УДК 53.072

Quantum modeling of text similarity judgments

Diosa Echeverri. J. C. (ITMO University),

Научный руководитель – Ph.D, Surov I. (ITMO University)

- In this work a new method for quantum behavioral modeling is proposed. The method allows to account for contextuality of semantics by generalizing classical semantic differential to the quantum domain.

- The novelty of the method is to use text similarity instead of actual decision probabilities. The method would streamline experiments and information analysis, and the reason for this is because it is a cheaper and faster method.

- We report the idea of the method and preliminary positive results.

Введение.

In 1992, the researchers Tversky and Shafir through some experiments showed how individuals can change their decisions depending on the level of information or the context of each decision. Later, the same experiment was carried out in other countries with different languages and cultures. The results obtained by the researchers were similar in different cases, confirming the phenomenon in which the results are similar regardless of changes in the variables of the experiment (among others).

This experiment has been modeled on different occasions using the model of human behavior based on quantum theory, finding a kind of parameters called phase parameters. If this phase parameter is similar for the same experiment (that is, there is quantum stability) in different contexts, it could be used for other purposes.

However, the main problem with quantum cognitive modeling is the inability to determine the phase parameters before the experiment. This makes it impossible to use quantum behavioral models as a predictive tool.

Here we aim to solve this problem with a new approach allowing to extract decision probabilities from text similarity judgments.

Основная часть:

We propose to use some text similarity measures (Leacock and Chodorow, Wu and Palmer, Resnik, Lin and Jiang and Conrath) as probabilities of choice for different events and show how the quantum phase behaves.

For this new approach we resort to the use of the classical semantic differential, which refers to the constitution of a measure and / or representation of a framework under which a word or a symbol can be measured or identified.

This semantic differential is generally obtained by conducting surveys with which the representation framework of a concept is built. In other words, the constitution of a frame of reference consumes large amounts of resources and time. However, since this measure can change depending on the context in which the term is used, we propose the use of the quantum semantic differential. This new differential is defined as the measure or frame of representation of a concept based on the context that accompanies it. In this way we go from analyzing a concept as unique to analyzing it in multiple contexts.

The quantum model has advantages for modeling human decisions since it captures the contextuality of human cognition and allows to explain questions that for classical theory are irrational.

This time we compare words and their synonyms (such as game or gamble) with the final choice text of the game, “play again” and its synonyms. We carry out the same comparison with the variants of the game, win or lose and the final choice of the game and we obtain the probabilities. In this way we obtain 3 probabilities that allow us to find the quantum phase parameters and observe their behavior. To calculate the probabilities, we create a matrix where all the words from text 1 (for example “game”) are related to all the words from the final choice or text 2, in this case “play”. These relationships are calculated using the WordNet tool and it is different for each of the algorithms mentioned above. In this way, the resulting matrix is expanded to the number of synonyms that each word of the first text has (or synset in the WordNet tool) compared to the number of synonyms that each word has in the second text or final choice, in this case resulting a matrix of size n * m for each algorithm used.

Once these relationships have been calculated, we select those relationships that have the highest value for each of the words in text 1 with the words in text 2 and vice versa. In this way we simplify our matrix and obtain two matrices, the first with the relations from text 1 to text 2 and the second with the relations from text 2 to text 1.

The idea is to relate these data to what happens in real life, this is the reason way we calculate for each of the words in text 1 and text 2 an indicator known as Term frequency – Inverse document frequency or TF-IDF, which is the importance of a word in a set of documents. In this case we will use the TF-IDF of the Reuters corpus that contains more than 10 thousand documents and relates real content. Once we have this data, we use the text similarity equation and obtain a single value that we will use as probability to calculate the quantum phase.

This approach has several advantages. The first is that it is a cheaper way to conduct experiments. The second is that it is a faster way to perform experiments and obtain the quantum phases to make estimates and the third is that it does not require conducting an experiment in real life.

Выводы.

It is possible to explain the quantum phase phenomenon by using text analysis.

A way to find the quantum phase in a faster and cheaper way is presented.

The results of different experiments are presented.

Diosa Echeverri J. C. (автор) Подпись

Surov I. A. (научный руководитель) Подпись