Title
An interpretable wine text analysis framework

I want to submit an abstract for:
Conference Presentation

Corresponding Author
Jing Cao

E-Mail Corresponding Author
jcao@smu.edu

Affiliation
Southern Methodist University

Co-Author/s
<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chenyu Yang</td>
<td>Southern Methodist University</td>
</tr>
</tbody>
</table>

E-Mail/s Co-Author/s
<table>
<thead>
<tr>
<th>E-Mail</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:chenyuy@smu.edu">chenyuy@smu.edu</a></td>
</tr>
</tbody>
</table>

Keywords
interpretable text analysis; wine review; deep learning model; attention mechanism

Research Question
How to use wine text review to provide interpretable data analysis results?

Methods
Statistical framework combined with a deep learning component

Results
interpretable text analysis

Abstract
There is an ongoing debate on whether wine reviews provide meaningful information on wine properties and quality. However, few studies have been conducted aiming directly at comparing the utility of wine reviews and numeric measurements in wine data analysis. Based on data from close to 300,000 wines reviewed by Wine Spectator, we investigate whether wine reviews are useful in predicting a wine’s quality classification. We group our sample into one of two binary quality brackets, wines with a critical rating of 90 or above and the other group with ratings of 89 or below. This binary outcome constitutes our dependent variable. WE compare logistic regression models based on numerical covariates such as the price and age of wine and deep learning models applied to wine review texts. By comparing the explanatory accuracy of the models, our results suggest that deep
Learning models using wine review texts are more accurate in predicting binary wine quality classifications than are various numerical covariates—including the wine's price.

In the first part of the study, we have demonstrated that deep learning models in text analysis are capable of yielding more accurate results than traditional data analysis using numerical variables. However, it is not clear how those results are being produced. This is not uncommon, since it is well known that most of the deep learning models are often described as “black box” algorithms due to their complex structures and lack of interpretability. This lack of interpretability can undermine trust in the results and prevent applications of the models in areas where interpretability is as important as predictive accuracy. In the second part of the study, we develop a statistical framework which incorporates a deep learning component in text analysis. The statistical framework provides greater transparency and reasoning of the underlying processes, the deep learning component helps ensure that the classification accuracy is still maintained. By doing so, we can combine the best of two worlds: the interpretability of a statistical model and the high predictive performance of deep learning models.

Specifically, using the wine review data, we assume that words in a review can be categorized as either sentiment words or function words. Sentiment words are associated with either a positive or a negative sentiment, describing emotion or experience that is pleasant or desirable (e.g., delightful, smooth, pleasant) or unpleasant or undesirable (e.g., sour, unpleasant, corked). On the other hand, function words are words that are used to structure the sentence and convey meaning without emotional implications, such as prepositions, conjunctions, articles, pronouns, and auxiliary verbs, etc. The proposed model is able to recognize sentiment words, estimate sentiment score at the word-level, and determine the overall sentiment of the review by combining the sentiment scores of individual words, thus, providing interpretable results in text analysis.

The model can also be applied to authorship identification. There are altogether 9 wine experts providing wine reviews for Wine Spectator. We show that the proposed model can accurately identify the corresponding author of the wine reviews. The results are accurate and interpretable, in that the usage of particular words, the writing styles of the wine experts can be inferred from the model.

In summary, the interpretable deep learning model can provide a deeper insight on the text used in wine reviews, from the sentiment carried by individual words, to the different wording styles exhibited by different wine reviewers. In addition, the model is flexible to incorporate variables other than the text information. If they contain unique information, the analysis results can be further improved. The techniques used in this model share some of the common features embedded in the most updated language modeling – ChatGPT.

**Consent**

✔ I agree to the privacy policy.

You find the link to our privacy policy at the bottom of the page. In the privacy policy you find a link for exporting and/or erasing your personal data stored in our database.