Lexical Relations follow from Frame Semantics

CS 497 Report

Supervisor - Dr. Amitabha Mukerjee

Pushpraj Shukla
Computer Science and Enginnering
Indian Institute of Technology, Kanpur
Kanpur, India 208016
Email: praj@cse.iitk.ac.in

***Abstract*— We present our work aimed at realizing how semantic correlations dictate the other lexical features at the sentence level. We discuss the importance of the task in view of the SensEval task of Semantic Role Labeling and a review of the various schemes being employed for semantic similarity and clustering. Lastly, we discuss how we can automatically induce classes or clusters from semantic correlations obtained from FrameNet data.**

I. Introduction

Bootstrapping semantics from text is one of the most challenging tasks in natural language learning. But it is also the most important and natural task to be done, because semantic constraints guide the way we perceive language. A very naive example could be the fact that the meaning of an unknown word can often be inferred from its context. Consider the following (slightly modified) example given in [1]:
A bottle of *tezguino* is on the table. Everyone likes *tezguino. Tezguino* makes you drunk. We make *tezguino* out of corn.

The contexts in which the word *tezguino* is used suggest that *tezguino* may be a kind of alcoholic beverage made from corn mash. *What semantic correlations guided us to make the decision?* The answer to the above question could guide many information extraction and decision making systems to better levels of efficiency.

If it correct, even if intuitive, that lexical relations follow from the embedded semantics? These are some questions we tried to look for in our study.

The report follows the approach we adopted during the course of our study. The first part suggests a motivation to the task of realizing semantic constraints, based on the framework of the Role tagging and the SensEval task. That is followed, by a review of the work on similarity measures and creating semantic classes or clusters. Later, we discuss our attempt at the task of automatically creating semantic classes from the FrameNet[2] and the semantic correlations it encodes within its tagged corpus.

II. Automatic Labeling of Semantic Roles

The aim of the exercise is to assign semantic roles to (syntactic) constituents of a sentence. This will allows one to recognize semantic arguments of a situation even when expressed in different syntactic configurations.

A. Motivation

Semantic Role Labeling (SRL) has recently received a lot of attention in the NLP community, which is also evident from its inclusion in SensEval - the Annual International Workshop on the Evaluation of Systems for the Semantic Analysis of Text. The SRL task in SensEval uses the FrameNet corpus[2]: given a sentence instance from the corpus, a systems job is to identify the phrase constituents and their corresponding role.

FrameNet introduces *medium* roles to constituents of a sentence which describe abstract actions or relationships, along with their participants. They are not the domain specific *deep* nor are they too shallow like Doer and Done. See the Appendix for some details on FrameNet roles.

A shallow semantic interpreter based on this roles could play an important role in information extraction. For example, known that the verbs *send* and *receive* would share the semantic roles (sender, recipient, goods etc.) defined with respect to a common *TRANSFER* frame. Such a common frame might allow a Q/A system to take a question like *Who sent the letters to the customers?* and discover that "All the customers received letters from the operations research branch of the company."

B. Past Work

Gildea and Jurafsky [3] were the first to apply a statistical learning technique to the semantic role labeling on FrameNet data. They describe a model for determining the most probable role for a constituent given the frame, the target word, with well defined frame boundaries. Thompson et al[4] improved upon the design to give a probabilistic generative model in which the constituents are generated by the semantic roles. Since being included in SensEval, there have been many learning approaches used for the task [5] [6]

C. The Role Tagging Process

The Role Tagging Process is based on training a classifier on a labeled training set and testing out on a held-out portion

of data. The system is trained by first using a syntactic parser to analyze the training sentences, matching annotated frame elements to parse constituents and extracting various features from the string of words and the parse tree. During testing, the parser is run on the test sentences and the same features are extracted. Probabilities for each possible semantic role are then computed from the features.

The features used by most systems are syntactic, like Phrase type, Grammatical Function, Parse Tree Path, Position, Voice, Head Word etc. The question is **whether the syntactic data is enough to attain a sufficient level of accuracy in semantic tagging**. Consider the following set of sentences -
1.Ram and Shyam argued about politics.
2.Ram and Shyam argued on the state of affairs.
3.Ram and Shyam argued on the telephone.
4.Ram and Shyam argued in French.

Here the FRAME is Conversation and the Target Word is Argue. The underlined phrases are all prepositional phrases. Consider the task given to a syntax-based system to distinguish politics and state of affairs as TOPICS and Telephone and French as MEDIUM.

There is definitely scope for semantic knowledge added to this decision. This can be in the form of identifying semantic constraints that follow a construct or the semantic relatedness between words that serve a particular role for a Frame.

III. Towards Learning Semantic Constraints

Following from the previous section, we would like to realize constraints to the semantic construction of a particular sentence or we would like to have a semantic relatedness measure that distinguishes two words of different meanings, capable of being assigned ambiguous roles.

Some suggested similarity measures in literature are that of Semantic Distance and Gloss Overlap[7] from Wordnet. The general approaches to finding sematic similarity have been to gather statistical data from a corpora and to reason about similarity based upon the context. Some NLP systems use a hybrid approach information from a hand crafted lexical Knowledge Base, such as Wordnet is used. [8] came up with a Information Content metric for finding semantic similarity which is derived solely from Wordnet without the use of external statistical data or corpus.

A slightly different way to look at constraints and relatedness to look at groups of words which exhibit similar semantic properties and co-occurrences. This is the thought that we shall follow in the remainder of our sections.

A. Semantic classes

Many applications dealing with textual information require classification of words into semantic classes (or concepts). Text mining systems often convert text into a set of features, many of which are defined in terms of semantic classes.

Also, lexical-semantic resources, including thesauri and WORDNET, are difficult and expensive to create and maintain, and "their usefulness has been severely hampered by their limited coverage, bias and inconsistency" [9]. Automated and semi-automated methods for developing such resources are therefore crucial for further resource development and improved application performance.

(A) multiple sclerosis, diabetes, osteoporosis, cardiovascular disease, Parkinson's, rheumatoid arthritis, heart disease, asthma, cancer, hypertension, lupus, high blood pressure, arthritis, emphysema, epilepsy, cystic fibrosis, leukemia, hemophilia, Alzheimer, myeloma, glaucoma, schizophrenia, ...

(B) Mike Richter, Tommy Salo, John Vanbiesbrouck, Curtis Joseph, Chris Osgood, Steve Shields, Tom Barrasso, Guy Hebert, Arturs Irbe, Byron Dafoe, Patrick Roy, Bill Ranford, Ed Belfour, Grant Fuhr, Dominik Hasek, Martin Brodeur, Mike Vernon, Ron Tugnutt, Sean Burke, Zach Thornton, Jocelyn Thibault, Kevin Hartman, Felix Potvin, ...

Fig. 1. The semantic classes shown by Lin et al.

The idea is to come with clusters of words that frequently follow the same semantic relations, for example an argument(role) for a particular verb. Clustering has been recently used in an attempt to augment syntactical data in order to improve results of role taggers.

[10] use a Maximum Entropy learner, augmented by EM-based clustering to model the fit between a verb and its argument candidate. The instances to be classified are sequences of chunks that occur frequently as arguments in the training corpus.

In [1] and [11], Lin and Pantel propose a methodology for automatically extracting semantic classes. Their classes consist of clustered instances like the ones shown in Figure 1.

Their algorithm, UNICON (for UNsupervised Induction of CONcepts) can to classify words with low frequency counts and can cluster a large number of elements in a high-dimensional space using the CLIQUE [12] clustering algorithm, specifically designed to handle high-dimensional spaces. [13] give an extension to this scheme by automatically labelling the extracted semantic classes.

We follow Lin's approach to finding clusters, but our approach would differ from the distributional hypothesis he uses in ways and for reasons given in the next section.

IV. Automatic Induction of Semantic Classes from Semantic Correlations

A. Motivation

Systems that extract thesauri often identify similar words using the distributional hypothesis that similar words appear in similar contexts. This approach involves using corpora to examine the contexts each word appears in and then calculating the similarity between context distributions.

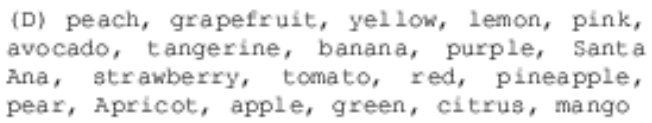


Fig. 2. Problem with polysemous words

But, distributional similarity is at best an approximation to semantic similarity. [9] quantifies the notion in his thesis that some events in the context distribution are more indicative of meaning than others. For instance, the object-of-verb context wear is far more indicative of a clothing noun than get.

Existing distributional techniques, including [1] do not effectively utilize this information. [9] proposes a context-*weighted* similarity metric for the purpose. It is natural to think that these "eights" could be automatically assigned if "semantic correlations" are themselves used as the metric. FrameNet roles could be a good measure for such a similarity metric.

Further, there are occasions where the distributional approaches fail. This is very common in polysemous words, where it cannot distinguish between their different senses. For example, Lin's approach give the top-20 similar words of *orange* as shown in Figure. 2. Here, the *color* and the *fruit* senses of *orange* are mixed up. We expect such issues to be automatically resolved if we consider semantic relations, with the corpora giving us the *semantic context* each word appears in, rather than syntactic features giving us context.

B. The Implementation

We have tried to come up with a word similarity measure based on the semantic distribution rather than selectional information based on syntax. Our work is based on the tagged FrameNet corpus[2], containing about 130,000 annotated sentences with approximately 6,000 lexical items(words).

1) Feature Set: Every word (extracted from the tagged FrameNet corpus) that needs to be classified, has a feature vector associated with it that represents the word's correlations with other words and the semantic *frames* in the corpus.

For any new tagged sentence, for example
The commander ordered [the troops *Agent*] not to $retaliate_{frame:Revenge}$ [against the rebels *Victim*]
The following new features are added (or the value incremented, if the feature is already present)-
for TROOP - Agent-Revenge
for REBEL - Victim-Revenge
for retaliate - Agent-troop, Victim-rebel

Thus, the words which take on the roles(Frame-Elements) are assigned a feature set consisting of all possible roles it served with respect to every possible frame. With about 400 frames and 4 frame elements for each the feature vector comes to around 1200 features for each words.

The Target words (usually the verbs, retaliate here) have a feature set of the words the roles they played with the target word as the frame. The feature vector could be large, but on an average shows a length of 1000.

The feature set can be modified in our design according to try out various results. The further information that can be added from the tagged data itself, are POS tags and the head word relationships.

Our feature vector would thus look something like - location-of-Action, manner-Of-Action, medium-Of-Communication, time-of-Event, place-Of-Event
instead of the feature set like - object-of-buy, object-of-like, object-of-sell, modifier-of-deed, etc.. used by Lin.

We expect to achieve semantic clusters that can distinguish let us say Movie-stars and Politicians from a pool of person names or Bank names from Country names. We also expect to achieve a fine granularity when keeping the Target Words for a particular frame together. For example, words like *blame, disapprove, dispute* mean opposite to *admire* and *appreciate*, though all of them come under the same frame *Judgement*.

2) The Parsing: FrameNet provides POS tags with each annotated sentence. Since we, need the parsing only to get the head word if a phrase and not a single word serves a particular role, the tags like NN1, referring to the main Noun (refer to the FN annotations) give the Head Word information for NPs and PPs. We do not employ any other parsers.

Word stemming is needed to unite various morphological variants of a word and have a single feature set for it. We have used KTEXT, a text processing program that uses the PC-KIMMO parser[14].

3) The Similarity Measure and Clusters: In our algorithm, for each element e, we construct a frequency count vector $C(e) = (c_{e1}, c_{e2}, \ldots, c_{em})$, where m is the total number of features and c_{ef} is the frequency count of feature f occurring in element e. We construct a **mutual information** vector $MI(e) = (mi_{e1}, mi_{e2}, , mi_{em})$, where mi_{ef} is the mutual information between element e and feature f, which is defined as:

$$mi_{ef} = log \frac{\frac{c_{ef}}{N}}{\frac{\sum_i c_{if}}{N} \times \frac{\sum_j c_{ej}}{N}}$$

where $N = \sum_i \sum_j c_{ij}$ is the total frequency count of all features of all elements.

We compute the similarity between two elements e_i and e_j using the *cosine coefficient* of their mutual information vectors:

$$sim(e_i, e_j) = \frac{\sum_f mi_{e_i f} \times mi_{e_j f}}{\sqrt{\sum_f mi^2_{e_i f} \times \sum_f mi^2_{e_j f}}}$$

From this, we obtain a similarity matrix for the elements. Using a greedy heuristic, similar to one presented in [1], we compute a set of cliques[1] for each element.

We then sort the collection of cliques in descending order of clique size and average similarity among clique members.

[1]A set of elements is a clique if each element in the set is among the top n most similar to each other, where n is a user defined parameter

Following this, we remove the cliques in C that have overlap with higher ranked cliques.

The result is a set of clusters.

V. Conclusion

Our study thus focussed on various issues related to finding semantic constraints that dictate lexical relations. We tried to approach it from the framework of the Role tagging and the SensEval task. We saw how semantic constraints can be realized by finding similarity measures and creating semantic classes or clusters.

Finally, we attempted the task of automatically creating semantic classes from the FrameNet and the semantic correlations it encodes within its tagged corpus. The task shows promise, in theory and considering preliminary results we achieved.

VI. Future Work

There can be substantial work done on the clustering front. The Clique approach gives us a number of small clusters, which can be easily merged by calculating similarities between *centroids* of cliques in a similar way as we calculated similarities between elements. Issues relating to low-frequency words need to be resolved.

The use of the clusters or chunks of words with semantic roles can be used to enhance the performance of Automatic Role Taggers. Accuracies achieved with this have to be estimated.

It would be interesting to compare the thesaurus generated from FrameNet, with other existing thesauri and those generated from Wordnet.

References

[1] D. Lin and P. Pantel, "Induction of semantic classes from natural language text," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. San Francisco, California: ACM Press, 2001, pp. 317 – 322.

[2] C. Johnson and C. J. Fillmore, "The framenet tagset for frame-semantic and syntactic coding of predicate-argument structure," in *Proceedings of the first conference on North American chapter of the Association for Computational Linguistics*. Morgan Kaufmann Publishers Inc., 2000, pp. 56–62.

[3] D. Gildea and D. Jurafsky, "Automatic labeling of semantic roles," *Comput. Linguist.*, vol. 28, no. 3, pp. 245–288, 2002.

[4] *A Generative Model for Semantic Role Labeling*, ser. Lecture Notes in Computer Science, vol. 2837. Springer, 2003.

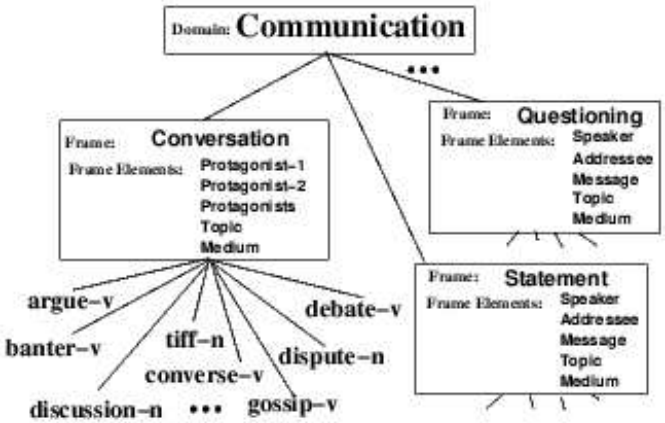
[5] U. Baldewein, K. Erk, S. Pado, and D. Prescher, "Semantic role labelling with similarity-based generalization using EM-based clustering," in *Proceedings of Senseval'04*, Barcelona, 2004. [Online]. Available: http://www.coli.uni-sb.de/ erk/OnlinePapers/senseval.ps

[6] R. S. Swier and S. Stevenson, "Unsupervised semantic role labelling," in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 2004, pp. 95 – 102.

[7] S. Banerjee and T. Pedersen, "An adapted lesk algorithm for word sense disambiguation using wordnet," in *Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing*. Springer-Verlag, 2002, pp. 136–145.

[8] N. Seco, T. Veale, and J. Hayes, "An intrinsic information content metric for semantic similarity in wordnet," in *Proceedings of ECAI 2004, the 16th European Conference on Artificial Intelligence*, 2004.

[9] J. R. Curran, "From distributional to semantic similarity," Ph.D. dissertation, University of Edinburgh. College of Science and Engineering. School of Informatics, 2003. [Online]. Available: http://hdl.handle.net/1842/563



[10] U. Baldewein, K. Erk, S. Padó, and D. Prescher, "Semantic role labelling with chunk sequences," in *Proceedings of CoNLL-2004*. Boston, MA, USA, 2004, pp. 98–101.

[11] P. Pantel and D. Lin, "Document clustering with committees," in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM Press, 2002, pp. 199–206.

[12] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, "Automatic subspace clustering of high dimensional data for data mining applications," in *Proceedings of the 1998 ACM SIGMOD international conference on Management of data*. ACM Press, 1998, pp. 94–105.

[13] P. Pantel and D. Ravichandran, "Automatically labeling semantic classes," in *HLT-NAACL 2004: Main Proceedings*, D. M. Susan Dumais and S. Roukos, Eds. Boston, Massachusetts, USA: Association for Computational Linguistics, May 2 - May 7 2004, pp. 321–328.

[14] Antworth and L. Evan, *PC-KIMMO: a two-level processor for morphological analysis*, ser. Occasional Publications in Academic Computing. Dallas,TX: Summer Institute of Linguistics, 1990, no. 16.

Appendix

(Taken from the documents accompanying the FrameNet Package)

The Berkeley FrameNet project is an online lexical resource for English, based on frame semantics and supported by corpus evidence.The aim is to document the range of semantic and syntactic combinatory possibilities (valences) of each word in each of its senses, through manual annotation of example sentences and automatic capture and organization of the annotation results

A **lexical unit** is a pairing of a word with a meaning. Typically, each sense of a polysemous word belongs to a different **semantic frame**, a script-like structure of inferences that characterize a type of situation, object, or event. In the case of **predicates** or governors, each annotation accepts one word in the sentence as its **target** and provides labels for those words or phrases in the sentence which fill in information about a given instance of the frame. These phrases are identified with what we call **frame elements** (FEs) - participants and props in the frame whose linguistic expressions are syntactically connected to the target word. A frame semantic description of a predicating word, derived from such annotations, identifies the frames which underlie a given meaning and specifies the

ways in which FEs, and constellations of FEs, are realized in structures headed by the word.

The annotated sentences are the building blocks of the database: marked up in XML, they form the data from which the lexical entry descriptions are derived. This format supports searching by lexical unit, frame, frame element, and combinations of these.