Commentary on Application of Data Mining in Fruit Quality Evaluation

Jinjian Hou, Dong Wang, Wenshen Jia, Ligang Pan

HAL Id: hal-01614200

https://inria.hal.science/hal-01614200

Submitted on 10 Oct 2017

Commentary on application of data mining in fruit quality evaluation

Jinjian Hou[1,2,a], Dong Wang[2,3,b], Wenshen Jia[2,3,c], Ligang Pan[1,2,3,d,*]

[1]China Three Gorges University, Yichang, 443002; [2] Beijing Academy of Agriculture and Forestry Sciences, Bei Jing, 100097; [3]Risk Assessment Laboratory for Agro-products of the Ministry of agriculture, Beijing, 100097

[a]hou_jinjian@163.com, [b] nirphd@163.com, [c]jiawenshen@163.com, [d]panlg@nercite.org.cn

Abstract. In order to provide reference for the fruit quality research and fruit selective breeding, in this paper, data mining methods of fruit quality in recent years, including fuzzy comprehensive evaluation method, analytic hierarchy process method, gray correlation degree analysis method and so on, which were compared for the characteristics of advantages and disadvantages. Furtherly, the main evaluation factors of the common fruits were summarized. Finally, the research on data mining methods of the fruit quality was summarized and prospected. This review indicated that data mining methods could evaluate multi-index of fruit quality comprehensively, which will provide reference for rapid detection of fruit quality and cultivation of the excellent species. Meanwhile, it will be a new direction in the field of fruit quality research by studying more main factors of fruit quality and simplifying the evaluation procedures in the near future.

1   Introduction

Fruit quality is one of the most important factors that dictates the economic value and directly affects the market competitiveness of the fruit. China firmly remains as the world's superpower in terms of fruit trees, fruit yield, and planting area [1-2].However, the proportion of fruit exports trail behind those of other countries [3].A study found that, on one hand, given the country's lagging fruit quality evaluation and sorting technology [4], grading of fruit quality is disordered, and resulting quality is uneven. Single fruit indices, such as strawberry [5] and apple [6] soluble solids, sweet orange titratable acid [7], and kiwi fruit hardness [8],   which

---

[1] **Author profile:** Jinjian Hou(1988-), male, master, Mainly engaged in the study of fruit quality. Yichang Hubei, Three Gorges University

* **Communication author:** Ligang Pan(1964-), male, researcher, doctor, Mainly engaged in agriculture Product quality safety and related technology research. Beijing, Beijing Academy of Agriculture and Forestry Sciences, 100097. Email：panlg@nercite.org.cn

were analyzed by scholars. However, single indicators can only evaluate the quality of one aspect of the fruit but fails to meet the requirements of fruit quality evaluation, because these indicators cannot be used in comprehensively evaluating fruit quality and present certain limitations. On the other hand, fruit quality evaluation contains numerous factors, including the internal and external quality factors of fruit, are involved in evaluating fruit quality, and each factor present close correlation and relative independence, resulting in difficulty in conducting fruit quality evaluation and grading work. In view of the above problems, the fruit quality data mining analysis of the indicators of hierarchical method and classification system, and relevant evaluation methods can be used to simplify quality indicators, extract the main evaluation factors, and simplify the evaluation process. The use of data mining study on comprehensive evaluation methods for fruit quality has become a research hotspot in recent years, scholars have used data mining methods on fruit quality of Nanfeng tangerine[9], apple[10], pineapple[11], pear[12] and other fruit. Results show that data mining method can be used to effectively evaluate fruit quality. At present, the literature on fruit quality data mining is rarely reported. In this paper, data mining methods applied on fruit quality in recent years were reviewed and analyzed. Finally, the main evaluation factors of common fruits were consolidated for evaluating fruit quality research and providing a reference.

2   Main fruit quality indicators and access methods

Fruit quality includes the appearance and the intrinsic qualities. The main evaluation indices include fruit shape index, fruit weight, fruit color, fruit firmness, soluble solids, vitamin C, and others, as shown in Table 1. These indicators are representative of the different aspects of fruit characteristics. Close relationships exist among these indicators, such as total sugar, including soluble solid matter represented by sucrose and other reducible carbohydrate carbonyl components, which denote different attributes that are also related. At present, methods for obtaining quality indices mainly involve chemical and instrument measurement methods, but the difficult quantitative analysis indices of fruit flavor can be obtained only by depending on expert scoring.

Table 1. Main index and access methods of fruit quality

main indexes of fruit quality	access methods
fruit shape index	fruit vertical diameter divided by transverse diameter
fruit flavor	evaluation score method
fruit weight	electronic balance measurement
fruit color	colorimeter measurement
edible rate	Peel fruit weight divided by the weight after to go nuclear
fruit hardness	fruit hardness tester
soluble solid content	refractometer determination
titratable acid content	potentiometric titration or NaOH titration method
VC content	automatic potentiometric titration or 2,6-dichlorindophenol

	titration method
soluble sugar content	anthrone colorimetry
sugar acid ratio	Soluble total sugar content divided by content of soluble acid
Solid acid ratio	soluble solids content divided by titratable acid content
carotene	Spectrophotometric method, thin layer chromatography, high performance liquid chromatography

3 Data mining overview

3.1 Simple mathematical method

For simple-featured and small amounts of data, existence of the unknown and potential information can be handled by simple mathematical processing method, such as mean, percentage, classification method. A simple data processing method can mine the data set of potential, valuable information.

3.2 Mathematical statistics method

Statistical analysis is mainly used to complete knowledge summary and relational knowledge mining. For some data, a function or relationship that cannot be expressed in a function exists. At this point, implicit data information can be excavated by using mathematical statistics method. Common methods include regression analysis, correlation analysis, and principal component analysis.

3.3 Artificial intelligence method

For large amounts and particularly complex data sets, a general data mining method cannot obtain the data set of implicit information. At this point, we can use artificial intelligence method of data mining, which is extremely complex. The main methods include fuzzy evaluation, association rules, and clustering analysis.

4 Fruit quality data mining method

4.1 Single evaluation method

(1)Fuzzy evaluation method

Fuzzy evaluation method is influenced by numerous factors so as to conduct a comprehensive evaluation of a highly effective multi-factor decision method. One characteristic of this method is that, instead of an absolutely positive or negative evaluation result, fuzzy sets are used to represent the results [13]. The advantages of

fuzzy evaluation are that we can quantify several qualitative indices, overcome the disadvantages of qualitative analysis, and objectively and accurately evaluate the pros and cons of varieties [14]. Its disadvantages include information duplication problem, which is caused by the unresolved correlation between the evaluation indexes. Thus, the confirmation of membership function and fuzzy correlation matrix, among others should be studied in the present research [15].This method is mainly used in fruit quality identification and breeding of good varieties, and is presently applied in the quality evaluation of longan [16], persimmon [17], and other fruits.

（2）Analytic hierarchy process method

The analytic hierarchy process is a multi-objective decision analysis method that combines qualitative and quantitative analysis methods [18]. The main concept of this method is to decompose the complex problem of fruit quality evaluation into several levels and factors. Comparison between two indices is essential for judgment. The judgment matrix is established by computing the largest eigenvalue in the matrix, and corresponding eigenvectors can indicate the different degrees of importance weights and provide a basis for selecting the optimal evaluation index. One advantage of the analytic hierarchy process is that not only the weight coefficient of each evaluation index is obtained but simultaneous filtering by accidental factors determines the perception of differences and the different dimension of factors in a unified evaluation system with high reliability and small error. On the other hand, one disadvantage is the limited number of fruit indicators, with the maximum generally being 9. This method has been applied in the cultivation of good varieties of jinxixiaozao [19], pear [20], and other fruits.

（3）Gray correlation degree analysis

Correlation analysis is the main tool in grey correlation analysis method using the grey system theory for the comprehensive evaluation of the research object. The correlation coefficient and correlation between the sequence of numbers and the reference sequence are compared to determine the primary and secondary factors and their correlation degree [21]. This method offers the advantages of simplicity, ease of operation, and intuitiveness. On the other hand, its disadvantages include strong subjectivity and difficulty in determining certain optimal values. This method is mainly used in the situations where in the index correlation between is too high. Grey correlation degree analysis method has performed an important function in the comprehensive evaluation on muskmelon [22], peach [23], amomum[24], and other fruits.

（4）Principal component analysis

The goal of principal component analysis is to secure the data under the principle of minimum information loss and convert the more original data and related indicators into new, fewer data at smaller orthogonal transforms to each other or comprehensive indices with slight correlation to simplify the evaluation process[25–26].Principal component analysis presents advantages of calculating the comparison standard, capability of being realized on the computer and using special software for analysis. Its disadvantage is that the new comprehensive index is difficult to explain, and the general method of combining clustering is used. Principal component analysis method is mainly used for more quality indicators, and the correlation among the indices of strong case and multiple correlated stochastic variables according to the main component of the contribution rate are simplified into several variables to avoid

traits and related traits caused by error evaluation [29].At present, the analysis is used for the comprehensive evaluation for selecting fruit quality evaluation factors and fruit quality [27-28].

4.2 Hybrid evaluation method

（1）Principal component cluster analysis method

For the multi-index evaluation of sorting fruit quality, the variance contribution of the first principal component F1 rate is not sufficiently high. In other words, the first principal component expression of original data information is not large enough, only the first principal component scores for evaluating the sample sort are one-sided. At this point, the two methods of combining principal component analysis and clustering analysis are combined to form "principal component clustering analysis method". As an advantage, the method can extract multiple indicators simultaneously with most of the information, prevent the artificial selection evaluation factor of subjectivity, and provide a true reflection of varieties of comprehensive characteristics so as to offer an objective basis for breeding materials [29].One disadvantage is clustering difficulty when the data is too large. Principal component cluster analysis method can effectively extract the main quality factors, simplify the fruit quality evaluation work, and provide theoretical basis for fruit speed measurement. The method has been used for tomato [30] and Lee apricot [31] quality rapid detection.

（2）Rationalization-satisfaction degree and multiple value method

The so-called "reasonable–satisfaction" refers to fruit varieties that demonstrate the characteristics of satisfaction that people need. The reasonable degree is1 if a characteristic species is in full compliance with the "rule". If not in line with "rule", then the reasonable degree is 0 [32]. The advantages of the algorithm are simplicity, ease of calculation, and the ability to distinguish between good quality and poor quality. Its disadvantage is larger algorithm error. The algorithm objectively and accurately reflects the people's needs and satisfaction degree of fruit quality. The method can not only be used as a method to identify fruit quality but also can be used as a reference value of fruit tree breeding species, especially for the breeding of commercial varieties. At present, the method has been used on pear [33] and other fruits for the cultivation of good varieties.

（3）Principal component cluster combined with rationalization-satisfaction multidimensional value analysis theory of merger rules

The algorithm presents new ideas and methods of comprehensively evaluating fruit quality in combination with principal component analysis, cluster analysis, and multidimensional value theory "reasonable–satisfaction" composite evaluation method. This method can be used to extract the main factors of common fruit so as to simplify the evaluation process. Moreover, the method can be used for fruit breeding. Its disadvantage is that the method computation is trivial, complex, and requires large amounts of calculation. This combined method has been applied in mango [34] fruit quality assessment factor selection and simplify the work of mango fruit quality evaluation.

4.3 Parts of comprehensive fruit evaluation factors

Table 2 summarizes the evaluation factors and the use of the method of data mining for fruit quality after the main evaluation factors of fruit in certain literature. Numerous fruit quality indicators, the presence of fruit quality evaluation using a single index presents certain limitations and evaluation of all indicators inevitably requires too much work. Fruit quality data mining methods can effectively reduce fruit evaluation indices and simplify the evaluation process. Jiyun Nie, et al [35]. used principal component analysis to select five indicators of the contribution rate of more than 95.75% of the previous four components reflecting apple quality as the main evaluation factors. Haying Zhang, et al[36], simplified 19 peach quality indicators for five items according to principal component analysis, clustering analysis, and the national standard of GB - 10653-1989 regarding "the fresh peach" requirements indicators. Table 2 shows that the use of data mining methods can effectively reduce the evaluation index, provide good evaluation of fruit quality, and solve the problems of limited single index evaluation and hefty workload of multi-index evaluation. The approach provides new ideas and methods for the evaluation of fruit quality.

Table 2. Part of the fruit of evaluation factors and main evaluation factor

fruit	evaluation factors	main evaluation factor
apple	soluble solid content、soluble sugar content、solid acid ratio and other seven quality index	fruit hardness、soluble sugar content、titratable acid content、Sugar acid ratio、VC content
pear	fruit weight、fruit shape index、fruit hardness and other eight quality index	fruit weight、soluble solid content、Sugar acid ratio、titratable acid content、fruit hardness
peach	soluble solid content、titratable acid content、Solid acid ratio and other nineteen quality index	fruit weight、fruit hardness、moisture content、solid acid ratio and flavor
mango	fruit weight、fruit hardness、soluble solid content and other nineteen quality index	Sugar acid ratio、VC content、fruit hardness、Carotene and other nine quality index

5   Conclusion and prospect

As people's living standards continue to improve, the demand for fruit quality keeps growing. The search for rapid, simple methods of evaluating fruit quality has become a hot topic in the field of fruit quality analysis. The composition of fruit quality evaluation factors is too numerous, and different degrees of correlation and relative independence exist among and between different quality factors. The use of single quality index to evaluate the quality of fruit exist certain limitations. Moreover, a single indicator can only explain the quality of fruit in a certain aspect but cannot evaluate the overall quality of fruit. Through the use of data mining methods can combine multiple quality metrics for the comprehensive evaluation of fruit quality to

obtain a comprehensive and objective assessment. In actual fruit quality assessment process, the use of relevant data mining method to determine the main fruit quality evaluation factors of common fruits can substantially reduce the workload of fruit quality appraisal.

Data mining method provides a new thinking and approach to the selection and breeding of fruit. First, data mining can identify the good traits of prominent fruit varieties, which can provide hybrid parent reference for the improvement of the fruit quality. Second, for single specific varieties, data mining methods can distinguish between fruit quality and provide a basis for directional breeding, further improvement of fine varieties, thereby yielding more excellent varieties. For the comprehensive evaluation of overall poor quality or general varieties and given the highly prominent individual quality, selection of a specific function and strong varieties not only will aid in improving the level of comprehensive utilization of fruit but is also conducive to determining the different uses and maximizing the performance of fruits on the basis of quality characteristics.

In recent years, with the development of cloud platforms, computers, massive databases, networking, and other technologies, data mining will perform a more important function in fruit production, distribution, sales, and consumption sectors. Although various studies have thorough, all kinds of fruit quality assessment method of research are more mature, but the existing methods are still hard to meet the needs of actual production and consumption. Therefore, finding a rapid and easy method for evaluating fruit quality remains a hot topic in the field of fruit quality analysis. The application of data mining method to more fruits and to extract major evaluation factors and simplify the evaluation process will become a new direction in fruit quality research.

Acknowledgment

References

1. Zhang Jiguo, Hu Jilian, Zhang Zhaoxin. Research on the fruit industry development in China[J]. Journal of Shandong Agricultural University(Social Science), 2002, 4(3):31-34.
2. Chen Minglin. Nondestructive Detection of Fruit Internal Quality Based on Portable Near Infrared Spectrometer [D]. Hangzhou: Zhejiang University, 2010.
3. Li Junliang. Research on the Quality Grading of Fruit Based Machine Vision and NIRS [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2011.
4. Li Guangmei, Wei Xinhua, Li Luxing, et al. Reserching Actuality and Development of Friut Grader [J]. Journal of Agricultural Mechanization Research, 2009(9):20-23.

5. María-Teresa Sánchez, María José De la Haba, Miriam Benítez-López, et al. Non-destructive characterization and quality control of intact strawberries based on NIR spectral data[J]. Journal of Food Engineering, 2012, 110(1):102-108.

6. Ouyang Aiguo, Xie Xiaoqiang, Liu Yande. Selection of NIR Variables for Online Detecting Soluble Solids Content of Apple[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(4):220–225.

7. Sun Qi, Xie Rangjin, Deng Lie, et al. Analysis of Fruit Quality and Acid Constituents of Three Bred Cultivars Selected from Jiangjin Sweet Orange[J]. Food Science, 2015, 36(6):124–129.

8. Liu Hui, GuoWenchuan, Yue Rong. Non-destructive Detection of Kiwi fruit Firmness Based on Near-infrared Diffused Spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(3):145–149.

9. Ni Zhihua, Zhang Sisi, Gu Qingqing, et al. Evaluation of fruit quality of Nanfeng Tangerine based on multivariate statistics[J]. Journal of Fruit Science, 2011, 28(5):918-923.

10. Dong Yueju, Zhang Yugang, Liang Meixia, et al. Selection of Main Indexes for Evaluating Apple Fruit Quality[J]. Acta Agriculturae Boreali-Sinica, 2011, 26(Supplement):74–79.

11. Lu Xinhua, Sun Dequan, Evaluation on Fruit Quality of 12 Pineapple Cultivars Introduced from Thailand[J]. Chinese Journal of Tropical Crops, 2011, 32(12):2205-2208.

12. Tian Rui, Hu Hongju, Yang Xiaoping, et al. Selection of Factors for Evaluating Pear Fruit Quality[J]. Journal of Yangtze University(Nat Sci Edit), 2009, 6(3):8–11.

13. Li Lingjuan, Shen Lingtong. An improved multilevel fuzzy comprehensive evaluation algorithm for security performance [J]. The Journal of China Universities of Posts and Telecommunications, 2006, 13(4):48-53.

14. Shen Qingqing, Zhu Jianhua, Peng Hongxiang, et al. Fuzzy Comprehensive Evaluation of Fruit Characteristics in Early Mature Seedling Litchi Resource of Southwestern Guangxi[J]. Southwest China Journal of Agricultural Sciences, 2011, 24(4):508–515.

15. Meng Lihong, Chen Yaning, Li Weihong, et al. Fuzzy comprehensive evaluation model for water resources carrying capacity in Tarim River Basin, Xinjiang, China [J]. Chinese Journal of Geographical Sciences, 2009, 19(1):89-95.

16. Zhu Jianhua, Yu Pingfu, Huang Fengzhu, et al. Quantifying analysis of main fruit characters of Longyan germsplasm in Guangxi [J]. Southwest China Journal of Agricultural Sciences, 2006, 19(2):283-286.

17. Deng LiBao, He XinHua, Xu JiongZhi, et al. Diversity analysis of fruit traits and fuzzy comprehensive evaluation of persimmon germplasm resources in Guangxi[J]. Guihaia, 2013, 33(4):508–515.

18. Thomas L.SAATY. Decision making – the analytic hierarchy and network processes (AHP/ANP) [J]. Journal of Systems Science and Systems Engineering, 2004, 13(1):1-35.

19. LiuNing, ZhangYigong, ZhaoZhihui, et al. Quality Evaluation of Jinsixiaozao Fruits and Its Relationship with Soil Body Configuration[J]. Journal of Agricultural Mechanization Research, 2009(12):125–129.

20. Zhao Jiali, Qiao Jinchun, Li Huizhuo, et al. Comprehensive Evaluation on the Internal Quality of Yali' Pear Fruit from the Main Production Area in Hebei Province[J]. Northern Horticulture, 2010(16):33–35.

21. Li Xiaoyu, Wang Wei. Estimation of apple storage quality properties with mechanical property based on grey system theory [J]. Transactions of the Chinese Society of Agricultural Engineering, 2005(2):80-86.

22. Lin Biying, Gao Shan, Lin Feng, et al. Grey relational grade analysis on yield characteristics of inbred melons[J]. Fujian Journal of Agricultural Sciences, 2008, 23(2):178-181.

23. Zhang Xuan, Sun Ya, Wang Yuning, et al. Determination of Freezing Point Temperature of Different Yellow Peach Cultivars and Correlation Analysis of Impact Factors[J]. Journal of Food Science and Technology, 2013, 31(4):37–41.

24. Zhang Wei, Wei Xiang, Long Yunhui, et al. Gray Correlation Degree Analysis on Main Agronomic Traits of Amomum tsao-ko Fruits[J]. Journal of Honghe University, 2013, 11(2):52–53.

25. Sai T.Y, L J.S, Wang S.B. Principal component analysis in construction of 3D human knee joint models using a statistical shape model method [J]. Computer Methods in Biomechanics and Biomedical Engineering, 2015, 18(7):721-729.

26. Xu Chenshan, Gao Dongsheng. Comprehensive evaluation on fruit quality of peach cultivars in green house based on principal component analysis[J]. Science and Technology of Food Industry, 2014(23):84–94.

27. Qin Hongyan, Xu Peilei, Ai Jun, et al. Diversity of Fruit Quality and Phenotypic Traits of Actinidia arguta Planch Germplasm Resources and Their Principal Component Analysis[J]. Chinese Agricultural Science Bulletin, 2015, 31(1):160-165.

28. Wang xuan, Bi Jingfeng, Liu xuan, et al. Different Origin Fuji Apple Quality Evaluation Factors Choice[J]. Journal of Nuclear Agricultural Sciences, 2013, 27(10):1501–1510.

29. Ding Chunmei. Principal component analysis of water quality monitoring data in XiaSha region [J]. Environment and Transportation Engineering, 2011:2321-2324.

30. Zhao Sidong, Yuan Deyi, Zhang Lin, et al. Study of the Screening of Sand Pear (Pyrus pyrifolia) Cultivars[J]. Journal of Central South University of Forestry & Technology (Natural Science), 2007, 27(1):30–34.

31. Han Zequn, Jiang Bo. A Study on Comprehensive Evaluation of the Processing Tomato Varieties Multiple Traits[J]. Scientia Agricultura Sinica, 2014, 47(2):357-365.

32. Niu Junli, Liu Maoxiu, Peng Qiumei. Analysis of Main Index Components and Cluster of Plumcot "Weihou" during Different Harvest Periods[J]. Xinjiang Agricultural Sciences, 2015, 52(1):33-36.

33. Ayiguli Tiemuer, Yusufu Abulitifu, Patiman Abudureheman, et al. Analysis and Evaluation on of Fruit Quality of Main Pear Varieties in Xinjiang[J]. Xinjiang Agricultural Sciences, 2014, 51(3):417 -422.

34. Xin Ming, Zhang Ezhen, He Quanguang, et al. Selection of evaluation factors for mango fruit quality[J]. Journal of Southern Agriculture, 2014, 45(10):1818-1824.

35. Nie Jiyun, Li Zhixia, Li Haifei, et al. Evaluation Indices for Apple Physicochemical Quality[J]. Scientia Agricultura Sinica, 2012, 45(14):2895–2903.

36. Zhang Haiying, HanTao, Wang Younian, et al. Selection of factors for evaluating peach(Prunus persica) fruit quality[J]. Transactions of the CSAE, 2006, 22(8):235–239.