Word Sense Disambiguation using ILP

Lucia Specia[1], Ashwin Srinivasan[2,3], Ganesh Ramakrishnan[2], Maria das Graças V. Nunes [1]

[1] ICMC – University of São Paulo, Trabalhador São-Carlense, 400, São Carlos, 13560-970, Brazil
{lspecia, gracan}@icmc.usp.br
[2] IBM India Research Laboratory, Block 1, Indian Institute of Technology, New Delhi 110016, India
[3] Dept. of Computer Science and Engineering, University of New South Wales, Sydney, Australia
{ashwin.srinivasan, ganramkr}@in.ibm.com

Abstract. We investigate the use of ILP for the task of Word Sense Disambiguation (WSD) in two different ways: (a) as a stand-alone constructor of models for WSD; and (b) to build interesting features, which can then used by standard model-builder such as SVM. Experiments examining a multilingual WSD task in the context of English-Portuguese machine translation of 7 highly ambiguous verbs showed promising results: our ILP-based standalone approach outperformed the results of propositional algorithms and the ILP-generated features yielded improvements on the accuracy of standard propositional algorithms when compared to their use with low level features only.

1 Introduction

Word Sense Disambiguation (WSD) aims to identify the correct sense of an ambiguous word in a sentence. Sometimes described as an "intermediate task"— that is, not an end in itself — it is necessary in most natural language tasks like machine translation, information retrieval, and so on. That is extremely difficult, possibly impractical, to solve completely is a long-standing view [1] and accuracies with state-of-the art methods are substantially lower than in other areas of text understanding (part-of-speech tagging accuracies, e.g., are now over 95%, while the best WSD results are still well below 80%). It is generally thought that WSD should benefit significantly by adopting a "deep approach" in which access to substantial body of world knowledge could assist in resolving ambiguities. This belief is based on the following observation. While it is true that statistical methods like support vector machines using shallow features referring to the local context of the ambiguous word have usually yielded the best results to date, the accuracies obtained are low and significant improvements do not appear to be forthcoming. The incorporation of large amounts of domain knowledge has been hampered by the following: (a) access to such information in electronic form suitable for constructing models; and (b) modeling techniques capable of utilizing diverse sources of domain knowledge, even when they are available. The first of these difficulties is now greatly alleviated by the availability in electronic form of very large semantic lexicons like WordNet, dictionaries, parsers, grammars and so on. In addition, there are now very large amounts of "shallow" data in the form of electronic text corpora from which statistical information can be readily extracted. Using these diverse sources of information is, however, beyond the capabilities of existing general-purpose statistical methods that have been used for WSD resulting in the development of various *ad hoc* techniques for using specific sets of information for particular tasks. Arguably, ILP systems provide the most general-purpose framework for dealing with such data: there are explicit provisions made for the inclusion of background knowledge of any form; the representation language is powerful enough to capture the contextual relationships that arise; and modeling is not restricted to being of a particular form. In this paper, using a task in Machine Translation (MT) as a test bed and 7 different sources of background knowledge, we investigate the use of ILP for WSD in 2 ways: (a) the induction of disambiguation models; and (b) the construction of interesting features to be used by propositional algorithms such as SVM, which have presented reasonably good results with low level features in previous work.

2 Empirical Study

Aim. We investigate whether the use of an ILP system equipped with substantial background knowledge can significantly improve WSD accuracies for a task in English-Portuguese MT.
Materials.
*Data***.** Data consist of 7 highly frequent and ambiguous verbs and a sample corpus of around 200 English sentences for each verb with the verb translation automatically annotated. The verbs are (numbers of possible translations in our corpus in brackets): *come* (11), *get* (17), *give* (5), *go* (11), *look* (7), *make* (11), and *take* (13).
Background Knowledge. We exploit knowledge from 7 syntactic, semantic and pragmatic sources: (1) Bag-of-words of ±5 words surrounding the verb; (2) Part-of-speech tags of ±5 content words surrounding the verb; (3) Subject and object syntactic relations with respect to the verb; (4) 11 collocations with respect to the verb; (5) Selectional restrictions and semantic features; (6) Phrasal verbs possibly occurring in the sentence; (7) Overlapping words in dictionary definitions for the possible verb translations and the surrounding words in the sentence. Refer to [3] for details about representation of these knowledge sources.
Algorithm. We used the ILP system Aleph [4] to implement the construction of disambiguation models and features for statistical model construction.
Method. For reasons of space, we refer the reader to [4] for details of how models and features are constructed by Aleph. We follow standard methodology to construct and test models (i.e., cross-validation on training data to select the best models; and testing on unseen data).
Results.
Disambiguation Models. We evaluated the results according to the metrics usually employed for WSD, namely accuracy on the positive examples. We used Aleph constraint's mechanism to create a rule to classify all the cases that have not been classified by other rules according to the majority class (the most frequent translation in the corpus). Table 1 shows the accuracy achieved by the test-bed previously mentioned, according to a 10-fold cross-validation strategy, together with the accuracy of two propositional algorithms that usually perform well on the WSD task, C4.5 and SVM, here having the relational features pre-processed in order to allow an attribute-value representation. ILP results are significantly better (t-test, $p < 0.05$).
Feature Construction. The numbers of features constructed by the ILP engine for each verb are shown in Table 2. Together with 23 original low-level features, these new ILP features were then used to test SVM's performance. The enhanced set of features improved SVM's accuracy for 2 verbs ("come" and "go"), with other verbs unaffected. It is not evident at this stage whether this due to (a) the small number of examples; (b) inadequate relational features;

(c) inadequate background knowledge; or (d) inadequate model construction by the statistical method. We are conducting experiments to shed further light on these questions.

Table 1. Verbs and their possible translations in the sample corpus

Verb	Accuracy		
	ILP	**C4.5**	**SVM**
come	0.82	0.53	0.62
get	0.51	0.36	0.26
give	0.96	0.96	0.98
go	0.88	0.76	0.74
look	0.83	0.57	0.79
make	0.81	0.74	0.74
take	0.81	0.31	0.44
Average	**0.80**	**0.6**	**0.65**

Table 2. Accuracy achieved by SVM with and without ILP features

Verb	# ILP features	Accuracy: original feature set	Accuracy: enhanced feature set
come	483	**0.58**	**0.71**
get	329	0.31	0.31
give	19	0.98	0.98
go	174	**0.69**	**0.71**
look	677	0.79	0.79
make	4122	0.76	0.76
take	411	0.47	0.47

3 Concluding Remarks

The results reported here suggest that ILP could play a useful role in WSD. As stand-alone constructor of WSD models, ILP yielded particularly good results, significantly outperforming propositional algorithms on the same data. This is mainly due to the real hybrid nature of the approach and the rich set of knowledge sources employed. Regarding the use of ILP to construct features, our findings suggest that: addition of background does improve the accuracy of WSD models in some cases. Features can be constructed efficiently: usually, the time taken for feature-construction was comparable, or often less, than to that taken to build models. In both cases, the results strongly support the undertaking a substantially larger study with more data.

References

1. Bar-Hillel, Y. The Present Status of Automatic Translation of Languages. Advances in Computers, 1 (1960) 91-163
2. Mooney, R.J. Inductive Logic Programming for Natural Language Processing. 6th International Inductive Logic Programming Workshop (1997) 3-24
3. Specia, L. A Hybrid Relational Approach for WSD – First Results. Coling-ACL (2006) 55-60
4. Srinivasan, A. The Aleph Manual. See: http://web.comlab.ox.ac.uk/oucl/research/areas/machlearn/Aleph/