Sentiment prediction from twitter sentiment classifier

Anonymous

1 Introduction

The main goal of this report is to develop twitter sentiment classifier which takes the input as a large set of tweets.For each tweet,the corresponding sentiment will be predicted(positive or negative).Two datasets are used in the experiment which are training data used for training each model and development data set used for model fitting and tuning. In this report, the embedding sentence transformer is used to convert each twitter into a 384-dimensional vector(Blodgett el al.,2016).These vectors capture each Tweet's "meaning," allowing comparable Tweets to be clustered together in the 384-dimensional space.The research question is about exploring bias in twitter sentiment classification as well as finding out any possible solutions to close the performance gap in models.

2 Literature Review

The first paper is about learning and evaluating classifiers under sample selection bias.The paper(Zadrozny,2004) explores the bias problem in different classifier and investigate how it affects a variety of well-known classifier learning approaches both analytically and experimentally.A bias correction method is also shown in the paper which is particularly relevant for classifier evaluation under sample selection bias.This paper identifies two categories(global and local) and examines both of them to get conclusion that bias affects global classifiers instead of local classifiers

.The second paper is about characterizing bias in classifiers using generative models.This paper(McDuff&Song&Kapoor,2019) explores a simulation-based method for systematic interrogation of classifier using generative adversarial models.It examines Bayesian optimization to efficiently interrogate independent face image classification systems and a progressive conditional generative model to synthesis lifelike facial images.This paper demonstrates how this method can quickly identify bias in racial and gender.

3. Method

3.1 Model

3.1.1 Baseline- Naive Bayes

Naive Bayes uses conditional probability from the Bayes theory which assumes the training set's characteristic conditions are all independent. The maximum posterior probability is obtained by learning the joint probability distribution from input to output based on the given training set. The naive Bayes is not sensitive to irrelevant features.In addition,it is easy to scale to many feature dimensions amd data sizes which fits well with our datasets.These are reasons why naive Bayes is chosen as the first model for experiment.

3.1.2 Model 1-Logic Regression

The baseline chosen in experiment is logic regression.Logic regression is a kind of binary classifier.It uses probabilistic discriminative since it contains P(y|x) instead.It models the

probability of one event occurring by using log odds function.It assumes the event is a linear combination of independent variables.The logic regression has no restrictive assumptions on features,and it is very suitable to frequency-based features(NLP) which fits well with our datasets.These are reasons why logic regression is chosen for experiment.

3.1.3 Model 2-KNN

In statistic,the KNN algorithm is actual the non-parametric model.The basic principle behind the KNN algorithm is that it will find the first k samples and see which label does majority of them belongs to.KNN model firstly store all training example.Then in the testing part, it then calculates the distance between test data and train data.It tries to find the the first K samples which means nearest neighbors.At last, it computes target concept for the test instance based on labels of the training instances.There are multiples ways for calculating the distance such as cosine distance,Euclidean distance and hamming distance in KNN.The K-nearest neighbors algorithm has no assumptions and supports both classification and regression.These are reasons why KNN model is chosen for the second model for experiment.

3.2Evaluation metrics

There are four evaluation metrics used in the experiment.They are accuracy,precision,recall and F-score.All of them are calculated by 4 elements(TP,TN,FP,FN) where TP means true positive,TN means true negative,FP means false positive,FN means false negatives.We use accuracy since accuracy is the most straightforward way to demonstrate the performance of the model.Use it is a direct way of showing the gap between two different demographic groups.We use precision and recall since both precision and recall must be examined to fully evaluate the effectiveness of a model.

3.2.1 Accuracy

Accuracy is calculated by (TP+TN/TP+TN+FP+FN).It is the fraction of predictions our model got right.In other word,it quantifies how frequently the classifier is correct.In our experiment, we calculate the accuracy for both baseline and machine models in the first step.Furthermore,we also use it to explore the research question for comparing the accuracy gap between AAE and SAE in each model.

3,2,2 Precision

Precision is used to measure how often are we correct as well as when we predict that an instance is interesting.It is calculated by(TP/TP+FP) which means a model that produces no false positives has a precision of 1.0.For our experiment,the precision is used in the first step for showing the performance of both baseline and machine models.

3.2.3Recall

Recall is used to measure what proportion of actual positives is identified correctly.It is calculated by (TP/TP+FN) which means a model that produces no false negative has a recall of 1.0.For our experiment,the recall is used in the first step for showing the performance of both baseline and machine models.

4.Results

Confusion Metrics/Model	Logic Regression	Naive Bayes	KNN
Accuracy	0.69825	0.61475	0.621
precision	0.6985	0.6181	0.621
recall	0.69825	0.6147	0.621

Table 1 The confusion metrics for each model

In the first step of experiment,we firstly train three models and get the confusion metrics results from each of them.The

above table shows the accuracy,precision and recall for each model.This shows that the baseline logic regression model has the highest value of all confusion metrics compared to other two models.

Logic Regression Model	AAE	SAE
Accuracy	0.6645	0.732

Table 2 The accuracy of AAE and SAE group in logic regression

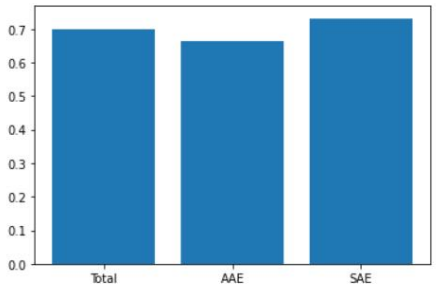

Image 1 The accuracy of three groups in logic regression

Naive Bayes Model	AAE	SAE
Accuracy	0.5575	0.652

Table 3 The accuracy of AAE and SAE group in Naive Bayes

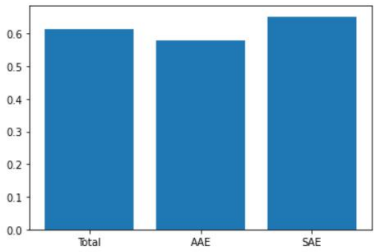

Image 2 The accuracy of three groups in Naive Bayes

KNN Model(K=1)	AAE	SAE
Accuracy	0.5935	0.6485

Table 4 The accuracy of AAE and SAE group in KNN

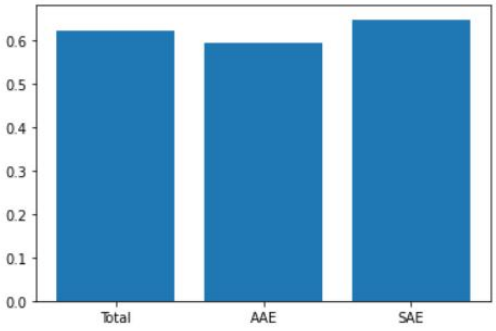

Image 3 The accuracy of three groups in KNN

The three tables and bar charts above show the accuracy comparison for each model grouped according to AAE and SAE.This shows that for each model,the accuracy for SAE group is higher than the one in AAE group.In other word,the bias can be found in each model.In the next discussion part,we will explain why the gap exists in each model and explore some effective ways to close the gap.

5.Discussion

5.1 Gap in each model

From the tables and bar charts in the result part,we examines the accuracy for AAE group and SAE group.In doing so,it points to the gap in logic regression is lower than the one in baseline(naive Bayes),and the KNN model has the smallest gap.What this means is logic regression has lower bias than naive Bayes and the KNN in the experiment has the lowest bias compared to other two models.To explain the reason of that,firstly we compare logic regression and Naive Bayes.There are some similarities between these two models.For example,both classifiers are linear models.While the difference is that when using logical Regression model,minimize an error function is doing by adjusting the model's weights.On the other hand,the weights of Naive Bayes are determined by the conditional probabilities from the features in train data,regardless of how erroneous the conclusion is.Because bias means the amount of error in any

models,it implies logic regression has lower bias than naive Bayes since the error function can minimize the error.In addition,the algorithm becomes more biased as it makes more assumptions. In Naive Bayes, features are assumed to be independent, which is not true in practise. Therefore, Naive Bayes is biased more.Naive Bayes makes the assumption of features being independent which Having carried out these points,we conclude that the logic regression has lower bias in comparison to Naive Bayes.Next,to explain why the KNN has lowest gap,it is noticeable that the value of k selected in the experiment is 1.The KNN model is non-parametric model whereas LR is a parametric model.The value of K can impact the model complexity and the performance of the KNN model. When the small K value is applied,it means that the model is easier to capture noise.It implies that the KNN model will perform over-fitting when k is one.This is significant since over-fitting means model has low bias but high variance.This is the main reason why the KNN model has the lower gap than other two models when k value is one in the experiment

.

5.2Explore the effect of adjusting value of K in KNN model on gap

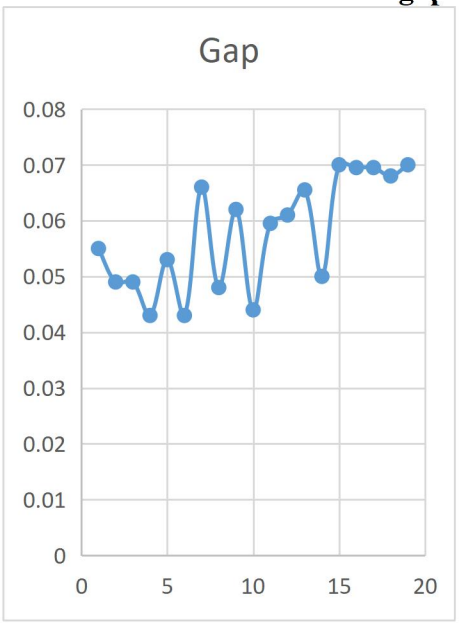

Figure 1 The change of gap based on the change of K value

The above figure illustrates that the gap between two groups fluctuates up as the value of K increases.This tells us that selecting a lower value of K such as 4 or 6 is a good method to close the gap.To explain the reason behind the data,the effect of K value need to be considered.It is noticeable that when the K value is low,the noise is easily to be captured which means the model is easier to be over-fitting.This is important because over-fitting can lead to low bias and high variance.However,as the value of K increases,the danger of grouping together unrelated classes can increase.The model is easier to be under-fitting which points to high bias and low variance.This is the reason why as shown in the figure,the higher value of K can increase the accuracy gap between AAE and SAE group.Having carried out these points,we conclude that selecting a lower value of K is a good option in KNN model to close the gap(Reduce the bias).

5.3 Explore the effect of adjusting value of C on gap in Logic Regression

Table 5 The change of gap based on the change of C value

C value	1	2	3	4	5	6	7	8	9
Accuracy Gap	0.06755	0.06755	0.0667	0.06655	0.0666	0.06655	0.06655	0.0666	0.06655

The above table illustrates that changing value of C doesn't give any help on closing the gap.To explain the reason, firstly we need to identify the meaning of C parameter here.C indicates the regularization strength,the smaller value of C means the stronger regularization.While the regularization

is used for fitting a model onto our test set correctly.It assists us prevent over fitting and obtain the optimal model.When the strength of regularization changes,the gap doesn't change since the over-fitting still hasn't happened. Indeed,the parameter λ in regularization controls the impact on bias and variance.It is the tuning parameter that determines how much we want penalize our model's flexibility. Change value of parameter C doesn't change the value of . For example,as the value decreases, the value of coefficients decreases, lowering the variance. This rise in is helpful up to a degree because it merely reduces variance without sacrificing any significant characteristics in the data. However, after a certain value, the model begins to lose crucial properties, resulting in bias and under-fitting. As a result, the value of the λ should be carefully chosen in order to control the bias in our experiment.On the contrary,the value of C in logic regression doesn't impact that.

6. Conclusion

In this article,we firstly evaluate the performance with different supervised model including baseline Naive Bayes and KNN model and Logic regression by using confusion metrics.In addition,we have proposed two methods which can close the performance gap between AAE and SAE group in a specific group.In addition, we also have illustrated why change value of parameter C cannot lead to a lower gap.While adjusting value of λ can control the bias.It can be found that adjusting the parameter inside each classifier is a good option to reduce the bias in twitter sentiment classifier.

Reference

Zadrozny, B. (2004, July). Learning and evaluating classifiers under sample selection bias. In *Proceedings of the twenty-first international conference on Machine learning* (p. 114).

McDuff, D., Ma, S., Song, Y., & Kapoor, A. (2019). Characterizing bias in classifiers using generative models. *Advances in Neural Information Processing Systems*, *32*.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R. (2011). Sentiment analysis of Twitter data.
In Proceedings of the Workshop on Language in Social Media (LSM 2011), pages 30–38, Portland, Oregon.
Association for Computational Linguistics.

Blodgett, S. L., Green, L., and O'Connor, B. (2016). Demographic dialectal variation in social media: A case study of African-American English. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 1119–1130, Austin, Texas. Association for Computational Linguistics.

Blodgett, S. L. and O'Connor, B. (2017). Racial disparity in natural language processing: A case study of social media african-american english. arXiv preprint arXiv:1707.00061. Elazar, Y. and Goldberg, Y. (2018).

Adversarial removal of demographic attributes from text data. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 11–21, Brussels,Belgium. Association for Computational Linguistics.

Kiritchenko, S. and Mohammad, S. (2018). Examining gender and race bias in two hundred sentiment analysis systems. In Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, pages43–53, New Orleans, Louisiana. Association for Computational Linguistics.

Reimers, N. and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks.

In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 3http://academichonesty.unimelb.edu.au/policy.html

9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Sch utze, H., Manning, C. D., and Raghavan, P. (2008). Introduction to information retrieval, volume 39. Cambridge University Press Cambridge.