Quantum Logic and Natural Language Processing

Ilya Makarov, Anastasia Frolenkova and Ivan Belov

National Research University Higher School of Economics,
School of Data Analysis and Artificial Intelligence,
125319, 3 Kochnovskiy Proezd, Moscow, Russia

Abstract. The paper presents a short summary on the applications of the quantum logic categorical constructions to the natural language processing. We give a brief overview on the topic of quantum logic in general, and in natural language processing, in particular. As a result, we discuss comparison of sentences and their representation in quantum logic formalism. The examples of using quantum diagrams are considered in order to understand text analysis in terms of quantum logic techniques.

Keywords: Natural Language Processing, Compositional Distributional Model of Meaning, Quantum Logic, Similarity Measures

1 Introduction

The main goal of our work is to provide an overview on combining words and grammar in the natural language processing using quantum logic categorical methods.

In [1], the authors created a new model of categorical formalism inspired by the quantum teleportation protocol. In terms of quantum information, quantum teleportation is a way of sharing information between processes. In [2], a framework of diagrammatic calculus was proposed for reasoning on quantum information processes. The main proposed advantage of the model is that it takes into account grammatical structure of a sentence as well as meanings of individual words.

2 Sentence Representation

We consider the natural language processing problem of forming the meaning of a sentence from the meanings of its parts. There are two core approaches to this issue: distributional models and formal semantics.

The former one does not take into account the interaction between syntactically linked words and is based only on the meaning of words, from which the sentence was composed. The word meaning is usually represented as vector, calculated from the context, in which this word occurs. One word is considered to be in context of another word, if it is in the n-word window of this word in the text. For instance, in the sentence 'The cat ate the mouse' words 'the' and 'ate' are in window of size one of the word 'cat'. The sentence meaning can be then determined as certain simple functions like addition or element-wise multiplication of vectors of constituting words. The main drawback of this method is its non-compositional nature. For example, the sentences 'The cat ate the mouse' and 'The mouse ate the cat' will have the same meanings in such type of models.

The latter one, in opposite, is concerned only with syntax. The meaning of a sentence is represented as a function derived from the grammatical structure of the sentence. The major aim of this approach is to represent natural language sentences as logical expressions. For this reason, the only thing we can say about the meaning of a sentence, is its truth or falsity. Thus, formal semantics deal with qualitative scale, and we cannot numerically estimate similarity of meaning of two sentences. In [1] the authors draw an analogy between quantum teleportation and natural language processing problem and apply diagrammatic calculus to grammatical structure of sentences.

Each of the two approaches has its own shortcomings, which are more or less solved in the other one. In [3], Coecke, Sadrzadeh and Clark presented so-called compositional distributional model unifying these two approaches of sentence representation. They described grammar structure of sentence using algebra of pregroups as a type-categorial logic, assigned different representations to words with different types, and took tensor product and inner product as operations that "glue" parts of a sentence together.

3 Similarity Analysis

One of the main applications of sentence representation models is to compute how close are the meanings of two given sentences. If a model can define, which sentences are similar and which are not, then we could apply this model for paraphrasing problem and improving search engines answers to queries taking into account answers for similar queries. We can also use similarity analysis as a model indicator: if model correctly determine meaning of any sentence, then it can identify which of them are close.

In [4], the authors gave the examples how to compare meanings of sentences using the model from [3]. Here, we present a short description of their method.

In distributional models all meanings of words are represented as vectors, that is why we cannot *apply* the verb to its subject and object. In order for the model to take it into account, all words are divided into atomic type and compound functional type. Nouns have atomic type, and verbs, adjective phrases, prepositional phrases, adverbs have compound types.

The authors of [4] consider the vector space $\mathbb{N}$, and the bases of $\mathbb{N}$ are annotated with 'properties' obtained by combining dependency relations with nouns, verbs and adjectives.

Nouns are assigned to vectors with elements equal to the number of times they have been in relations with the bases in the corpus of text.

Verbs are assigned with matrices in the following way: element i, j is equal to the number of times a noun with property i has been subject of the verb, and a noun with property j has been object of the verb in the corpus of text. In a similar way they define vectors for other syntax types.

For example, basis vectors might be associated with properties such as 'arg-hungry', denoting the argument of the adjective 'hungry', 'subj-eat' denoting the subject of the verb 'eat', 'arg-tasty' denoting the argument of the adjective 'tasty'. If a noun has occurred as an argument of 'hungry' 5 times, a subject of 'eat' 6 times and argument of 'tasty' 4 times, its vector is $(5, 6, 4)$. If there are 5 occasions when something hungry buys something tasty, the value in the cell $(1, 3)$ of the 'buy' matrix is 5.

Using the described algorithm one can compute the inner product and cosine similarity of two sentences. The main feature of this methodology consists of its ability to compare sentences with different grammatical structure due to reduction of sentences to one type.

Let us consider some simple text as an example how to get representations of nouns and verbs:

> 'Hungry predators eat tasty victims. Fast predators like tasty food. Small victims like tasty food. Big predator frighten tasty victims. Small victims don't like big predators. Fast predators eat fast victims. Hungry victims eat tasty food. Hungry predators frighten big victims. Small predators eat hungry victims.'

Let the bases are 'arg-big', 'arg-small', 'arg-tasty', 'arg-hungry', 'arg-fast'. The vector for the noun 'predator' is $(2, 1, 0, 2, 3)$, the vector for the noun 'victim' is $(1, 2, 2, 2, 1)$. The matrix representation for the verb 'eat' with respect to a given text corpus:

$$\begin{bmatrix} 0\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,2\,0\,0 \\ 0\,0\,0\,0\,1 \end{bmatrix}$$

In large text datasets, these matrices are not so sparse. Let us manually define example of non-trivial distribution for nouns and matrix for 'eat':

	Predators	Victims	Wolves	Rabbits
Big	6	1	5	0
Small	0	5	0	6
Tasty	0	3	0	4
Hungry	5	2	7	2
Fast	6	3	6	6

C_{ij}	Big	Small	Tasty	Hungry	Fast
Big	2	6	7	0	4
Small	0	3	5	0	0
Tasty	0	0	0	0	0
Hungry	4	6	8	0	3
Fast	2	3	3	0	3

Table 1. Distributions of parameters for the nouns and the verb 'eat'.

With the help of these representations we can compare meanings of the two sentences:

$$\langle\overline{S_1} \mid \overline{S_2}\rangle = \langle\overline{predators\ eat\ victims} \mid \overline{wolves\ eat\ rabbits}\rangle =$$
$$= \sum_{ij} \{\langle\overline{predators} \mid \overline{n_i}\rangle \cdot C_{ij}^{eat} \cdot \langle\overline{n_j} \mid \overline{victims}\rangle \times$$
$$\times \langle\overline{wolves} \mid \overline{n_i}\rangle \cdot C_{ij}^{eat} \cdot \langle\overline{n_j} \mid \overline{rabbits}\rangle \} = 148470, \quad (1)$$

where n_i are standard basis vectors with $n_i[j] = 1$ if $j = i$ and 0, otherwise.

Normalising it by the product of the lengths of both sentence vectors leads us to the cosine similarity value

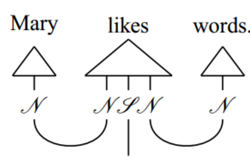
$$\frac{\langle\overline{S_1} \mid \overline{S_2}\rangle}{\sqrt{\langle\overline{S_1} \mid \overline{S_1}\rangle} \cdot \sqrt{\langle\overline{S_2} \mid \overline{S_2}\rangle}} = 0.958,$$

which is close to 1 and, thus, our sentences have close meanings.

4 Quantum Logic in Diagrams

In the proposed compositional distributional model the meaning of a sentence is obtained by applying some function of the tensor products of the words' meaning vectors. This function is a morphism corresponding to the grammatical structure of the sentence in the category of finite dimensional vector spaces of meanings. Basically it is a linear map of sentence's structure and words' meanings to the meaning of the sentence.

Such a linear map for the grammatical type reduction can be represented as a diagram. For example, lets look at the examples presented in [7]. The sentence *Mary likes words* can be represented along with grammatical types of the words as the following diagram:



The following pregroup type reduction corresponds to the diagram above:

$$n^{-1} \cdot n \cdot s \cdot n^{-1} \cdot n \leqslant 1 \cdot s \cdot 1 \leqslant s.$$

The idea behind the usage of diagrams is that we treat quantum logic as a theory of types of systems, processes and their interactions, and all the properties are also specified in terms of processes and their compositions. The types of systems are the objects in consideration, e.g. words in a sentence. Processes are morphisms of types, such as compound types and type reduction. And composition of morphisms form sequential application on processes. Then the diagrammatic calculus framework can be applied to reason on information processes.

5 Evaluation on NLP Problems

The abstract categorical model of Coecke et al.[3] was implemented by authors of [5]. It was evaluated on a word disambiguation task for transitive and intransitive sentences against a benchmark dataset provided by Mitchell and Lapata (2008). For both tasks datasets with potentially ambiguous verbs were provided in [6]. In these datasets, every verb is given with a context of potentially disambiguating nouns, subject nouns for intransitive verbs and pairs of subject and object nouns for transitive verbs. The task was to compose given verbs with corresponding nouns and evaluate similarity of different meanings of the verb.

The authors showed that on the task with intransitive verbs the categorical method performs on par with the other approaches. The other approaches are addition an multiplicative models, which basically are applications of meaning vectors addition and multiplication, correspondingly. There are still doubts on the small size of the context to be fully representative, however, considering more complex grammatical structure showed better performance in comparison to the other models.

6 Discussion

There are several directions of quantum logic study in NLP area. We aim to compare the diagrammatic reasoning versus semantic modelling.

We first plan to implement the model and run it on available corpus data in order to evaluate performance of the model and indicate possible complexity issues for further optimisation. As authors of [3] pointed out, a straightforward implementation of the model leads to efficiency issues that should be taken into consideration, investigated and dealt with optimisation techniques.

We would also try to extend the notion of the sentence meaning to a definition of the paragraph meaning. On can use the concept of syntactic thickets described, for example, in [8] for determination of paragraph grammar structure. One could think about considering further applications on identifying nonsense [10] or ambiguous [11] sentences. Another direction will be study of the theoretical foundations of computational complexity in terms of significant complexity constraints on Lambek Calculus as one of the important grammar extensions for NLP [9].

Acknowledgements. The work is supported by the Russian Science Foundation under grant 17-11-01294 and performed at National Research University Higher School of Economics, Russia.

References

1. Clark, S., Coecke, B., Grefenstette, E., Pulman, S.G., Sadrzadeh, M. (2013). A quantum teleportation inspired algorithm produces sentence meaning from word meaning and grammatical structure. CoRR abs/1305.0556.

2. Abramsky, S., Coecke, B. (2004) A categorical semantics of quantum protocols. In: Proceedings of 19th IEEE conference on Logic in Computer Science, IEEE Press, pp. 415425.
3. Coecke, B., Sadrzadeh, M., Clark, S. (2010). Mathematical foundations for a compositional distributional model of meaning. arXiv preprint arXiv:1003.4394.
4. Grefenstette, E., Sadrzadeh, M., Clark, S., Coecke, B., Pulman, S. (2014). Concrete sentence spaces for compositional distributional models of meaning. In Computing meaning, Springer Netherlands, pp. 71-86.
5. Grefenstette, E,. Sadrzadeh, M. (2011). Experimental Support for a Categorical Compositional Distributional Model of Meaning. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pp. 13941404.
6. Mitchell, J., Lapata, M. (2008). Vector-based models of semantic composition. Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, pp. 236244.
7. Zeng, W., Coecke, B. (2016). Quantum Algorithms for Compositional Natural Language Processing. Workshop on Semantic Spaces at the Intersection of NLP, Physics and Cognitive Science (SLPCS16) EPTCS 221, 2016, pp. 6775.
8. Galitsky, B., Ilvovsky, D.I., Kuznetsov, S.O., Strok, F. (2013). Matching sets of parse trees for answering multi-sentence questions. In RANLP, pp. 285-293.
9. Kanovich, M., Kuznetsov, S., Scedrov, A. (2016). Undecidability of the Lambek Calculus with a Relevant Modality. FG 2016, pp. 240-256.
10. Ferguson, T.M. (2015). Logics of nonsense and Parry systems. Journal of Philosophical Logic, vol. 44 (1), Springer Netherlands, pp. 65-80.
11. Wurm C., Lichte T. (2016). The Proper Treatment of Linguistic Ambiguity in Ordinary Algebra. In: Foret A., Morrill G., Muskens R., Osswald R., Pogodalla S. (eds) Formal Grammar. FG 2016, FG 2015. Lecture Notes in Computer Science, vol 9804. Springer, Berlin, Heidelberg, pp. 306-322