Natural Language Processing with Distributional Compositional Models

Jean Maillard

Supervised by: Dr Stephen Clark, Computer Laboratory, University of Cambridge

UNIVERSITY OF CAMBRIDGE

Natural Language Processing is the field of Articial Intelligence concerned with getting computers to perform useful tasks involving human language. Traditionally, the field has been divided into two "camps": compositional and distributional.

Distributional Models

Distributional models are best summarised with a quote by John Firth, who said *"You shall know a word by the company it keeps"*. The principle is that the meaning of a word can be captured by the words which appear in its contexts. This information is used to assign a vector to each word. For example, suppose we need to characterise the words *cat*, *dog* and *snake*.

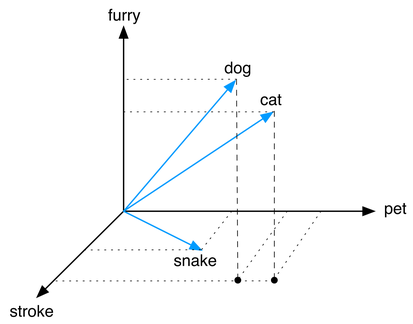

Figure 1: *A simple vector semantic model: the words cat, dog and snake are represented here as vectors in a three-dimensional vector space.*

A distributional model would analyse any large corpus of text—such as Wikipedia, or the archives of the New York Times—and observe the following:

- cats and dogs are often described as *pets*, but snakes not as often
- snakes are very unlikely to be described as *furry*, cats are often furry and dogs perhaps even more so
- all three can be found as the object of the verb *stroke*

In this simple example, each one of these three characteristics is assigned to one dimension of a vector field. The frequency with which a characteristic occurs for a certain word determines the magnitude of the corresponding vector component. Therefore, the words could be represented as in Fig. 1.

Advantages The advantage of these models is that vectors give us a notion of distance, so it is easy to compute how close in meaning two words are. In the example above for instance, *cat* and *dog* have a lot in common, while *snake* is quite different.

Disadvantages While these models have been shown to be successful at understanding the meaning of individual words, they fail to compute the meaning of sentences.

Compositional models

Compositional models are an older approach, based on classical ideas from logic. They can be summarised by Frege's *principle of compositionality*, which states that:

> The meaning of a phrase is a function of the meanings of the parts of the phrase and how those parts are put together.

These models treat human languages as programming languages, which are compiled down to some formal language such as lambda calculus. The way the sentence is parsed determines how the individual objects are combined together in order to yield a well-formed formula.

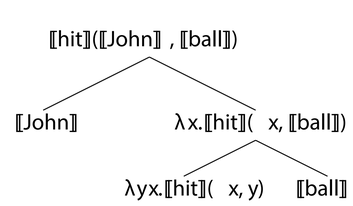

Figure 2: *A parse tree for the simple sentence "John hit the ball".*

Advantages By analysing the grammatical structure and the interaction between words, compositional models are able to obtain the logical meaning of a sentence.

Disadvantages The compositional approach is mostly concerned with how the meanings of words combine, but is not concerned with their individual meanings. Moreover, the complex rules for the logical interpretation of natural language can be prohibitively difficult to learn.

Distributional Compositional models

Distributional Compositional models are a new approach, which combine the strengths of boths Distributional and Compositional models.

Like in Distributional models, nouns are represented by vectors living in the *noun space*, denoted $\mathbb{R}^n$. Furthermore, sentences are described by vectors in a *sentence space*, $\mathbb{R}^m$. We will then be able to compare the meaning of two sentences using the familiar dot product.

From this we can deduce the representation of all other syntactic categories, as follows.

Example "John hit the ball" is a full sentence, and therefore it should be represented as a vector in $\mathbb{R}^m$. Similarly, *John* and *ball* are noun vectors.

$$\text{John,ball} \in \mathbb{R}^n,$$
$$\text{John hit the ball} \in \mathbb{R}^m.$$

Combining a noun and a determiner yields a *noun phrase*. These objects have similar properties to nouns, therefore we expect "the ball" to be also described by a noun vector. It follows that the determiner *the* is a function:

$$\text{the} : \mathbb{R}^n \to \mathbb{R}^n,$$
$$\text{the} : \text{ball} \mapsto \text{the ball}.$$

Similarly, a transitive verb with subject and object forms a sentence:

$$\text{hit} : (\mathbb{R}^n, \mathbb{R}^n) \to \mathbb{R}^m,$$
$$\text{hit} : (\text{John,the ball}) \mapsto \text{John hit the ball}$$

This gives us a recipe to obtain the meaning of a sentence, by applying the functions of the various parts of speech to the vectors of the nouns.

Tensors One further assumption we can make is that these functions between vector spaces are multilinear. In that case, each part of speech will be associated to a specific tensor:

- Nouns are vectors in $\mathbb{R}^n$.
- Transitive verbs are third-order tensors in $\mathbb{R}^m \otimes \mathbb{R}^n \otimes \mathbb{R}^n$.
- Determiners are second-order tensors in $\mathbb{R}^n \otimes \mathbb{R}^n$,

and so on. Therefore, the specific form of the tensors for each syntactic category is known and fixed. Their components are found using machine learning techniques, and depend on the specific task the model is trying to achieve.

Conclusions

Vector space models of meaning have been shown to be effective at a number of natural language processing tasks. The Distributiona Compositional model expands these models by incorporating ideas from Compositional models. This allows the model to obtain not only the meaning of individual words, but also of full sentences.

References

- Coecke B, Sadrzadeh M, Clark S (2011) *Mathematical Foundations for a Compositional Distributional Model of Meaning*, Linguistic Analysis, 36
- Grefenstette E (2013) *Towards a Formal Distributional Semantics: Simulating Logical Calculi with Tensors*, arXiv:1304.5823

Acknowledgements: the PMC Bursary organisers and sponsors, St John's College and the Computer Laboratory.