

Review of the Voice-for-the-Voiceless Classifier

Ajay Biswas

220CS2184

Supervisor: Dr. Tapas Kumar Mishra

November 30, 2021



Contents

- 1 Introduction
- 2 Literature survey
- 3 Voice-for-the-Voiceless Classifier
- 4 Conclusion

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.

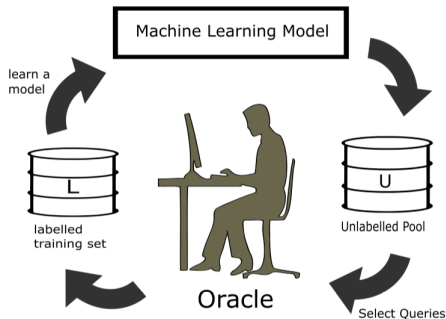


Figure: Pool-based Active Learning

- Collecting large amount of unlabeled data is easier but manually labeling them is tough.
- Active Learning reduces this effort by taking help from user and training on remaining unlabelled points.

Introduction

Active Learning Scenarios

- Membership Query Synthesis
- Stream-Based Selective Sampling
- Pool-Based Sampling

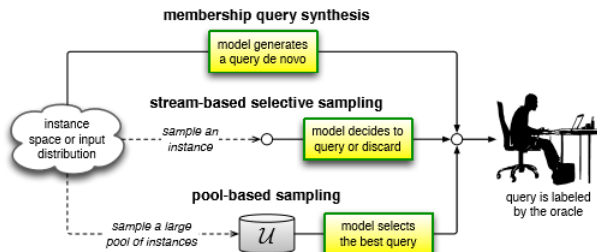


Figure: Diagram illustrating the three main active learning scenarios [12]

i. **Enhanced LSTM for Natural Language Inference**

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. The model has an accuracy of 88.6 %.

ii. **Semantic sentence matching with densely-connected recurrent and co-attentive information**

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. **Active Learning Using Pre-clustering**

Nguyen and Smeulders [7] 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. **Active Sentence Learning with AUSDS**

Ru et al. [10] 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. **Active Learning via Membership Query Synthesis for Semi-supervised Sentence Classification**

Schumann and Rehbein [11] 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 competitive performance to pool-based AL strategies while substantially reducing annotation time.

vi. **Adversarial Active Learning based Heterogeneous Graph Neural Network for Fake News Detection (AA-HGNN)**

Ren et al. [9] 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. **Detecting Offensive Tweets via Topical Feature Discovery over a Large Scale Twitter Corpus**

Xiang et al. [16] 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, 

viii. **Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection**

Watanabe et al. [15] 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. **Voice for the Voiceless: Active Sampling to Detect Comments Supporting the Rohingyas**

Palakodety et al. [8] proposes a classifier which can classify comments supporting the Rohingyas. This is done by building a corpus from YouTube comments and applying multiple AL strategies based on nearest-neighbors in the comment-embedding space. The classifier provided an accuracy of 75.38% with SVM(n gram) and 77.71% with SVM(n gram + embedding).

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.

Voice-for-the-voiceless Classifier

[illegible]

Figure: Dataset Snapshot

Active Learning Steps

As illustrated in figure 4, 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.

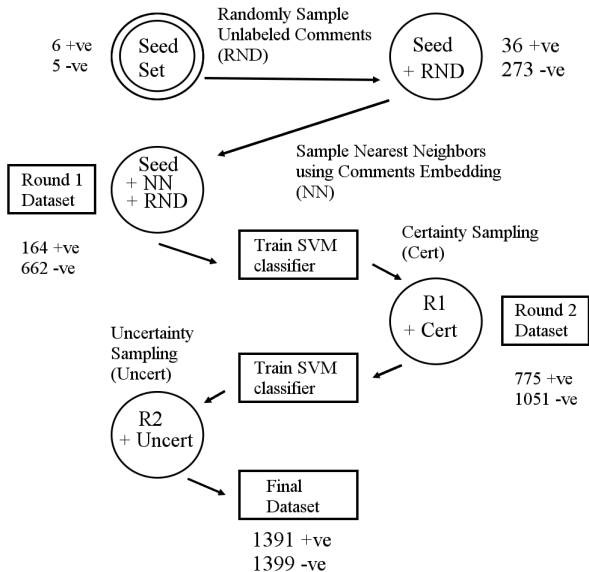


Figure: System Diagram

Performance Analysis

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 1.

Voice-for-the-voiceless Classifier

Table: 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 \pm 9.90%	71.27 \pm 5.23%	73.65 \pm 3.45%	73.65 \pm 3.45%	76.49 \pm 3.51%
Recall	32.35 \pm 7.65%	72.52 \pm 4.23%	79.39 \pm 3.72%	79.39 \pm 3.72%	80.30 \pm 3.73%
Accuracy	82.04 \pm 2.34%	75.95 \pm 3.10%	75.38 \pm 2.76%	75.38 \pm 2.76%	77.71 \pm 2.56%
F1 score	43.02 \pm 7.90%	71.75 \pm 4.32%	76.34 \pm 2.77%	76.34 \pm 2.77%	78.28 \pm 2.71%
AUC	83.61 \pm 2.88%	83.64 \pm 2.84%	83.67 \pm 2.61%	83.67 \pm 2.61%	85.91 \pm 2.32%

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 2. 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.

Voice-for-the-voiceless Classifier

Table: 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

Implementation

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 [6] or comments supporting the farmer protest [14].

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 and performed few analysis. In our future work, we will be following their approach to classify comments supporting the Indian farmers protest.

References I



Dana Angluin.

Queries and concept learning.

Machine learning, 2(4):319–342, 1988.



Ashraful Azad and Fareha Jasmin.

Durable solutions to the protracted refugee situation: The case of rohingyas in bangladesh.

Journal of Indian Research, 1(4):25–35, 2013.



Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, Hui Jiang, and Diana Inkpen.

Enhanced lstm for natural language inference.

arXiv preprint arXiv:1609.06038, 2016.

References II



Seonhoon Kim, Inho Kang, and Nojun Kwak.

Semantic sentence matching with densely-connected recurrent and co-attentive information.

In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 6586–6593, 2019.



David D Lewis and William A Gale.

A sequential algorithm for training text classifiers.

In *SIGIR'94*, pages 3–12. Springer, 1994.



Devika Misra and Catherine Viens.

The citizenship amendment act (caa): the struggle for india's soul.

REVISTA LÜVO, page 25, 2020.

References III



Hieu T Nguyen and Arnold Smeulders.

Active learning using pre-clustering.

In *Proceedings of the twenty-first international conference on Machine learning*, page 79, 2004.



Shriphani Palakodety, Ashiqur R KhudaBukhsh, and Jaime G Carbonell.

Voice for the voiceless: Active sampling to detect comments supporting the rohingyas.

In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 454–462, 2020.



Yuxiang Ren, Bo Wang, Jiawei Zhang, and Yi Chang.

Adversarial active learning based heterogeneous graph neural network for fake news detection.

In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 452–461. IEEE, 2020.



Dongyu Ru, Jiangtao Feng, Lin Qiu, Hao Zhou, Mingxuan Wang, Weinan Zhang, Yong Yu, and Lei Li.

Active sentence learning by adversarial uncertainty sampling in discrete space.

arXiv preprint arXiv:2004.08046, 2020.



Raphael Schumann and Ines Rehbein.

Active learning via membership query synthesis for semi-supervised sentence classification.

In *Proceedings of the 23rd conference on computational natural language learning (CoNLL)*, pages 472–481, 2019.



Burr Settles.

Active learning literature survey.
2009.



Burr Settles, Mark Craven, and Lewis Friedland.

Active learning with real annotation costs.

In *Proceedings of the NIPS workshop on cost-sensitive learning*, volume 1. Vancouver, CA:, 2008.

References VI



Amar Shankar.

Indian agriculture farm acts: 2020.

International Journal of Modern Agriculture, 10(2):2907–2914, 2021.



Hajime Watanabe, Mondher Bouazizi, and Tomoaki Ohtsuki.

Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection.


IEEE access, 6:13825–13835, 2018.



Guang Xiang, Bin Fan, Ling Wang, Jason Hong, and Carolyn 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*, pages 1980–1984, 2012.

-  Xiaojin Jerry Zhu.
Semi-supervised learning literature survey.
2005.

Thank you!!