Review of the Voice-for-the-Voiceless Classifier

1st Ajay Biswas
Department of Computer Science and Engineering
National Institute of Technology
Rourkela, India
220cs2184@nitrkl.ac.in

2nd Tapas Kumar Mishra
Department of Computer Science and Engineering
National Institute of Technology
Rourkela, India
mishrat@nitrkl.ac.in

***Abstract*—Machine Learning based text classification provides high degree of accuracy, however, requires huge manual annotation effort if the data-set size is humongous. Active learning (AL) aims to reduce manual annotation effort as well as tries to maintain high training accuracy. In this paper, we surveyed ongoing researches in the field of active learning and reviewed the voice-for-the-voiceless classifier. We also identified the potential areas for the improvement of this classifier as well as highlighted its use in other domains too.**

***Index Terms*—active learning, rohingya, svm, voice-for-the-voiceless**

I. Introduction

Active learning is a technique for reducing manual annotation effort during training phase of machine learning. The annotation is done by a human (called oracle) which helps AL systems to achieve high accuracy with few labelled instances. For problems having large collection of unlabeled data, pool-based sampling is used [1]. Figure 1 shows working of pool-based AL. AL is very useful in classifying comments which involves a person's opinion, belief or political interest, as it's too complicated to be dealt with plain unsupervised machine learning. Also, there are numerous challenges to be dealt with before any classification could take place. Some of the challenges are (i) Dealing with multiple languages having different levels of grammatical accuracy, (ii) Un-structured data, (iii) ambiguous sentences, (iv) Unrelated comments, etc.

The organization of this paper is as follows: Section 1 provides brief introduction briefly describes about classification of comments through active learning; Section 2 provides a brief summary on the related works; Section 3 describes the work done in [2]. It describes how they have built the data-set containing comments fetched from YouTube. Also it briefly summarizes how they applied active learning to classify comments; and finally, Section 4 concludes our paper.

II. Literature Survey

Previously various research were conducted in the field of Active Learning and Sentence Classification. We are focusing on those researches which are related to our work.

i. Chen et al. [3] proposed a state-of-the-art result on the Stanford Natural Language Inference Dataset using Long Short-Term Memory (LSTM). They employed Bi-directional LSTM (BiLSTM) as one of the building blocks. Later it is used to perform inference composition to construct the final prediction.

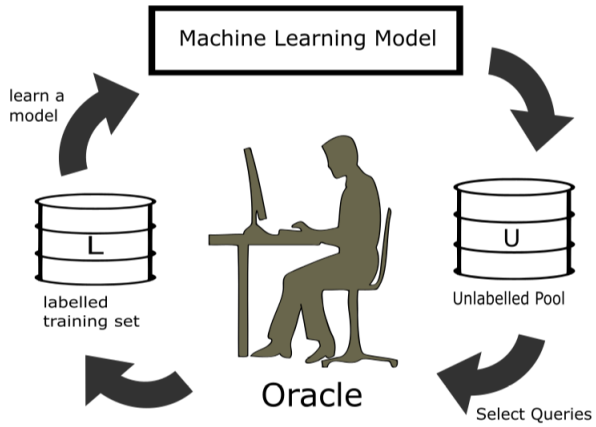

Fig. 1. Pool-based Active Learning

ii. Kim et al. [4] propose a densely-connected co-attentive recurrent neural network to find semantic relation between sentences. To overcome the problem of ever-increasing size of feature vector due to densely connected networks, they also have propose an autoencoder after dense concatenation.

iii. Nguyen and Smeulders [5] incorporated clustering into active learning. The algorithm first constructs a classifier on the set of the cluster representatives, and then with the help of a local noise model, it passes the classification decision to the other samples. The model allows selecting the most representative samples as well as avoids labelling samples in the same cluster. The paper focuses on discriminative models including logistic regression and Support Vector Machines (SVM) which are less sensitive to training data and hence, good for active learning.

iv. Ru et al. [6] propose adversarial uncertainty sampling in discrete space (AUSDS) which retrieves informative unlabeled samples more efficiently and is 10x faster when compared to typical uncertainty sampling method for active learning.

v. Schumann and Rehbein [7] showed that it is possible to use Membership Query Synthesis [5] for generating AL queries for natural language processing, using Variational Autoencoders for query generation, and provides com-

petitive performance to pool-based AL strategies while substantially reducing annotation time.

vi. Ren et al. [8] propose a novel fake news detection framework "Adversarial Active Learning based Heterogeneous Graph Neural Network for Fake News Detection (AA-HGNN)", which employs a novel hierarchical attention mechanism to perform node representation learning in the HIN. In this paper, the authors model the news content and related entities as a News-HIN. The AA-HGNN utilizes both structural information as well as News-HIN to identify fake news.

vii. Xiang et al. [9] proposes a novel approach which exploits linguistic regularities in profane language via statistical topic modeling on a huge Twitter corpus, and detects offensive tweets using these automatically generated features. This approach works with various Machine Learning models such as J48 decision tree learning, SVM, logistic regression (LR) and random forest (RF).

viii. Watanabe et al. [10] proposes a pragmatic approach to collect hateful speech. The proposed approach uses unigram and patterns that are automatically collected from training dataset. Accuracy of 87.4% was achieved on detecting whether a tweet is offensive or not (binary classification) and 78.4% accuracy when detecting a tweet is hateful, offensive, or clean (ternary classification).

ix. Palakodety et al. [2] proposes a classifier which can classify comments supporting the Rohingyas [11]. This is done by building a corpus from YouTube comments and applying multiple AL strategies based on nearest-neighbors in the comment-embedding space.

III. Voice for the Classifier

This section provides details of the work done in [2]. The main goal of their work is to detect comments defending the Rohingyas among large number of comments.

A. Dataset

In this work, the authors have constructed a substantial corpus of YouTube video comments using high-frequency search queries from 19 different countries consisting of 263,482 comments from 113,250 users in 5,153 relevant videos. To fetch the comments, the authors have used the publicly available YouTube API. The corpus was more than 50% written in English, the rest of the corpus comprised of German, Hindi, Bengali, Malay, Urdu, French and Arabic comments.

B. Active Learning Approach

As illustrated in figure 2, their approach consists of the following steps.

- Constructed a seed set of positive and negative comments.
- Expanded the seed set by Random Sampling comments from the unseen corpus.
- Obtained real valued embeddings for the comments and found Nearest Neighbors of the seed set.
- Further expand using minority-class certainty sampling.
- Performed final expansion using uncertainty sampling.

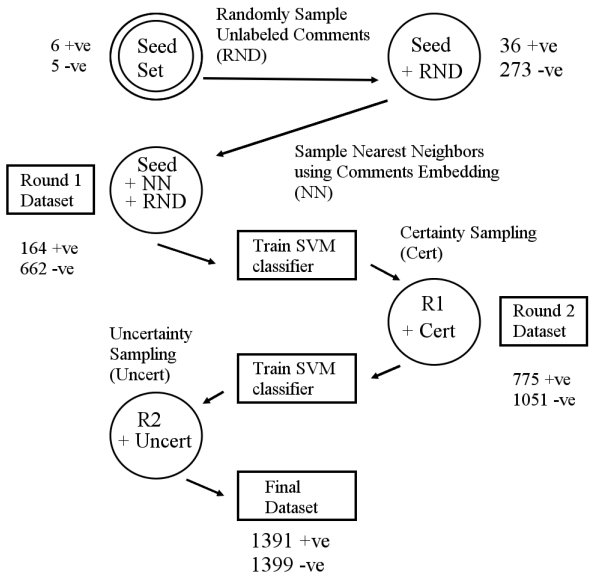

Fig. 2. System Diagram

C. Performance Analysis

The following results are claimed by the authors. The classification was performed using SVM classifier having a 90/10 train/test split. Token n-grams (n up to 3) as taken as features and evaluated performance on the test set. The experiment was repeated 100 times on 100 randomly chosen test-train splits. The results are shown in table I.

D. Implementation

Since we couldn't find any publicly available dataset for this paper, we tried to build the dataset using YouTube API. We fetched 19228 comments out of which 18758 were written in English characters. The comments were generated by the help of the query set generated by providing term 'Rohingya' in Google Trends mentioned in table II. The size of the dataset is small because less number of 'Rohingya' related terms are being searched in 2021 and also many videos have comment section disabled.

From the results we have observed that the voice-for-the-voiceless classifier was able to identify comments supporting Rohingyas with good accuracy. This classifier has the potential to be used in other domains also like classifying comments supporting CAA-NRC [12] or comments supporting the farmer protest [13].

IV. Conclusion

In this paper we provided survey of researches in the field of active learning and also reviewed the Voice-for-the-voiceless classifier. We also managed to build the dataset using the Youtube API. In our future work, we will be following their approach to classify comments supporting the Indian farmers protest.

TABLE I
VOICE-FOR-THE-VOICELESS CLASSIFIER PERFORMANCE

Performance measure	**Seed set + random sampling + NN in the embedding space**	**Certainty Certainty Sampling**	**Uncertainty Sampling**	**SVM(n gram)** (Improved Model)	**SVM(n gram + embedding)** (Improved Model)
Precision	67.17 ± 9.90%	71.27 ± 5.23%	73.65 ± 3.45%	73.65 ± 3.45%	76.49 ± 3.51%
Recall	32.35 ± 7.65%	72.52 ± 4.23%	79.39 ± 3.72%	79.39 ± 3.72%	80.30 ± 3.73%
Accuracy	82.04 ± 2.34%	75.95 ± 3.10%	75.38 ± 2.76%	75.38 ± 2.76%	77.71 ± 2.56%
F1 score	43.02 ± 7.90%	71.75 ± 4.32%	76.34 ± 2.77%	76.34 ± 2.77%	78.28 ± 2.71%
AUC	83.61 ± 2.88%	83.64 ± 2.84%	83.67 ± 2.61%	83.67 ± 2.61%	85.91 ± 2.32%

TABLE II
SEARCH QUERIES FOR DATASET BUILDING

Search Queries
the rohingya crisis
where is rohingya
rohingya map
rohingya crisis
rohingya charity
rohingya genocide
myanmar rohingya
muslim genocide
are rohingya terrorist
who are rohingya
deport rohingya
kill rohingya

REFERENCES

[1] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," in *SIGIR'94*. Springer, 1994, pp. 3–12.

[2] S. Palakodety, A. R. KhudaBukhsh, and J. G. Carbonell, "Voice for the voiceless: Active sampling to detect comments supporting the rohingyas," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 454–462.

[3] Q. Chen, X. Zhu, Z. Ling, S. Wei, H. Jiang, and D. Inkpen, "Enhanced lstm for natural language inference," *arXiv preprint arXiv:1609.06038*, 2016.

[4] S. Kim, I. Kang, and N. Kwak, "Semantic sentence matching with densely-connected recurrent and co-attentive information," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 6586–6593.

[5] H. T. Nguyen and A. Smeulders, "Active learning using pre-clustering," in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 79.

[6] D. Ru, J. Feng, L. Qiu, H. Zhou, M. Wang, W. Zhang, Y. Yu, and L. Li, "Active sentence learning by adversarial uncertainty sampling in discrete space," *arXiv preprint arXiv:2004.08046*, 2020.

[7] R. Schumann and I. Rehbein, "Active learning via membership query synthesis for semi-supervised sentence classification," in *Proceedings of the 23rd conference on computational natural language learning (CoNLL)*, 2019, pp. 472–481.

[8] Y. Ren, B. Wang, J. Zhang, and Y. Chang, "Adversarial active learning based heterogeneous graph neural network for fake news detection," in *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2020, pp. 452–461.

[9] G. Xiang, B. Fan, L. Wang, J. Hong, and C. Rose, "Detecting offensive tweets via topical feature discovery over a large scale twitter corpus," in *Proceedings of the 21st ACM international conference on Information and knowledge management*, 2012, pp. 1980–1984.

[10] H. Watanabe, M. Bouazizi, and T. Ohtsuki, "Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection," *IEEE access*, vol. 6, pp. 13 825–13 835, 2018.

[11] A. Azad and F. Jasmin, "Durable solutions to the protracted refugee situation: The case of rohingyas in bangladesh," *Journal of Indian Research*, vol. 1, no. 4, pp. 25–35, 2013.

[12] D. Misra and C. Viens, "The citizenship amendment act (caa): the struggle for india's soul," *REVISTA LÜVO*, p. 25, 2020.

[13] A. Shankar, "Indian agriculture farm acts: 2020," *International Journal of Modern Agriculture*, vol. 10, no. 2, pp. 2907–2914, 2021.