Interpretation of Semantically Tagged Data using Fuzzy Linguistic 2-Tuples

Mohammed-Amine Abchir[†], Isis Truck and Anna Pappa
LIASD – EA4383, Université Paris 8, 2 rue de la Liberté
Saint-Denis, 93526, France

Keywords: Fuzzy Logic, Knowledge Representation, Semantic Interpretation, PoS Tagging, Geolocation.

Abstract: We propose a natural language interface with interpretation of partially tagged semantically data in closed question/answering domain (geolocation) using fuzzy linguistic 2-tuples. The interface is a tool of configuration tasks such as alerts definition and modification, alerts messages, and other man-machine dialogue. The aim is to respond with precision to user's query, expressed in natural language, taking into account imprecision and vagueness. The combination of NLP techniques and fuzzy logic to interpret linguistic variables helps elicitation of business-level objectives avoiding useless and costly computation of middleware information. This paper introduces a methodology that deals with contextual fuzzy semantics in natural language interfaces.

1 INTRODUCTION

We start with some brief definitions about the linguistic notions mentioned in this paper to help a better understanding of the NLP techniques used, we continue with a brief description of geolocation issues, and finally with a detailed description of our interface explaining the method with examples of the geolocation domain.

We borrow from the Introduction in *The Philosophy of Language* (Martinich, 1996) the definition of Semantics as the study of the meanings of linguistic expressions. The term "meaning" is vague and ambiguous since one could give different kinds of meaning as being part of the same semantics. Linguists also refer to Pragmatics as a semantic notion which does mostly with context dependent features of language.

In Fuzzy Semantics, where semantics is combined with fuzzy logic, an interesting approach about what a fuzzy set represents in a theory of natural language semantics could be the meaning of a vague expression.

Semantic Interpretation (SI) for textual data is the process of analyzing a tagged text to a representation of its meaning, where the input is a syntactically parse tree (Hirst, 1987) and the output the meaning of that tree. Recently a novel method for fine-grained semantic interpretation of unrestricted natural language texts has been proposed (Gabrilovich and Markovitch, 2009). In nowadays SI is mostly used to develop tools for speech recognition (see SISR version 1.0 by W3C®) (Tichelen, 2007), and is the process of representing and describing the meaning of natural language utterance. Alternatives of semantic interpretation is the model theory with ontologies, where according to different propositional attitudes we find different ontologies such as sense constructive ones with or without cognitive agent (Hausser, 2001).

In Artificial Intelligence the research in Natural Language Processing has long been to endow machines with "understanding" ability, and the difficulty has always been how to represent human semantics for machines. Most approaches are based on manually encoded text data helped by statistical techniques to create lexical knowledge, without solving the problems of polysemy and synonymy.

The geolocation applications, mostly concern troubleshooting of delivery rounds (optimization problems), fleet and vehicle tracking and also personal tracking or child location. Usually there is one server (a kind of *hub*) that coordinates geoinformation on a single platform in order to be able to track devices (vehicles, persons, mobiles or tracking devices themselves are all considered as *devices* in this paper). Ideally, clients should configure themselves the hub either through a Web

[†] Corresponding author: maa@ai.univ-paris8.fr

Abchir M., Truck I. and Pappa A..
Interpretation of Semantically Tagged Data using Fuzzy Linguistic 2-Tuples.
DOI: 10.5220/0004158304290432
In *Proceedings of the 4th International Joint Conference on Computational Intelligence* (FCTA-2012), pages 429-432
ISBN: 978-989-8565-33-4

interface or directly on the telephone, having a phone conversation with a virtual assistant. For the moment this is quite hard to do since it implies an important expertise to be able to translate the needs expressed in a natural language into a set of Forth scripts and programs written in other languages. Our method permits to create semantic dependencies in both clearly explicitly stated expressions and vague ones according to user's geo-information needs.

This paper is organized as follows: in next section we give some works in NLP and fuzzy semantics, then we explain our method and describe the interface, we finally present a use case and highlight the interest of this work.

2 SEMANTIC AND FUZZY LOGIC ANALYSIS

At the first place to make discourse analysis, we can use part-of-speech tagging (PoS) to (try to) disambiguate words (*e.g.* “cross” can be a noun, an adjective or a verb) (Winograd, 1971). However these techniques permit to “understand” sentences without ambiguity in a closed domain context but they don't consider any imprecision or vagueness in the meaning. The first approaches to deal with this come from Zadeh when he introduced in 1965 the fuzzy set theory, the fuzzy logic and the concept of linguistic variables (Zadeh, 1965). The fuzzy sets could be employed to integrate vagueness throughout the relational structure of meaning including both the concept of structure and reference that a term denotes.

Since 1965, many models have been proposed, mainly based on the empirical or possibility theory which handling incomplete information (Zadeh, 1978). But recently, one seems the most appropriate in our case: the 2-tuple fuzzy linguistic model [9] because it deals with words and uses a simple internal representation of them. Indeed the idea is to deal only with words or linguistic expressions in translating them into a linguistic pair (s_i,α) where s_i is a triangular-shaped fuzzy set and α a symbolic translation. If α is positive then s_i is reinforced else s_i is weakened. If the information is perfectly balanced (*i.e.* the distance between words is exactly the same, then all the s_i values are equally distributed on the axis). But if not – that may happen when talking about distance, for instance, “almost arrived” and “close to” are closer to each other than “near” and “out of the route” – the s_i values may not be equally distributed on the axis. That is why another model has been proposed by the same team to deal with such information that they call multi-granular linguistic information (Martínez et al., 2010) for a deeper review of these models.

In next sections we explain the methodology with a use case to show the interest of the approach.

3 LINGUISTIC 2-TUPLES MODEL AND OUR NLP APPROACH

In recent papers, it has been shown that despite its advantages, the 2-tuple model or unbalanced linguistic term sets doesn't fit our needs perfectly especially when one (or more) linguistic expression is far away from its next neighbor (Abchir and Truck, 2011). The new model we propose fully takes advantage of the symbolic translations α that become a very important element to generate the data set.

Our 2-tuples are twofold. Indeed, except the first one and the last one of the partition, they all are composed of two half 2-tuples: an upside and a downside 2-tuple. The choice of our 2-tuple model is relevant since the linguistic terms used in the geolocation context are usually unbalanced.

The methodology we use to deal with imprecision inside the natural language is inspired by the Parts of Speech (PoS) recognition and tagging (Pappa, 2009). We simplify the analysis using semantic tags because the context (geolocation software) is known. Here is an example: “I want to create an alert when the truck gets very close to the warehouse” (see below).

```
<tokens>
    <token gram="PRON">I</token>
    ...
  <token gram="NOUN" sem="ALERT">alert</token>
    ...
  <token gram="VERB" sem="ZONE_ENTRY">gets</token>
  <token gram="ADV" sem="FUZZY_MODIF_+">very</token>
  <token gram="ADJ" sem="DISTANCE">close</token>
</tokens>
```

A tree using a simplified tree-adjoining grammar (TAG)-based is then created, where each leaf node represents the semantic tag of a token from the lexicon. This grammar describes the components of a geolocation alert that can be created by the end user:

```
ALERT=TYPE,MOBILE,PLACE,NOTIFICATION
TYPE=ZONE_ENTRY|ZONE_EXIT|CORRIDOR
...
PLACE=TOWN|ADDRESS|POI|ZOI
```

Once we defined the lexicon (list of tagged tokens)

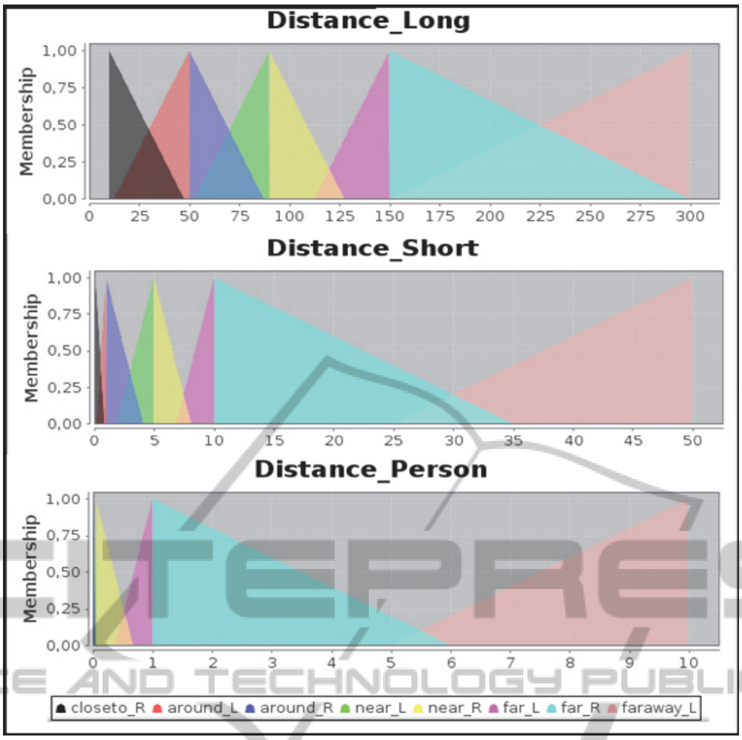

Figure 1: The three partitions for Distance.

and the grammar of the target domain, we use them in the natural language interface to parse, tag and analyze each user answer.

4 FUZZY SEMANTICS IN NLP

In order to fit with the user's needs, a semantic interpretation of his words is necessary all along the NLP process. Important business data is modelized as fuzzy partitions using linguistic 2-tuples described in Fuzzy Control Language (FCL) scripts. Thanks to the jFuzzyLogic (Abchir, 2011) library (a Java FCL specifications IEC 61131-7 implementation), these FCL scripts are then used in the semantic interpretation process. Thus, we are able to create various FCL scripts for the same data and we choose automatically at runtime the most appropriate fuzzy partitioning. The choice of a fuzzy partitioning depends on several criteria as the type of the mobile, the type of alert, the global distance of the route... We also support the use of semantic fuzzy modifiers such as *very*, *extremely*, *highly*, *really*... to take fully into account the users preferences. These modifiers act on the symbolic translation e of the linguistic 2-tuples (s_i,α) to modify their semantic value. For example, «far», «very far» and «extremely far» don't have the same “meaning” semantically.

To illustrate the adaptive fuzzy partition selection, we consider three mobile types: a car in the city, a long distance delivery truck and a child who gets home from school. For these three mobile types, the expert of the domain chooses five terms to qualify the distance measurements: close to, around, near, far, faraway. If we consider these two sentences: «notify me when my child is around home» and «notify me when the truck is around Paris», the term «around» will be associated to two different linguistic terms having two different semantic values. Thereby, we create three fuzzy partitions in three different FCL scripts each one corresponding to a mobile type.

Figure 1 shows the three different partitions for the distance: Distance_Long the partition for long distance routes, Distance_Short is the one for short distance routes as city driving, city mail devilery... and Distance_Person is used for human being following as for children location, marathon runners following...

5 CONCLUSIONS

In this paper we have presented a methodology to deal with natural language interfaces when data are incomplete or vague. We mix NLP techniques with a 2-tuple representation model to express data within their imprecision. The interpretation of the partially semantic-tagged data provides the "closest" meaning which helps avoiding useless and costly computation. In a second part, we presented an application of this methodology to the geolocation domain using FCL scripts.

In our future works, we will explore further the use of the fuzzy linguistic 2-tuples model in the definition of word's semantic.

REFERENCES

A. P. Martinich, "The Philosophy of Language", ed., *Third Edition Oxford Univerity Press*, (1996).

G. Hirst, "Semantic interpretation and the resolution of ambiguity", *Cambridge University Press*, (198)7.

E. Gabrilovich and Sh. Markovitch, "Wikipedia-based Semantic Interpretation for Natural Language Processing, in *Journal of Artificial Intelligence Research 34*, pp. 443-498, (2009).

Luc Van Tichelen, "Semantic Interpretation for Speech Recognition" (SISR) version 1.0, by eds: *Nuance Communications, Dave Burke, Voxpilot, in W3C Recommendation* of april (2007).

R. Hausser, "The four basic ontologies of semantic interpretation", in *Information Modeling and Knowledge Bases XII*, H. Jaakkola et al. (Eds.) IOS Press, pp. 21-40, (2001).

T. Winograd, "Procedures as a Representation for Data in a Computer Program for Understanding Natural Language", *MIT AI Technical Report* 235, (1971).

L. A. Zadeh, "Fuzzy sets", Information and Control 8 (3): 338–353 (1965).

L. Zadeh, "Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst*., Vol 1, pp. 3-28 (1978).

F. Herrera and L. Martínez, "A 2-tuple fuzzy linguistic representation model for computing with words". *IEEE Transactions on Fuzzy Systems*, 8(6):746–752 (2000).

F. Herrera, E. Herrera-Viedma, and L. Martínez, "A fuzzy linguistic methodology to deal with unbalanced linguistic term sets." *IEEE Transactions on Fuzzy Systems*, 354–370 (2008).

L. Martínez, D. Ruan, and F. Herrera, "Computing with words in decision support systems: An overview on models and applications". *International Journal of Computational Intelligence Systems*, 3(4):382–395 (2010).

M.-A. Abchir and I. Truck, "Towards a New Fuzzy Linguistic Preference Modeling Approach for Geolocation Applications". *In Proc. of the EUROFUSE Workshop*, 413–424 (2011).

A. Pappa, "Constructing lexicon with morpho-syntactic features from untagged corpora", in ECC'09 *Proceedings of the 3rd international conference on European computing conference*, (2009).

M.-A. Abchir, "A jFuzzyLogic Extension to Deal With Unbalanced Lin-guistic Term Sets". *Book of Abstracts*, 53–54 (2011)