Semantic Quantum Correlations in Hate Speeches

Francesco Galofaro

Department of Philosophy and Education Sciences (DFE), University of Turin
Centro Universitario Bolognese di Etnosemiotica (CUBE), University of Bologna

The intervention shows the first results of a research conducted on a corpus of 7000 posts collected on the Reddit social network during the 2016 American presidential campaign. The research is the result of a collaboration between Berkeley DLab, who shared the corpus, LSI - CentraleSupélec and CUBE. Thanks to funding from the Anti-Defamation League, the corpus has been labelled to apply Machine Learning techniques: 400 posts have been labelled as "hate speech" by human analysts. Galofaro, Toffano and Doan applied to both sub-corpora (hate and non-hate speeches) an analysis technique inspired by Greimas's structural semantics, Eco's semiotics, and Quantum Information Retrieval (van Rijsbergen). Each text was formalized as a semantic network using the HAL technique. We then measured the semantic similarity between two key words formalized as two-word vectors with the classical measure of cosine-similarity. We compared it with the degree of quantum correlation between them measured with the Born rule. This correlation, linked to the co-occurrence of the word vectors in the same contexts, extracts from the latter useful information to characterize the considered semantic relationships ("presence of correlation", "absence of correlation" or "presence of anticorrelation"). In this manner, the new technique allows to overcome some critical aspects of the Machine Learning techniques currently in use, being based on the meaning of the text and not on the way in which the human analyst labels the corpus.

Galofaro, F., Toffano, Z. and Doan, B.-L., "Quantum Semantic Correlations in Hate and Non-Hate Speeches", Proceedings CAPNS 2018, arXiv:1811.02701

Galofaro, F., Toffano, Z. and Doan, B.-L. (2018), "A quantum-based semiotic model for textual semantics", Kybernetes, Vol. 47 No. 2, pp. 307-320. https://doi.org/10.1108/K-05-2017-0187