction of discourse-driven dynamic inference, which may modify the featural structures of words and categories dynamically in order to make the correct inference. Third, information from the lower levels such as pronunciation of words, recognition confidence and n-best alternatives of the recognizer output can be incorporated to assist the making of correct inferences.

References:

[1] Bolinger, D. Syntactic Blends and Other Matters. *Language* 37. 1961, pp.366-81.

[2] Bolinger, D. *Generality, Gradience, and the All-Or-None*. The Hague: Mouton, 1961.

[3] Castro, J.L. & Zurita, J.M., An Inductive Learning Algorithms in Fuzzy Systems, Technical Report, 1995.

[4] Comrie, B. *Language Universals and Linguistic Typology*, Oxford: Blackwell, 1989.

[5] Corrigan R., Eckman F. & Noonan, M. (eds.) *Linguistic Categorization*, Amsterdam/ Philadelphia; Benjamins, 1989.

[6] Dagan, I., Lee L. & Pereira, F.C.N., Similarity-Based Models of Word Co-occurrence Probabilities, *MLJ34*(1-3), 1999.

[7] Dahl, D., Bates, M., Brown, M., Fisher, W., Hunicke-Smith, K., Pallett, D., Pao, C., Rudnicky, A. & Shriberg, E., Expanding the scope of the ATIS task: the ATIS-3 corpus. *Proc. ARPA Workshop on Human Language Technology*, 1994, pp. 43-48.

[8] Fellbaum, C. (Editor), *WordNet: An Electronic Lexical Database*, MIT Press, 1998.

[9] Finch, S. & Mikheev, A. A Workbench for Finding Structure in Texts. Applied Natural Language Processing, Washington D.C., 1997.

[10] Fraser, N. *Dependency Parsing.* PhD, University College London, 1993.

[11] Geutner, P., Fuzzy Class Rescoring: A Part-Of-Speech Language Model, *Proceedings of Eurospeech*, 1997.

[12] Givon, T., *Syntax: A Functional Typological Introduction*. Vols. I and II, John Benjamins, 1984 and 1990.

[13] Herrera, F.; Lozano, M. & Verdegay, J.L., Generating Fuzzy Rules from Examples using Genetic Algorithms, *Fuzzy Logic and Soft Computing*, B. Bouchon-Meunier, *et al*. Eds. 1995.

[14] Hipp, J., Guntzer U. & Nakhaeizadeh, G., Algorithms for Association Rule Mining - A General Survey and Comparison, *SIGKDD Explorations*, 2000, Volume 2, Issue 1, pp 58-64.

[15] Sun J., Togneri , R. & Deng, L., A Robust Speech Understanding System Using Conceptual Relational Grammar, *ICSLP*, 2000.

[16] Jones, R., McCallum, A., Nigam, K. & Riloff, E., Bootstrapping for Text Learning Tasks, *IJCAI-99 Workshop on Text Mining: Foundations, Techniques and Applications*. 1999.

[17] Lakoff, G. *Women, Fire and Dangerous Things: What Categories Reveal About The Mind*. University of Chicago Press, 1987.

[18] Langacker, R. W., *Foundations of Cognitive Grammar*, Vols. I and II. Stanford U Press, 1987, 1991.

[19] Langacker, R. W. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Mouton de Gruyter, 1991.

[20] Mangu, L., Brill, E. & Stolcke, A., Finding Consensus in Speech Recognition: Word Error Minimization And Other Applications of Confusion Networks, *Computer Speech and Language*, 14(4): pp.373-400, 2000.

[21] Matthews, P. H., *Syntax*, Cambridge University Press, 1981.

[22] McKelvie, D., Brew, C. & Thompson, H., Using SGML as a Basis for Data-Intensive NLP. *Applied Natural Language Processing*, Washington D.C., 1997.

[23] Medina, J.M., Vila, J.C. & Pons O., Towards the Implementation of a Generalized Fuzzy Relational Database Model, *Fuzzy Sets and Systems*, vol. 75, pp.273-289, 1995.

[24] Mel'cuk, I. A., *Dependency Syntax: Theory and Practice*. Albany: State University Press of New York, 1988.

[25] Meyer, C.F., *Apposition in Contemporary English*. Cambridge University Press, 1992.

[26] Palmer, D., Ostendorf, M. & Burger, J.D., Robust Information Extraction from Spoken Language Data, *Proceedings of Eurospeech*, 1999.

[27] Riloff, E., An Empirical Study of Automated Dictionary Construction for Information Extraction in Three Domains, *AI Journal*, vol. 85, 1996.

[28] Riloff, E. & Schmelzenbach, M., An Empirical Approach to Conceptual Case Frame Acquisition, *Proceedings of the Sixth Workshop on Very Large Corpora* , 1998.

[29] Riloff, E., & Jones, R., Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping, *Proceedings of the Sixteenth National Conference on Artificial Intelligence* (AAAI-99) ,1999, pp. 474-479.

[30] Riloff, E., Automatically Generating Extraction Patterns from Untagged Text, *Proceedings of the Thirteenth National Conference on Artificial Intelligence* (AAAI-96), 1996, pp. 1044-1049.

[31] Sekine, S., Automatic Sublanguage Identification for a New Text, *Second Annual Workshop on Very Large Corpora*, Tokyo, 1994, pp. 109-119.

[32] Sleator, D. & Temperley, D., Parsing English with a Link Grammar, Carnegie Mellon University Computer Science technical report CMU-CS-91-196, October 1991.

[33] Subasic, P. & Huettner, A., Affect Analysis of Text Using Fuzzy Semantic Typing, *FUZZ-IEEE* 2000, San Antonio.

[34] Taylor, J.R., *Linguistic Categorization: Prototypes in Linguistic Theory*, Oxford University Press, 1995.

[35] Vijay-Shanker, K. & Weir D.J., Parsing Some Constrained Grammar Formalisms, *Computational Linguistics*, vol. 19:4, 1993.

[36] Vila, M.A. *et al.* A Logic Approach to Fuzzy Relational Databases, *International Journal of Intelligent Systems*, vol.9, 1994 pp.449-461.