Wineinformatics: The Evaluation of the Computational Wine Wheel in the Past Decades

Long Le[1]; Bernard Chen[2]*

[1]Department of Mathematics, University of Central Arkansas, Conway, AR 72034, USA

[2]Department of Computer Science and Engineering, University of Central Arkansas, Conway, AR 72034, USA

*Corresponding author: Bernard Chen

Department of Computer Science and Engineering, University of Central Arkansas, Conway, AR 72034, USA.
Tel: +1 501 450 3308
Email: bchen@uca.edu

Abstract

The Computational Wine Wheel (CWW) emerged in 2014 as a response to the limitations of the traditional Wine Aroma Wheel. This innovative tool, blending the concepts of wine aroma classification and natural language processing, introduced a novel approach to analyzing wine attributes. Initially developed with a focus on the top 100 wines from Wine Spectator in 2011, the CWW underwent successive iterations, culminating in the latest version, CWW 3.0. With expanded categories and subcategories, as well as the inclusion of reviews from multiple sources including Robert Parker's Wine Advocate, the CWW has evolved into a comprehensive resource for wine analysis. Through the creation of significant datasets like the Elite Bordeaux dataset and the Big dataset, the CWW has facilitated extensive research on wine attributes and trends. The ongoing development of the CWW underscores its importance as a dynamic tool in the field of Wineinformatics, promising continued advancements in wine analysis and understanding.

Keywords: Wineinformatics; Computational Wine Wheel; Wine Reviews; Natural Language Processing

Introduction

Wine is one of the most popular kinds of beverage in the world. Mankind has been fermented fruits, such as grapes, peaches, or berries, into wine for thousands of years. Red wine, white wine, and sparkling wine are a few popular types of wines consumed around the globe. Wine is distinguished based on many characteristics: the fruit used in its preparation, the year it was produced, the region at which the fruit is grown. Furthermore, a wine is also characterized by its sweetness, tannins, color, and aroma.

The complication in wine and the art of wine making requires an expertise level to understand. The study and science of wine and wine making is called oenology, or enology. The role of an oenologist is to perform wine analysis, monitoring quality control parameters, and make decisions during the winemaking process based on analytical and sensory descriptions of a wine [1]. A viticulturist is an expert in growing grapes, in particular for winemaking. In particular, a viticulturist is in charge of pest control, fertilizing, pruning the vines, monitoring the development of the fruits including deciding when to harvest. The role of a sommelier, on the other hand, is to taste and make recommendations as a form of wine reviews to consumers based on the quality of a wine.

In recent years, new technology has been utilized in oenology and viticulture [2]. Wineinformatics, a field of study that employs digital technology to gather and transform large amounts of wine review data into useful knowledge through various ma-chine learning algorithms [3], proves more beneficial to winemakers than analyzing wine's physicochemical composition, encompassing acidity, residual sugar, alcohol content, and other pertinent parameters [4-7]. Figure 1 provides an example of a wine evaluation from In recent years, new technology has been utilized in oenology and viticulture [2]. Wineinformatics, a field of study that employs digital technology to gather and transform large amounts of wine review data into useful knowledge through various machine learning algorithms [3], proves more beneficial to winemakers than analyzing wine's physicochemical composition, encompassing acidity, residual sugar, alcohol content, and other pertinent parameters [4-7]. Figure 1 provides an example of a wine evaluation from both perspectives.

In consumers' perspective, using wine reviews is considered more approachable to learn about the quality and the characteristics of a wine than using physicochemical laboratory data. Laboratory data is often difficult to obtain due to its associated cost and in general not for consumers. Wine reviews, on the other hand, are much more available. There are hundreds of wine review magazines and websites, such as Wine Spectator, Robert Parker, or wine.com. Each of these sources has hundreds of thousands of reviews by wine experts for a wide range of wines.

However, structured data, such as physicochemical laboratory records, can be easily interpreted and analyzed by computers. In contrast, unstructured data, like wine reviews showed in

figure, require natural language processing techniques to enable computers to understand human language-based reviews. With a vast repository of millions of wine reviews sourced from diverse outlets, often available at minimal expense, there lies immense potential to uncover valuable insights beneficial to a wide audience. Therefore, the Computational Wine Wheel (CWW) was developed started from 2014 to focus on processing wine reviews into computer understandable format so that data mining techniques, such as classification, clustering, association rules and regression, can be applied to the data collected from the reviews various wine related knowledge discovery [8-14]. In this review, the evolution of Computational Wine Wheel in the past decade is described and discussed, transitioning from single-source to multi-source evaluation.

The evolution of the Computational Wine Wheel (CWW)

In order to use words in wine reviews to classify wine scores, the very first Computational Wine Wheel (CWW) has developed in [15]. It is essentially a filter, or a sieve, that collects important words, or attributes, from a review. These attributes, then, are used in classification algorithms to determine the quality of the wine. For example, below is the review of Dow's Vintage Port 2011 from Wine Spectator:

Powerful, refined and luscious, with a surplus of dark plum, kirsch and cassis flavors that are unctuous and long. Shows plenty of grip, presenting a long, full finish, filled with Asian spice and raspberry tart accents. Rich and chocolaty. One for the ages. Best from 2030 through 2060.

The bolded words are the attributes that can be extracted from the review. These attributes can be of different types: savory (chocolaty, tart), body (long), or adjective (powerful, refined). In order to create the Computational Wine Wheel, a technique in natural language processing called "Bag of Words" is used.

Bag of Words

From a collection of reviews, "Bag of Words" extracts the attributes by tokenizing, removing the stop-words, normalizing the tokens, and creating the master dictionary. The first step, tokenizing, is done by breaking the sentences into unique words and phrases. Then, the stop-words, such as articles (a, an, the) or prepositions (to, for), are removed from the list of words. Token normalization condenses words with similar meaning into a single representation, such as "red" and "reddish" into "red." Finally, all the words that are still in the list at the end form a

master dictionary, or a sieve, to be used to extract attributes from reviews.

Note that attributes can contain more than one word. For example, "raspberry tart" is considered to be one attribute. Another note is that similar words are not always necessarily normalized into a single attribute. For example, "apple" and "fresh apple" are condensed into "apple," but "green apple" is considered a separate attribute, because in wine, green apple gives a distinct flavor. This is where domain knowledge is very important when it comes to natural language processing.

Wine Aroma Wheel

One of the earlier attempts to create a master dictionary for wine reviews is the Wine Aroma Wheel, developed by Ann. C. Nobel [16]. It contains words that describe fragrance and flavors and consists of 12 categories, each with subcategories that maps to different taste, scent and aromatic qualities of red and white wines. While the Wine Aroma Wheel is useful to study wine, one of its limitations is that it does not include adjective and wine body attributes. For example, if the Wine Aroma Wheel is applied to the review of Dow's Vintage Port 2011, the only attributes extracted are dark plum, kirsch, cassis, Asian spice, raspberry tart, and chocolaty, which are all savory attributes.
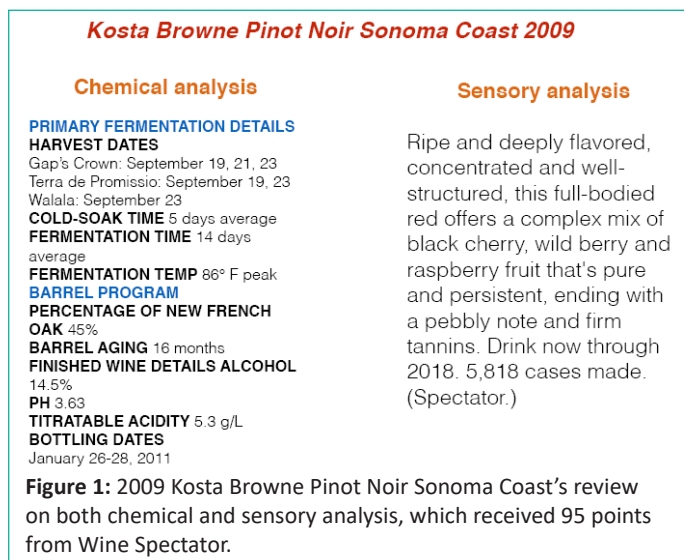
Wine Reviews

Quality data with minimal noise is essential for successful data science research. Therefore, the cornerstone of this research lies in high-quality wine reviews. According to the Wine School of Philadelphia, there exist five prominent wine review platforms offering extensive databases of professional wine critiques.

Wine Spectator [23], a renowned wine magazine since 1979, meticulously evaluates over 15,000 wines worldwide. It stands as a pinnacle in the field and a primary source for previous Wineinformatics research. Wine Enthusiast [24], established in 1988, annually reviews a staggering 24,000 wines. Its reviews are accessible for free with a simple email registration. Beyond wine critiques, the magazine encompasses a broad-spectrum including wine accessories, storage solutions, education, food pairing, and lifestyle features.

Antonio Galloni's Vinous [25], established in 2012, boasts a team of top-tier wine critics offering online wine evaluations. It enjoys high esteem in the wine trading realm, emerging as a trailblazer in contemporary wine literature. Robert Parker's Wine Advocate [26], arguably the most eminent wine critic, revolutionized the industry with his Wine Advocate magazine in 1978. Introducing the widely adopted 100-point scale, his influence has been profound, shaping the wine trade landscape significantly. Decanter [27], founded in London in 1975, not only offers comprehensive wine reviews but also delves into wine producers and regions. While their critiques are detailed and focused, they occasionally diverge from the perspectives of other aforementioned magazines.

Among various prestigious wine magazines, Wine Spectator can be considered as an easier data source to start aggregating wine reviews because of their strong on-line wine review search database and consistent wine reviews. These reviews are mostly comprised of specific tasting notes and observations while avoiding superfluous anecdotes and non-related information [3].

Kosta Browne Pinot Noir Sonoma Coast 2009

Chemical analysis

PRIMARY FERMENTATION DETAILS
HARVEST DATES
Gap's Crown: September 19, 21, 23
Terra di Promissio: September 19, 23
Walala: September 23
COLD-SOAK TIME 5 days average
FERMENTATION TIME 14 days average
FERMENTATION TEMP 86° F peak
BARREL PROGRAM
PERCENTAGE OF NEW FRENCH OAK 45%
BARREL AGING 16 months
FINISHED WINE DETAILS ALCOHOL 14.5%
PH 3.63
TITRATABLE ACIDITY 5.3 g/L
BOTTLING DATES
January 26-28, 2011

Sensory analysis

Ripe and deeply flavored, concentrated and well-structured, this full-bodied red offers a complex mix of black cherry, wild berry and raspberry fruit that's pure and persistent, ending with a pebbly note and firm tannins. Drink now through 2018. 5,818 cases made. (Spectator.)

Figure 1: 2009 Kosta Browne Pinot Noir Sonoma Coast's review on both chemical and sensory analysis, which received 95 points from Wine Spectator.

Computational Wine Wheel

In order to overcome the Wine Aroma Wheel's limitations, [17] introduced the Computational Wine Wheel (CWW) for the first time in 2014. In that research, top 100 wines from Wine Spectator in 2011 was studied and utilized with bag of words technique to create the master dictionary. Merging the concept of wine aroma wheel and natural language processing methods, not only were flavor attributes included, other physical attributes, such as acidity or tannins, were also present. The end result was the first version of the CWW, with 12 categories and 28 subcategories that contained 547 original tokens and 376 normalized tokens. Both hierarchical clustering and association classification algorithms were applied on the dataset contained 1000 wine reviews processed by the first CWW and received satisfactory results [17].

Later on, [18] improved the original CWW by using Wine Spectator's reviews of the top 100 wines in a 10-year period, from 2003 to 2013. The newer version, CWW2.0, was extended to include 14 categories with 34 subcategories. There are 1932 original tokens and 986 normalized tokens for a more comprehensive list of attributes. From 2016~2022, numerous dataset were generated and processed by the CWW2.0 to discover different types of information related to wine: In [13, 20-21], a large dataset contains more than 100,000 Wine Spectator wine reviews with vintage 2006-2015 were collected and processed thorough CWW2.0 to study the ranking of wine reviewers [20], regression on wine price and grade [21] as well as multi-label and multi-target methods in Wineinformatics [13]; a smaller dataset targeted on Bordeaux with more than 14,000 wine reviews and an even smaller focused on elite Bordeaux with 1359 wine reviews were proposed and studied to know more about 21st century Bordeaux wines [12]. The large Bordeaux dataset is currently publically available through IEEE data port [22].

In 2022, Robert Parker's reviews were included in the Wineinforamtics researches and performed a direct comparison with Wine Spectator's review [11]. In 2023, the Computational Wine Wheel 2.0 was revised into 3.0 by including reviews from Robert Parker's Wine Advocate and adopts Neural Networks into Wineinformatics researches [19]; 513 of Robert Parker's elite Bordeaux reviews were included in the creation of the new Computational Wine Wheel. More attributes were added to the new version CWW3.0 to increase the number of original tokens to 2589 that were condensed to 1191 normalized tokens. Below is a table that compares all 3 versions of the Computational Wine Wheel. The subcategories with no word count in the CWW columns were not included in the original wheel. Since this review is focus on the evolution of the Computational Wine Wheel, how to utilize the CWW is clearly described in [3 and 19]. Generally speaking, the more tokens stored in the Original (3rd column in table1), the more words used by human language can be picked up by the Computational Wine Wheel; meanwhile the more tokens used by Normalized (4th column in table1), the more attributes can be provided to the machine learning algorithms. Therefore, table1 clearly suggests that the evolution of the Computational Wine Wheel has made it far more robust compared to its state a decade ago.

Two major datasets were developed through the CWW3.0 [19]: an Elite Bordeaux dataset and a big dataset. The Elite Bordeaux dataset contains 513 elite Bordeaux wines with BOTH Wine Spectator's and Robert Parker's wine reviews. The Big dataset contains BOTH Wine Spectator's and Robert Parker's wines reviews from Bordeaux (2341 wines), Italy (3198 wines),

Table 1: Number of tokens in each subcategory comparison between three versions of the Computational Wine Wheel.

Category	Subcategory	Original			Normalized		
		CWW1.0	*CWW2.0*	*CWW3.0*	*CWW1.0*	*CWW2.0*	*CWW3.0*
Caramel	Caramel	9	71	97	7	40	56
Chemical	Petroleum	3	9	11	1	5	6
	Sulfur		11	11		10	10
	Pungent		4	4		3	4
Earthy	Earthy	18	72	128	2	31	47
	Moldy		2	2		2	2
Floral	Floral	15	61	87	15	39	45
Fruity	Berry	18	49	84	15	28	39
	Citrus	11	37	56	11	23	35
	Dried Fruit	21	67	76	21	60	65
	Fruit	5	22	42	4	9	16
	Other	7	25	22	7	18	9
	Tree Fruit	12	39	55	9	31	40
	Tropical Fruit	15	48	67	11	27	36
Fresh	Fresh	15	41	75	12	29	44
	Dried	6	25	50	6	21	39
	Canned/Cooked	7	16	18	7	15	17
Meat	Meat	1	25	36	1	13	21
Microbiological	Yeasty	3	5	5	3	4	4
	Lactic	3	14	14	2	6	6
Nutty	Nutty	3	25	27	3	15	20
Overall	Tannins	24	90	124	3	4	6
	Body	17	50	61	10	23	17
	Structure	9	40	51	2	2	2
	Acidity	14	40	61	3	3	4
	Finish	50	184	233	6	5	13
	Flavor/Descriptors	217	649	889	179	432	467
Oxidized	Oxidized		1	2		1	2
Pungent	Hot		3	3		2	2
	Cold		1	1		1	1
Spicy	Spice	26	83	85	21	44	53
Wood	Resinous	6	24	31	6	9	12
	Phenolic	1	6	6	1	4	5
	Burned	11	47	51	8	26	28

and California (4180 wines); therefore, the big dataset comprises 10,232 wines with a total of 20,464 wine reviews. It includes the name, vintage, score, and wine reviews for each wine, providing a comprehensive overview of all the wine reviews collected. Currently, three major datasets that contains more than 200,000 wine reviews are under development as the latest projects: all wines review available in Wine Spectator after year 2000; all wines reviews available in Robert Parker's Wine Advocate after year 2000; and all wines reviews available in Both

Wine Spectator and Robert Parker's Wine Advocate after year 2000. These datasets will be heavily studied to discover possible methods for merging wine reviews to understand more about quality wines by analyzing the end product and deconstructing the sensory attributes of the wine; this process is similar to reverse engineering in the context of wine in order to study and improve the winemaking techniques employed.

Conclusion

In conclusion, the evolution of the Computational Wine Wheel has been a remarkable journey, marked by significant advancements in its capabilities and scope. From its inception in 2014 to the latest iteration in 2023, the wheel has undergone transformative enhancements driven by rigorous research and innovation. Initially introduced to overcome the limitations of the traditional Wine Aroma Wheel, the Computational Wine Wheel amalgamated the concepts of a new data science application in wine and natural language processing techniques, resulting in a pioneering tool capable of analyzing not only flavor attributes but also physical characteristics like acidity and tannins.

Subsequent versions of the Computational Wine Wheel, such as CWW2.0 and CWW3.0, expanded upon the original framework by incorporating larger datasets spanning multiple years and wine reviewers. This expansion facilitated diverse analyses ranging from wine reviewer rankings to regression on wine price and grade, as well as multi-label and multi-target methods in Wineinformatics. Additionally, the inclusion of reviews from esteemed critics like Robert Parker broadened the wheel's scope and enriched its attributes, leading to more comprehensive insights into wine characteristics and trends.

The evolution of the Computational Wine Wheel represents an ongoing endeavor. While this research focused on two renowned wine magazines, there exists a multitude of others awaiting inclusion to enrich the breadth and depth of the CWW. Furthermore, with each magazine and wine expert generating tens of thousands of reviews annually, the challenge lies in determining the optimal strategy for expanding the CWW to accommodate this vast influx of data.

In summary, the evolution of the Computational Wine Wheel stands as a testament to the power of interdisciplinary collaboration and technological advancement in the field of Wineinformatics. As researchers continue to push the boundaries of data analysis and interpretation, the Computational Wine Wheel remains at the forefront, empowering wine enthusiasts and data scientists with valuable insights into the world of wine.

Author Statements

Author Contributions

Conceptualization, B.C.; Investigation, L.L. and B.C.; Writing—original draft, L.L. and B.C.; Writing—review & editing, L.L. and B.C. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Zoecklein Bruce, Kenneth C Fugelsang, Barry H Gump, Fred S Nury. Wine analysis and production. Springer Science & Business Media. 2013.

2. Tisseyre Bruno, Hernan Ojeda, James Taylor. New technologies and methodologies for site-specific viticulture. J Int Sci Vigne. 2007; 41: 63-76.

3. Chen Bernard. Wineinformatics: A New Data Science Application. Springer Nature. 2023.

4. Cortez P, Cerdeira A, Almeida F, Matos T, Reis J. Modeling wine preferences by data mining from physicochemical properties. Decis. Support Syst. 2009; 47: 547–553.

5. Chen Mu-Chen, Long-Sheng Chen, Chun-Chin Hsu, Wei-Rong Zeng. An information granulation based data mining approach for classifying imbalanced data. Information Sciences. 2008; 178: 214-3227.

6. Capece Angela, Rossana Romaniello, Gabriella Siesto, Rocchina Pietrafesa, Carmela Massari, Cinzia Poeta, et al. Selection of indigenous Saccharomyces cerevisiae strains for Nero d'Avola wine and evaluation of selected starter implantation in pilot fermentation. International journal of food microbiology. 2010; 144: 187-192.

7. Edelmann A, Diewok J, Schuster KC, Lendl B. Rapid method for the discrimination of red wine cultivars based on mid-infrared spectroscopy of phenolic wine extracts. J Agric Food Chem. 2001; 49: 1139–1145.

8. McCune Jared, Alex Riley, Bernard Chen. Clustering in wine informatics with attribute selection to increase uniqueness of clusters. Fermentation. 2021; 7: 27.

9. Dong Zeqing, Travis Atkison, Bernard Chen. Wine informatics: using the full power of the computational wine wheel to understand 21st century Bordeaux wines from the reviews. Beverages. 2021; 7: 3.

10. Kwabla William, Falla Coulibaly, Yerkebulan Zhenis, Bernard Chen. Wine informatics: can wine reviews in bordeaux reveal wine aging capability?. Fermentation. 2021; 7: 236.

11. Tian Qiuyun, Brittany Whiting, Bernard Chen. Wineinformatics: Comparing and Combining SVM Models Built by Wine Reviews from Robert Parker and Wine Spectator for 95+ Point Wine Prediction. Fermentation. 2022; 8: 164.

12. Dong Zeqing, Xiaowan Guo, Syamala Rajana, Bernard Chen. Understanding 21st century bordeaux wines from wine reviews using naïve bayes classifier. Beverages. 2020; 6: 5.

13. Palmer James, Victor S Sheng, Travis Atkison, Bernard Chen. Classification on grade, price, and region with multi-label and multi-target methods in wineinformatics. Big Data Mining and Analytics. 2019; 3: 1-12.

14. Chen Bernard, Valentin Velchev, Bryce Nicholson, Joey Garrison, Moani Iwamura, Ryan Battisto. Wine informatics: Uncork Napa's Cabernet Sauvignon by Association Rule Based Classification. In 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). IEEE. 2015; 565-569.

15. Chen Bernard, Christopher Rhodes, Aaron Crawford, Lorri Hambuchen. Wine informatics: Applying data mining on wine sensory reviews processed by the computational wine wheel. In 2014 IEEE International Conference on Data Mining Workshop. IEEE. 2014; 142-149.

16. Noble Ann C, Rich A Arnold, John Buechsenstein, E Jane Leach, Janice O Schmidt, Peter M Stern. Modification of a standardized system of wine aroma terminology. American journal of Enology and Viticulture. 1987; 38: 143-146.

17. Chen Bernard, Christopher Rhodes, Aaron Crawford, Lorri Hambuchen. Wineinformatics: Applying data mining on wine sensory reviews processed by the computational wine wheel. In 2014 IEEE International Conference on Data Mining Workshop. IEEE. 2014; 142-149.

18. Chen Bernard, Christopher Rhodes, Alexander Yu, Valentin Velchev. The computational wine wheel 2.0 and the TriMax tri-clustering in wineinformatics. In Advances in Data Mining. Applications and Theoretical Aspects: 16th Industrial Con-ference, ICDM 2016, New York, NY, USA. Proceedings. Springer International Publishing. 2016; 16: 223-238.

19. Le Long, Pedro Navarrete Hurtado, Ian Lawrence, Qiuyun Tian, Bernard Chen. Applying Neural Networks in Wineinformatics with the New Computational Wine Wheel. Fermentation. 2023; 9: 629.

20. Chen Bernard, Valentin Velchev, James Palmer, Travis Atkison. Wineinformatics: A quantitative analysis of wine re-viewers. Fermentation. 2018; 4: 82.

21. Palmer James, Bernard Chen. Wineinformatics: Regression on the grade and price of wines through their sensory at-tributes. Fermentation. 2018; 4: 84.

22. https://ieee-dataport.org/open-access/wineinformatics-21st-century-bordeaux-wines-dataset (Assessed on 4/7/2024)

23. https://www.winespectator.com/ (Assessed on 4/7/2024)

24. https://www.wineenthusiast.com/ (Assessed on 4/7/2024)

25. https://vinous.com/ (Assessed on 4/7/2024)

26. https://www.robertparker.com/ (Assessed on 4/7/2024)

27. https://www.decanter.com/ (Assessed on 4/7/2024)