

Classification of Essential Oil Groups with Ensemble of Binary SVM Decision Trees

Y. Nozaki¹, D. Prasetyawan¹, T. Debnath¹, T. Nakamoto¹,

¹ Tokyo Institute of Technology, 4259 Nagatsuta, Midori, Yokohama, Kanagawa, Japan, nakamoto.t.ab@m.titech.ac.jp

Abstract:

Essential oils are aromatic complex mixture from natural sources. This paper presents the method to classify and to predict the category of essential oils by using mass spectra. Ensemble of Binary Support Vector Machine Decision Tree was used for the classification model. The model showed a sufficient performance on classifying essential oils to three categories.

Key words: Classification, Essential oil, Mass spectrometry, Support vector machine, Decision tree.

Introduction

Essential oils are aromatic liquid compounds from natural sources. Typically, they are complex mixtures of hundreds of individual chemical compounds. There are more than one hundred of essential oils that are commonly used in our daily life and they are divided into several categories, such as 'Herb', 'Citrus', 'Floral', and so on [1]. It is easy to classify essential oils even if they are similar, however, it is not easy to classify groups of essential oils such as herb, citrus, and floral ones.

In this paper, we present the method to classify and to predict the category of an essential oil from its mass spectrum by using ensemble of Binary Support Vector Machine Decision Tree (SVM-BDT), which are previously proposed by Madjarov et al [2]. The mass spectrum is one of the representative physicochemical properties of chemical substances. Based on earlier studies, which clarified the relationship between the scent of a chemical and its chemical structure [3], mass spectra of essential oils can be regarded as useful property for classification of essential oils.

Data

58 mass spectra of essential oils of three categories, namely, "Herb", 'Citrus' and 'Floral', were collected by Gas chromatography -mass spectrometry (Agilent Technologies, 7890B GC System and 5977B MSD). 22 herbal essential oils, 20 citrus essential oils, and 16 floral essential oils are collected. Intensities between 50-250 [m/z] were used for the experiments.

The mass spectra were normalized by dividing by the maximum value in the dataset.

Method

As mentioned, essential oils are highly complex mixture thus their mass spectra are not easy to be separated in the subspace created by linear transformation such as Principal Component Analysis (PCA). Fig.1 shows the two-dimensional principal subspace created by PCA for the essential oil data. As shown in Fig.1, three categories can not to be separated with linear hyperplane as individual categories are widely overlapping each other. Therefore, it is necessary to use a SVM with non-linear kernel function for a better classification. In this study, Gaussian kernel was used.

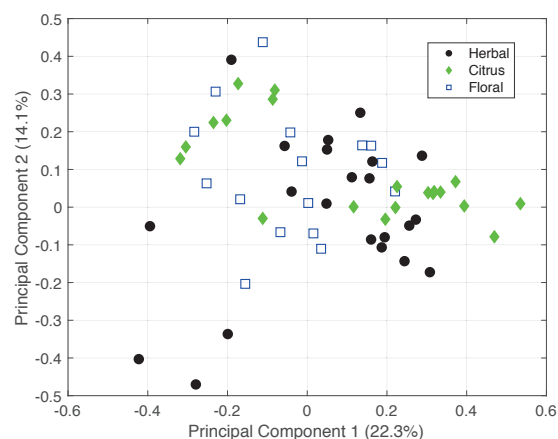


Fig. 1. Two-dimensional principal subspace for the 58-essential oils

Two SVMs form the decision tree as shown in Fig.2. Each SVM makes binary decision. The first SVM, SVM A, separates 3 original categories into 2 classes, namely 'category 1(Herb) or 3(Floral)' and '2(Citrus)'. Then the second SVM, SVM B separates samples, which are not fallen into the category of 2(Citrus).

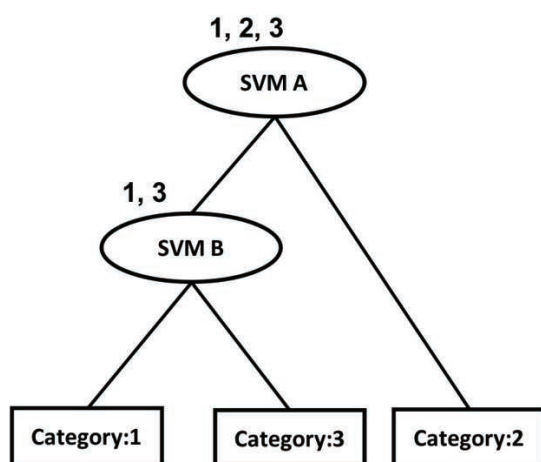


Fig. 2. SVM-BDT architecture

Result

The performance of the SVM-BDT model was evaluated by 10-fold cross validation. The result of classification on validation sets are summarized in Fig.3. As shown in Fig.3, the model was able to predict the correct category of unknown samples with accuracy of >77.6% over three categories.

	1	2	3	
1	15 25.9%	3 5.2%	1 1.7%	78.9% 21.1%
2	5 8.6%	16 27.6%	1 1.7%	72.7% 27.3%
3	2 3.4%	1 1.7%	14 24.1%	82.4% 17.6%
	68.2% 31.8%	80.0% 20.0%	87.5% 12.5%	77.6% 22.4%
	1	2	3	
	Target Class			

Fig. 3. The performance of a SVM-BDT on a dataset of essential oils of 3 categories: class 1, 2, and 3 indicate 'Herb', 'Citrus', and 'Floral', respectively

Conclusion

In this paper, we studied the method for classifying multi category of essential oils by using mass spectra. It is possible to roughly classify the groups of essential oils. The results of the experiment also indicate that the SVM-BDT model is useful even when the samples are widely overlapping in the subspace created by linear transformation. As future work, we plan to extend the empirical evaluation for a dataset of larger size and larger number of categories.

Acknowledgement

This work was partially supported by JST (Japan Science and Technology agency).

References

- [1] W. Sellar, "The Dictionary of Essential Oils", C W Daniel, ISBN: 978-0852073469, England, 1992, (Japanese translation: ISBN: 978-4-938344-32-0)
- [2] G. Wadjarov, D. Gjorgjevikj, and T. Delev, "Ensembles of Binary SVM Decision Trees", *ICT Innovations 2010*, Web Proceedings, ISSN 1857-7288, 2010
- [3] T. Nakamoto, M. Ohno, and Y. Nihei, "Odor Approximation Using Mass Spectrometry", *IEEE Sensors Journal*, 12(11), 3225-3231, 2012