AN A.I. FOR WINE?

USING EXPERT REVIEWS TO PREDICT WINE CHARACTERISTICS

Jeff Sheng, Andy Tsao, and Vincent Sheu, CS 221 Final Project Poster (Project mentor: Steve Mussmann)

"Lush, plush, and absolutely delicious, this shows a range of softly approachable flavors including blackberry, cherry, currant, anise, cocoa, mineral, buttery, oak, and spice. The tannins are thick but finely ground. It's firmly set in the modern cult style and may not be an ager but it sure impresses now."

- Constant 2009 Cabernet Sauvignon, Points: 94, Price $130, Napa Valley, California, United States, Cabernet Sauvignon

Project Summary:.

• Can we use various artificial intelligence/machine learning methods to make predictions of a wine's score, price, variety (grape) and country of origin?
• We use text-based feature extraction, regression/loss, sentiment analysis, and deep learning to do prediction tasks based on brief wine reviews
• Preliminary results show that sentiment analysis and price are strong predictors for wine score.
• Furthermore, a unigram/bigram model with SVD for dimensionality reduction can predict grape variety and country of origin very well.
• Initial tests with a bidirectional RNN using the reviews (in the form of GloVe 200 Dimensional Word Vectors) to predict scores show that deep learning models may also be a promising, but currently do not outperform other methods. In our final paper, we will explore refinements of each of these models.

"A rough, pinchy, nose with leather and funk is no way to begin. The palate stays hard and spiky with burnt, rustic, murky flavors. Roasted and mossy on the finish and not very good overall."

- Cruz Alta 2007 Grand Reserve Malbec Mendoza, Points: 80, Price: $20, Mendoza, Mendoza Province, Argentina, Malbec

Model 1: Sentiment Analysis and Predictions

• When wine shopping, customers often use price as a proxy for quality (here, represented by score). Regression on 100,000 wines shows that price is a reasonably-good predictor of quality (here, represented by a wine's score/points). Given that the 100-point wine rating system's bands are typically 5 points wide, a mean squared error of 7.87 corresponds to a mean deviation of 2.80 points—meaning that price, alone, often predicts the correct band.

• Using Google's cloud platform for natural language processing, we conducted sentiment analysis on wine reviews. Here, we show that review sentiment (as defined by Google) also appears to be a strong predictor of score/points. Also as expected, score/points & review sentiment together provide a marginally stronger prediction than either alone.

• In the future, we'd like to run sentiment analysis using only wine reviews as a training set—Google's blackbox analysis probably trains on general text from the Internet, and we feel limiting the space to wine reviews may improve the predictive power of review sentiment. We also plan on segmenting data out to see if predictive accuracy is particularly low/high for wines in a given band of scores.

Regression	Mean Squared Error	Slope	Intercept	n
Score (points) v. Price ($) [base]	7.87	0.0297	87.3959	100,377
Score (points) v. Review Sentiment & Price ($)	7.34	Sentiment: 2.3915 Price: 0.0296	86.2731	100,377
Score (points) v. Review Sentiment	8.98	2.4273	87.3564	100,377
Price ($) v. Review Sentiment	1879.18	1.2079	36.6007	100,377

Data

Our data comes from a publicly available dataset courtesy of Zach Thoutt, posted on Kaggle.com, and is a scraping of the Wine Enthusiast website on November 22, 2017. Fields include:

Points: Wine Enthusiast rating on a scale of 80-100.
[e.g. "Classic: a great wine" is 95-100; "Good: a solid, well-made wine" is 80-84]
Variety: Type of grapes used to make the wine (i.e. Pinot Noir)
Review: Sommelier writing about the wine's taste, smell, look, feel, etc.
Country: Country that the wine is from
Province: Province or state that the wine is from
Region 1: Wine growing area in a province or state (i.e. Napa)
Winery: Winery that made the wine
Price: Cost per bottle

*From the original dataset, we cleaned by removing wines (with varietals) that only had a small number of data points, taking out the 10% rarest wines. In addition, we also performed some sanitation—date ranges with "â€"" were replaced with a hyphen ("-") (i.e. 2017â€"2019 was replaced with 2017-2019). Finally, for price-related analysis, any wines without price data were removed from the dataset as well.

Stanford University

Figure 1: Score v. Price Baseline (Model 1: Sentiment Analysis)

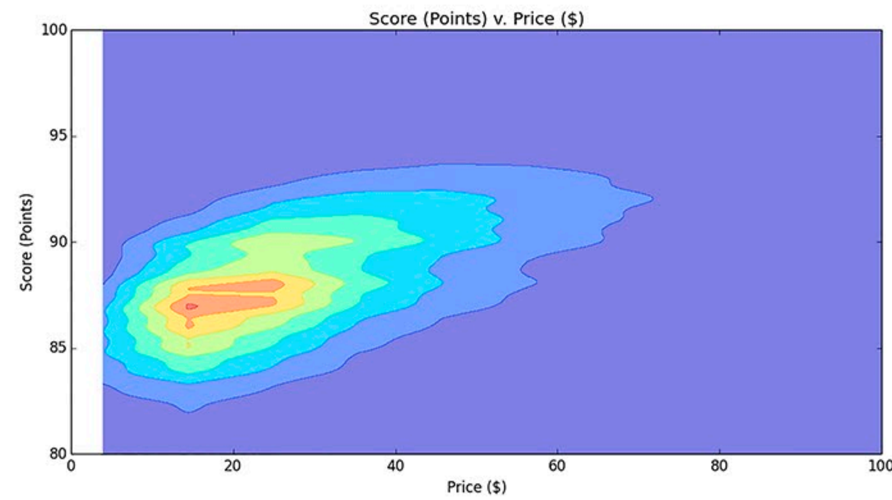


Figure 2: Score v. Sentiment (Model 1: Sentiment Analysis)

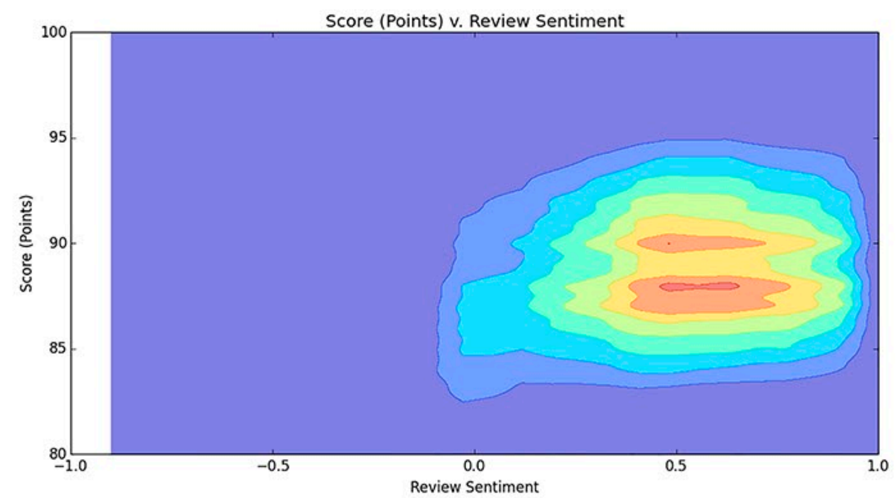


Figure 3: Price v. Sentiment (Model 1: Sentiment Analysis)

Model 2: Predicting Categorical Responses Using Word Count Features

• We extracted from each review a set of words with weights corresponding to how frequently the word occurs in the review relative to the entire corpus. More explicitly, the weight of each word is given by

$$tf(t,\ d) \cdot \log\left(\frac{1+n_d}{1+df(d,t)}\right)+1$$

where *tf(t, d)* is the frequency of the word in the review and *df(d, t)* is the number of reviews that contain the word. Intuitively, a word is weighted highly if it occurs in a particular review but is rarely seen in other reviews.

• Each feature vector is sparse, since each review only utilizes a small fraction of the total number of words. We therefore decided to use singular value decomposition to create a low-rank estimate of the feature matrix.

• When reducing the features to a 200-dimensional subspace, we were able to achieve a prediction accuracy of 59% for grape varietals and 81% for country of origin.

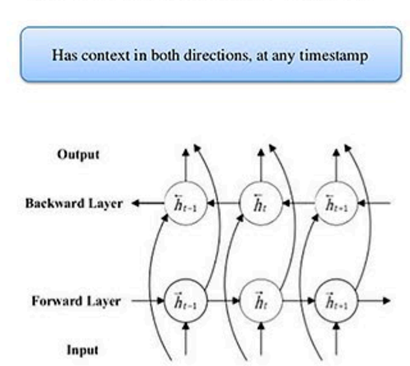
Model 3: Deep Learning with RNNs

• We began experimenting with recurrant neural networks in the form of bidirectional RNNs, to use reviews (as GloVe 200 Dimensional Word Vectors) to predict scores.

• On the bottom right, we show the results after each epoch of the training loss and dev loss using 10 training epochs. On the last epoch, we see a training loss of 16.44.

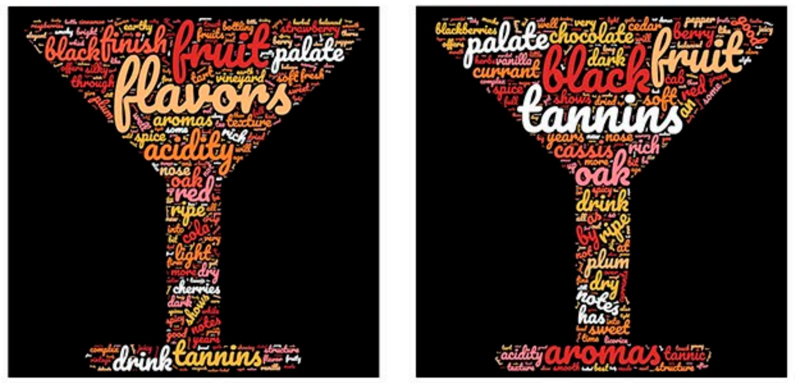
• While not as good as sentiment analysis, the results are promising and can be fine tuned with experimenting with different models and inputs for different characteristic predictions which we plan to do for our final paper and final project results.

Bidirectional RNN



Comparison of two popular white wines. On the left is Chardonnay, which is characterized by an oaky, fruity taste, while the Sauvignon Blanc on the right is typically crisper and lighter.

Comparison of two popular red wines. On the left is Pinot Noir, which is often described as being lighter and more acidic. On the right is Cabernet Sauvignon, which offers a heavier, fuller taste.



Training/Dev Loss Results from Bidirectional RNN
Reviews (as 200-dim GloVe word vectors), Predicting Score

```
0
Training loss:  369.452255785
Dev error:  [ 68.13703616]
1
Training loss:  36.4793592662
Dev error:  [ 66.11164354]
2
Training loss:  31.5654218058
Dev error:  [ 29.8079938]
3
Training loss:  25.2020678571
Dev error:  [ 27.83696981]
4
Training loss:  24.3824201522
Dev error:  [ 26.11824104]
5
Training loss:  23.3223425541
Dev error:  [ 21.65741938]
6
Training loss:  22.7056102823
Dev error:  [ 22.77718536]
7
Training loss:  22.7576878577
Dev error:  [ 18.97570778]
8
Training loss:  21.5252120498
Dev error:  [ 20.12926612]
9
Training loss:  21.1015245368
Dev error:  [ 16.44133518]
```