Towards the Construction of a Multi-agent Approach for Discovering the Meaning of Natural Language Collaborative Conversations

Luis Casillas
University of Guadalajara, Department of Computer Sciences. Av. Revolucion 1500, 44840 Guadalajara, Mexico
luis.casillas@red.cucei.udg.mx

Thanasis Daradoumis
Open University of Catalonia, Department of Computer Sciences. Rambla Poblenou 156, 08018 Barcelona, Spain
adaradoumis@uoc.edu

Abstract

On the one hand, natural language is the main communication media for humans. It has a complex construction, based on the diversity of meaning for words and expressions according to the context. On the other hand, computers are not prepared to handle this ambiguity. The present work aims at presenting a multi-agent approach for dealing with the problem of discovering the meaning of expressions written in Spanish, based on a flexible recovery system and Bayesian principles. At a first stage, agents are supposed to identify the role of the words composing a sentence. At a second stage, a second set of agents is supposed to coordinate among them in order to assemble a meaning. Our research forms part and contributes to the analysis of collaborative conversations among participants in a web-based collaborative learning environment.

1. Introduction

Natural language is the main communication media for humans. It has a complex construction, based on the diversity of meanings for words and expressions according to the context. Moreover, computers are not prepared to handle this ambiguity. Humans are capable of this kind of understanding due to common sense and reasoning; meanwhile computers are not able to perform that reasoning. Therefore, natural language processing (NLP) implies the construction of specific algorithms oriented to discover the meaning stored in certain sets of words. NLP is a field of the artificial intelligence, devoted to producing or understanding natural language.

The present work aims at presenting a multi-agent approach for dealing with the problem of discovering the meaning of expressions written in Spanish, based on a flexible recovery system and Bayesian principles. At a first stage, agents are supposed to identify the role of the words composing a sentence. At a second stage, a second set of agents is supposed to coordinate among them in order to assemble a meaning.

This mechanism will be used over discussion forums, in order to perform an analysis of conversations. The goal is to build a collaboration network based on the interaction acts detected by the conversation's analyzer.

2. Related work

NLP is an ancient task in Artificial Intelligence (AI). The first stages of AI were devoted to manage the task of achieving in computers a behavioral-profile comparable to the behaviors shown by humans. This romantic-goal was set to reach an electronic brain, powerful as human's brain. Nevertheless the complexity of human behaviors versus the simplicity of the computer's capabilities, led to several dead ends; which conducts to an impassable wall: handling the ambiguity present in the reality.

Different efforts have been conducted in order to deal with NLP, most of them related with the English as object language. Science fiction has presented diverse computers maintaining conversations with humans, although the computer H.A.L. 9000; shown in the movie "2001: a space odyssey" (directed by Stanley Kubrick and based on the novel of Arthur C. Clarke), is a remarkable example of an advance computer-treatment for NLP (and other high profiled behaviors for a machine!). In the reality, the most recognized project involving NLP is the classic ELIZA [11]; which is a clever system capable to support a conversation in an interactive fashion. The project ELIZA has inspired a significant number of interactive agents based on the Web. This kind of systems support simulations of conversations, but is common to arrive to dead-ends in which the conversation is blocked in empty argumentations.

0-7695-3109-1/08 $25.00 © 2008 IEEE
DOI 10.1109/CISIS.2008.40

A practical way to approach the NLP is through the analysis of word-roots. The Spanish language has a structure based in common roots, which direct to different words with slightly different meanings. Hence, in order to avoid machine confusions, all the discourse elements (words) will be analyzed in order to bind them under a common meaning. Thus, a set of agents will perform an analysis over the words forming the sentences of the discourse. This analysis is performed using retrieve-trees [1] for discovering common roots and the presence of the words in the vocabulary. The retrieve-trees for roots analysis were used successfully by [9] over Portuguese language, which is too similar to the Spanish. Besides, during the analysis we will use the Theorem of Bayes as in [4] to measure the probability for every word as member of specific word categories (article, substantive, adjective, verb and etcetera).

The agents have been also considered to support the efforts for handling the discovery of meaning from written expressions presented in natural language. The study [6] offers a framework for allowing users to instruct agents regarding the tasks to perform, this model is restricted for certain interaction forms to avoid ambiguity; unfortunately this approach restricts the communication to a query language model. The project [8] is also oriented to the development of specific interfaces for supporting the communication with users through agents devoted receive and process queries from the users. The solution proposed in [9] involves also the usage of agents, in this case the use of agents-technology is oriented to solve the problem of discovering morphology roles for every word, which is precisely the approach we are considering in our solution; nevertheless, the project shown in [9] has an unknown continuity.

The search for a semantic Web involves the NLP. The work [10] provides an approach for Web-search using complex expressions, based on the sub-products: NL-Page, a parser generator and MIN, a multi-agent framework. This machinery provides a powerful basement for supporting interactive agents during web searches, although our approach is not intended to the interaction between humans and agents, our agents will work over the messages written by humans when they interact; which is a most complex interaction model, away from our goals.

3. Problem description

A computer supported cooperative environment (CSCW: Computer Supported Cooperative Work, CSCL: Computer Supported Collaborative Learning) could include forums and chat-rooms. Our research is approached to the analysis of collaborations among participants of such environments. The first stage of such research was oriented to the quantitative analysis of the collaboration events, in order to produce a social network; which provides high level knowledge regarding the collaboration acts. The second stage, in which we are working on now, is devoted to the performance of a qualitative analysis of the conversations inside the forums and chat rooms. Such task involves NLP.

Our analysis must discover the intentions in the messages and the consequent reactions from the other participants in the same space (forum or chat-room), as well as the intentions in the respective reactions. Hence, an interaction network is built; now modeling the intentions in every direction.

NLP is a very specific task regarding the language involved. Although common-sense follows concrete rules (common to every human, regardless of his/her origins), we agree to think that every culture has developed a language machinery according to specific social development and depending on context conditions; these facts imply different morphological analysis for every language and context in which it is used. As we wrote before, our task will be performed over the Spanish language. Hence, our NLP will be implemented over messages written in Spanish language and with the specific goal of produce a simplified representation, oriented to recover a synthetic semantics.

4. Solution proposed

We are not looking for a deep discovery of meaning from the conversations. A synthetic meaning will be enough to discover the collaboration acts among participants of the conversation. In order to achieve this treatment over the analyzed expressions, we have modeled agent-based machinery. Figure 1 sketches our approach for meaning discovery. The Word-Agents (smiling faces in fig. 1) are supposed to identify words belonging to specific categories in the sentence. Word-Agents use retrieve-trees [1] for a fast data-recovery and for providing the skill of binding two or more words with common roots (a frequent phenomenon in Spanish), as sketched in figure 2.

Word-agents are also prepared for producing a probability value for every word in each category. Such probability is calculated according to the Theorem of Bayes [4]. The probabilities for calculating the Bayesian model are based on the impressing study [2], which provides a rich data-bank of frequency of

words in the Spanish language. These elements of information allow the Morphology-Agents (Clouds in fig. 1) to coordinate in order to produce the synthetic semantics following different steps.

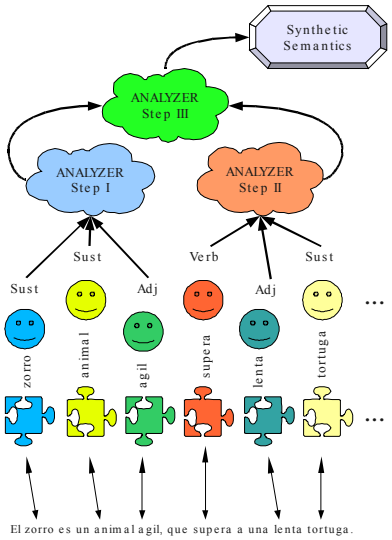

Figure 1. Agent-based model for meaning discovery during NLP working over the Spanish sentence for: "The fox is an agile animal that passes a slow turtle".

The Bayesian principle (see formula 1) for learning is approached as a way to gather the certainty of the hypothesis related bind a word to certain category in the sentence. The Spanish word "bien" is an adverb and a substantive at the same time. Such word in Spanish is among the 175 most used. The average probability to use it as an adverb is 0.0013873025 and the probability for substantive is 0.0001328026. The probability to use that word as adverb is ten times bigger to use it as substantive. These values are used to calculate the certainty of every word, for every category; the higher probabilities will imply the main meaning bound to the word.

$$P(B|A)=(P(A|B)*P(B))/P(A) \qquad (1)$$

Morphology-Agents use their delegates (Word-Agents) to acquire a primary meaning for every word in the analyzed sentence, as well as the probability for being correctly settled as article, substantive, adjective, verb and etcetera; according to the responsibility assign every Word-Agent.

5. Conclusions and future work

NLP is a complex task. Our effort is not oriented to fulfil all the pending tasks in such field. We are developing a tool to handle a subset of the semantics underlying conversations in the spaces devoted to short communications among users in CSCW and CSCL environments. The current paper is approached to order the work developed up to the moment in this task. Fully-functional algorithms (implemented in Mono-C# over Linux-Ubuntu 7.10) are currently working, in order to activate every piece of the whole model, and we are developing the mechanisms to bind these pieces. When all the pieces are bound, we will be enabled to process messages written by humans and discover the synthetic-semantics. At this point, the compact meaning discovered will be compared to some reduced explanation for the very same messages; but these last will be process by humans, thus a qualitative comparison will allow the measuring of the success of the model.

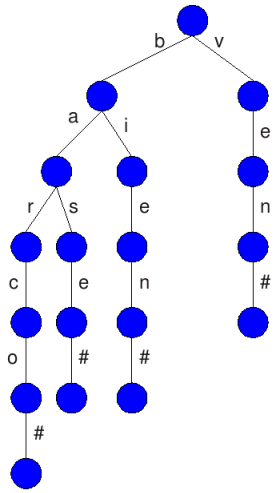

Figure 2. Retrieve-Tree (Trie-Tree) storing the Spanish-words: "barco", "base", "bien" and "ven". Adapted from [1].

Acknowledgements. This work has been partially supported by the Spanish MCYT project TSI2005-08225-C07-05.

10. References

[1] Aho, A. V.; Hopcroft, J. E.; Ullman, J. "Data Structures and Algorithms". Ed. Addison-Wesley, USA. 1982.

[2] Almela, R. "Frecuencias del español, diccionario y estudios lexicos y morfologicos". Ed. Universitas, Spain. 2005.

[3] Bianchi, D.; Poggi, A. *Ontology based automatic speech recognition and generation for human-agent interaction*. 13th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises; Volume 0, Issue 0, pp 65 – 66. 2004.

[4] Escolano, F.; Cazorla, M. A.; Galipienso, M. I.; Colomina, O.; Lozano, M. A. "Inteligencia Artificial: Modelos, Técnicas y Áreas de Aplicación". Ed. Thomson, Spain. 2003.

[5] Fum, D.; Guida, G.; Tasso, C. *A distributed multi-agent architecture for natural language processing*. International Conference On Computational Linguistics; Proceedings of the 12th conference on Computational linguistics - Volume 2; pp 812-814. Hungry. 1988.

[6] Kemke, C. *An Architectural Framework for Natural Language Interfaces to Agent Systems*. Proceedings of Computational Intelligence. San Francisco, California, USA. 2006.

[7] Keselj, V. "Multi-Agent Systems for Internet Information Retrieval using Natural Language Processing". Graduate Thesis, University of Waterloo. Canada. 1998.

[8] Rubin, S. H.; Dai, W. *Natural Language Enabled Interface Agent*. IEEE/WIC/ACM International Conference on Intelligent Agent Technology; pp 544-547. USA. 2004.

[9] Strube, V. L.; Carneiro, P. R.; Filho, I. S. M. *Distributing linguistic knowledge in a multi-agent natural language processing system: re-modelling the dictionary*. Procesamiento del lenguaje natural, ISSN 1135-5948, Nº. 23, 1998 , pp. 104-109. Spain. 1998.

[10] Vlado, K. "Multi-agent Systems for Internet Internet Information Retrieval using Natural Language Processing". Ph.D. Thesis, Department of Computer Science; University of Waterloo, Ontario, Canada. 1998.

[11] Weizenbaum, J. *ELIZA: A Computer Program For the Study of Natural Language Communication Between Man and Machine*. Massachusetts Institute of Technology; Department of Electrical Engineering. Communications of the ACM Volume 9, Number 1, pp 36-35. USA. 1966.

[12] Yang, M. S.; Yang, C. C.; Chung, Y. M. *A natural language processing based Internet agent*. IEEE International Conference on Systems, Man, and Cybernetics; Computational Cybernetics and Simulation; Volume 1, pp. 100-105. USA. 1997.