8th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2017

Analyzing Weak Semantic Map of Word Senses

Alexander A. Eidlin[1], Maria A. Eidlina[1] and Alexei V. Samsonovich[1,2]

[1]Department of Cybernetics and BICA Lab, Institute for Cyber Intelligence Systems National Research Nuclear University "Moscow Engineering Physics Institute" Kashirskoe Shosse 31, Moscow, 115409, Russian Federation
[2]George Mason University, Fairfax, VA 22030, USA.
a.aidlin@gmail.com, mashaeidlina@mail.ru, asamsono@gmu.edu

Abstract

A weak semantic map, as opposed to a strong semantic map, allows for a choice of coordinates that are characterized by definite semantics: e.g., valence, arousal, dominance. Weak semantic maps of words can be built from synonym-antonym dictionaries, by pulling synonyms together and antonyms apart. Polysemy is one of the problems with this approach. Indeed, typically one and the same word has multiple meanings, while it has to be represented by only one point on the map. To solve this problem, it seems natural to use word senses rather than words as the map elements. In this work, we consider a semantic map of word senses built from the thesaurus of Microsoft Word for the Russian language. To determine senses of words with the purpose of its further representation on the built map of word senses is developing a method, the result of which is presented in this article. The main conclusion is that semantic maps of word senses have an advantage over semantic maps of words, since they allow for a context-specific evaluation of word meaning. This approach removes the restriction caused by polysemy - the multiplicity of senses of one word.

Keywords: semantic space, weak semantic map, semantic map of word senses, sentiment analysis, semantic search

1 Introduction

A semantic map, or semantic space, is an embedding of a set of concepts (represented by words or otherwise) into a metric space, the geometry of which captures the semantics of relations among concepts. Weak semantic maps [1] are characterized by semantic interpretability of their coordinates. In this article we consider a vector-space-based weak semantic map of word senses, built using the Microsoft Word Thesaurus for the Russian language, where each sense is represented by a set of words. The method of weak semantic map construction is described in our previous works [2-4].

The main hypothesis addressed here is that a semantic map of word senses has an advantage over semantic maps of words, since it removes the restriction caused by polysemy - the multiplicity of senses of one word. This should be the case, if (a) coordinates of individual word senses can be

defined by the map more precisely than coordinates of words with polysemy, and (b) word senses can be effectively disambiguated. To test the part (b), we develop a method for disambiguating the sense of a selected word in text using the semantic map of word senses.

Usually, a person disambiguates the relevant sense of a word with a polysemy based on the context in which the word is used. In Russian texts it is customary to denote a certain sense implied in the text by using several words-synonyms having this sense in common. Therefore, it seems logical to assume that one could correctly identify the relevant sense of a word with a polysemy as the sense, the words defining which occur in the text most frequently. This, in essence, is the idea of our approach.

2 Materials and methods

2.1 Semantic map of word senses

The Microsoft Office Russian Thesaurus was used to construct a weak semantic map of word senses. The general method of weak semantic map construction is described in [2]. Here the same method was applied to the synonym-antonym graph of word senses, that was constructed first. To do this, we had to re-process the available Thesaurus. In the Microsoft Office Thesaurus, each word has a set of senses, and under each sense it has a list of synonyms. Each unique sense is therefore implicitly defined in this Thesaurus by a synset: i.e., a maximal set of words, all of which have one sense in common. Then, any two overlapping synsets can be considered (likely) synonyms. When defined in this sense, synonymy means semantic relatedness rather than semantic identity. Defining antonym relations among synsets is a bit trickier, because antonyms of individual words are not separated according to senses in this Thesaurus; instead, a (possibly empty) list of antonym words is given for each word. Therefore, it was decided that two synsets can be considered (likely) antonyms, if more than one word antonym pairs can be found, such that one of the words of the pair belongs to the first synset, and another – to the second.

2.2 Method of word sense disambiguation

The essence of the developed method for the determination of a sense of a word is as follows. At the first stage all those senses which contain the chosen word in its set of defining words are selected from the dictionary of the semantic map. Since a word with a polysemy was selected, a list of several senses is formed at this stage (the element of the list is the sense, i.e. the list element is a set of words which has a selected word). The next stage considers texts in which the chosen word is used, and in different texts this word is used in different senses. Each text is treated as a "bag of words", and each word is sequentially viewed from the text. From the list of senses selected at the previous stage, the sense is selected, the words of which are most often found in the text.

A flowchart of determining a sense of a word in one text according to the developed method is shown in Figure 1.