SSNCSE_NLP@DravidianLangTech-EACL2021: Meme classification for Tamil using machine learning approach

B. Bharathi & Agnusimmaculate Silvia A
Department of CSE
Sri Siva subramaniya Nadar College of Engineering
Kalavakkam - 603110
`bharathib@ssn.edu.in`
`agnusimmaculate18011@cse.ssn.edu.in`

Abstract

Social network is the place for the creation or sharing of information, ideas or other forms of expression among people. This exchange of information may contain offensive, trolling or malicious contents targeting users or communities. The memes is one of the way to include trolling. A meme is an image or video that represents the thoughts and feelings of a specific audience. The challenge of dealing with memes is that they are region-specific and their meaning is often obscured in humour or sarcasm. A meme is a form of media that spreads an idea or emotion across the internet. The multi modal nature of memes, postings of hateful memes or related events like trolling, cyberbullying are increasing day by day. Memes make it even more challenging since they express humour and sarcasm in an implicit way, because of which the meme may not be offensive if we only consider the text or the image. In this paper we proposed a approach for meme classification for Tamil language that considers only the text present in the meme. This work explains the submissions made by SSNCSE_NLP in DravidanLangTech-EACL2021 task for meme classification in Tamil language. We achieve F1 scores of 0.50 using the proposed approach using the test-set.

1 Introduction

More recently, as memes and GIFs dominate the social feeds; typo-graphic visual content has become a considerable element of social media. 'Meme' is a viral image or video often altered by internet users for humorous effect. These convey human expressions but with a wide range of emotions, and often require context to fully understand humor and sarcasm. "Sarcasm is a type of sentiment where people express their negative feelings using positive or intensified positive words in the text (Bharti et al., 2016)". The meme classification is handled by considering both the linguistic features as well as visual features. This task will be considered as multi modal approach. The approaches of natural language processing and computer vision are to be used for the meme classification task. Using this idea, a model has to be designed which deals with the challenges associated with the semantic nature of text and image, and performs classification on the emotional responses obtained.

Tamil language is official language of Tamil Nadu, Singapore, and Sri Lanka (Chakravarthi, 2020b,a; Chakravarthi and Muralidaran, 2021). From the 6th century BCE to the 3rd century CE, the Sangam (koodal) period encompassed ancient Tamil Nadu, Kerala, and portions of Sri Lanka (then known as Tamilakam). It was named after the illustrious Madurai-based Sangam (koodal) academies of poets and scholars. Tamil was the first Indian classical language to be listed as classical language, and it is one of the world's oldest classical languages still in use by common people. Inscriptions in the Tamil language written in Tamili (Tamil-Bharmi) script have been found in Sri Lanka, Thailand, and Egypt. This work explains the submissions made by SSNCSE_NLP in DravidanLangTech-EACL2021 task for meme classification in Tamil language.

The remainder of the paper is organized as follows. Section 2 discusses the related work on meme classification and related tasks using text and images. The dataset about the shared task is described in Section . Section 4 outlines the features and machine learning algorithms used for this task. Results are discussed in Section 5. Section 6 concludes the paper.

2 Related work

Researchers experimented with few approaches used in the text sentiment classification task as well

as emotion classification task using images. Sentiment analysis from the text is essential in understanding the idea behind the meme. The authors of (Wang and Wen, 2015) concentrate on generating meme descriptions through non parametric estimation approaches which uses lexical features, POS features, named-entity features, frame-semantics features and dependency triples. In (Amalia et al., 2018), the memes are classified by applying image processing and OCR Tesseract which are combined with Naïve Bayes Algorithm. The survey of using visual linguistic information used for meme classification, image capationed are discussed in (Afridi et al., 2020). The MultiOFF multimodal meme data set for offensive content detection data set was developed by (Suryawanshi et al., 2020a) (Suryawanshi et al., 2020b). Most of the research work on meme classification uses both linguistic and visual features for better performance. In our proposed approach, only text based features were used. The proposed work have been described in the next section.

3   Data-set Analysis

The data set given by the shared task organizers contains the image file and transcriptions present in the image file. All the image files are in jpeg format and have text embedded on them in the Tamil language (Suryawanshi et al., 2020c),(Suryawanshi and Chakravarthi, 2021). Later, each meme has been annotated by three annotators (native speakers) in troll or not_troll class. The class label (troll or not_troll) of the data is embedded in the filename. E.g. A meme with filename "troll_0.jpg" has a label "troll", similarly, "Not_troll_0.jpg" has label Not_troll. Further, the transcription of captions for each file is given in the data set. The data distribution of train and test set is given in Table 1

Table 1: Data distribution for meme classification in Tamil

Class label	Train set	Test set
Troll	1282	395
Not troll	1018	272

4   Experimental setup and features

From the given text transcription, n-gram and TFIDF and BERT embeddings are extracted as features. As the content of the memes is a mix of Dravidian language grammar in Roman lexicons along with English grammar, it becomes challenging to find pre-trained models for this context. So a simple n-gram approach is considered. Also, the advancements done by the transformer model for pre-training and the availability of multilingual trained models encourage to experiment with BERT pre-trained embeddings.

The extracted features are used to train machine learning models such as Multi-Layer Perceptron, Random Forest, K-nearest neighbour, and the performance of these models are compared. For shared task for meme classification is provided only with training set. To analyze the performance of the proposed approach, 5 fold cross validation is used. The metrics such as accuracy and weighted average F1 score are analyzed. The machine learning model implementations and the metrics of comparison is used from Scikit-learn [1].

4.1   Count and TFIDF n-grams

As the data-set consists of sentences in Tamil as well few colloquial words. So the n-gram model trained from scratch is considered for the tasks and this strategy has shown good results with HI-EN and BN-HI-EN datasets (Mishra et al., 2018). Basic count-based and Term Frequency Inverse Document Frequency-based models are compared for feature extraction. The char TFIDF is calculated as explained in equation 1.

For each char t in a document d from the document set D TD-IDF is calculated as:
N is the total number of documents.

$$tfidf(t,d,D) = tf(t,d).idf(t,D) \qquad (1)$$

where

$$tf(t,d) = log(1 + freq(t,d)) \qquad (2)$$

$$idf(t,D) = log(Ncount(d\epsilon D : t\epsilon d)) \qquad (3)$$

Different character n-gram are constructed with n-gram range varying from 1 to 7. Out of which n-gram range of 1-5 showed good results for meme classification task in Tamil.

4.2   BERT Embedding

The memes are given in Tamil language with few colloquial words. For experimental purpose BERT multilingual model is used for this task. BERT multilingual model trained on a large corpus from 104 different languages (Devlin et al., 2018), which

---

[1] https://scikit-learn.org/stable/

includes Malayalam, Tamil, and English could be considered. A fixed dimension embedding is generated by sentence-transformers, with the pre-trained base-multilingual-cased model implementation as explained in (Reimers and Gurevych, 2020).

These embeddings are then used to classify as Troll or Not-troll by training a machine learning model. Different models such as Random Forest, K-nearest, and Multi Layer Perceptron (MLP) are compared, out of which the MLP with 512 hidden layers, generated the best results.

5 Observations

The output from n-gram embedding is a sparse matrix of high dimension. And as the total number of training samples is limited, it is observed that the Random Forest model and MLP model is producing the same performance for this task. Whereas the embedding generated by the BERT model is of 512 dimensions for each meme. The model comparison based on train-set evaluation for the meme classification and cross-validation with k = 5 is shown in table 2. As we can see that the overall performance of the TFIDF model is slightly better than the sentence transformers. This maybe is due to the difference in the base corpus the BERT model is trained on, which is conflicting with the given colloquial Tamil corpus. Whereas the TFIDF and the Count vectorization yielded similar results because of the short length of the meme and containing varied usage of words by different users. The performance of the cross validation scores for the different machine learning approaches are tabulated in Table 2.

Table 2: Cross validation scores of the meme classification task using different machine learning approaches

Features	Classifiers	Acc
Tfidf	MLP	89
Tfidf	Random forest	89
Tfidf	K-nearest	88
Count-vec	MLP	89
Count-vec	Random forest	87
Count-vec	K-nearest	86
Sent. Transformer	MLP	85

From the Table 2, it has been noted that Tfidf and n-gram features giving almost the same performance. The sentence transformer is slightly lower performance than the others, this may be because of the colloquial words used in the transcription.

Performance of the test set is given in Table 3. In Table 3, the results of other participating teams are listed.

Table 3: Performance of the meme classification task using test set

Team	Precision	Recall	F1
Codewithzichao	0.57	0.60	0.55
IIITK	0.56	0.59	0.54
NLP@CUET	0.55	0.58	0.52
SSNCSE_NLP	0.58	0.60	0.50
Simon_work	0.53	0.58	0.49
TrollMeta	0.45	0.41	0.48
UVCE-IIITT	0.60	0.60	0.46
cean	0.53	0.57	0.43
HUB	0.50	0.54	0.40
iiit_dwd	0.52	0.59	0.30

6 Conclusion

A meme is a form of media that spreads an idea or emotion across the internet. Memes make it even more challenging since they express humour and sarcasm in an implicit way, because of which the meme may not be offensive if we only consider the text or the image. In this paper we proposed a approach for meme classification for Tamil language that considers only the text present in the meme. Tfidf and n-gram features given slightly better performance than the sentence embeddings. We achieve F1 scores of 0.50 using the proposed approach using the test-set. Meme classification performance can be increased by considering visual features also.

References

Tariq Habib Afridi, Aftab Alam, Muhammad Numan Khan, Jawad Khan, and Young-Koo Lee. 2020. A multimodal memes classification: A survey and open research issues.

A. Amalia, A. Sharif, F. Haisar, D. Gunawan, and B. B. Nasution. 2018. Meme opinion categorization by using optical character recognition (ocr) and naïve bayes algorithm. In *2018 Third International Conference on Informatics and Computing (ICIC)*, pages 1–5.

S.K. Bharti, B. Vachha, R.K. Pradhan, K.S. Babu, and S.K. Jena. 2016. Sarcastic sentiment detection in tweets streamed in real time: a big data approach. *Digital Communications and Networks*, 2(3):108–121. Advances in Big Data.

Bharathi Raja Chakravarthi. 2020a. HopeEDI: A multilingual hope speech detection dataset for equality, diversity, and inclusion. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 41–53, Barcelona, Spain (Online). Association for Computational Linguistics.

Bharathi Raja Chakravarthi. 2020b. *Leveraging orthographic information to improve machine translation of under-resourced languages*. Ph.D. thesis, NUI Galway.

Bharathi Raja Chakravarthi and Vigneshwaran Muralidaran. 2021. Findings of the shared task on Hope Speech Detection for Equality, Diversity, and Inclusion. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Pruthwik Mishra, Prathyusha Danda, and Pranav Dhakras. 2018. Code-mixed sentiment analysis using machine learning and neural network approaches.

Nils Reimers and Iryna Gurevych. 2020. Making monolingual sentence embeddings multilingual using knowledge distillation. *arXiv preprint arXiv:2004.09813*.

Shardul Suryawanshi and Bharathi Raja Chakravarthi. 2021. Findings of the shared task on Troll Meme Classification in Tamil. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Shardul Suryawanshi, Bharathi Raja Chakravarthi, Mihael Arcan, and Paul Buitelaar. 2020a. Multimodal meme dataset (MultiOFF) for identifying offensive content in image and text. In *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, pages 32–41, Marseille, France. European Language Resources Association (ELRA).

Shardul Suryawanshi, Bharathi Raja Chakravarthi, Mihael Arcan, and Paul Buitelaar. 2020b. Multimodal meme dataset (MultiOFF) for identifying offensive content in image and text. In *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, pages 32–41, Marseille, France. European Language Resources Association (ELRA).

Shardul Suryawanshi, Bharathi Raja Chakravarthi, Pranav Verma, Mihael Arcan, John Philip McCrae, and Paul Buitelaar. 2020c. A dataset for troll classification of TamilMemes. In *Proceedings of the WILDRE5– 5th Workshop on Indian Language Data: Resources and Evaluation*, pages 7–13, Marseille, France. European Language Resources Association (ELRA).

William Yang Wang and Miaomiao Wen. 2015. I can has cheezburger? a nonparanormal approach to combining textual and visual information for predicting and generating popular meme descriptions. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 355–365, Denver, Colorado. Association for Computational Linguistics.