Review of Analysis and Automatic Recognition of Extremism in Online Texts written by Bartłomiej Balcerzak

Lisa Kaati
Uppsala 2018-02-27

I have reviewed the thesis Analysis and Automatic Recognition of Extremism in Online Texts written by Bartłomiej Balcerzak. My recommendation is to accept the thesis. The review follows the suggested methods and evaluation criteria for reviewers of PhD theses submitted to the Faculty of Informatics of PJAIT.

What are the research problems and objectives considered in the thesis, and have they been sufficiently clearly described by the author?

The research described in the thesis is about how machine learning can be used to identify and analyze online extremism in written text. The problem is difficult since it deals with problematic concepts such as extremism and radicalization. These concepts lack clear boundaries and can have different meanings depending on who uses them and in what context. Extremism in this context is defined using three factors: fringe, supremacy, and violence.

Detecting extremism is done using four different types of linguistic perspectives: semantics, structure, sentiment, and narrative. The objective of the thesis is to 1) find a feature set that can be used to identify extremism, 2) find linguistic features that can be used to identify narratives, 3) test the proposed sets of features on different ideologies and different kind of data and 4) use document semantics to find relevant features.

The author has done an excellent job in breaking down his research into more manageable research objectives. The objectives are clearly described.

Does the thesis contain an appropriate analysis of state of the art (based on global scientific literature, current knowledge and applications in industry)? Does the analysis of related work demonstrate sufficient expertise of the author? Have the conclusions of the review of related work been sufficiently clearly stated?

The contribution of this thesis is highly interdisciplinary and related work can be found in many areas - in computer science research as well as in social science. I think that the author shows an understanding of the issues related to definitions of extremism. The author also shows that there is related research in many different fields. The conclusion of the review of related work is clearly stated.

Does the research described in the thesis use a correct scientific methodology?

The research uses a correct scientific methodology. Relevant data is collected; different algorithms are used to train models using a satisfactory approach. The models are tested and standard metrics are used to describe the performance. The author also provides reasoning and explanation of the results.

What are the original and innovative contributions of the author, and what is the position of these contributions compared to the state of the art?

As noted by the author, previous research that is related to detecting extremism has focused on detecting propaganda from a single ideology or terrorist group. The innovation of this research lies in the approach of identifying extremism independent of ideology. I also think that the idea of combining different feature sets (semantics, sentiment and lexical features) is innovative and creative.

How do you evaluate the publication record of the candidate?

Bartłomiej Balcerzak has published a large number of conference papers in different areas. Several of his publications are in the area of applied machine learning and text analysis. I think he has enough publications for a PhD but in the future, he should focus more on journal publications.

Did the author present his results correctly and convincingly? (Please evaluate the clarity, conciseness, correctness of the thesis or presented research articles).

The author provided his results in a clear and correct way and did also provide a discussion on the limitations of his approach. All results are presented in the text and with more exact numbers in the appendix.

The arguments and motives for different choices are convincing. However, one thing that is not clear to me is the datasets. The author states that identifying data containing extremism is hard and requires a great deal of consideration. I would have liked to see a list of the groups that were selected to express extremism. I would also like to see some kind of description of how the selection was done from a linguistic perspective. Are there certain cues in the text that makes a page qualify as an extremist page? I also would like to see some motivation on why the far left was used as a case. I would have guessed that the far left in US are not considered violent (similarly to many right-wing movements that are still in favor of democracy and does not believe in violence to change society). Instead of considering the far left, I think that it could have been interesting to include extreme groups such as violent animal rights groups, violent anti-abortion movement, and anti-fascistic movements.

What are the weak and strong points of presented research results?

The strong point of the research is that the idea and the approach are interesting and provide new insights into linguistic markers of extremism. The research presented in this thesis is well described and many different approaches to solve the problems have been considered.

I would say that the weak point in this research is the data. The author's definition of extremism includes a factor of violence, which I think is very important. I would even suggest changing the use of extremism to violent extremism since the term extremism can be misinterpreted (I think many people consider themselves to be extreme in but they do not support the use of violence – I would guess that the interest in this work is to find violent extremism). Since the factor violence is included in the definition of extremism it is more a matter of choosing suitable terminology. To me, it is hard to understand if the definition of extremism is fulfilled in all data sources. I would have liked to see a list of groups from where data was collected and also a more detailed description of how the assessment of each document was done. Data is the foundation of this research and it is important to be transparent with what data that is used.

Using the terminology far left and far right might also be misleading (see my comment in the previous section). When the term far right is used what kind of organizations are included? Do all of them encourage violence or are some of the just very far to the right politically? I would prefer to use the term violent right-wing extremism, right-wing extremism or radical nationalistic movements. A discussion on the use of terminology and what organizations/sites that are included in each category would be interesting.

The low annotator agreement shows that it is difficult to find, in particular, group references and enemy references. The sample size of 65 webpages is mentioned but how large was each website? My understanding is that parts of sentences describing each narrative feature were identified. How many sentence samples were annotated in total?

I am missing a discussion on how this kind of models should (or should not?) be used. This kind of research may pose a threat towards freedom of speech and privacy and a discussion about that would be interesting to read.

Can the models that are trained on one type of dataset be used to identify violent extremism from a different ideology? A test of the models “in the wild" would be very interesting to get an understanding of how well this kind of technologies would work.

What is the contribution of the thesis to the discipline of information technology?

The contribution of this thesis is that information technology is used to solve difficult problems in social science. Using machine learning to detect extremism and as a mean to study common factors in expressions of violent extremism is very interesting and it can provide researcher with more understanding of the driving factors for violent extremism. A new set of features for detecting extremism is presented in this thesis as well as a study of the usefulness of these features.

Are the presented achievements of the author sufficient to grant him/her a doctoral degree in the field of technical sciences in the discipline of computer science or software engineering?

Yes. I think the author has studied a relevant problem and used appropriate methods. The results are presented in a satisfactory way with references to relevant literature. The results justify the conclusions.