Predicting and Visualizing Wine Characteristics Through Analysis of Tasting Notes From Viewpoints

Flanagan, Brendan
Graduate School of Information Science and Electrical Engineering, Kyushu University

Wariishi, Nao
Department of Electrical Engineering and Computer Science, Kyushu University

Suzuki, Takahiko
Research Institute for Information Technology, Kyushu University

Hirokawa, Sachio
Research Institute for Information Technology, Kyushu University

http://hdl.handle.net/2324/1498216

Predicting and Visualizing Wine Characteristics Through Analysis of Tasting Notes From Viewpoints

Brendan Flanagan[1], Nao Wariishi[2], Takahiko Suzuki[3], Sachio Hirokawa[3]

[1] Graduate School of Information Science and Electrical Engineering, Kyushu University
b.flanagan.885@s.kyushu-u.ac.jp
[2] Department of Electrical Engineering and Computer Science, Kyushu University
wariishi.nao.141@s.kyushu-u.ac.jp
[3] Research Institute for Information Technology, Kyushu University
{suzuki, hirokawa}@cc.kyushu-u.ac.jp

When describing complex characteristics of a specific genre, specialist expressions are often used. This can become quite a problematic situation for an inexperienced person, as expressions not used in everyday language are difficult to understand. This is particularly apparent when trying to describe wines, known as winespeak, as a range of specialist expressions are used in a subjective manner. In this paper, we propose that the descriptions of wines can be analyzed from various points of view to automatically predict and visualize the characteristics described within the expressions as a radar chart. This would enable those not knowledgeable in winespeak to visualize and compare the complex descriptions often found in expert tasting notes.

There are many papers on research into the language that is used to describe wines, called winespeak. Some of this research is dedicated to analyzing wine tasting notes from different points of view. Paradis and Eeg-Olofsson (2013) examined tasting notes to identify expressions and words that are related to the viewpoints of vision, smell, taste, and touch. 39 typical phrases of these sensory expressions were identified. Caballero (2007) focused on how manner-of-motion verbs are used from the point of view of describing a wine's intensity and persistence, and collected 56 typical sentences that contain such verbs.

There is also related research into the visualization of wine tasting notes for linguistic analysis. Kerren et al. (2011) visualized wine tasting notes using word trees generated from parts of speech and words. Their system enables the analysis of linguistic patterns within single wine reviews or based on regions and varieties. However the system is highly specialized and not intended for general use. In previous research, we examined the relations of Winespeak expressions and visualized these as mindmaps (Hirokawa, 2014).

In this paper, we propose that tasting notes can be analyzed to predict the classification of wines from various points of view. We have collected 91,010 wine tasting notes from the Wine Enthusiast website. A portion of this data has been manually classified by hand from different points of view. This will then be used to train and evaluate the prediction performance of Support Vector Machine (SVM) models for the automatic classification of wines from each viewpoint. The results of the classifiers are then combined to form a viewpoint vector that will be used to visualize characteristics as a radar chart to provide an overview and comparison of wine features as seen in Figure 1.
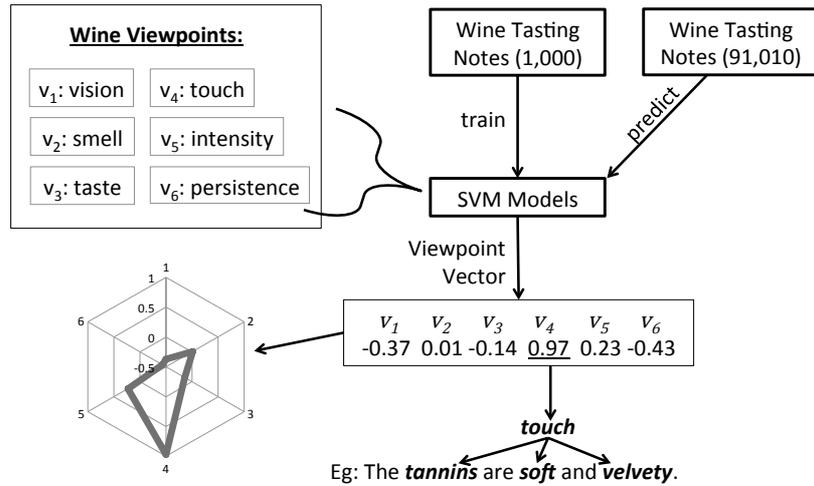
Fig. 1. An overview of the automatic prediction and visualization of wines from multiple viewpoints.

As a proof of concept, we conducted an experiment focusing on the *"touch"* viewpoint as described in Paradis and Eeg-Olofsson (2013). SVM training and testing data was prepared by manual classification of 1000 wine tasting notes from the viewpoint of touch. Initially, a model was created using all the data as training data for the purpose of scoring individual feature words. Each word w_i was evaluated using a classifier trained with all words to calculate a score of the word's relevance $score(w_i)$ to the model. The word score was used as a feature rank to search for an optimal feature selection. A search was conducted by selecting the top N words by the absolute value of the $score(w_i)$. The prediction performance at each stage of the search was evaluated using 5-fold cross validation with 80% training data and 20% testing data.
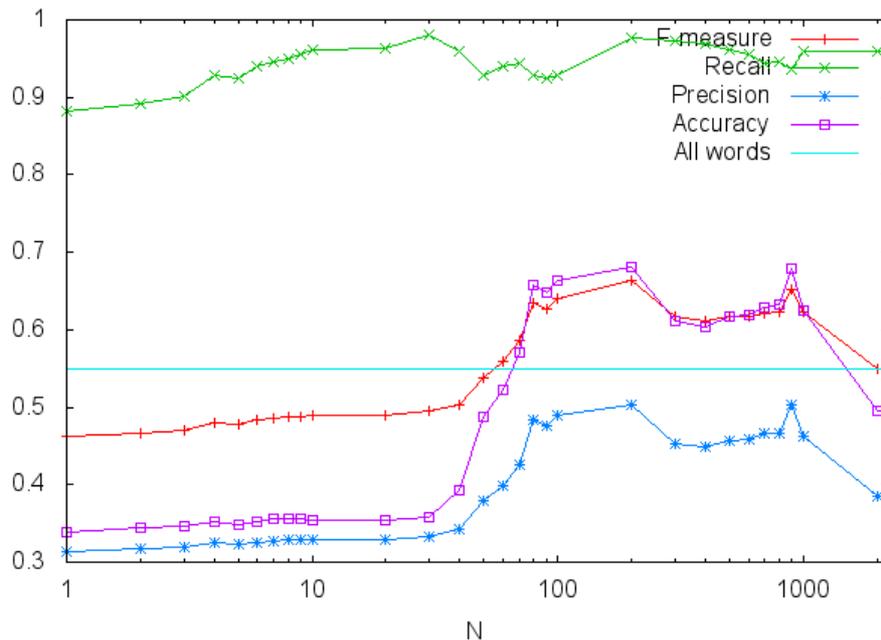

Fig. 2. Prediction performance of feature selection using the top N absolute value score words.

The results of the search for an optimal feature set are shown in Figure 2. The baseline prediction performance is an F-measure of 0.54 for a model created by analyzing all 2224 feature words from the sample tasting notes. Prediction performance peeked at an F-measure of 0.66 for a model created by analyzing 200 of the top scoring words. This indicates that the top 200 words are representative features for the *touch* viewpoint.

Top 10 Positive Score Words		Top 10 Negative Score Words	
Score(wi)	word	Score(wi)	word
1.2118	crisp	-0.2060	what
0.9906	mouthfeel	-0.2062	there
0.9653	soft	-0.2201	cranberry
0.7595	tannin	-0.2214	you
0.7129	sharp	-0.2252	fairly
0.6691	mouth	-0.2269	refreshing
0.6087	round	-0.2332	over
0.5885	weight	-0.2352	beautifully
0.5445	feel	-0.2356	cabernet
0.5350	dry	-0.2525	spice

Table 1. Top 10 positive and negative score words.

The top 10 positive and negative feature words are shown in Table 1. Most of the top 10 positive words are adjectives or nouns that are related to touch. Although the word "tannin" is not directly related to the *touch* viewpoint, it has a high word score, and as observed during the manual classification of testing data, this could be attributed to its frequent use with touch adjectives. An example of the use of touch related adjectives in conjunction with "tannin" can be seen in the Rey Sol 1999 Barbera wine tasting note (H, S. 2002): "It's rich, round, soft, lush and velvety, and very dry, with exquisite tannins". We hypothesize that through the analysis of other viewpoints, subsets of feature words can be identified and be used to score the representative characteristics of wines for automatic visualization.

References:

Caballero, R., Manner-of-motion verbs in wine description, Journal of Pragmatics 39(12), pp.2095-2114, 2007

H, S., Rey Sol 1999 Barbera Wine Tasting Note, 2002, Retrieved October 10, 2014, from http://buyingguide.winemag.com/wineries/rey-sol/varietals/barbera

Hirokawa, S., Flanagan, B., Suzuki, T., & Yin, C., Learning Winespeak from Mind Map of Wine Blogs, In *Human Interface and the Management of Information. Information and Knowledge in Applications and Services*, pp.383-393, 2014

Kerren, A. , Prangova, M., Paradis, C, Visualization of sensory perception descriptions, Proceedings of the International Conference on Information Visualization, pp. 135-144, 2011

Paradis, C., Eeg-Olofsson, M., Describing Sensory Experience: The Genre of Wine Reviews, Metaphor and Symbol, 28(1), pp.22-40, 2013