Multimodal Sentiment Analysis on Ideological Extremism Content, based on Social Cognitive Theories

Christopoulou Stavroula-Afroditi, 11412984

March 4, 2017

Abstract

In the century of vast technological developments and intellectual achievements, the emergence of numerous supporters of extreme doctrines is remarkable. The following thesis through sentiment analysis techniques, will approach the large spectrum of extremism by examining relations between members of extremist social networks, and deep causes of accession of those members in aforementioned groups.

1 Personal details

My email `mailto:evitaams@gmail.com`

My supervisors email `mailto:s.rudinac@uva.nl`

My github account wiki `https://github.com/StavroulaChristopoulou`

2 Research question

An extremist group, is a group of individuals whose values, ideals and beliefs differ significantly of what society considers normal. They operate on the edge of conventional political groups in society, however the problem with defining a group as extremist is that the group does not view itself as being extremist.

Internet has allowed violent extremist groups to disseminate their propaganda messages more easily and has become the ideal platform to communicate with one another and with their followers[1]. The interactive nature of these online platforms allows for dynamic research on intrinsic motivations and deeper correlations.

Having access to a data set with pictures of extremist group members of a well known social network, we can scrutinize closely the members that compose this social group and their special characteristics. Since the social composition is multivariate, interesting questions arise.

The main research question of this dissertation is "Whether it is possible to estimate "extremeness" of ideologies in extremist groups, based on automatic, multi-modal sentiment analysis".

Additionally, the corresponding sub questions to this topic are the following: Which of the groups present higher levels of ideological "extremeness"? Which topics evoke extreme sentiments? Which are the motives behind the participation and exposure of extremists on social media, given the fact that the possibility of social marginalization and labelling exists? Is it possible to correlate this accession with depressive feelings?

3 Related Literature

The related literature that will support this dissertation will be based both on technological and social research. Most of the literature will contain extraction and interpretation of information on sentiment analysis based on social networks. Furthermore, psychological approaches and theories of motivation will be apposed. Literature: [2], [3], [4], [5], [6], [7], [8], [10], [11], [12] .

4 Methodology

The data set contains pictures of extremists that are members of a specific social network. They are both men and women, and they come from various cultural background, socioeconomic level, educational level etc.

In order to answer the research question "Whether it is possible to estimate "extremeness" of ideologies in extremist groups, based on automatic, multimodal sentiment analysis", the following methodology will be implemented. Firstly, the implementation will take place in two stages, the sentiment analysis on visual content, and the sentiment analysis on textual content that corresponds to the pictures.

The pictures will be classified as positive or negative content pictures, based on pre-determined features. After this implementation the second step will be to assign the textual content to images, and determine whether it is negative or positive. This will be performed with Sentiment analysis methods that are used to detect polarity in thoughts and opinions of users in on line social media.

The software that will be used, SentiWordNet 3.0, will automatically annotate all WordNet synsets according to their degrees of positivity, negativity, and neutrality (with the three scores summing up to 1). This procedure will include four substeps: (1) seed set expansion, (2) classifier training, (3) synset classification, and (4) classifier combination.

The evaluation of the accuracy of SentiWordNet 3.0 will consist of the comparison of a small, manually annotated subset of WordNet against the automatic annotations of the same synsets as from SentiWordNet.

WordNet sense mappings will be derived using a number of heuristics. Each mapping will have a confidence value associated to it, ranging from 0 (lowest confidence) to 100 (highest confidence). Heuristics used for the determination of mappings include the comparison of sense keys, similarity of synset terms, and relative tree location (comparison of hypernyms).

Alternatively, for each Micro WordNet synset that can not been mapped by the previous method, the computation will be implemented through the similarity between its gloss and the glosses of all WordNet3.0 synsets, where

a gloss will be represented by the set of all character trigrams contained in it. Similarity is computed via the Dice coefficient: $QS = \frac{2|X \cap Y|}{|X| + |Y|}$ where higher Dice(X, Y) value means stronger similarity.

In order to evaluate the quality of SentiWordNet it will be tested how well it ranks by positivity (resp., negativity) the synsets in Micro WordNet 3.0. The comparison of rankings will be implemented with the p-normalized Kendall, between the standard rankings and the predicted rankings $\tau = \frac{(\text{nd}) - (\text{nu})}{n(n-1)/2}$.

[12] .

Furthermore iFeel, a Web application that provides an open API, will allow the detection of sentiments and will give access to SentiWordNet. Therefore, it will be useful as a comparison tool of the strengths and weaknesses of our sentiment analysis and provide graphical visualization.

5 Risk assessment

A possible issue is that Sentiment analysis approaches based on low-level features have the limitation of low interpretability. In case that the first research question will not flourish, as a back up plan will be the implementation of the second research question : "Which of these groups have higher sentiment levels", where the methodology and implementation will be followed as explained above.

Furthermore, multiple tools will be available for the suppleness of the interpretation on sentiment evaluation, such as the Circumplex model of affect and the Whissel dictionary of affect in language.

6 Project plan

The thesis implementation will be completed in 14 fourteen weeks:

1-3 week: Literature review and extraction of social-psychological theories that will support research question and sub-questions.

2-5 week: Data collection, textual and visual(image) processing

3-6 week: Data Analysis, Sentiment analysis

7 week: Scores processing and interpretation of the outcomes

8 week: WordNet map and visualisation of the outcomes

9-10 week: Interpretation of the outcomes and sociological evaluation

10-12week: Thesis authorship

13 week: Thesis defence

References

[1] Anne Aly, Stuart Macdonald, Lee Jarvis, Thomas Chen *Violent Extremism Online* 2016: New Perspectives on Terrorism and the Internet

[2] Bing Liu *Sentiment Analysis and Opinion Mining* 2012: Morgan and Claypool Publishers)

[3] Federico Alberto Pozzi Elisabetta Fersini Enza Messina Bing Liu *Sentiment Analysis in Social Networks* 2016: 1st Edition

[4] Christian Rohrdantz, Lars-Erik Haug, Daniel A. Keim, Umeshwar Dayal, Ming C. Hao*Feature-based Visual Sentiment Analysis of Text* 2012: Publication 3, 2

[5] Soujanya Poria, Erik Cambria, Newton Howard, Guang-Bin Huang, Amir Hussain *Neurocomputing* 2016: Fusing audio, visual and textual clues for sentiment analysis from multimodal content

[6] Stefan Siersdorfer, Enrico Minack, Fan Deng, Jonathon Hare *Analyzing and Predicting Sentiment of Images on the Social Web*

[7] Jianbo Yuan, Quanzeng You and Jiebo Luo *Sentiment Analysis Using Social Multimedia* 2015: Switzerland

[8] Jana Machajdi, Allan Hanbury *Affective image classification using features inspired by psychology and art theory* 2010: Firenze, Italy

[9] Xia Hu, Lei Tang, Jiliang Tang, Huan Liu*Exploiting Social Relations for Sentiment Analysis in Microblogging* 2013

[10] Fei Yan, Krystian Mikolajczyk*Deep Correlation for Matching Images and Text* 2015: University of Surrey

[11] Sandra Graham, Bernard Weiner*Theories and Principles of Motivation*: University of California

[12] Stefano Baccianella, Andrea Esuli, Fabrizio Sebastiani*SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining* Pisa, Italy

Document Streams