A generic evaluation of a categorical compositional-distributional model of meaning
An MSc thesis proposal

Jiannan Zhang
Department of Computer Science
University of Oxford
Supervisors: Bob Coecke, Dimitri Kartsaklis

April 12, 2014

Contents

1 Introduction

Expressing meaning is one of the core tasks in Computational Linguistics, because representation of meaning is the slate for many important applications, including paraphrase identification, question-answering systems, text summarization, information retrieval, machine translation and so on.

There have been for a long time two "camps" in Computational Linguistics. The pure compositional approach can be summarized by Frege's idea about the principle of compositionality: the meaning of a sentence is the functional composition of the meaning of the words. On the other hand, distributional semantics is based on the idea that the meaning of an expression depends on its context. As the web emerged, building vector spaces for English words became possible because billions of web pages contain a large amount of text . Vector space based approaches have achieved great successes during these years, and they have been applied to almost all applications. However, there is one problem of distributional semantics: it is hard to find a particular sentence in any corpus, even though a big corpus is used.

Although a sentence vector can not be created from the corpus directly, it can be constructed by composing the word vectors within the sentence. This is an approximation of sentence vector if such a vector exists. Some recent developments of compositional-distributional models obtained promising results. One important model to incorporate compositionality into vector space is Coecke et al.'s categorical model of meaning [3], and its effectiveness has been verified with simple sentences [5][9].

This proposal is about a more generic evaluation of this categorical compositional-distributional model of meaning. In particular, relative pronouns will be the main topic of concern in this research.

2 Compositional and Distributional Models

This section gives a brief review of previous developments in this field, including compositional and distributional semantics.

2.1 Models based on sentence structure

The tradictional way to represent sentence meaning is to use the syntax-semantics structure. Syntax can be represented based on a set of grammars, such grammars can be modeled in various ways, such as CFG (Context-free grammars), CCG (Categorial combinatory grammars), PCCG (Probabilistic CCG), PCFG (Probabilistic CFG) [19][4]. The semantics can be represented by logical expressions, one widely used way is to use lambda calculus expressions combined with higher-order logic [13][8]. Usually, the sentence to

process is parsed based on some grammars first, then a logical expression can be assigned based on the syntactic structure.

Some research further makes use of the topological structure to resolve this problem, especially in the applications that similarity comparison is important. In these methods, a sentence is firstly parsed by some dependency parser, the similarity of two sentences depends on the common edges of the trees (word-overlap) [18]. The model in [1] also takes synonyms into consideration. Remarkably, Vasile Rus, et al (2008) used a graph based model [14]. In his approach, text is firstly converted into a graph according to some dependency-graph formalism (text – dependency graph – graph). The graph is the representation of the sentence. Then, the sentence similarity problem is reduced to a graph isomorphism search (graph subsumption or containment). Rus' model obtained very positive results for paraphrasing.

2.2 Distributional models

The Distributional Hypothesis [7] tells us the meaning of a word is determined by the company it keeps (J. R. Firth), the implication is that if we can get the context of a word/phrase, we get the meaning. As a large volume of text on the Web becames available, this idea could be implemented. One early example is Turneys (2001) corpus-based model for synonym mining from the web [17]. One of the base lines of this approach is Mihalcea et al.'s model [11], in which both word similarity and word specificity are taken into account. The base-line accuracy of this work is 65.4%, whereas a random guess can achieve 51.3%. One such model to handle more complex and longer sentences is Mitchell and Lapata's work (2008) [**?**].

Usually the procedure can be summarized as follows: first parse large corpus, then build word vector spaces by counting co-occurrence of words in a fixed window. After this, every word can be represented by a vector, the basis of the vector space is a subset of the available words in the corpus. When a sentence comes, the meaning of a sentence can be constructed by adding/multiplying all the word vectors in the sentence. It proved itself to be simple and useful (e.g. in thesaurus extraction task (Grefenstette et al. [6]).

The limitation of pure vector space models is that the word order within sentences is lost, therefore it has the bag-of-words problem (e.g. "dog chases man" has the same vector as "man chases dog"). In recent years, people started to consider possible solutions to incorporate word order into vector space models.

2.3 Compositional-distributional models

One elegant idea is to represent sentences as tensor products of words (Smolensky, 1990, [12]). In addition, Clark and Pulman further proposed the

idea to embed word types into the representation of vectors [2]. Since tensor product is non-commutative, it preserves the word order. But long sentences produce higher dimensional vectors, and sentences of different length will fall into vector spaces of different dimensions.

More feasible models have been proposed in recent years. The Deep Learning based model developed by Richard Socher (2010) gained promising results. In particular, the model uses recursive neural networks (RNN), it works particularly well on identifying negations. His recursive autoencoder model has been applied to paraphrase detection,which also obtained very good results [16]. Coecke et al. (2010) proposed a categorical framework of meaning which takes advantage of the fact that both a pregroup grammar and a finite dimensional vector space share a compact closed structure. The construction of a sentence vector from word vectors is performed by tensor contraction. The framework preserves the order of words, and more importantly, does not suffer from the dimensionality expansion problems of other tensor-product approaches, since the tensor contraction process guarantees that every sentence vector will live in a basic vector space. It has been experimentally verified by Grefenstette et al. [5] and Kartsaklis et al. [9]. In the latter experiment, they used Oxford Concise School Dictionary and WordNet to extract some words with simple definitions (mostly phrases or subj-verb-obj sentences), then performed a classification task on the definitions regarding to the words.

3 Thesis Proposal

This MSc project is about one step further of the evaluation of the categorical model of meaning. In addition to the simple sentence structures, this research will build practical models for subject/object relative pronouns, and incorporate propositions and conjunctions based on the previous experience from previous research papers.

3.1 Relative pronouns

The previous evaluations only considered very simple sentence structures, while a very important part to construct more complex sentences is relative pronouns. A recent development of the categorical model gave a theoretical solution to model subject and object pronouns [15], where a noun phrase is modified by the relative clause. The model is based on Frobenius Algebra formalisms, subject/object relative pronouns are thought as objects that pass information from the relative clause to the noun phrase they modify.

3.2 Evaluation

The evaluation can be accomplished by the following steps:

- Build word vector space
  Based on some large corpus (British National Corpus / UKWAC / GIGAWORD), a basic word vector space can be constructed. Usually a 2K dimensional vector space is used in this word vector space. But using singular vector decomplisition (SVD), the vector space can be reduced to 300-dimensional.

- Build verb vectors
  From the word vector space, verbs can be constructed. A simple and effective way is to build matrixes for verbs instead of cubes, as originally described in [5], and later also used in [9] (named CPSBJ and CPOBJ).

- Parse selected sentences and assign vectors to words
  A given sentence can be parsed according to certain pregroups grammars. The sentences for evaluation is selected from the Oxford Junior Dictionary, following Kartsaklis et al.'s approach [9].

- Build sentence vectors
  Based on the types of words in the sentence, a sentence vector can be constructed using tensor contraction.

- Evaluation of accuracy
  The evaluation task is to calculate cosine distance between a definition (a sentence) and the words. A sentence will be assigned to the word which has smallest cosine distance. A correct classification should assign a definition to the correct word.

3.3 Resources and tools

The word vector space is consistent with the previous experiments, provided by Dimitri Kartsaklis and will be used as the basic word vector space in this experiment. To create the verb vectors and the whole model, Python's libraries will be used for experiments, including Shelve, NumPy, SciPy, PyBrain. C&C parser will probably be used to build pregroups parse.

3.4 Expectations of experiments

The core of this project is to build a practical categorical compositional-distributional model and perform evaluation on this model as described above. The model is supposed to handle more complex sentences and produce reasonably good results for the classification task.

4 Project Time Table

A tentative schedule is following:

Date	Task completed
May 8	Verb vectors constructed (CPSBJ/CPOBJ)
May 20	Selection of words from Oxford Junior Dictionry
May 31	Proper parse of selected definitions (sentences), build sentence vectors
June 9	Classification results (first round)
June 20	Refine results
June 30	Paper first draft
Afterwards	Refinement and further work

5 Conclusion

This research proposal completed a brief survey of the development of compositional and distributional models of meaning, and outlined the envisioned plan in the next 2-3 months. All planned tasks will be subject to change based on the ongoing process of the research.

References

[1] I. Androutsopoulos and P. Malakasiotis. A survey of paraphrasing and textual entailment methods. *arXiv preprint arXiv:0912.3747*, 2009.

[2] S. Clark and S. Pulman. Combining symbolic and distributional models of meaning. In *AAAI Spring Symposium: Quantum Interaction*, pages 52–55, 2007.

[3] B. Coecke, M. Sadrzadeh, and S. Clark. Mathematical foundations for a compositional distributional model of meaning. *arXiv preprint arXiv:1003.4394*, 2010.

[4] R. Ge and R. J. Mooney. A statistical semantic parser that integrates syntax and semantics. In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, pages 9–16. Association for Computational Linguistics, 2005.

[5] E. Grefenstette and M. Sadrzadeh. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1394–1404. Association for Computational Linguistics, 2011.

[6] G. Grefenstette. *Explorations in automatic thesaurus discovery*. Springer, 1994.

[7] Z. S. Harris. Distributional structure. *Word*, 1954.

[8] D. Kartsaklis. Compositional operators in distributional semantics. *arXiv preprint arXiv:1401.5327*, 2014.

[9] D. Kartsaklis, M. Sadrzadeh, and S. Pulman. A unified sentence space for categorical distributional-compositional semantics: Theory and experiments. In *In Proceedings of COLING: Posters*. Citeseer, 2012.

[10] C. Leacock, G. A. Miller, and M. Chodorow. Using corpus statistics and wordnet relations for sense identification. *Computational Linguistics*, 24(1):147–165, 1998.

[11] R. Mihalcea, C. Corley, and C. Strapparava. Corpus-based and knowledge-based measures of text semantic similarity. In *AAAI*, volume 6, pages 775–780, 2006.

[12] J. Mitchell and M. Lapata. Composition in distributional models of semantics. *Cognitive science*, 34(8):1388–1429, 2010.

[13] R. Montague. Universal grammar. *Theoria*, 36(3):373–398, 1970.

[14] V. Rus, P. M. McCarthy, M. C. Lintean, D. S. McNamara, and A. C. Graesser. Paraphrase identification with lexico-syntactic graph subsumption. In *FLAIRS conference*, pages 201–206, 2008.

[15] M. Sadrzadeh, S. Clark, and B. Coecke. The frobenius anatomy of word meanings i: subject and object relative pronouns. *Journal of Logic and Computation*, 23(6):1293–1317, 2013.

[16] R. Socher, E. H. Huang, J. Pennington, A. Y. Ng, and C. D. Manning. Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In *NIPS*, volume 24, pages 801–809, 2011.

[17] P. Turney. Mining the web for synonyms: Pmi-ir versus lsa on toefl. 2001.

[18] S. Wan, M. Dras, R. Dale, and C. Paris. Using dependency-based features to take the para-farce out of paraphrase. In *Proceedings of the Australasian Language Technology Workshop*, volume 2006, 2006.

[19] L. S. Zettlemoyer and M. Collins. Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. *arXiv preprint arXiv:1207.1420*, 2012.