Estimating Question Quality: Basic Approach

Summary

This report aimed to find the best metric for estimating the quality of questions based on the problem statement of the NeurIPS competition. In preprocessing, I added confidence, grouped, and dealt with N/A values. In ordering and evaluation, I tried several metrics, including confidence and correct answer ratio. The results showed that the quality of the question is quite related to confidence and correct answer ratio, and their product metric produced the best results.

Introduction

The NeurIPS competition (and SSU ML class 2023) presented a challenge: how can we estimate the quality of questions? In response, this report presents a methodology for identifying the best metric for assessing question quality.

Methods

During the preprocessing stage, I added confidence, grouped, and handled missing data. The 'answer_metadata_task_3_4.csv' dataset was merged to use confidence metadata. Dataframe grouping was then done by QuestionId to estimate each question's quality. To deal with the missing data, I initially dropped rows with N/A values, but the resulting metrics were unsatisfactory. To address this, I filled in the missing values with the mean value.

In the ordering and evaluation stage, several metrics were experimented with, including confidence, correct answer ratio, and their product. The results showed that the product of confidence and correct answer ratio was the most effective metric for determining question quality. The report suggests that lower confidence and correct answer ratio are better indicators of question quality.

You can view all of my works in GitHub(https://github.com/tuxedcat/question_quality_assessment).

Results

Metric	evaluated
Confidence ascending	0.76
Confidence descending	0.36
Correct answer ratio ascending	0.76
Correct answer ratio descending	0.44
Confidence * CorrectRatio ascending	0.8
Confidence * CorrectRatio descending	0.4

Discussion

The results indicate that the quality of questions is highly related to both confidence and correct answer ratio. By multiplying these two metrics, the product metric produced the best estimate of question quality. These findings have practical implications for various domains where question quality is important, such as education and online forums.

Conclusion

In conclusion, this report presents a comprehensive methodology for estimating the quality of questions using several metrics. The results showed that the product of confidence and correct answer ratio produces the best estimate of question quality. These findings have practical implications for question quality assessment in various domains, including education, online forums, and question-answering platforms. Future work can extend this methodology by (cross-)entropy, incorporating additional features and employing more sophisticated machine learning algorithms.

References

https://proceedings.mlr.press/v133/wang21a.html
https://pandas.pydata.org/docs/

Estimating Question Quality: An Unsupervised Learning Approach

Summary

This report presents follow-up study of an exploration of unsupervised learning methods for estimating question quality in the NeurIPS competition. Initially, KNN was considered but I couldn’t find how to apply it. Instead, I considered to use Principal Component Analysis (PCA) as a well-known unsupervised learning method for dimension reduction. Additionally, the impact of normalization, specifically Min-Max scaling, was examined. The findings indicated that normalization did not contribute to performance improvements, likely because both confidence and correct answer ratio are proportion-based metrics without specific units.

Introduction

The NeurIPS competition posed a challenge in estimating the quality of questions, necessitating an investigation into appropriate unsupervised learning methods. This report details the exploration of using KNN and subsequently Principal Component Analysis (PCA) as a means to address this challenge. Furthermore, the impact of normalization techniques on the performance of the models is examined.

Methods

Initially, KNN was considered but I couldn’t find how to apply it. Instead, I considered to use Principal Component Analysis (PCA) as a well-known unsupervised learning method for dimension reduction. But the observations already reduced in 2D representation enough (confidence, correct ratio). So I decided to apply Min-Max scaling to get better result.

```
from sklearn import preprocessing
data['Confidence']/=100
scaler = preprocessing.MinMaxScaler().fit(data)
data = pd.DataFrame(scaler.transform(data),columns=['Confidence','IsCorrect'])
```

[Picture 1] Min-Max scaling code

Results

Previous result: 0.8

Current result: 0.76

Discussion

The project utilized unsupervised learning techniques to estimate the quality of questions, focusing on PCA for dimension reduction. However, the attempt to improve performance through Min-Max scaling, did not yield positive results. This outcome may be attributed to the nature of the features used, as both confidence and correct answer ratio are proportion-based metrics without specific units.

Conclusion

While the normalization technique of Min-Max scaling was applied, it did not contribute to performance improvements. The findings highlight the unique characteristics of confidence and correct answer ratio as proportion-based metrics without specific units, which may limit the efficacy of certain normalization techniques. Future work can involve investigating alternative unsupervised learning algorithms or considering additional features to enhance question quality assessment in various domains.