A Lexicon Based Approach to Detect Extreme Sentiments

Special Session along with ICIMP 2020
The Fifteenth International Conference on
Internet Monitoring and Protection
September 27, 2020 to October 01, 2020
Lisbon, Portugal
https://www.iaria.org/conferences2020/ICIMP20.html

Authors:
Sebastião Pais - sebastiao@di.ubi.pt
Irfan Tanoli - irfan.khan.tanoli@ubi.pt
Miguel Albardeiro - miguel.albardeiro@ubi.pt
João Cordeiro - jpaulo@di.ubi.pt

Presenter

- Irfan Khan Tanoli
  - Post-Doc Researcher at University of Beira Interior.
  - Receive PhD Degree at Gran Sasso Science Institute, L'Aquila, Italy.
  - Received the M.S degree in software engineering from Technical University of Madrid, Madrid, Spain.
  - Current Research Interest
    - Natural Language Processing.
    - Controlled Natural Processing.
    - Semantic Analysis.
    - Software Engineering.
    - Machine Learning.
  - Current Research Project
    - Moves Project (http://moves.di.ubi.pt/).

Co-presenter

- Sebastião Pais
  - Professor at the Computer Science Department, the University of Beira Interior.
  - Researcher at NOVA-LINCS and GREYC Laboratory.
  - Current research and teaching interests:
    - Artificial Intelligence.
    - Statistical Natural Language Processing.
    - Lexical Semantics.
    - Machine Learning.
    - Unsupervised and Language-Independent Methodologies.
  - Current Research Project
    - Moves (http://moves.di.ubi.pt/).
    - C4 - Cloud Computing Competence Centre (c4.ubi.pt).

Outline

- Introduction
- Proposal
- Approach
- Experimental Setup
- Results
- Discussion
- Conclusion
- Future Work

Introduction

- Extreme sentiment analysis is to detect and classify extreme sentiment(s):
  - represent(s) the most negative and positive sentiments.
- An extreme sentiment is the worst or the best view, judgment, or appraisal formed in one's mind about a particular matter or people.

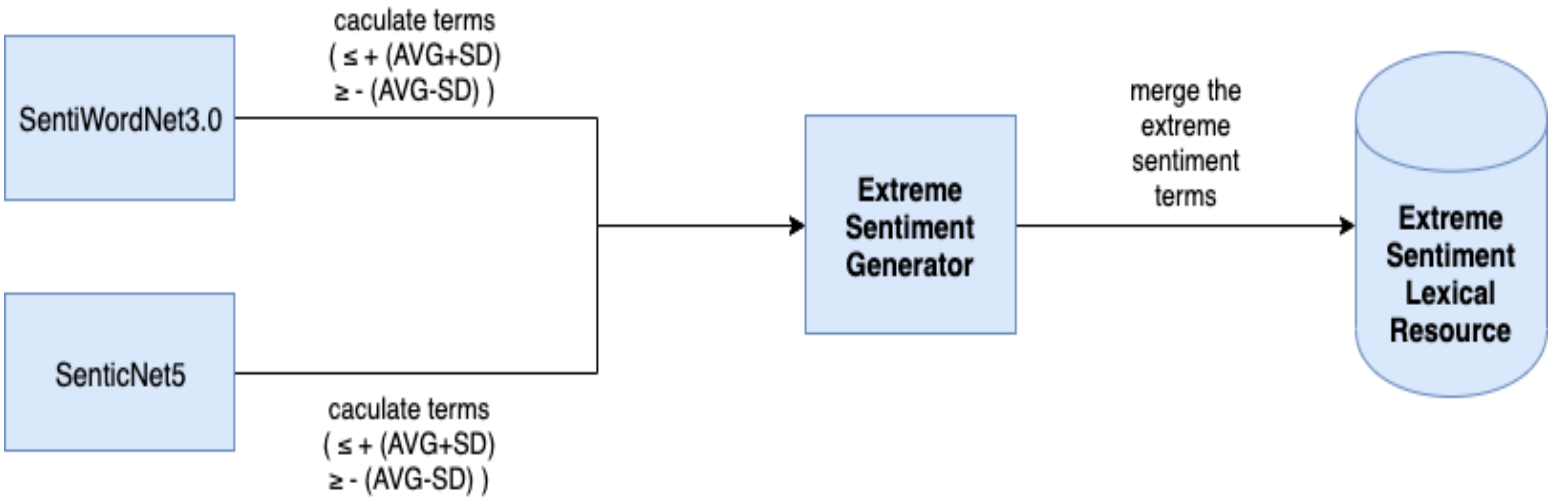
Proposal

- An interesting unsupervised and language-independent approach for detecting people's extreme sentiments on social platforms.
- Analyze two standard corpora:
  - SENTIWORDNET 3.0.
  - SenticNet 5.
- Extract extreme words having a high negative and positive polarity, reflecting people's extreme sentiments.

Approach

- Design and develop a prototype system called Extreme Sentiment Analyzer (ESA) composed of two different components:
  - Extreme Sentiment Generator (ESG).
  - Extreme Sentiment Classifier (ESC).
- ESG is based on statistical methods.
- Apply it on SENTIWORDNET 3.0 and SenticNet 5 to generate a standard lexical resource known as ExtremeSentiLex.
- Embed ExtremeSentiLex in the ESC
  - run on the compilation of five different datasets, which are constituted of social network and media posts.

ARCHITECTURE



Extreme Sentiment Collection Process.

Approach

The approach based upon two steps:

1. Define Extreme Polarity.
2. Generating Extreme Sentiment Lexicon.

Approach:
1. Define Extreme Polarity

- Develop a python application "Extreme Sentiment Generator (ESG):
  - Calculate the average and standard deviation of terms from the original lexical resources,
  - Filter and save it into a new lexical resource.
- Define two conditions in ESG:

if Tp > Average + StandardDeviation then

The term is classified as Extreme Positive

end if

if Tn < Average StandardDeviation then

The term is classified as Extreme Negative

end if

Approach:
1. Define Extreme Polarity

- SENTIWORDNET 3.0
  - Average for positive terms: 0.366
  - Standard Deviation for positive terms: 0.211
  - Extreme polarity for positive terms: 0.577
  - Average for negative terms: 0.412
  - Standard Deviation for negative terms: 0.230
  - Extreme polarity for negative terms: 0.642
- For Example:
  - SENTIWORDNET 3.0 terms generated by ESG:
  - ultrasonic 0.375 (non positive extreme)
  - selfless 0.875 (positive extreme)
  - thrash 0.125 (non negative extreme)
  - abduction 1 (negative extreme)

Approach:
1. Define Extreme Polarity

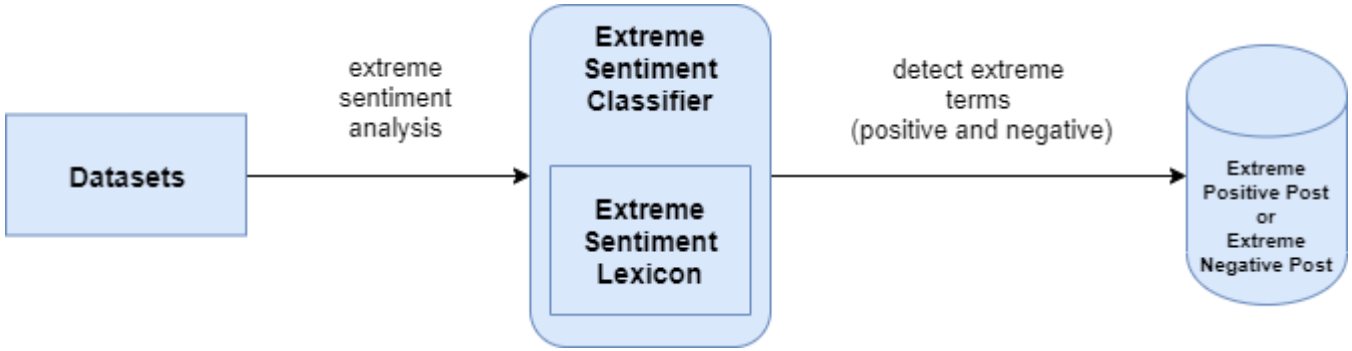
- Senticnet 5
  - Average for positive terms: 0.504.
  - Standard Deviation for positive terms: 0.362.
  - Extreme polarity for positive terms: 0.866.
  - Average for negative terms: -0.616.
  - Standard Deviation for negative terms: 0.306.
  - Extreme polarity for positive terms: -0.922.
- Example:
  - grace 0.79 (positive non extreme).
  - pioneer 0.97 (positive extreme).
  - anemic -0.918 (negative non extreme).
  - traffic -0.97 (negative extreme).

2. Generate Extreme Sentiments Lexicon

- Merge both files obtained from SENTIWORDNET 3.0 and SenticNet 5 terms stored in CSV files.
- Example:

Term	Score
Absolutely	+0.88
Accept	+0.93
Acknowledgeable	+0.95
Acne	-0.96

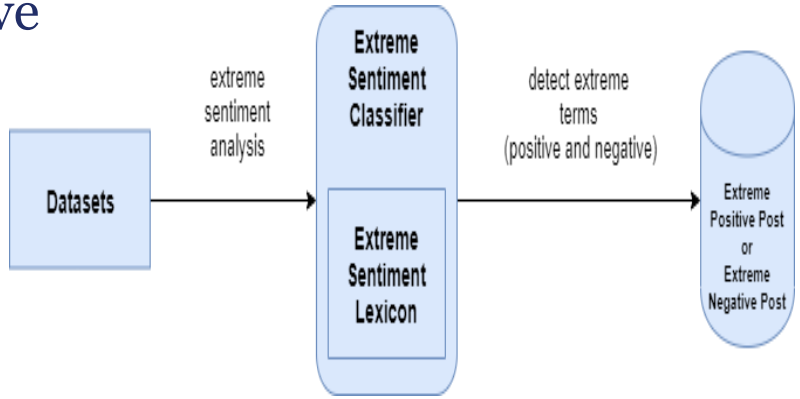
Experimental Setup



Performance testing of Extreme Sentiment Classifier

Experimental Setup

- Develop Extreme Sentiment Classifier (ESC) having ExtremeSentiLex embed in it.
- We perform the experiments five datasets:
  - TurnToIslam.
  - Ansar1.
  - RT-Polaritydata
  - T4SA.
  - Sentiment 140.
- We use confusion matrix to analyze the performance computing:
  - Recall.
  - Precision .
  - Accuracy.
  - F1-score.



Results

	RT-polarity	Sentiment 140	T4SA
Recall EP	91%	95%	98%
Recall EN	21%	50%	45%
Precision EP	59%	65%	89%
Precision EN	65%	92%	86%
F1 score EP	72%	77%	93%
F1 score EN	32%	65%	59%
Accuracy	60%	72%	89%

Table: Indicators Of System Efficiency

Discussion

- The overall status of acquired results are quite satisfactory.
- In some evaluation measures, for certain datasets, we have more than 90%.
- The results of Sentiment 140 and T4SA are really prominent, where none of the values is less than 45%.
- RT-polarity, there appear some low values on negative terms, i.e., recall and F1 score for EN.
- The measure of accuracy for all data resources is equal to or greater than 60%:
  - indicating the overall performance of the approach is better.
- The calculation of recall, precision, f1 score and accuracy for Ansar1 and TurntoIslam is not performed:
  - Since these datasets are directly referred to as 'Correction of common misconceptions about Islam'.
  - There is a possibility of radical participants may occasionally show their support for extremist fundamentalist militant groups.

Conclusion

- Presented an unsupervised and language-independent approach for people's extreme sentiments detection on social media platforms.
- Our approach is based upon:
  - Defining extreme polarity for terms.
  - Generating extreme sentiments lexicon by relying upon two standard lexical resources:
    - SENTIWORDNET 3.0 and SenticNet 5.
- Experimented with five different social networks and media data .
- We provided a standard lexicon:
  - Useful other researchers to exploit it for sentiment analysis studies as well
  - Antiextremism authorities to identify people's extreme sentiments
    - On social networks and can prevent violent extremism.

THANK YOU