

Abstract

A set of features for the detection of extremist materials online is proposed in this thesis. Extremism is a phenomenon, which can be detrimental to social stability and safety, and with the spread of the Internet its effect has grown even more significantly. Thus, creating effective filters of such content is a timely task. The proposed features are based on the idea of extremism understood as a multifaceted phenomenon based on varying aspects of language, such as semantics, structure, sentiment and narrative. In the course of this thesis it has been found that implementing these four aspects into a feature set (dubbed the linguistic features) provided robust and general results, which outperformed features derived from bag of words or simple lexicon based methods, as well as purely structural approaches. The thesis is organized as follows. In the introduction, there are established the research objectives as well as terms and tools used in the analysis. Afterwards, the linguistic features are introduced and tested in detecting a single type of ideology. In the next step the linguistic features are tested on a new ideology in a different context (social media). In this step, word embeddings are also introduced as a basis for a method of computing features representing semantics. Finally, a combination of the computed semantic features and linguistic features are tested against a benchmark in classification tasks related to three different ideologies. In all cases, this combination of features achieved results significantly better than the benchmark.

Streszczenie

Ekstremizm nie jest nowym zjawiskiem społecznym, jednakże wraz z rozwojem nowoczesnych technik informatycznych, oraz ułatwionego do nich dostępu, efekty jakie propagacja ekstremizmu ma na społeczeństwo stały się bardziej doniosłe. W związku z tym, konieczne staje się znajdowanie rozwiązań mających na celu skuteczne monitorowanie, oraz wykrywanie treści ekstremistycznych w Internecie. Poniższa rozprawa przedstawia propozycję zdefiniowania, oraz opracowania technik wydobywania cech z tekstów, które mogą być wykorzystane do skutecznego wykrywania treści ekstremistycznych. Przedstawione w rozprawie cechy reprezentują wiele aspektów języka, aby móc lepiej odwzorować wieloaspektowe zjawisko, jakim jest język politycznego ekstremizmu. Cechy przedstawione w rozprawie można podzielić na dwie kategorie: cechy lingwistyczne, reprezentujące takie właściwości tekstu, jak jego struktura, nacechowanie emocjonalne, lub powtarzające się elementy narracji, oraz semantyczne category, przedstawiające najczęściej pojawiające się tematy w dokumencie. Wkładem w rozwój informatyki, jaki oferuje ta rozprawa, jest zarówno zestaw owych cech, jak i metody pozwalające na automatyczne ich wydobywanie. Cechy poddano testowi w zadaniu klasyfikacyjnym, związanym z odkrywaniem języka ekstremizmu w tekście. zadanie przeprowadzono na dokumentach reprezentujących trzy różne skrajne ideologie, oraz porównano z dotychczas stosowanymi metodami. Zaproponowane zestawy cech osiągnęły lepsze wyniki, niż dotychczas stosowane metody.