MJFAS Malaysian Journal of Fundamental and Applied Sciences
ISSN:2289-5981 e-ISSN:2289-599X
UTM UNIVERSITI TEKNOLOGI MALAYSIA

Classification of perfume products using gas chromatography-mass spectrometry and pattern recognition

Tengku Fazarina bt Tengku Ab Ghani*, Hasmerya Maarof, Mohamed Noor Hassan

²Department of Chemistry, Faculty of Science, UTM, 81310 UTM Skudai, Johor, Malaysia
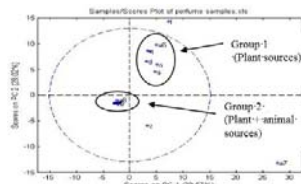*Corresponding Author: tgfazarina@gmail.com (Tengku Fazarina bt Tengku Ab Ghani)

GRAPHICAL ABSTRACT

ABSTRACT

Perfume is a mixture of fragrant essential oils and aroma compounds, fixatives, and solvents used to give human body, objects, and living spaces a pleasant odour. Perfume compositions are mixture of natural and synthetic sources which consist of 100 or more components. The difference in composition of perfume is the main characteristics which can classify the perfume into several groups such as animal and plant sources. Therefore, correct identification of each perfume product composition is crucial in order to do the classification. The perfume products have been analyzed using gas chromatography mass spectrometry and chemometric technique. There are 34 samples of perfume products used in this article which is taken from different brand available around Skudai. Chemometric technique was applied in this research to simplify data obtained by allowing them to form classes to their respective groups. Two unsupervised pattern recognition technique namely principal component analysis (PCA) and cluster analysis were used. The perfume product samples investigated in this research can be classified into two main groups namely animal and plant sources. Principal component analysis (PCA) and cluster analysis were successfully used in classifying the perfume product according to their respective groups based on the compositional information from the chromatogram of each samples.

Keywords: perfume products, gas chromatography-mass spectrometry, chemometric technique, principal component analysis, cluster analysis

1. INTRODUCTION

From last thousands of years, the Egyptians used perfume from plants, gums and resins for purposes like religious rites but nowadays consumer use perfume to feel good and making their home feel pleasant. Perfume is a mixture of fragrant essential oils and aroma compounds, fixatives, and solvents used to give human body, objects, and living spaces a pleasant odour [1]. The word perfume was derived from the Latin "perfumum", meaning through smoke. The art of making perfumes starts in ancient Mesopotamia and Egypt but was developed and more refined by the Romans and Persians. The French later gave the name "parfum" to the pleasant odour that drift through the air from burning incense. Egyptians took pride in making beautiful containers to kept their perfumes and these containers were also things of great beauty. The containers were made from materials such as alabaster, glass, ebony, and porcelain. Some perfume bottles were also made from gold and stone.

Natural perfume is also referred as an essential oil. It can be extracted from flowers, fruits, roots, leaves, woods, barks, resin and also from seeds. Aromatherapists still believe that essential oil is characteristic of essences extracted from plants. Some perfumes can be extracted from animal glands and organs. Meanwhile, synthetic type of perfume can be obtained from chemicals by synthesis process. Mitsui (1998) stated that natural perfume is in high demand while it is difficult to produce and this makes high cost of natural perfume. However, the synthetic perfume does not easily mixed together to get new fragrance chemicals because it has side effects and produce unpleasant odour. [2].

An aroma compound, also known as odorant, aroma, fragrance, or flavor, is a chemical compound that has a smell or odour. Mixtures of aromatic chemicals and essential oils are used to produce perfume and also natural aromatic oils. But now synthetic, raw material and other substances are also used in making good perfume. Substantially, different perfumes are formed from different types of sources [1,3]. Each perfume consists of hundreds of aromatic chemicals that provide own odour and aroma characteristic [3]. There are various brands in the market produced locally and internationally that are worth buying. Perfume compositions are an important part of many industries ranging from the luxury goods sectors, food services industries and to manufacturers of various household chemicals. The purpose of using perfume or fragrance compositions in these

industries is to attract customers through their sense of smell and entice them into purchasing the perfume or perfumed product. As such there is a significant interest in producing a perfume formulation that people will find aesthetically pleasing.

In perfume industry, the most commonly used analytical instruments are gas chromatography (GC) and mass spectrometry (MS) because most ingredients are polar and semi-volatile in nature. Besides that, perfumes are one of the fine fragrances which are ~~is~~ also used in creams, lotions, detergents and almost 100 ingredients or more contained fine fragrances [4, 5]. GC-MS can also reveal that some of the formula for a particular perfume is due to the presence of essential oils and other ingredients consisting of complex chemical mixtures.

The result of GC-MS analysis was analyzed by using chemometric method. In order to process large data sets like chromatogram, multivariate analysis (MVA) techniques can be applied [3]. Chemometric is a mathematical and statistical methods to improve the understanding of chemical information and to correlate quality parameters or physical properties to analytical instrument data [6]. Patterns in the data are modelled; these models then routinely applied in future to predict the same quality parameters. The result of the chemometric approach is to gain efficiency in assessing product quality. The only requirements are an appropriate instrument and software to interpret the patterns in the data.

After that, pattern recognition method was applied to illustrate the compositional and distribution data of perfume. Methodologies in pattern recognition can be divided into three types that is statistical, syntactic, and hybrid [7]. It aims to classify data (patterns) based on either a priori knowledge or on statistical information extracted from the patterns. The patterns are usually shows the group of measurements or observations, defining points in an appropriate multidimensional space. The application of principal component analysis (PCA) and cluster analysis will be used to group the perfume and classify the unknown sample to its specific groups of the family [4]. PCA is a common method that reorganizes information in a data set of samples. New principle component (PC) will be built by replacing the original variables and it will reduce the uncorrelated variables [5].

2. EXPERIMENTAL

2.1 Sample collection

A total of 34 perfume samples were used in this study. The perfume samples were purchased from nearby Skudai market, different types of sources, different brand concentration and suppliers in order to get a variety of perfume samples. The perfume samples that are used in this study were listed in Table 1. Numerous samples were used to ensure that grouping can be done effectively for varieties of perfumes in the classification using principal component and cluster analysis.

2.2 Sample preparation

About 1.5mL of perfume samples were diluted first with 3mL ethanol as a solvent. Then all the samples were kept in refrigerator before analysis procedure carried out.

2.3 Analysis of Perfume Sample

There are various type of instruments that can be used to get the data to perform analysis. Nowadays, instrument like gas chromatoraphy, mass spectrometry, nuclear molecular resonance and atomic absorption spectrometry are usually used in analysis. In this study, gas chromatography-mass spectrometry was used to get the composition of perfume product. About 1μL of the sample was taken and injected into the injector port. The detector temperature was maintained at about 310°C with ramping time of 10°C/min and the injector temperature is 310°C.

2.4 Variable Selection

The main chemometrics processing and data manipulation were performed using MATLAB software. MATLAB is an acronym for MATrix LABoratory which use as computational environment especially suited to problems in matrix. Before application of chemometric analysis, variable selection is important in order to get the sample grouping. First, retention time of all peak area of each sample was compared to animal and plant standard perfume. After comparison, the peak area of chemical compound with same retention time will be choosen as a variable. After that the chosen variable digital data of the chromatogram will convert into data matrix that is readable by MATLAB. Then all data was standardized in Microsoft Excel to make sure that all the data was in the same unit before performing Principal Component Analysis (PCA) and cluster analysis classification [6].

2.5 Chemometric Analysis

The perfume chromatogram data was ~~is~~ used for chemometric analysis. Two unsupervised method of pattern recognition used are principal component analysis and cluster analysis. Principal component analysis can be considered as "the mother of all method in multivariate analysis" and known as statistical data reduction [8]. Cluster analysis is an exploratory data analysis tool for solving classification problems [9]. The technique has ability to determine relationship between objects (samples) without using prior information about these relationships. The mean-centred and normalized data set was applied to cluster analysis [10]. Before both techniques are applied, area of selected peaks of chromatogram data will be pre-processed and normalized [6]. Principal component analysis (PCA) and cluster analysis were carried out after the complete data

obtained. It was done by dividing the sample with standard peak area. Then MATLAB 7.6 software and PLS_Toolbox 6.2 were used to accomplish the work. PLS_Toolbox is a collection of essential and advanced chemometrics routines that works within the MATLAB computational environment [6]. It contains the tools needed by chemical engineers, analytical chemist and other scientist to explore their data and build predictive models.

Table1: Perfume product samples, brands and its concentrations

BRAND	NAME OF PERFUME	TYPE OF CONCENTRATION	SAMPLE LABEL
Elizabeth	-	EdP	a
Next Simply Sparkle	-	EdT	b
Playboy	-	EdT	c
Unknown	Banana		d
Dashing	For Men	EdT	e
Urban Flower	-	EdP	f
Active Man	-	EdP	g
Johnson Baby	-	EdT	h
Adidas	Flora	EdT	i
Body Shop	Kistna	EdP	j
Avon	Far Away	EdP	k
Avon	Celelore	EdT	l
Body Shop	Vanilla	EdT	m
CK	-	EdP	n
Enchaunter	Putih	EdT	o
CK For Him		EdP	p
Body Shop	Soft Musk	EdT	q
Avon	Little White Dress	EdP	r
Slip Into	-	EdP	s
Avon	Butterfly	EdT	t
Body Shop	Cherry	EdP	u
Simply Scent	Lily	EdT	v
Simply Scent	Lavender	EdT	w
Body Shop	Aqua Lily	EdT	x
Simply Scent	Papermint	EdT	y
Simply Scent	Mango	EdT	z
Body Shop	Lychee Blossom	EdT	a1
Body Shop	White Musk	EdP	a2
Simply Scent	Jasmine	EdT	a3
Silky Girl	Spice	EdT	a4
Enchaunter	Kuning	EdT	a5
Silky Girl	Sugar	EdT	a6
Juicy De Countour	-	EdP	plant standard
Unknown	KasturiPutih	EdP	animal standard

3. RESULTS & DISCUSSION

The chromatogram peak from GC-MS indicates different type of chemical compound which exist in perfume products. All the chromatogram sample was compared with the standards based on their sources either originating from animal or plant to identify the retention time of the compounds. *Juicy De Contour* was choosen as a standard perfume for plants source group among all the samples. It is because the sample has flora and fruity scents as a main accord. From its chromatogram, 20 peaks with higher value of percentage area were choosen as variables for plant standard perfume. These peaks might be the significant compounds present in the perfume samples. After the comparison, only one sample has 9 peaks that were similar with the plant standard perfume. It shows that *banana* is derived from plant source. Next, among all the perfume samples, *kasturiputih* was choosen as a standard perfume for animal source group. It is because the sample has a deer

scent as a main accord. It also has 20 peaks with higher values of percentage area which might be the significant compound as variables because the other 16 peaks were similar to plant standard perfume peaks. This proved that *kasturiputih* originates from a mixture of two sources.

Chemometric techniques with principal component analysis and cluster analysis method were utilized to perform the classification. The classification was based on the compounds existing in the perfume product as the variables. PCA is an unsupervised method because it does not consider sample types during the analysis. Figure 1 and 2 depicts the score and loading plot for 34 samples and 24 variables respectively. It can be seen from Figure 1, the PC1 and PC2 explained 39.57% and 29.02% of the total variance respectively. The analysis showed that there are were two main groups present in this study which is from plant and animal sources. Group 1 originated from a plant source. It is because one of the samples is *banana*, labelled as *d*, was derived from chemical compound which act as variables for plant standard perfume. In group 1, there is another four samples, *a, n, q* and *a6* which can be classified as the same group perfume which derived from plant source similar to banana. On the other hand, the score plot also illustrates that the perfume sample labelled *a7* which is *Juicy De Countour* has the highest purity of chemical compounds compared to samples in group 1. Because of its high purity, *Juicy De Countour* perfume has the highest price from the other perfume samples.

Besides this, there are three samples which were classified together as group 2. Group 2 was derived from animal source. This is due to the presence of *kasturiputih* which originated from animal source. Besides *kasturiputih*, perfume sample labelled as *a3* and *a5* are also in group 2. These two samples of perfume also contain the chemical compounds as a present in *kasturiputih.* Next, the other 23 samples are classified into group 2 which comes from animal source. The label of the samples cannot clearly be seen due to overlapping of samples in the score plot. Most of the samples were classified together in this group because most of the variables used in this study influence the classification of sample in group 2. It might be that all samples have same concentration of responsible component. In this study, sample labelled *t* which is *butterfly* type perfume was not in the range of the group. This seemed to be outlier. This happened because of the slightly difference in peak which might be caused by the technical error during the injection. PCA was shown to be a good statistical method for discriminating all the samples and for establishing the relationship between chemical composition in perfume and perfume products.

Loading plots as shown in Figure 2 is plotted to identify the variables that contribute the most variations in the data. From the loading plot, it is shown that 19 variables were clustered together in one place and other 5 variables were separated from each other. Therefore it is unable to identify the variables that can contribute the most variations in the data. From the loading plot, variables number 3 which is *lily aldehyde* could influence the classification of the plant

source perfume. It can be proven by the presence of the *lily aldehyde* component in all the samples in group 1.
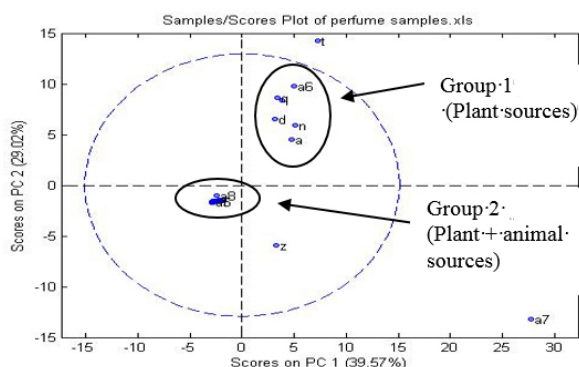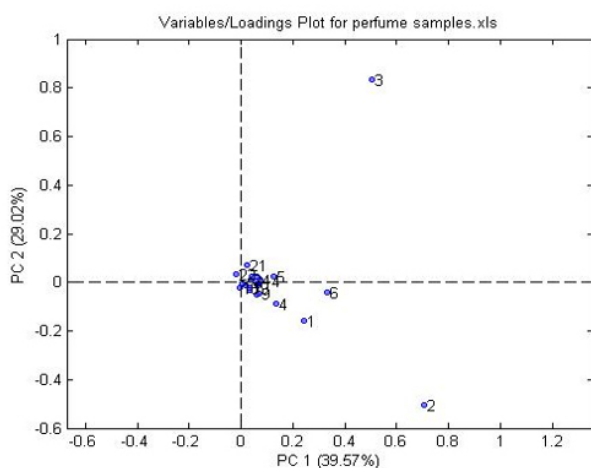

Fig. 1 Scores plot of PC2 versus PC1

Fig. 2 Loading plot of perfume based product data

The aim of this exploratory technique, cluster analysis is to search for "natural" groups among observations. Figure 3 show the dendogram thus computed which is demonstrating the similarity characteristic of 34 perfume samples that could be revealed more clearly. The similarity between 34 samples was closely related with 2 standard perfumes. Similar to the PCA results, samples could be classified into two main groups which demonstrate the differentiation of sources. The results KNN dendogram revealed that cluster 1 and cluster 2 are same as PCA. Cluster 1 show that 6 samples were clustering together which are originating from plant source. Sample labelled *t* also in clustering group 1 which might be due to nearest distance between these samples. Cluster 2 contains the majority (25 samples) of the samples which originated from plant and animal sources. KNN dendogram for cluster 2 shows all the samples compared to PCA results. In PCA result for group 2, only some samples are clear and the others are overlapping. Sample labelled *z* also consider as group 2 because their distance was near with group 2

samples. Similar to result in PCA, sample labelled *a7* considered as a separate cluster according to KNN. Therefore, it was observed that exploration techniques, as cluster analysis, are good statistical tools to detect natural group of the population studied. By considering the results of both PCA and cluster analysis methods, it was ~~we~~ concluded that all the samples are divided into 2 groups which are samples originating from plant and animal sources.
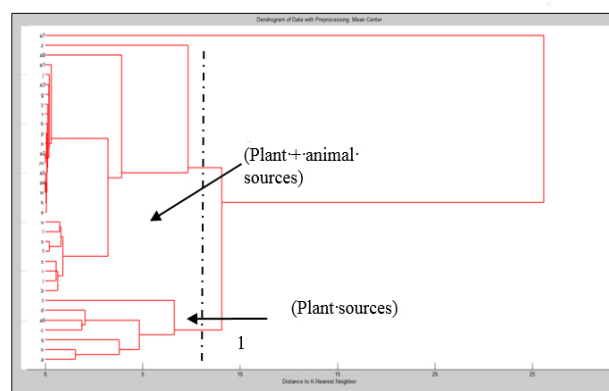

Fig. 3 KNN dendogram of 34 perfume samples

4. CONCLUSION

The present study demonstrated that the chemometric approach using exploration data techniques such as cluster analysis and PCA based on GC-MS data of chemical compounds in perfume allows to classify the perfume product based on their original sources. GC-MS is a simple and rapid method for quantifying chemical compounds in perfume. The GC-MS chromatogram give large data and more variables which have been used by the principal component analysis (PCA) and cluster analysis. The samples have their own components that serve as the variables for the analysis. GC-MS was utilized to detect the components in the samples in which the chromatogram was used to classify the perfume products. The data from the chromatogram was used to identify the composition of compounds existing in the samples.

Pattern recognition techniques have been explored for its effective use in complex data matrices. Principal component analysis (PCA) and cluster analysis were examined to construct and simplify model for perfume product. The chromatography data from the GC-MS and pattern recognition technique using principal component analysis (PCA) and cluster analysis were successfully use to classify the samples into two groups which were originated from different types of perfumes. The technique explored are faster and more reliable that visual interpretation of chromatograms providing investigators a powerful tool in chemometric analysis. Hence, pattern recognition method can be used to classify the perfume product.

ACKNOWLEDGEMENT

REFERENCES

[1]   F. Mutelet, G. Ekulu, M. and Rogalski, J. Chromatogr. A. 969 (2002) 207.

[2]   M. Gvierson, J. Barber, Hunting Anthony, L.L. (1989). Natural Ingredients in Cosmetics. Cranford, USA: Micelle Press. 12-26.

[3]   H. Justwan, B. Dahl, and G.H. Isaksen, Marine and petroleum geology.23 (2006) 213.

[4]   K. Urdal, N.B. Vogt, S.P. Sporstol, R.G. Lichtenthaler, H. Mostad, K. Kolset, S. Nordenson, and K. Esbensen, Marine Pollution Bulletin.17 (1986) 366.

[5]   Davies, A. M. C. (2005). The principle of principal component analysis.Department of Statistical Science, University College London, UK.1, p. 35-38.

[6]   Brereton, R.G. (2009). Chemometrics for Pattern Recognition. Bristol, UK: Wiley. p. 1-13.

[7]   Beebe, K.R, Pell, R.J., Seasholtz, M.B. (1998). Chemometrics:A Practical Guide. New York. John Wiley &Sons.Inc, p. 61-65.

[8]   P. Yu, J. Agric. Food Chem. 53 (2005) 7115.

[9]   Wise. B.M., Gallagher, N.B., Bro, R., Shaver, J.M., Windig, W., Koch, R.S. (2004). PLS_Toolbox 3.5 for use with MATLAB^{TM}. Eigenvector Res. Inc. p. 1-6.

[10]   Ph. Burg, J.L. Selves, and J.P. Colin, Anal. Chimica Acta. 317 (1995) 107.