Wine Recommendation System

Marlee Bryant
College of Engineering
The University of Alabama
Tuscaloosa, Alabama, United States
mabryant4@crimson.ua.edu

***Abstract*—**

I. Introduction

A wide variety of applications utilize recommendation systems, which analyze the user's preferences, with or without their specific input, to tailor their digital experience, ideally improving satisfaction and engagement. Users have come to expect and seek out this kind of personalization, especially when a wide variety of options are available and there exists a large degree of uncertainty regarding their quality. Wine selection certainly falls into this category, given the unique vocabulary related to flavors and aromas as well as the elitism surrounding having good taste. The average consumer requires a simple means of determining which wines they will enjoy based on prior knowledge of their experiences.

Wine selection typically begins with the decision between the three broad categories of Red, White, or Rosé, then narrows to more specific categories like Chardonnay, Riesling, or Pinot Grigio. Unfortunately for the consumer, there is still an overwhelming number of options at this point which vary greatly in taste and aroma. Turning to expert or consumer ratings is a possible solution for narrowing these options and machine learning algorithms exist for this kind of analysis, but taste in wine is highly subjective.

The proposed algorithm in this paper will account for these factors, as well as the preferences of the user. For beginners, these preferences can be elicited from other food, drink, and scent preferences, or for more experienced wine consumers, comparisons can be made to wines they already know they enjoy. Additionally, when users try recommended wines more data can be gathered from their specific ratings and reviews. Wine recommendations can also be filtered by objective features, such as price and category. By accounting for individual taste in addition to general and expert taste, consumer satisfaction with wine purchases can be increased, providing greater comfort for wine exploration.

II. Related Works

Specific research on wine recommendation systems is sparse, and instead wine-related machine learning research is focused on determining the general quality of wine [1]. As a result, this section will discuss food recommendation systems instead, which have many similar components to a wine recommendation system, in that they utilize user preferences such as ingredients, cuisines, and diet styles to generate new recommendations. Some approaches also utilize similar user's preferences by including data from ratings and reviews. This aligns well with a wine recommendation system utilizing flavor profile, wine categories, and consumer and expert ratings.

The majority of food recommendation systems used a content-based approach, where the user's preferences are compared to the characteristics of the foods to determine their degree of similarity [2].

Nilesh et al. use a content-based approach for Indian food recommendation by requesting ingredients or recipes that the user enjoys and finding new recipes containing similar ingredients [3]. They do so by analyzing the text of recipes to group together ingredient keywords and compare them. Their approach produced inconsistent results across the set of different ingredients, as some ingredients might be very common, like rice, and fail to yield highly personalized results. They also make the assumption that user's want to consume recipes with the same ingredients, when instead analyzing the flavor category of each ingredient might yield better results. This is because a recipe containing a certain ingredient, for example chicken, could have many different flavor profiles, such as spicy, creamy, or salty. This problem is somewhat improved when their algorithm is provided with more ingredients as input, since this provides a better description of the flavor the resulting recipe will have. A solely content-based approach, especially one based on so little input data, will likely fail to be sufficiently specific to each individual user.

David Massimo et al. attempt to take a somewhat more collaborative approach to home-cooked recipe recommendation by using both content-based comparison, and user tags and ratings [4]. After the user enters their preference information by adding recipes they like, they are prompted with a variety of recipe options and asked to rate them then provide tags explaining their reasoning behind their rating. This information is used to train a model which will then provide tailored results from both the content-based and collaborative analysis. They tested their algorithm in the form of an android application and compared results of the hybrid approach to those of only content-based and only collaborative approaches, determining that the hybrid approach had the best performance. The addition of user tags is important to the performance of their algorithm, but is provided as an optional choice, so for a lazy user who may choose not to include tags, the quality of the algorithm performance would decrease. The training set for a collaborative model needs to be robust in order to ascertain an appropriate model for recommendations.

III. Approach

Currently, machine learning related to wine is primarily oriented towards determining its general quality based on expert's qualifications [1]. The proposed algorithm will instead focus more on user-specific taste to produce more satisfactory representations for the average consumer. This algorithm will more closely resemble those used in food recommendation systems, which combine the data of general ratings and reviews with user-provided data in order to personalize the results.

The three main categories of recommendation systems are content-based, collaborative, and hybrid [2]. A content-based approach to wine recommendation would compare features of a specific wine to the user's indicated preferences and recommend wines which closely align with these. A collaborative approach would compare the user's indicated preferences to other user's preferences and suggest wines liked by this other user. The proposed algorithm uses a hybrid approach, which combines the elements of both content-based and collaborative filtering, weighing their results to produce a more accurate recommendation.

A. Data Collection

This algorithm will require three different collections of data: the user's preferences, the features of each wine, and other user's preferences for comparison.

The user's preferences will be derived in two different ways depending on whether they qualify themselves as a new or experienced wine consumer. For new consumers, they will be asked a series of questions regarding their preferences of other drinks, foods, and aromas. Experienced wine drinkers will be asked questions related to which categories of wine they consume the most and what flavor notes they enjoy. Both groups will be given the opportunity to input wines they have tried and provide a rating of these, allowing the opportunity for the dataset of wine ratings to begin to be populated.

Data related to the wine will be scraped from a variety of websites, searching for information related to type of wine, flavor notes, and ratings and reviews. This will require processing of the text in descriptions of the products and reviews. Ideally, the dataset will include each individual wine only once and combine data ascertained from different websites.

Other user's preferences will be emulated by scraping of expert and consumer blogs and reviews, which will also require processing the text for keywords. Additional data can be provided through other user's real or simulated experience of the algorithm, including their preferences profile and ratings of wines.

B. Data Analysis

For the content-based portion of the analysis, the data will be grouped into vectors representing its characteristics with numerical values representing the degree to which it aligns with each characteristic. For the wine vectors, some characteristics will simply be 1 or 0, such as the category, and others will be represented on a scale from -1 to 1, such as sweetness or boldness. For the user vectors, all characteristics will be on the scale from -1 to 1, because even for categories that are either existent or non-existent for each wine, like whether it is white or red, the user's preference may not be equally all or nothing, like if they prefer white wine but occasionally enjoy red. The wine and user vectors will have the same structure otherwise, which will be useful for comparison in the algorithm.

For the collaborative portion, the data will be used to create a training set, with the users preference information as the input and the degree to which they liked each wine as the output.

C. Algorithm Details

The algorithm approach for the content-based portion of the recommendation system simply involves taking the dot product of the user vector with each of the wine vectors to produce a similarity metric [2].

The algorithm approach for the collaborative portion will utilize a learning decision tree generated from the training set data. Since the input and output values are continuous, they will be subdivided using split points. The decision tree structure will be determined by a greedy divide-and-conquer function which begins with data deemed to be most important. A calculation of the information gain of each characteristic will be used to determine its importance, and the tree will be optimized to provide most information gain with the least complexity. The tree will then be pruned using a significance test to ensure that all nodes provide relevant information.

Since the collaborative portion of the approach is significantly more time-consuming than the content-based portion, the collaborative algorithm will be performed on only a portion of the results generated from the content-based algorithm. This will allow the wines already deemed as similar to the user's preferences to be further curated, narrowing the results to a manageable amount of more accurate recommendations.

References

[1] S. Aich, A. A. Al-Absi, K. Lee Hui and M. Sain, "Prediction of Quality for Different Type of Wine based on Different Feature Sets Using Supervised Machine Learning Techniques," 2019 21st International Conference on Advanced Communication Technology (ICACT), PyeongChang KwangwoonDo, Korea (South), 2019, pp. 1122-1127.

[2] A. Kumar, P. Tanwar and S. Nigam, "Survey and evaluation of food recommendation systems and techniques," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 3592-3596.

[3] David Massimo, Mehdi Elahi, Mouzhi Ge, and Francesco Ricci. 2017. "Item Contents Good, User Tags Better: Empirical Evaluation of a Food Recommender," Session 1: Late - breaking Results and Demo,UMAP '20 Adjunct, July 14–17, 2020, Genoa, Italy 75 System. In Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization. ACM, 373–374.

[4] N. Nilesh, M. Kumari, P. Hazarika and V. Raman, 'Recommendation of Indian Cuisine Recipes Based on Ingredients," 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW), Macao, Macao, 2019, pp. 96-99.