Textual Entailment Recognition on Large Datasets

Subalakshmi Shanthosi S (186001008)
ME CSE, Semester 3

Dr. Aravindan Chandrabose
Supervisor

Project Review: 1 (14 August 2019)

Department of Computer Science and Engineering
SSN College of Engineering

1 Introduction

The intent of Textual entailment is to identify whether one piece of text can be plausibly inferred from another It is a major generic core problem in Natural Language Understanding(NLU).

The great potential of integrating (monolingual) TE recognition components into NLP architectures has been reported in several areas, such as question answering, information retrieval, information extraction and document summarization.

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.

Crowdsourcing services have been recently used for creation of NLP resources is that the acquisition and annotation of large datasets, needed to train and evaluate NLP tools and applications, can be carried out in a cost-effective manner.

The accuracy of pretrained model decreases on increase in the size of dataset used for training.This counter-intuitive result is due to Hetrogenity and increased corpus size.

2 Motivation

In Natural language processing field, Recognising textual entailment (RTE) is of paramount importance.All the complex NLU problems have discern entailment as it's sub NLP task. Image recognition and Computer Vision see rely on RTE to improve results obtained applying visual techniques alone.

Applications dealing with text needs a semantic framework for applied semantics and RTE may provide such framework.

Textual Entailment is used for modelling language variability in NLP Tasks which are given below.
• Variability of semantic expression : Same meaning can be inferred from different texts.
• Ambiguity in meaning of words : Different meanings can be inferred from same text.
Textual entailment is also used for Machine Translation Evaluation.Applying such entailment phenomenon on MT evaluation provides the well formedness of the output sentence generated by the translation system.
Textual entailment (TE) in natural language processing is a directional relation between text fragments.The relation is directional because even if ”t entails h”, the reverse ”h entails t” is much less uncertain.
Role of Knowledge source for TE is very crucial as success of the entailment system heavily depends on the background knowledge. Background knowledge includes facts, conventions, peculiar language features such as certain metaphors, idioms, proverbs, common beliefs etc. Challenges in Textual Entailment:

• Paraphrasing: Same meaning can be inferred from different texts.
T and H contains same fact but expressed with different words.
Example: ”the cat devours the mouse” is a paraphrase of ”the cat consumes the mouse”
• Strict Entailment: T and H carry same facts such that one can be inferred from another.
Example: Yahoo bought Overture and Yahoo owns Overture.

3 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.

The three shared tasks are as follows:

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

4 Literature survey

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

In this paper,OLID[6] dataset is used which contains a collection of annotated tweets and an annotation indicating the level of Offensive language severity assesment and categorisation done.Traditional Machine Learning and Deep Learning techniques are employed to identify offensive languages. Deep Learning methods uses Bi-LSTM to derive vectorised tweets and uses attention mechanism to map the offensive slang words to a named a named entity - ”GRP”.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.

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

In this paper,the proposed system works on OLID dataset which 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 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 even when it is modelled to work on large datasets.

4.3 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.Thus,PTE plays an important role in different NLP applications

like Text summarisation and Question answering by reducing redundant information.
In this paper,contributions are as follows:

1. Extending classical TE by including two categories of partial entailment for Bengali tweets.
2. Manual creation of PTE annotation on Bengali Social Media Text corpus.The corpus includes total 5916 numbers of tweet pairs.

Emperical Definition of Partial Entailment: Defines four categories of PTE.

1. PTE-I : Preserverance of original entailment relationship and the relationship is bi-directional. i.e) H entails from T and T entails from H.
2. PTE-II : This category has two conditions to follow.
   - Condition I:If H entails from the whole meaning of T and have additional information, then it is a category of PTE-II and represents as,

     $$(X_H \text{ Entails from T}) + Y_H$$

   - Condition II: If T entails from the whole meaning of H and have additional information, then it is also a category of PTE-II and represents as,

     $$(X_T \text{ Entails from H}) + Y_T$$

3. PTE-III: If a portion of H entails from a portion of T or vice verse, then it is a category of PTE-III and represents as,

   $$(X_H \text{ Entails from T}) + Y_H$$
   $$(X_T \text{ Entails from H}) + Y_T$$

4. PTE-IV: If T or H does not entail from H or T, then its a category of PTE-IV and represents as Non-entailed.

Sequential Minimal Optimization(SMO) based PTE recognition approach is used on Social Media Text for partial matching. Future work is to make this system robust to handle code-mixed tweets.

4.4 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.

4.5 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. This paper reports the results of the first Shared Task on Aggression Identification which was organised jointly with the First Workshop on Trolling, Aggression and Cyberbullying (TRAC - 1) at COLING 2018. The aim of this shared task is Classification of Social Media Text as overt aggression, covert aggression and nonaggression. The dataset considered has subjective inaccurate annotation and contains code-mixed texts which are noisy and need to be carefully filtered.
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.

5 Existing 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].

6 Proposed system

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.

References

[1] 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.

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

[3] 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.

[4] Bogdan Lazarescu, Christo Lolov , Silvia Sapora, *Absit invidia verbo: Comparing Deep Learning methods for offensive language.*,arXiv:1903.05929v3 1– 5,2019.

[5] 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.

[6] 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.