Textual Entailment Recognition on Large Datasets

Presented By,
SUBALAKSHMI SHANTHOSI S
(186001008) M.E (CSE)

Supervised By,
Dr. Aravindan Chandrabose

REVIEW-1
AUGUST 14,2019

Outline

- Introduction
- Motivation
- Problem statement
- Literature Survey
- Existing System
- Proposed Work
- Dataset Description
- References

Introduction

Textual Entailment:

- Textual entailment (TE) in natural language processing is a directional relation between text fragments. The relation holds whenever the truth of one text fragment foll ows from another text.
- In other words, Textual entailment is the task of determining whether a hypothesis is true, given a premise.
- Textual Entailment relations:

  text: If you help the needy, God will reward you.

  1. A *positive TE*(text entails hypothesis):
     hypothesis: Giving money to a poor man has good consequences.
  2. A *negative TE*(text contradicts hypothesis):
     hypothesis: Giving money to a poor man has no consequences.
  3. A *non TE*(text does not entail nor contradict):
     hypothesis: Giving money to a poor man will make you a better person.

Introduction(contd...)

Textual Entailment on large corpus:

- Significant decrease in accuracy when large datasets are used to train textual entailment recognition systems.
  There are two major reasons for this counter-intuitive result:
  - Homogeneity: Hand curated datasets which are aligned to a specific task by combining different sources data which have distinct characteristics. Thus, these datasets do not always capture the kind of entailment queries that naturally arise in an end task.
  - Corpus size: Larger dataset of higher magnitude are needed to capture the complex properties characterizing entailment.
- Social media texts which are publicly avaliable is voluminous and possible partial relationship as similarity between the texts(T-H) pair is relaxed and found either of the two to lacks or contains additional information than the other text.

Motivation

- Textual Entailment is predominantly dependent on high quality,huge annotated corpus. However, until now, the scarcity of such data on one hand, and the costs of creating new datasets of reasonable size on the other, have represented a bottleneck for a steady advancement towards achieving the state-of-the-art performance.
- It is essential to derive a mechanism for adopting relaxed relation aka. Partial Textual Entailment(PTE) on social media text as either one of the texts lack/suffice with evidence.
- Also the source of knowledge base plays a pivotal role in maximising the accuracy(F1 score).
- Research on Partial Entailment is an unexplored paradigm.

Problem Statement

- Given an collection of annotated tweets using an annotation model that encompasses following levels indicating the depth of detail on Offensive Language Identification and Categorisation task.Our goal is to find the presence of offensive language and the severity of its existance during impact assessment.

  Sub-Tasks identified for this problem are:

  1. Sub-task A - Offensive language identification
  2. Sub-task B - Automatic categorization of offense types
  3. Sub-task C - Offense target identification.

Literature Survey

SSN_NLP at SemEval-2019 Task 6: Offensive Language Identification in Social Media using Traditional and Deep Machine Learning Approaches[1]

- Offensive Language Identification Dataset (OLID), a new large-scale dataset of English tweets with high-quality annotation of the target and type of offenses is used for training and testing models[6].
- Traditional Machine Learning and Deep Learning techniques are employed to identify offensive languages.
- In Deep Learning approach dataset is preprocessed by using Bi-LSTM to vectorise the tweets and CNN is used for detecting aggression in text.
- In Traditional Machine Learning approach feature vectors are constructed using TF-IDF scoring and Multinomial Naive Bayes (MNB) and Support Vector Machine (SVM) with Stochastic Gradient Descent optimizer to build the models.

Literature Survey (Contd...)

An Exploration of State-of-the-art Methods for Offensive Language Detection[2]

- OLID dataset is partially preprocessed to annotate user as @USER and URL's as URL.Words are transformed to lower case and removing alphanumeric symbols leaving behind only letters,digits and underscore as acceptable characters.
- Word2Vec is used for generating word embeddings. Rather than using pre-trained models such as scraped Wikipedia pages, a combination of transforming vectors (max,averaging) is used for generating multi-word single vector for further processing.
- Auto-Keras was used to train a pre-trained BERT representation.But,BERT regards capitalized and syntactically incorrect statements as noise thus failing to categorise the level of abusive nature in that particular sentence.
- FastText trained with random search for fine tuning the existing pretrained model with scraped Wikipedia pages.

Literature Survey (Contd...)

Recognition of Partial Textual Entailment for Indian Social Media Text[3]

- Partial Textual Entailment in NLP is used for defining partial entailment relationship between T-H pair.
- Emperical Definition of Partial Entailment: Defines four categories of PTE.
    1. PTE-I :Original textual entailment relationship with bi-directional correspondence.
    2. PTE-II : If either of T or H entails whole meaning of another and contains additional information.
    3. PTE-III:If a portion of H entails from a portion of T or vice verse.
    4. PTE-IV: If T or H does not entail from H or T.
- Sequential Minimal Optimization(SMO) based PTE recognition along with other Machine Learning Techniques like Random forest (RF), Decision tree (DT), Logistic regression (LR) is used.

Literature Survey (Contd...)

Absit invidia verbo: Comparing Deep Learning methods for offensive language[4]

- Bag-of-words model is used as dataset initially and then word2idx for neural network model.
- Extensive use of PyTorch , Keras, scikit-learn, and Natural Language Toolkit(NLTK) is observed.
- PyTorch is selected to implement CNN, Keras for RNN and Linear Regression using scikit-learn.
- For Offensive language identification Logistic Regression,LSTM and B-LSTM outperforms other models.
- Each tasks is trained with 90% of samples and 10% of samples are used for testing.L2 regularisation is used for optimising results.

Literature Survey (Contd...)

Benchmarking Aggression Identification in Social Media[5]

- In this work,a dataset of 15,000 aggression-annotated Facebook Posts and Comments each in Hindi (in both Roman and Devanagari script) and English are provided for training and validation. For testing, two different sets - one from Facebook and another from a different social media - were provided.
- The aim of this shared task is Classification of Social Media Text as overt aggression, covert aggression and nonaggression.
- Multilingual lexicon of aggressive words. The lexicon is obtained by automatic translation from an handmade lexicon of offensive words in Italian, with minimal human supervision. The original words are expanded into a list of their senses. The senses are manually annotated to filter out senses that are never used in an offensive context.
- Even LSTM pretained FastText vector performed better than conventional Neural network models.

Exisiting System

- The existing approaches have used Deep Learning and pre-trained models like BERT,FastText,CNN or Conventional Machine Learning techniques like Naive Bayes and Stochastic Gradient Descent for identification and categorisation of Offensive language in Social media texts[1][2].
- Partial Textual Entailment can be used for Offensive Language Categorisation as it aims at finding SMO for partial matching and reducing reduntant information.[3]
- Deep learning approach have been predominantly used and shows promising results even on Multi-Lingual datasets.[5].

Proposed Work

- In our work, the offensive language usage can be identified in social media text by defining PTE rules by using Sequential Minimal Optimization(SMO) method.
- Increasing the dataset population by using Semantic textual similarity for determining paraphrases of offensive slang sentences.
- Finding means to incorporate Transfer Learning approach by using pre-trained models like XLM ,BERT and XLNet.

Dataset Description

- Forum of Information Retrieval Evaluation in 2017(Fire2017 IRLeD)
- Precedence Retrieval:
- The dataset consists of 200 current case which is formed by removing the links to the 2000 prior case and the prior case which have been cited by the case in current case.
- Catchphrase Extraction:
- The dataset consists of 100 documents and their corresponding gold standard catchphrases for training and the test set consists of 300 separate documents whose catchphrases were to be found.

References

- D. Thenmozhi, B. Senthil Kumar, Chandrabose Aravindan, S.Srinethe,*SSN NLP at SemEval-2019 Task 6: Offensive Language Identification in Social Media using Traditional and Deep Machine Learning Approaches.*Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019) 739 – 744,2019.

- Harrison Uglow,Martin Zlocha,Szymon Zmyslony,*An Exploration of State-of-the-art Methods for Offensive Language Detection.*arXiv:1903.07445 1– 5,2019.

- Dwijen Rudrapal , Amitava Das , Baby Bhattacharya ,*Recognition of Partial Textual Entailment for Indian Social Media Text.*Computacin y Sistemas, Year 23, Vol. 23 143 – 152,2019.

References

- Bogdan Lazarescu, Christo Lolov , Silvia Sapora, *Absit invidia verbo: Comparing Deep Learning methods for offensive language.*,arXiv:1903.05929v3 1– 5,2019.
- Ritesh Kumar , Atul Kr. Ojha , Shervin Malmasi , Marcos Zampieri ,*Benchmarking Aggression Identification in Social Media* ,Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying, 1 – 11 , 2018.
- Zampieri, Marcos and Malmasi, Shervin and Nakov, Preslav and Rosenthal, Sara and Farra, Noura and Kumar, Ritesh,*Predicting the Type and Target of Offensive Posts in Social Media* , Proceedings of NAACL,2019.

THANK YOU