Model-Based Clustering for Sensory Data and Liking

by

Moyi Li

A Thesis

Presented to

The University of Guelph

In partial fulfilment of requirements

for the degree of

Doctor of Philosophy

in

Applied Statistics

Guelph, Ontario, Canada

© Moyi Li, June, 2014
ABSTRACT

MODEL-BASED CLUSTERING FOR SENSORY DATA AND LIKING

Moyi Li
University of Guelph, 2014

Advisors:
Paul McNicholas
Ryan Browne

To gain insight into the characterization of products by clustering methods, we introduce a matrix normal mixture model using sensory data that consist of products, attributes, and records of intensity based on human senses (i.e., sight, smell, taste, etc.). Four parsimonious mixture models based on the mean of the distribution are also introduced. Moreover, we propose a matrix normal mixture model-based factor analysis to identify structures within and among groups and to facilitate visualization. Estimation of parameters in all models is conducted using the expectation-conditional maximization (ECM) algorithm. These models are applied to the analysis of data for the sensory properties of white bread.

Additionally, a new methodology to connect consumer liking scores and sensory data collected from trained panelists is introduced. We assume the conditional distribution of the liking scores is multivariate normal, with the mean being a function of the sensory data. The ECM is used to estimate the parameters of the mean function and the covariance matrix. Based on the results at the E-step (i.e., conditional estimation of the sensory matrices given by consumers’ liking scores), we
produce external preference mapping with the ideal location of the products denoted using a circular model and perform the multinomial logistic regression considering liking scores as ordinal data. Applications of these methods are illustrated on two real data sets (i.e., cupcakes and meat). Consequently, we find the ideal location on the preference map for the cupcake and also learn the mean probability for each liking score for each product across all consumers for both data sets.
This dissertation is dedicated to my mom, Fengce Guo; my dad, Shucheng Li; my husband, Pushp Awasthi; my advisors, Dr. Paul McNicholas and Dr. Ryan Browne. I would like to thank the University of Guelph Department of Mathematics and Statistics for their support in every aspect of my academic career. I am thankful for the valuable suggestions provided by Dr. Chris Findlay and John Castura during my research. I must mention and respect the sacrifice and the patience of my husband and family when I was away from them. This dissertation is wholeheartedly devoted to all of the above mentioned people and the University of Guelph.
Table of Contents

List of Figures vi

List of Tables vii

1 Introduction 1
  1.1 Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
  1.2 Model-based Approaches . . . . . . . . . . . . . . . . . . . . . . . . 2
  1.3 Preference Mapping . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
  1.4 Overview of the Thesis . . . . . . . . . . . . . . . . . . . . . . . . . 5

2 Background on Sensory Evaluation 6
  2.1 History of Sensory Evaluation . . . . . . . . . . . . . . . . . . . . . 6
  2.2 Definition of Sensory Evaluation . . . . . . . . . . . . . . . . . . . . 8
  2.3 Errors in Sensory Evaluation . . . . . . . . . . . . . . . . . . . . . . 10
  2.4 Measurement . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
  2.5 Panelists and Panel . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
  2.6 Sensory Tests . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
    2.6.1 Discrimination Tests . . . . . . . . . . . . . . . . . . . . . . . 15
    2.6.2 Descriptive Analysis . . . . . . . . . . . . . . . . . . . . . . . 17
    2.6.3 Affective Tests . . . . . . . . . . . . . . . . . . . . . . . . . . 20
  2.7 Sensometrics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
    2.7.1 Block Design . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
    2.7.2 Multivariate Analysis of Variance (MANOVA) . . . . . . . . . 23
    2.7.3 Principal Component Analysis (PCA) . . . . . . . . . . . . . . 24
    2.7.4 Generalized Procrustes Analysis (GPA) . . . . . . . . . . . . . 24
    2.7.5 Preference Mapping . . . . . . . . . . . . . . . . . . . . . . . 25
    2.7.6 Cluster Analysis and Factor Analysis . . . . . . . . . . . . . . 26
    2.7.7 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27

3 Matrix Normal Mixture Model-based Clustering on Sensory Matrices 28
  3.1 Matrix Normal Mixture Model . . . . . . . . . . . . . . . . . . . . . 28
  3.2 Parsimonious Models . . . . . . . . . . . . . . . . . . . . . . . . . . 30
    3.2.1 Model 1: $\Omega = \alpha 1'$ . . . . . . . . . . . . . . . . . . . . . 30
    3.2.2 Model 2: $\Omega = 1\beta'$ . . . . . . . . . . . . . . . . . . . . . 31
    3.2.3 Model 3: $\Omega = \alpha \beta'$ . . . . . . . . . . . . . . . . . . . . 31
3.2.4 Model 4: $\Omega = \text{SVD}'$ ........................................... 32
3.3 Factor Analysis Model ................................................... 32
3.4 Simulation Study ......................................................... 38
  3.4.1 Data ................................................................. 39
  3.4.2 Simulation Results ............................................... 39
3.5 Application ............................................................... 43
  3.5.1 Data ................................................................. 43
  3.5.2 Clustering Results ............................................... 45
  3.5.3 Visualization ..................................................... 45
  3.5.4 Hierarchical Clustering Results ............................... 49
3.6 Summary ................................................................. 54

4 A Connection Between Sensory Data and Consumer Liking Score 56
  4.1 Conditional Multivariate Normal Distribution of Liking Scores 56
  4.2 Multinomial Logistic Regression ................................... 60
  4.3 Simulation ............................................................. 61
    4.3.1 Data ............................................................. 61
    4.3.2 Simulation Results ........................................... 62
  4.4 Applications ......................................................... 67
    4.4.1 Cupcake Data .................................................. 68
    4.4.2 Meat Data ....................................................... 75
  4.5 Summary ............................................................. 78

5 Conclusions and Future Work ........................................ 80

Bibliography ................................................................. 82
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Factor plots about panelists for the simulation data set, when $G = 2$.</td>
<td>41</td>
</tr>
<tr>
<td>3.2</td>
<td>Factor plots about products for the simulation data set, when $G = 2$.</td>
<td>42</td>
</tr>
<tr>
<td>3.3</td>
<td>Factor plots for panelists, with $G = 4$.</td>
<td>47</td>
</tr>
<tr>
<td>3.4</td>
<td>Factor plots for products, with $G = 4$.</td>
<td>48</td>
</tr>
<tr>
<td>3.5</td>
<td>PCA plots for panelists, with $G = 4$.</td>
<td>52</td>
</tr>
<tr>
<td>3.6</td>
<td>PCA plots for products, with $G = 4$.</td>
<td>53</td>
</tr>
<tr>
<td>4.1</td>
<td>PCA performed on the correlation matrix for the simulation data.</td>
<td>63</td>
</tr>
<tr>
<td>4.2</td>
<td>Circular model-based preference mapping for simulation data.</td>
<td>64</td>
</tr>
<tr>
<td>4.3</td>
<td>The mean probability of each liking score based on the three groups of</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>consumers for the simulation data.</td>
<td></td>
</tr>
<tr>
<td>4.4</td>
<td>PCA performed on the correlation matrix for the cupcake data.</td>
<td>71</td>
</tr>
<tr>
<td>4.5</td>
<td>Circular model-based preference mapping for cupcake data.</td>
<td>72</td>
</tr>
<tr>
<td>4.6</td>
<td>The mean probability of each liking score based on the five groups of</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>consumers for the cupcake data.</td>
<td></td>
</tr>
<tr>
<td>4.7</td>
<td>The mean probability of each liking score based on the four groups of</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>consumers for the meat data.</td>
<td></td>
</tr>
</tbody>
</table>
List of Tables

3.1 Misclassification rate and average value of the Rand index for all models. 40
3.2 Forty-two attributes of white bread. .............................. 44
3.3 Forty-two attributes of white bread as clustered by the factor analysis model. ........................................ 46
3.4 Forty-two attributes of white bread as clustered by the Hierarchical model. ........................................ 50

4.1 Estimation of $\beta$ coefficients for the simulation data. ............ 62
4.2 The estimation of $\beta$ coefficients for the cupcake data. ............ 69
4.3 The estimation of $\beta$ coefficients for the meat data. ............ 76
Chapter 1

Introduction

1.1 Overview

This thesis is an exposition of matrix normal mixture model-based clustering on sensory data and the connections between sensory data and consumer liking scores. The purpose of studying sensory and consumer liking data is to investigate the properties of products and make it possible to achieve maximum consumer preference by creating or improving products. In sensory evaluation, the sensory data describe the intensity of product characteristics (i.e., attributes), with the result being a sensory profile of aroma, appearance, flavour, texture, and so on. The sensory data are represented as a matrix and consist of different brands of products and various sets of product attributes. Each panelist is required to record their perceptual intensity for a given product attribute and this process is usually repeated several times. The liking score data are provided by a number of consumers after they taste the given products and so only concern consumer evaluations or preference. The consumers are asked to rate the products based on a 9-point hedonic scale (Peryam and Girardot, 1952), i.e., successive integer values (1, 2, \ldots, 9) representing a scale from extremely dislike to extremely like.
The objective of this thesis is to cluster the sensory attributes based on the matrix normal mixture model. We also propose parsimonious models and extend the factor analysis model (Spearman, 1904) by adding another factor loading matrix. Moreover, to meet market demand and to find the optimal product to maximize the consumer liking, we propose a connection between liking data and sensory data via an assumption of multivariate normal distribution with a mean function.

1.2 Model-based Approaches

As far as a product is concerned, there are many external and internal properties that are of interest to consumers. This inevitably results in many similar or repeatable attributes being used. Due to the presence of trivial attributes, it is hard for researchers to precisely understand all product information. Moreover, large data sets burden the calculations and create technical problems. Therefore, grouping attributes is one of the most important tasks in the current sensory analysis.

One practical technique in statistics is clustering to group a set of physical or abstract objects into classes of similar objects. Model-based clustering, pioneered by Wolfe (1963), is a popular clustering method based on the idea that the observed data come from a population with several subpopulations. This method manages to model each of the subpopulations separately and the overall population as a mixture of these subpopulations using finite mixture models, that is

\[ f(x) = \sum_{g=1}^{G} \pi_g f_g(x), \]
where $\pi_g$ is the proportion of the population in the $g$th group and $f_g(\cdot)$ is the probability density function for the $g$th group. Maximum likelihood estimation is usually used in the model-based clustering method to find the best estimation of the parameters of the model.

Wolfe (1963) proposed the use of Gaussian mixture models for clustering analysis. Banfield and Raftery (1993), Celeux and Govaert (1995), Fraley and Raftery (1998), and Fraley and Raftery (2002) provided parsimonious mixture models using an eigenvalue decomposition of the group covariance matrices to reduce the estimation of parameters only between one and $Gp(p + 1)/2$. McNicholas and Murphy (2008) developed a family of parsimonious Gaussian mixture models using latent Gaussian models; the mixture of factor analyzers model is a special case. Also, Ghahramani and Hinton (1997) presented an expectation maximization algorithm for fitting the parameters of this mixture of factor analyzers. However, all of these methods can only be applied to multivariate data, and are not fit for diminutional matrices. Moreover, Yu et al. (2011) developed a probabilistic dimensionality reduction framework for 2D and higher-order data using a variational Bayes EM algorithm, and Xie et al. (2008), with the assumption of fixed noise in each component, provided an extension of model-based factor analysis for two-dimensional data together with an expectation-conditional maximization algorithm.
1.3 Preference Mapping

To determine the characteristics of commercial products and match them with consumer segment preferences, current product research is strongly focused on the relationship between the sensory data given by a few trained panelists and the liking scores obtained from a larger number of consumers. To learn how product attributes drive consumer liking, the sensory data must be linked with the liking data. Although these two data sets are collected from the panelists and consumers separately, we can join them because the same products are tasted by both.

Currently, external preference mapping (Schlich, 1995) is a widely used method to select a set of product prototypes that may maximize consumer liking. The researchers first use principal component analysis (Pearson, 1901) (PCA) to reduce the product dimensions based on their sensory attributes, and then use some type of polynomial model (i.e., vector, quadratic, circular, elliptical) (Carroll, 1972) to regress the liking scores and project them onto the coordinates of the products in the sensory space. The models considered are special cases of the circular surface model (Schlich and McEwan, 1992), which can be expressed as

$$L = \alpha + \sum_i \beta_i PC_i + \sum_i \delta_i PC_i^2,$$

where $L$ is overall liking, $PC_i$ are the product scores on the $i$th principal component (PC), and $\alpha$, $\beta_i$, and $\delta_i$ are regression coefficients for the $i$th PC intercept, linear, and quadratic terms, respectively.

Additionally, Faber et al. (2003) explored the possibility of including a larger number of principal components (PCs) while restricting the model choice to the simplest polynomial function. They also developed a heuristic decision rule for determin-
ing the number of PCs to keep; however, this is not a formal decision rule because the
test statistic in an analysis of variance is not distributed as Fisher’s F under the null
hypothesis. Another ideal is proposed by Tenenhaus et al. (2005), who constructed
a huge data set containing consumer liking scores, intensity of physico-chemical at-
tributes, and the intensity of sensory attributes. They used partial least squares
regression (Wold, 1966) to achieve links between the block of hedonic variables and
the physico-chemical and sensory blocks in order to cluster the consumers into homog-
eneous groups with respect to their tastes.

1.4 Overview of the Thesis

Chapter 2 gives basic background information on sensory evaluation and
some widely applicable methodologies in sensory and consumer science.

Chapter 3 proposes the clustering of sensory attributes based on the matrix
normal mixture model as well as the development of four parsimonious models. More-
over, to observe the structures of each cluster, a matrix normal mixture model-based
factor analysis is proposed with the assumption of variate errors within each group.

Chapter 4 proposes a new method to link the sensory attribute data and the
consumer liking scores by assuming the consumer liking scores have a conditional mul-
tivariate normal distribution. The mean of this conditional distribution is a function
of the sensory data. Estimating the coefficients of the mean function allows the deter-
mination of the influence of each attribute on the mean of the liking score given the
sensory data for all products while accounting for variability in the sensory attributes.
Meanwhile, the conditional estimation of the sensory data given by consumer liking scores allows for the production of a preference map that shows the ideal location for the products. In addition, by considering liking scores as ordinal, multinomial logistic regression can be used along with conditional estimation of sensory data to obtain a plot of the probability for each liking score for consumers across all products.

Chapter 5 presents some concluding remarks and suggestions for future work.
Chapter 2

Background on Sensory Evaluation

2.1 History of Sensory Evaluation

Sensory evaluation was of considerable interest in the late 1940s and on into the 1950s. Credits belonged to the US government at that time who wanted to provide more acceptable food for the military (Peryam et al., 1953) as well as the private sector. For instance, the Arthur D. Little Company introduced a Flavor Profile Method (Caul, 1957) that was a qualitative form of descriptive analysis that was, to some extent, able to minimize dependence on technical experts. The complete procedure was based on a group of about six trained people who were responsible for arriving at an agreement. Even though this approach created some disagreement among experimental psychologists about the concept of a group decision and the potential influence of an individual in the group on consensus (Jones, 1958), this method was at the time a considerable innovation in sensory evaluation, creating new interest in the discipline and stimulating more research and development in all aspects of the sensory process.

By the mid-1950s, the University of California at Davis offered a series of courses on sensory evaluation. Many other universities — for example Oregon
State University, the University of Massachusetts, and Rutgers University — also began to offer similar courses in sensory evaluation. These developments promoted a considerable interest in sensory evaluation, resulting in work by Boggs and Hansen (1949), Harper (1950), Giradot et al. (1952), Baker et al. (1954), Pangborn (1964); these achievements stimulated and facilitated the application of sensory evaluation in the industrial environment.

From the mid-1960s to 1970s, a number of factors including the energy crisis, food fabrication, the cost of raw materials (Stone, 1972), competition, and the internationalization of the marketplace directly or indirectly created opportunities for sensory evaluation. For example, the search for substitute sweeteners not only directly stimulated the development of new techniques (Inglett, 1974) to measure perceived sweetness and time-intensity, but also indirectly stimulated interest in developing and using direct data entry systems to evaluate the sweetness intensity of various ingredients (Guinard et al., 1985).

Since the 1980s, new products have been developing at an unprecedented rate and the diversity of choices within a product category have been rapidly increasing. Companies expect new approaches for anticipating and measuring the potential for a product’s success in the marketplace; however, the introduction of new products comes with high risk and therefore some companies rely solely on brand and line extensions (Lieb, 1989) to survive in the marketplace.

Generally speaking, these changes should have accelerated the acceptance of sensory evaluation; however, this has not occurred to any great extent until recently. Companies are now more aware of the importance of sensory evaluation and, therefore,
are devoting more effort to studying how sensory resources should be structured so that they can function more effectively in the future in order to achieve their goals.

2.2 Definition of Sensory Evaluation

Sensory evaluation is a widely applicable tool in the food and consumer product industries, and it is gradually becoming recognized as a technique that can be employed to accurately measure human responses to foods. These techniques can facilitate product developers’ efforts to develop new products and innovate new methods through ingredient and process modification while also considering cost reduction, quality maintenance, and product optimization.

The following definition of sensory evaluation has been accepted and endorsed by sensory evaluation committees within various professional organizations, such as the Institute of Food Technologists and the American Society for Testing and Materials:

Sensory evaluation is a scientific discipline used to evoke, measure, analyze and interpret reactions to those characteristics of food and materials as they are perceived by the senses of sight, smell, taste, touch and hearing. (Stone and Sidel, 1993)

This definition emphasizes that sensory evaluation is based on the perceptual behaviour, i.e., sight, smell, touch, hearing, etc. The first step in sensory evaluation is evocation, with guidelines to be followed with respect to the preparation and serving of product samples under controlled conditions so that biasing factors are minimized as much as possible. For instance, people are trained before a sensory test so that they are able to give their own judgments and not be impacted by the opinions of
those around them. Moreover, samples are often presented in unbranded manner, and referred to by random numbers to reduce bias that may affect sensory perceptions.

The second step in sensory evaluation is measurement. Sensory evaluation is a quantitative science that aims to establish specific relationships between product characteristics and human perception based on large amounts of data. For example, analysts are instructed to record responses that correspond to the perceived intensity of a sample.

Analysis, as the third step, is an important and indispensable part of sensory evaluation. By applying an appropriate analysis method, we are able to gain insight into the inner structure of sensory data and discover the potential relationships among products and/or between products and consumer liking. Many sensory techniques have been developed, including the triangle test, time intensity analysis, consumer preference test and so on. Moreover, statistical techniques are also important with respect to making inferences about the products being tested. The most popular statistical methods include analysis of variance, principal component analysis, experimental design, clustering.

The last step is interpretation. An acceptable interpretation of the results of the analysis should consider the method, the limitations of the experiment, the background and contextual framework of the study, and suggestions from previous research. In addition, the sensory analysts should be aware of the limitations of the test procedure and be clear on inherent risks and limitations.

Overall, sensory evaluation mainly focuses on product sensory profile characterizations and their differences, product preferences, and the relationships among
sensory variables such as sweetness, color, texture, etc. Balancing the influence of consumer response (i.e., attitudes, purchase intent) is also an important part of sensory evaluation.

2.3 Errors in Sensory Evaluation

As mentioned above, sensory evaluation is considered a science of measurement. Therefore, it should be used with precision and accuracy, avoiding relevant psychological errors and statistical errors (Meiselman, 1993). In this section, we introduce the nine most important types of psychological errors that can occur in sensory evaluation.

1. Central tendency error.

   This error occurs when people are not familiar with a test method or with products, so they avoid giving extreme values and conservatively rate the products in the mid-range of the scale.

2. Time-order error.

   This error is regarded as order effects, first-sample effects, or position effects. For example, the product served first is usually evaluated higher than it would be if presented in any another position.

3. Expectation error.

   This error occurs when people have too much knowledge about the product, and non-sensory responses in specific attributes or differences based on that
knowledge.

4. Errors of habituation and anticipation.

These two errors result from a tendency to give the same response when a series of slowly increasing or decreasing stimuli are presented. Panelists tend to repeat the same scores and therefore neglect the change in perception.

5. Stimulus error.

This error is similar to expectation error; that is, panelists evaluate the product based on prior knowledge of the physical stimulus rather than their perception of the stimulus.

6. Logical and leniency errors.

These two errors occur due to a lack of specific test instructions. As a result, panelists will give scores via a self-determined process or their feelings about the experimenter may influence how products are rated. These errors can be overcome by using specific test instructions and well-trained panelists.

7. Halo effect.

This error (Guilford, 1954) occurs when more than one attribute of a product is evaluated and the ratings tend to influence each other. In a test regarding orange juice, for example, panelists may be asked to rate not only their overall liking but also specific attributes. As a result, a well-liked product will tend to have its attributes (e.g., sweetness, acidity, flavour strength, mouthfeel) highly rated as well. Conversely, if the product is not well liked, then the attributes
will be rated unfavourably.

8. Proximity error.

In this error, adjacent characteristics tend to be rated more similarly than those that are further apart. Thus, the correlations between adjacent pairs may be higher than if they are separated by other attributes.

9. Contrast and convergence errors.

Contrast and convergence errors are described in detail by Kamenetzky (1959). The contrast error arises when a product of ‘poorer’ quality is followed by a product of ‘higher’ quality, resulting in the higher quality product being scored much higher than if it were preceded by a product of closer quality. Convergence error is the opposite effect, in that it masks small differences between two or more products.

Additionally, we still need to attach great importance to statistical errors as they can be a source of potential risk in the decision-making process. Statistical errors include Type I and Type II errors. For example, after performing a paired t-test to compare products A and B in sensory analysis, we usually conclude that there is a difference between the products when \( p \leq 0.05 \), which means than there is only 5\% probability that the result was due to chance. In this case, \( \alpha \) is the probability of a Type I error; this means we reject the null hypothesis, i.e., product A is not the same as product B, when it was in fact the same. In the practice, the \( \alpha \) value can be changed according to the business risk.
On the other hand, if we conclude that there is no difference between products A and B when in fact there is a difference, a Type II error has been committed and the probability of this error is $\beta$. There is a reciprocal relationship between $\alpha$ and $\beta$; however, this relationship can be controlled by the size of the group of people, the magnitude of the difference between the products, and the sensitivity and reliability of the panelists.

2.4 Measurement

Precise and accurate measurement is a prerequisite of performing sensory analysis. Based on the test objective and product characteristics for a particular test, choosing an appropriate scale is one of the most important tasks before a test is organized. Generally speaking, an ideal scale should be meaningful, uncomplicated, unbiased, and sensitive to differences. Stevens (1951) proposed a classification system for scales, as listed below.

1. Nominal scales: used in classification or naming. An example is a check-all-that-apply question, in which a list of sensory attributes is given and the trained panelist is asked to indicate which of the attributes are present in the sample, or a consumer is asked to indicate which of the words/terms apply to the sample.

2. Ordinal scales: used in ordering or ranking with numbers or words. Examples include ‘high’ to ‘low’ and ‘most’ to ‘least’, with respect to some attribute of a product set.
3. Interval scales: used in measuring magnitudes and assuming equal distances between points on the scale. For example, line scales (Lawless, 1989; Lawless and Malone, 1986a,b) allow individual marks on a horizontal line to indicate the intensity of the particular characteristic. The two anchors of the horizontal line reflect a continuum from weak to strong intensity.

4. Ratio scales: Ratio scales data have the same properties as interval scale data but there is an absolute zero and a constant ratio between points on the scale. Moreover, there are some scales with special features that are used for consumer sensory evaluation of products.

1. Hedonic scale: A nine-point hedonic scale (Jones et al., 1955; Peryam and Pilgrim, 1957) is used to measure consumer liking. The consumers are asked to indicate the term that best represents their attitude about the product and their responses are converted to numerical values for computational purposes from ‘like extremely’, 9 to ‘dislike extremely’, 1.

2. Just-about-right scale (JAR): JAR is used to assess the appropriateness of specific sensory levels, and assist product developers with identifying potential shortcomings of the products tested. The scales commonly used have five categories. As an example, the sweetness of a sample can be rated on a JAR scale (Vickers, 1988) comprised of ‘Not nearly sweet enough’, ‘Not sweet enough’, ‘Just about right’, ‘Too sweet’, ‘Much too sweet’.
2.5 Panelists and Panel

In the field of sensory evaluation, panelists are defined as trained people who are able to describe their sensory experiences using words that are more detailed than those used by consumers. All of these words are generated in previous training sessions. A descriptive sensory panel consists of trained panelists using consensus terminology to profile products. The descriptive analysis data can be used to precisely describe the relationships between two or more products and track product changes over time. Moreover, as far as product development is concerned, data from the trained sensory panel can be used to identify opportunities for new products that fill gaps in existing market maps, determine if it is possible for consumers to notice changes, and understand the magnitude of changes that will get a particular consumer reaction, and so on.

2.6 Sensory Tests

Currently, the most important concern for the sensory researcher is ensuring the analysis method that they use is appropriate and that statistical analysis of results can answer the questions of interest. Methods for sensory evaluation are classified into three main types according to the primary purpose: discrimination tests, descriptive analysis, and affective tests.
2.6.1 Discrimination Tests

Discrimination tests are designed to determine whether a difference exists between two products. Based on the statistical frequencies and proportions, we are able to select the product that is different from a set of similar or control products by choosing appropriate values of the test parameters: $\alpha$ and $\beta$.

This method can help product developers to exploit cheaper ingredients as a substitute for expensive ingredients in the original product. The test involves giving consumers two samples that are chemically different in formulation with the expectation that they will not be able to detect a difference. When testing for similarity, researchers should choose a small value for $\beta$ to ensure that there is only a small chance of missing any difference that exists. Conversely, if the developers want to know whether the new formulation can improve the products, i.e., they are testing for a difference in the formulations, then a small value of $\alpha$ is selected.

In general, the most frequently used discrimination tests are the paired comparison test, duo-trio test, and triangle test.

1. Paired comparison and difference paired comparison test.

A paired comparison test is used to select the product that has single specified designated characteristic from two possible serving sequences (AB, BA). This order is required to be randomized across panelists, with an equal number of panelists receiving either sample A or sample B first. In the statistical analysis, a one-tailed binomial test can be used because the sample that has the higher intensity in the specific characteristic is known in advance.
If the developers want to know whether the two products are different without specifying the sensory attributes, then a difference paired comparison test is recommended. This test provides four possible serving sequences (AA, BB, AB, BA), which should be randomized across panelists, with each sequence appearing an equal number of times. The results can only show that the samples are perceptibly different, but do not indicate which attribute is different.

2. Triangle test.

The triangle test developed by Helm and Trolle (1946) is a three product test. In this test, the three products are presented to panelists in one of six randomized service orders (AAB, ABA, BAA, BBA, BAB, ABB); panelists are expected to determine which one is most different from the other two. The null hypothesis states that the long-run probability of making a correct selection when there is no difference between the products is $\frac{1}{3}$. The results after using a binomial test will indicate whether there is a difference among the products but not which specific product attribute differed.

3. Duo-trio test.

The duo-trio test (Peryam and Swartz, 1950) can be considered an alternative triangle test, allowing the panelists to choose which of two coded products is most similar to a third product that is identified as reference. This test is usually used for products that have a relatively intense taste or odor, which may impact sensitivity. In the statistical analysis, the null hypothesis states that the long-run probability of the population making a correct selection when there is
no perceptible difference between the products is $\frac{1}{2}$. A one-tailed binomial test can often be used, such as in a cases where there is an expectation that one of the products has a higher intensity in a specific characteristic based on its formulation. The conclusion is the same as the paired difference test; that is, it can only determine if the products are perceptibly different, but not which specific product attribute differed.

### 2.6.2 Descriptive Analysis

Descriptive analysis is considered the most sophisticated sensory analysis methodology because qualitative sensory aspects of a product (i.e., appearance, aroma, flavour, texture and sound, etc.) as well as quantitative data are given by well-trained panelists. The trained panelists are required to use specific vocabulary to describe and indicate the intensity of perceived sensory attributes and also be able to differentiate them.

Descriptive analysis can be used for a specific sensory attribute or a comparison among several products. In the marketplace, this method can be used to indicate how a competitor’s product is different in the sensory attributes. Moreover, it can also be used in product development to measure how close a prototype is to other products.

A commonly used technique is generic descriptive analysis (Lawless and Heymann, 1998), which includes three steps: training the panelists, determining the panelists’ reproducibility, and having the panelists evaluate the products.

In the training session, usually done in 1 hour, two methods can be used.
One is called ‘consensus training’, which asks the panelists to generate descriptors for a wide range of products in a specific category; the panelists themselves need to move toward initial consensus after see the total list of descriptors elicited. Another method is ‘ballot training’. The panelists are provided with not only a wide range of products within the category, but also a word list of possible descriptors and references that could be used to describe the products. These two methods can be used independently or dependently according to the research needs. When they are combined, panelists firstly derive some descriptors on their own through consensus, and then others are added through suggestions by the panel leader or from word lists. The panel leader can reduce redundant terms.

Subsequently, the evaluation of panelist reproducibility begins. A subset of samples is provided to the panelists in triplicate, and the sensory researchers study the significance levels of the interaction effects associated with panelists using the data collected from three sessions. The panelists are considered unreproducible if there are significant effects among the panelists; in this case, they will receive training sessions again until the data they generate is reproducible.

In the third step, panelists are asked to evaluate samples. During this phase, all samples are coded in advance and under ideal conditions, will be served randomly in a single session, with different sessions as the replicates. The data can be analyzed by analysis of variance (ANOVA). As supplementary methods, some appropriate multivariate statistical techniques can give additional useful information. If the conditions do not allow for distribution of all samples in a single session, then an appropriate experimental design, for example a Latin square or balanced incomplete block design.
(Cochran and Cox, 1957; Petersen, 1985), can be used to present the products to the panelists.

Additionally, the trademarked methods are actively used by many companies. For instance, Flavour Profile® that is trademarked to Arthur D. Little and Co., Cambridge, Massachusetts, Texture Profile® that was created by scientists working for General Foods, Sensory Spectrum® that was created by Gail Civille when she worked at General Foods in the 1970s, and Quantitative Descriptive Analysis® that was developed to correct some of the perceived problems with the Flavor Profile analysis (Stone et al., 1974; Stone and Sidel, 1993).

Another consideration in descriptive analysis is focusing on single attributes. For example, if we are only interested in the sweetness of a chewing gum, we will track the intensity of the sweetness over a period time, and learn how the sweetness changes from the onset of the sweetness release until the flavour disappears or experimental time finishes. This method is called Time-Intensity (TI) analysis. Because the intensity of perception varies within a time period, Larson-Powers and Pangborn (1978) and Overboshc (1986) proposed defining a characteristic of a product using a time-intensity curve of that attribute. A typical TI curve graphically displays attribute intensity up to peak intensity \(I_{max}\) and until its disappearance. The onset of perception of an attribute, the length of time \(T_{max}\) required to reach maximum intensity \(I_{max}\), and the duration of perception can be obtained from the TI curve. The increase and the decrease angles, which give an indication of the rate of onset and decline of the attribute, have been studied by Bloom et al. (1955). Before collecting TI data, some protocols should be
clearly defined, for example, the type of delivery, amount of product, time to hold in the mouth, type of manipulation, expectoration, swallowing, etc. Additionally, panelists must be well trained and learn all of the protocols necessary for a well-controlled TI study.

2.6.3 Affective Tests

Affective analysis is used to assess consumer response with respect to current or potential customer product or specific product characteristics. Consumer preference can be directly measured by comparison of two or more products. We can also use scaling methods to measure the degree of acceptance and then compute the preferences. There are two type of affective methods: qualitative and quantitative. Both are used to gather affective responses from consumers. These techniques can stimulate product developers to improve and optimize products so that they can deliver what the consumer is looking for and thus occupy a place in the market.

A qualitative affective test measures the selected consumer response to product characteristics as noted by consumers in an interview, diary, submitted photo, collages, etc. This method is able to assess consumers’ initial response to a product concept and a product prototype.

A quantitative affective test measures overall preference for a product or potential products, product characteristics, and consumer response to specify sensory attributes of a product. There are two main types of tests, selected according to the primary task: the preference test and the acceptance test.

The preference test concerns which sample is preferred or liked better. It
does not indicate whether any of the products are liked or disliked. There are four types of preference tests (Meilgaard et al., 1974) for answering these questions.

1. Paired preference: considering two products, the panelist chooses one sample over another. The Results are often analyzed using a binomial test.

2. Rank preference: considering three or more samples, give the relative order of preference of samples. The Results are often analyzed using a Friedman test (Friedman, 1955).

3. Multiple all paired preference: considering three or more samples, the preference is made by a series of paired samples with all samples paired with all others. We can consider a block design with ANOVA in this case.

4. Multiple selected paired preference: considering three or more samples, the preference is made by a series of paired samples with one or two selected samples paired with two or more others. We can consider a block design with ANOVA in this case.

The acceptance test aims to measure consumer liking or preference, so the nine-point hedonic scale is often used to indicate degrees of dislike to like. From the relative liking score, the product developers can infer whether their products are popular in the market. Usually, an ANOVA is performed or a random complete block design is used with this type of data.
2.7 Sensometrics

2.7.1 Block Design

Block designs can be used to statistically analyze the data from sensory tests. For example, sensory discrimination tests basically ask whether panelists can detect a difference between products, either overall or with respect to some attribute; therefore, ranking and rating within a random complete block design (RCBD) can be used for discrimination. However, ratings within a RCBD are used more often for descriptive purposes. For the affective test, an RCBD, CBD, or balanced incomplete block design (BIBD) can be applied depending on the purpose of the experiment. Approaches using either the RCBD or BIBD (Meilgaard et al., 1974) are given below.

1. RCBD

- Ranking test: a set of samples is presented to each panelist in a balanced and random order. Panelists are asked to rank the samples based on the attribute of interest.

- Rating test: this test is based on a total of \( k \) samples, where \( k \) is a number ranging from 3 to 6 or at most 8 so it is possible to compare all samples as one large set. The panelists are asked to rate the intensity of the selected attribute on a numerical intensity scale, and the results can be evaluated by ANOVA.

2. BIBD

- Ranking test: this test is based on a total of \( k \) samples, where \( k \) is a number
ranging from 6 to 12 or at most 16. All samples are presented in a number of smaller blocks instead of one large set, according to one of the designs of Cochran and Cox (1957). Panelists are asked to rank the samples for the specified attribute. A Friedman test can be used to analyze the rank data.

- Rating test: this test is used when the number of samples $k$ ranges from 6 to 12 or at most 16. It is similar to the BIBD ranking test analyzed using ANOVA, but the panelists are instead asked to rate the intensity of the attribute of interest on a numerical scale.

### 2.7.2 Multivariate Analysis of Variance (MANOVA)

The multivariate analysis of variance (MANOVA) is usually used to determine whether significant differences exist between two or more treatments when compared with respect to all dependent variables of interest (Stevens, 1986). In descriptive analysis, each panelist might evaluate several attributes. In this case, using MANOVA is preferable to ANOVA because MANOVA is able to analyze all attributes simultaneously. Moreover, MANOVA can measure homogeneity for the panel with the product × panel interaction term and discriminate the panel using a model of products and panelists and the product × panelist interaction term.

### 2.7.3 Principal Component Analysis (PCA)

Principal component analysis is widely used to simplify and/or describe interrelationships among multiple attributes and among products (Ghosh and Chat-
topadhyay, 2012; Chapman et al., 2001; Liu et al., 2004; Wang et al., 2011), and this method is often applied to descriptive data. For example, when panelists evaluate two attributes of a product, it is very possible that the attributes are redundant and both measure the same underlying characteristic (Heymann and Noble, 1989). Using PCA, we can eliminate these redundancies by extracting a few uncorrelated principal components (PC) that have largest possible variance. As a result, the PCA plot can graphically illustrate the relationships among attributes in the product sensory space. Moreover, the PCA plot can also help us to compare the sensory spaces described by different panels.

2.7.4 Generalized Procrustes Analysis (GPA)

In contrast to PCA, generalized procrustes analysis (Gower, 1975) uses a consensus configuration based on individual responses rather than panel mean values. Based on the evaluations given by different panelists’ terms, a mathematical consensus can be built using multidimensional rotation, translation and scaling. The analysis proceeds through an iterative process that minimizes the value of the Procrustes statistics, that is, the residual distance between the individual configurations and the consensus configuration after completion of the Procrustes analysis. As a result, GPA can provide plots that give information about attribute vectors and panelists (Meullenet et al., 2007).

This method has been used with sensory profiling data, particularly free choice profiling (Williams and Langron, 1984) that allows panelists to generate their own individual attribute list for a product set.
2.7.5 Preference Mapping

Preference mapping (Carroll, 1972) was designed to develop a deeper understanding of consumer acceptance of products and is able to assist product developers in selecting a single product that may maximize consumer liking. Moreover, this technique can represent and preserve the response for each consumer to identify the consumers segments that tend to like the same types of products or have similar expectations for the sensory characteristics of a product.

Internal preference mapping (MDPREF) (Chang and Carroll, 1968) is used to construct a multidimensional representation of products and consumers. Based on the liking scores given by consumers for each product, a preference space is produced by PCA performed on centred overall liking data. The direction of each vector in MDPREF represents the direction of increasing liking and the length of the vector is directly proportional to the amount of variance explained by the first two preference dimensions. The consumer segments can also be assessed by MDPREF.

External Preference Mapping (PREFMAP) concerns the multidimensional representation of products based on their sensory profile or a set of other external data, such as instrumental measures of colour, texture, flavour, etc. Again using PCA, the data matrix with products as rows and attributes as columns is decomposed and product scores on PCs are obtained. Because this method is limited within the sensory space, we therefore need to fit the consumer data in this space with some type of polynomial model (i.e., vector, circular, elliptical, quadratic) that is designed to regress the liking scores given to the product onto the coordinates of the product.
scores in the sensory space.

2.7.6 Cluster Analysis and Factor Analysis

Because consumer segmentation can help product researchers model differences in consumers (Moskowitz and Bernstein, 2000) and identify how sensory attributes drive liking, it is therefore necessary to use appropriated statistical methods to segment them into homogeneous groups. Moreover, liking for homogeneous subgroups of consumers is also interested.

The simplest method is hierarchal cluster analysis, which is based on the similarity or distance matrix. Finite mixture model-based cluster analysis is a more advanced way to group the consumers. An expectation-maximization (EM) algorithm (Dempster et al., 1977) is used for parameter estimation. EM algorithms are based on the complete-data likelihood, i.e., the likelihood of the observed and unobserved data, and consist of two steps, the E-step and the M-step, which are iterated until some stopping criterion is met. As another option, mixtures of factor analyzers can also achieve similar results by considering the common factors as additional latent variables in the model.

2.7.7 Summary

There are still many other statistical methods that can be applied to analyze sensory data and consumer liking data, and therefore sensory researchers need to clarify their goals and choose appropriate methods to get ideal results. Overall, a good sensory evaluation is able to provide information that is useful in management
business decisions about directions for product development, product changes and product quality, and this can help companies effectively avoid uncertainty and reduce competitive risk.
Chapter 3

Matrix Normal Mixture Model-based Clustering on Sensory Matrices

3.1 Matrix Normal Mixture Model

The matrix normal distribution (Pierre, 1999) is one of the most important matrix valued distribution. Let \( X \) be a \( p_1 \times p_2 \) random matrix, \( \Omega \) be a \( p_1 \times p_2 \) matrix called mean, \( \Sigma_1 \) be a \( p_1 \times p_1 \) matrix called the among-row covariance matrix, and \( \Sigma_2 \) be a \( p_2 \times p_2 \) matrix called the among-column covariance matrix. The covariance matrices \( \Sigma_1 \) and \( \Sigma_2 \) are required to be positive definite. A random matrix \( X \) has a matrix normal distribution with parameters \( \Omega, \Sigma_1, \Sigma_2 \), i.e., \( X \sim N_{p_1 p_2}(\Omega, \Sigma_1, \Sigma_2) \), if vec(\( X \)) \( \sim N_{p_1 p_2}(\text{vec}(\Omega), \Sigma_2 \otimes \Sigma_1) \). This definition involves the Kronecker product \( \otimes \) and the vec operator. The matrix normal probability density function is

\[
f(X) = \frac{1}{(2\pi)^{p_1 p_2} |\Sigma_1|^{p_2/2} |\Sigma_2|^{p_1/2}} \exp \left\{ -\frac{1}{2} \text{tr}[\Sigma_1^{-1}(X - \Omega)\Sigma_2^{-1}(X - \Omega)'] \right\}.
\]

which was introduced by Arnold (1981). A finite mixture model with matrix normal densities is given by

\[
g(X|\pi_g, \Omega_g, \Sigma_{1g}, \Sigma_{2g}) = \sum_{g=1}^{G} \pi_g f(X|\Omega_g, \Sigma_{1g}, \Sigma_{2g}),
\]
where \( g \) represents group. To apply the ECM algorithm (Meng and Rubin, 1993), we introduce latent variable \( Z_{ig} \), which is defined below as

\[
Z_{ig} = \begin{cases} 
1 & \text{if } x_i \text{ belongs to group } g \\
0 & \text{otherwise.}
\end{cases}
\]

Incorporating the latent variable, the complete-data log-likelihood function is

\[
l_c(X, Z_{ig} | \pi_g, \Omega_g, \Sigma_{1g}, \Sigma_{2g}) = \sum_{i=1}^{n} \sum_{g=1}^{G} Z_{ig} \log[\pi_g f(X|\Omega_g, \Sigma_{1g}, \Sigma_{2g})]. \quad (3.1)
\]

At the E-step, we obtain the expected complete-data log-likelihood by taking the expected value of the latent variable. This is given by

\[
\hat{z}_{ig} := E(Z_{ig} | X) = \frac{\pi_g f(X|\Omega_g, \Sigma_{1g}, \Sigma_{2g})}{\sum_{g=1}^{G} \pi_g f(X|\Omega_g, \Sigma_{1g}, \Sigma_{2g})}. \quad (3.2)
\]

At the M-step, we maximize the expected value of the complete-data log-likelihood function by taking the partial derivative with respect to \( \pi_g, \Omega_g, \Sigma_{1g}, \text{ and } \Sigma_{2g} \), separately. Then we set these derivatives equal to zero and solve them. As a result, we obtain the updates

\[
\hat{\pi}_g = \frac{n_g}{n}, \quad (3.3)
\]

\[
\hat{\Omega}_g = \frac{\sum_{i=1}^{n} \hat{z}_{ig} X_i}{n_g}, \quad (3.4)
\]

where \( n_g = \sum_{i=1}^{n} \hat{z}_{ig} \). The conditional updates for \( \Sigma_1 \) and \( \Sigma_2 \) are

\[
\hat{\Sigma}_{1g} = \frac{\sum_{i=1}^{n} \hat{z}_{ig} (X_i - \Omega_g) \Sigma_{2g}^{-1} (X_i - \Omega_g)'}{p_2 n_g}, \quad (3.5)
\]
and

\[ \hat{\Sigma}_{2g} = \frac{\sum_{i=1}^{n} \hat{z}_{ig}(X_i - \Omega_g)'\Sigma_{1g}^{-1}(X_i - \Omega_g)}{p_1n_g}. \] (3.6)

We iterate the E-step and M-step until the complete-data log-likelihood values converge.

### 3.2 Parsimonious Models

To obtain a concise and efficient model, we introduce four parsimonious mixture matrix normal models based on the \( \Omega \) matrix by reducing the estimation of the parameters.

#### 3.2.1 Model 1: \( \hat{\Omega} = \alpha 1' \)

In this case, \( \alpha \) is a \( p_1 \times 1 \) vector with different values and 1 is a \( p_2 \times 1 \) vector. As a result, \( \Omega \) will be reduced to a matrix with the same value within each row but varying values among rows. Based on this model, the number of estimators will be reduced to \( p_1 \). The expected value of the complete-data log-likelihood function is roughly the same as equation (3.1); accordingly, the \( \hat{z}_{ig} \) and \( \hat{\pi}_g \) refer to equations (3.2) and (3.3), respectively. These results are also used in Model 2 and Model 3.

In the M-step, the estimate of \( \alpha_g \) is

\[ \hat{\alpha}_g = \frac{\sum_{i=1}^{n} \hat{z}_{ig}X_i1}{n_g}, \]

and the \( \hat{\Sigma}_1 \) and \( \hat{\Sigma}_2 \) are obtained via (3.5) and (3.6) using \( \alpha 1' \) instead of \( \Omega \).
3.2.2 Model 2: $\hat{\Omega} = \beta'$

Because $\beta$ is a $p_2 \times 1$ vector containing $p$ different values and $1$ is a $p_1 \times 1$ vector, the $\Omega$ matrix has the identical values within each column but not among the columns. In this case, we only estimate $p_2$ parameters instead of $p_1 \times p_2$ parameters of the original $\Omega$ matrix.

Maximizing the expected value of the complete-data log-likelihood equation (3.1) with respect to $\beta_g$, $\Sigma_1$, and $\Sigma_2$ yields

$$\hat{\beta}_g = \sum_{i=1}^{n} \hat{z}_{ig} X_i' 1_{n_g},$$

and $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are obtained via (3.5) and (3.6) using $1\beta'$ instead of $\Omega$.

3.2.3 Model 3: $\hat{\Omega} = \alpha \beta'$

Now we set $\alpha$ to be a $p_1 \times 1$ vector and $\beta$ to be a $p_2 \times 1$ vector, so we can only estimate $p_1 + p_2$ parameters to represent $\Omega$. The algorithm is similar to the previous two models, and thus the maximum likelihood estimators of $\alpha_g$ and $\beta_g$ are obtained by

$$\hat{\alpha}_g = \frac{\sum_{i=1}^{n} \hat{z}_{ig} X_i \beta'_g}{n_g},$$

and

$$\hat{\beta}_g = \frac{\sum_{i=1}^{n} \hat{z}_{ig} X_i' \alpha_g}{n_g},$$

and $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are obtained via (3.5) and (3.6) using $\alpha \beta'$ instead of $\Omega$. 
3.2.4 Model 4: $\Omega = \text{SVD}'$

In this case, we use singular value decomposition to decompose the $\Omega$ matrix that is obtained from the full model. Therefore, the estimation of $\Omega$ is updated by principal eigenvectors of the rows and columns.

3.3 Factor Analysis Model

Factor analysis (Spearman, 1904) is a data reduction technique to find unobserved factors that explain the variability in the data. Because the attribute observations are composed of panelists and products as $p_1 \times p_2$ matrices, according to the model (Bartholomew and Knott, 1999), we must not only assume a $p_1$-dimensional random vector $X$ is modelled using an $m_1$-dimensional vector of latent factors, where $m_1 < p_1$, but also that a $p_2$-dimensional random vector $X$ is modelled using an $m_2$-dimensional vector of latent factors, where $m_2 < p_2$.

Therefore, the model is given by $X = \Lambda_1 U \Lambda_2' + E$, where $\Lambda_1$ is a $p_1 \times m_1$ matrix of factor loadings, $\Lambda_2$ is a $p_2 \times m_2$ matrix of factor loadings, the latent matrix $U \sim N_{m_1 m_2}(\text{vec}(\Delta), \Xi_2 \otimes \Xi_1)$ and $E \sim N_{m_1 m_2}(0, \Psi_2 \otimes \Psi_1)$, where $\Psi_1 = \text{diag}(\psi_{11}, \psi_{12}, \ldots, \psi_{1r})$, and $\Psi_2 = \text{diag}(\psi_{21}, \psi_{22}, \ldots, \psi_{2s})$. The distribution of $\text{vec}(X)$ follows a multivariate normal distribution defined by $\varphi(.)$, with mean $\text{vec}(\Lambda_1 \Delta \Lambda_2')$ and covariance matrix $\Lambda_2 \Xi_2 \Lambda_2' \otimes \Lambda_1 \Xi_1 \Lambda_1' + \Psi_2 \otimes \Psi_1$. 
The complete-data log-likelihood function is given by

\[
l(X, U) = \sum_{g=1}^{G} \sum_{i=1}^{n} z_{ig} [\log \pi_g + \log f(X_i | A_1 U_{ig} A_2', \Psi_{1ig}, \Psi_{2ig}) + \log f(U_{ig} | \Delta_g, \Xi_{1g}, \Xi_{2g})],
\]

(3.7)

where \( g \) represents groups and \( i \) represents observations. Hence, the expected complete-data log-likelihood function at the \( k \)th step is

\[
Q(\vartheta) = \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} (\log \pi_g + W_{1ig} + W_{2ig})
\]

\[
= \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} \log \pi + \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} W_{1ig} + \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} W_{2ig} \quad (3.8)
\]

where

\[
\hat{z}_{ig} = E[z_{ig} | X] = \frac{\pi_g \varphi[\text{vec}(X_i); \text{vec}(A_1 \Delta_g A_2'), \Lambda_2 \Xi_{2g} \Lambda_2' \otimes A_1 \Xi_{1g} A_1' + \Psi_{2ig} \otimes \Psi_{1ig}]}{\sum_{g=1}^{G} \pi_g \varphi[\text{vec}(X_i); \text{vec}(A_1 \Delta_g A_2'), \Lambda_2 \Xi_{2g} \Lambda_2' \otimes A_1 \Xi_{1g} A_1' + \Psi_{2ig} \otimes \Psi_{1ig}]};
\]

\[
W_{1ig} = E_{\vartheta_k} [\log f(x_i | A_1 U_{ig} A_2', \Psi_{1ig}, \Psi_{2ig}) | x_i, Z_{ig} = 1],
\]

\[
W_{2ig} = E_{\vartheta_k} [\log f(U_{ig} | \Delta_g, \Xi_{1g}, \Xi_{2g}) | Z_{ig} = 1],
\]

and \( \vartheta_1 = \{ A_1, A_2, \Psi_{1ig}, \Psi_{2ig} \} \), \( \vartheta_2 = \{ \Xi_{1g}, \Xi_{2g} \} \).

Now, we expand \( W_{\vartheta_2} \) in equation (3.8) to get

\[
W_{\vartheta_2} = \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} E \left\{ -\frac{m_1 m_2}{2} \log 2\pi - \frac{m_2}{2} \log |\Xi_{1g}| - \frac{m_1}{2} \log |\Xi_{2g}| - \frac{1}{2} \text{tr} [\Xi_{1g}^{-1} (u_{ig} - \Delta_g) \Xi_{2g}^{-1} (u_{ig} - \Delta_g')] | x_i \right\}
\]

\[
- \frac{1}{2} \text{tr} [\Xi_{1g}^{-1} (u_{ig} - \Delta_g) \Xi_{2g}^{-1} (u_{ig} - \Delta_g')] | x_i \right\}
\]

\[
= \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} \left\{ -\frac{m_1 m_2}{2} \log 2\pi - \frac{m_2}{2} \log |\Xi_{1g}| - \frac{m_1}{2} \log |\Xi_{2g}| - \frac{1}{2} \text{tr} [E (\Sigma_{1g}^{-1} u_{ig} \Sigma_{2g}^{-1} u_{ig})]
\]

\[
- \Xi_{1g}^{-1} E (u_{ig}) \Xi_{2g}^{-1} \Delta_g' - \Xi_{1g}^{-1} \Delta_g \Xi_{2g}^{-1} E (u_{ig}') + \Xi_{1g}^{-1} \Delta_g \Xi_{2g}^{-1} \Delta_g' | x_i \right\}.
\]
If we take the partial derivative of $W_{\theta_2}$ with respect to $\Delta_g$, and set that equal to zero, then we have

$$
\hat{\Delta}_g = \frac{\sum_{i=1}^n \hat{z}_{1g} E(U_{1g} | X_i)}{n_g}.
$$

Because the distribution of $\text{vec}(U)$ and $\text{vec}(X)$ follow the multivariate normal distribution, the conditional distribution of $\text{vec}(U)$ given $\text{vec}(X)$ is also multivariate normal. Therefore, the conditional mean is given by

$$
E[\text{vec}(U_{1g}) | \text{vec}(X_i), Z_{1g} = 1]
\begin{eqnarray}
&= &\text{vec}(\Delta_g) + (\Xi_{2g} \Lambda'_{2} \otimes \Xi_{1g} \Lambda'_1) [(\Lambda_2 \Xi_{2g} \Lambda'_2 + \Psi_{21g}) \otimes (\Lambda_1 \Xi_{1g} \Lambda'_1 + \Psi_{11g})]^{-1} \text{vec}(x_i - \Lambda_1 \Delta_g \Lambda'_2) \\
&= &\text{vec}(\Delta_g) + [\Xi_{2g} \Lambda'_{2}(\Lambda_2 \Xi_{2g} \Lambda'_2 + \Psi_{21g})^{-1} \otimes \Xi_{1g} \Lambda'_1 (\Lambda_1 \Xi_{1g} \Lambda'_1 + \Psi_{11g})^{-1}] \text{vec}(x_i - \Lambda_1 \Delta_g \Lambda'_2) \\
&= &\text{vec}(\Delta_g) + \text{vec} \left\{ \Xi_{1g} \Lambda'_1 (\Lambda_1 \Xi_{1g} \Lambda'_1 + \Psi_{11g})^{-1} (x_i - \Lambda_1 \Delta_g \Lambda'_2) [(\Lambda_2 \Xi_{2g} \Lambda'_2 + \Psi_{21g})' \Lambda_2 \Xi_{2g}']^{-1} \Lambda_2 \Xi_{2g}' \right\},
\end{eqnarray}
$$

and, thus

$$
\hat{U}_{1ig} = E(U_{1g} | \text{vec}(X_i), Z_{1g} = 1)
\begin{eqnarray}
&= &\Delta_g + \Xi_{1g} \Lambda'_1 (\Lambda_1 \Xi_{1g} \Lambda'_1 + \Psi_{11g})^{-1} (x_i - \Lambda_1 \Delta_g \Lambda'_2) [(\Lambda_2 \Xi_{2g} \Lambda'_2 + \Psi_{21g})' \Lambda_2 \Xi_{2g}']^{-1} \Lambda_2 \Xi_{2g}',
\end{eqnarray}
$$

(3.9)
where \( \hat{u}_{1ig} = \text{vec}(\hat{U}_{1ig}) \) and the conditional covariance matrix is
\[
\hat{U}_{2ig} = E[\text{vec}(U_{ig})'\text{vec}(U_{ig})'|\text{vec}(X_i), Z_{ig} = 1]
\]
\[
= \Xi_{2g} \otimes \Xi_{1g} - \left( \Xi_{2g}\Lambda'_2 \otimes \Xi_{1g}\Lambda'_1 \right) \left[ \left( \Lambda_2\Xi_{2g}\Lambda'_2 + \Psi_{2ig} \right) \otimes \left( \Lambda_{1g}\Xi_{1g}\Lambda'_1 + \Psi_{1ig} \right) \right]^{-1}
\]
\[
(\Xi_{2g}\Lambda'_2 \otimes \Xi_{1g}\Lambda'_1)' + \hat{u}_{1ig}\hat{u}_{1ig}'
\]
\[
= \Xi_{2g} \otimes \Xi_{1g} - \Xi_{2g}\Lambda'_2 (\Lambda_2\Xi_{2g}\Lambda'_2 + \Psi_{2ig})^{-1}\Lambda_2\Xi_{2g} \otimes \Xi_{1g}\Lambda'_1 (\Lambda_{1g}\Xi_{1g}\Lambda'_1 + \Psi_{1ig})^{-1}
\]
\[
\Lambda_{1g}\Xi_{1g}' + \hat{u}_{1ig}\hat{u}_{1ig}'
\]
\[
= \Xi_{2g} \otimes \Xi_{1g} - \Theta_{2ig} \otimes \Theta_{1ig} + \hat{u}_{1ig}\hat{u}_{1ig}'.
\]
(3.10)

where
\[
\Theta_{1ig} = \Xi_{1g}\Lambda'_1 (\Lambda_{1g}\Xi_{1g}\Lambda'_1 + \Psi_{1ig})^{-1}\Lambda_{1g}\Xi_{1g}'
\]
\[
\Theta_{2ig} = \Xi_{2g}\Lambda'_2 (\Lambda_{2g}\Xi_{2g}\Lambda'_2 + \Psi_{2ig})^{-1}\Lambda_{2g}\Xi_{2g}'.
\]

Now, taking the partial derivative of \( W_{\varphi_2} \) with respect to \( \Xi_{-1} \), we then have
\[
W_{\varphi_2} = \sum_{g=1}^{G} \sum_{i=1}^{n} \frac{m_2}{2} \Xi_{1g}' - \frac{1}{2} \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} \frac{\partial}{\partial \Xi_{1g}^{-1}} E[\text{tr}(u_{ig} \Xi_{-1}^{-1} u_{ig}' \Xi_{-1}')] - \Delta_{g}\Xi_{2g} \hat{U}_{1ig}' - \hat{U}_{1ig}\Xi_{2g}^{-1} \Delta_{g}' + \Delta_{g}\Xi_{2g}^{-1} \Delta_{g}'
\]
(3.11)

where
\[
\frac{\partial}{\partial \Xi_{1g}^{-1}} E[\text{tr}(u_{ig} \Xi_{-1}^{-1} u_{ig}' \Xi_{-1}')] = \frac{\partial}{\partial \Xi_{1g}^{-1}} \text{tr} \left\{ \left( \Xi_{2g}^{-1} \otimes \Xi_{1g}^{-1} \right) \hat{U}_{2ig} \right\}
\]
\[
= \frac{\partial}{\partial \Xi_{1g}^{-1}} \text{tr} \left\{ \Xi_{2g}^{-1} \otimes \Xi_{1g}^{-1} [\Xi_{2g} \otimes \Xi_{1g} - \Theta_{2ig} \otimes \Theta_{1ig} + \hat{u}_{1ig}\hat{u}_{1ig}'] \right\}
\]
\[
= \frac{\partial}{\partial \Xi_{1g}^{-1}} \left\{ \text{tr}[(\Xi_{2g}^{-1} \otimes \Xi_{1g}^{-1})] - \text{tr}[(\Xi_{2g}^{-1} \otimes \Xi_{1g}^{-1} \Theta_{1ig}')] + \text{tr}[(\Xi_{2g}^{-1} \otimes \Xi_{1g}^{-1}) \hat{u}_{1ig}\hat{u}_{1ig}'] \right\}
\]
\[
= \frac{\partial}{\partial \Xi_{1g}^{-1}} \left[ \text{tr}[(\Xi_{2g}^{-1} \Xi_{1g}^{-1})] - \text{tr}(\Xi_{2g}^{-1} \Theta_{2ig})\text{tr}(\Xi_{1g}^{-1} \Theta_{1ig}) + \text{tr}(\hat{U}_{1ig}\Xi_{2g}^{-1} \hat{U}_{1ig}' \Xi_{1g}^{-1}) \right]
\]
\[
= m_2 \times m_1 - \text{tr}(\Xi_{2g}^{-1} \Theta_{2ig})\Theta_{1ig}' + \hat{U}_{1ig}\Xi_{2g}^{-1} \hat{U}_{1ig}'.
\]
Therefore, we set equation (3.11) to zero and solve for \( \Xi_{1g} \) at the \( k + 1 \) iteration, which is

\[
\hat{\Xi}_{1g} = \frac{1}{m_2n_g} \sum_{i=1}^{n} \hat{z}_{ig} [m_2 \times m_1 - \text{tr}(\Xi_{1g}^{-1} \Theta_{2ig}) + \hat{U}_{1ig} \Xi_{2g}^{-1} \hat{U}_{1ig}' - \Delta_g \Xi_{2g}^{-1} \hat{U}_{1ig}]
\]

\(- \hat{U}_{1ig} \Xi_{2g}^{-1} \Delta_g' + \Delta_g \Xi_{2g}^{-1} \Delta_g']
\]

(3.12)

The same method is used to solve for \( \Xi_{2g} \), which is given by

\[
\hat{\Xi}_{2g} = \frac{1}{m_1n_g} \sum_{i=1}^{n} \hat{z}_{ig} [m_2 \times m_1 - \Theta_{2ig}' \text{tr}(\Xi_{1g}^{-1} \Theta_{1ig}) + \hat{U}_{1ig}' \Xi_{1g}^{-1} \hat{U}_{1ig} - \hat{U}_{1ig}' \Xi_{1g}^{-1} \Delta_g' - \Delta_g' \Xi_{1g}^{-1} \hat{U}_{1ig} + \Delta_g' \Xi_{1g}^{-1} \Delta_g']
\]

(3.13)

From the \( W_{\vartheta} \) function in equation (3.8), estimates of \( \Lambda_1, \Lambda_2, \Psi_{1ig}, \) and \( \Psi_{2ig} \) are obtained as follows:

\[
W_{\vartheta} = \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} E \left\{ - \frac{p_1p_2}{2} \log 2\pi - \frac{p_2}{2} \log |\Theta_{2ig}| - \frac{p_1}{2} \log |\Theta_{1ig}| - \frac{1}{2} \text{tr}[\Theta_{2ig}^{-1} (X_i - \Lambda_1 \Psi_{1ig} \Lambda_1') \Theta_{1ig}^{-1} (X_i - \Lambda_1 \Psi_{1ig} \Lambda_1')] \right\}
\]

\[
= \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} \left\{ - \frac{p_1p_2}{2} \log 2\pi - \frac{p_2}{2} \log |\Theta_{2ig}| - \frac{p_1}{2} \log |\Theta_{1ig}| - \frac{1}{2} \text{tr}[\Theta_{2ig}^{-1} X_i \Theta_{1ig}^{-1} X_i]
\]

\(- \Theta_{2ig}^{-1} \Lambda_1 \hat{U}_{1ig} \Theta_{1ig}' \hat{X}_i - \Theta_{2ig}^{-1} X_i \Theta_{1ig}^{-1} \Lambda_2 \hat{U}_{1ig}' \Lambda_1' + E(\Theta_{2ig}^{-1} \Lambda_1 \Psi_{1ig} \Lambda_2' \Theta_{1ig}^{-1} \Lambda_2 \Psi_{1ig} \Lambda_1')) \right\}
\]

By taking the partial derivatives of this equation with respect to \( \Lambda_1, \Lambda_2, \Psi_{11ij}, \) and
Ψ_{2ij}, and setting them equal to zero, we then have

\begin{align*}
\hat{\lambda}_1 &= \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} X_i \Psi^{-1}_{2ig} \Lambda_2 \hat{U}'_{1ig} \left\{ \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} \left[ \text{tr}(\Lambda_2' \Psi^{-1}_{2ig} \Lambda_2 \Xi_{2g} - \text{tr}(\Lambda_2' \Psi^{-1}_{2ig} \Lambda_2 \Theta_{2ig}) \Theta_{1ig} \right] ight. \\
&+ \left. \hat{U}_{1ig} \Lambda_2' \Psi^{-1}_{2ig} \Lambda_2 \hat{U}'_{1ig} \right\}^{-1}, \\
(3.14)
\end{align*}

\begin{align*}
\hat{\lambda}_2 &= \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} X_i' \Psi^{-1}_{1ig} \Lambda_1 \hat{U}_{1ig} \left\{ \sum_{g=1}^{G} \sum_{i=1}^{n} \hat{z}_{ig} \left[ \text{tr}(\Lambda_1' \Psi^{-1}_{1ig} \Lambda_1 \Xi_{1g} - \text{tr}(\Lambda_1' \Psi^{-1}_{1ig} \Lambda_1 \Theta_{1ig}) \Theta_{2ig} \right] ight. \\
&+ \left. \hat{U}'_{1ig} \Lambda_1' \Psi^{-1}_{1ig} \Lambda_1 \hat{U}_{1ig} \right\}^{-1}, \\
(3.15)
\end{align*}

\begin{align*}
\hat{\psi}_{1ig} &= \frac{1}{p_2} \text{diag}[X_i, \Psi^{-1}_{2ig} X_i' - X_i' \Psi^{-1}_{2ig} \Lambda_2 \hat{U}'_{1ig} \Lambda_2' - \Lambda_1 \hat{U}_{1ig} \Lambda_2' \Psi^{-1}_{2ig} X_i + \text{tr}(\Lambda_2' \Psi^{-1}_{2ig} \Lambda_2 \Xi_{2g}) \\
&- \Lambda_1' \Xi_{1g} \Lambda_1' - \text{tr}(\Lambda_2' \Psi^{-1}_{2ig} \Lambda_2 \Theta_{2ig}) \Lambda_1' \Theta_{1ig} \Lambda_1 + \Lambda_1 \hat{U}_{1ig} \Lambda_2' \Psi^{-1}_{2ig} \Lambda_2 \hat{U}'_{1ig} \Lambda_1'], \\
(3.16)
\end{align*}

and

\begin{align*}
\hat{\psi}_{2ig} &= \frac{1}{p_1} \text{diag}[X_i' \Psi^{-1}_{1ig} X_i - X_i' \Psi^{-1}_{1ig} \Lambda_1 \hat{U}_{1ig} \Lambda_1' - \Lambda_2 \hat{U}'_{1ig} \Lambda_1' \Psi^{-1}_{1ig} X_i + \text{tr}(\Lambda_1' \Psi^{-1}_{1ig} \Lambda_1 \Xi_{1g}) \\
&- \Lambda_2' \Xi_{2g} \Lambda_2' - \text{tr}(\Lambda_1' \Psi^{-1}_{1ig} \Lambda_1 \Theta_{1ig}) \Lambda_2' \Theta_{2ig} \Lambda_2 + \Lambda_2 \hat{U}'_{1ig} \Lambda_1' \Psi^{-1}_{1ig} \Lambda_1 \hat{U}_{1ig} \Lambda_2'], \\
(3.17)
\end{align*}

In the E-step, we obtain estimates of \( \hat{z}_{ig} \) and \( \hat{U}_{1ig} \) by setting initial values for \( \vartheta^{(0)} = \{ \Lambda_1^{(0)}, \Lambda_2^{(0)}, \Xi_{1g}^{(0)}, \Xi_{2g}^{(0)}, \Psi_{1ig}^{(0)}, \Psi_{2ig}^{(0)} \} \). In the M step, the system (3.12)-(3.17) at the
\( (k + 1) \)th iteration can be solved iteratively in the following order:

\[
\hat{\lambda}^{(k+1)}_1 \leftarrow \left\{ \Lambda_2^{(k)}, \Xi_1^{(k)}, \Xi_2^{(k)}, \Psi_1^{(k)}, \Psi_2^{(k)} \right\},
\]

\[
\hat{\lambda}^{(k+1)}_2 \leftarrow \left\{ \Lambda_1^{(k+1)}, \Xi_1^{(k)}, \Xi_2^{(k)}, \Psi_1^{(k)}, \Psi_2^{(k)} \right\},
\]

\[
\hat{\Xi}_1^{(k+1)} \leftarrow \left\{ \Lambda_1^{(k+1)}, \Lambda_2^{(k+1)}, \Xi_2^{(k)}, \Psi_1^{(k)}, \Psi_2^{(k)} \right\},
\]

\[
\hat{\Xi}_2^{(k+1)} \leftarrow \left\{ \Lambda_1^{(k+1)}, \Lambda_2^{(k+1)}, \Xi_1^{(k+1)}, \Xi_2^{(k+1)}, \Psi_1^{(k)}, \Psi_2^{(k)} \right\},
\]

\[
\hat{\Psi}_1^{(k+1)} \leftarrow \left\{ \Lambda_1^{(k+1)}, \Lambda_2^{(k+1)}, \Xi_1^{(k+1)}, \Xi_2^{(k+1)}, \Psi_1^{(k)}, \Psi_2^{(k)} \right\},
\]

and

\[
\hat{\Psi}_2^{(k+1)} \leftarrow \left\{ \Lambda_1^{(k+1)}, \Lambda_2^{(k+1)}, \Xi_1^{(k+1)}, \Xi_2^{(k+1)}, \Psi_1^{(k+1)} \right\}.
\]

We use the deterministic annealing technique introduced by Zhou and Lange (2010) to obtain estimates of parameters at the M-step. They proposed annealing algorithms that involve essentially trivial changes to existing optimization algorithms built on block relaxation or the EM or Majorize-Minimization/Maximization (MM) principle (de Leeuw J., 1994; Heiser, 1995; Becker et al., 1997; Lange et al., 2000; Hunter and Lange, 2004). For the multimode of log-likelihood function in the EM algorithm, this strategy significantly improves the chances of locating the global maximum by flattening the likelihood surface and gradually warping the substitute surface towards the original surface. Therefore, we can avoid multiple random starting points and also save computation time.
3.4 Simulation Study

The aim of performing this simulation is to test the feasibility of the matrix normal mixture model, four parsimonious models, and the mixture factor analysis model. We want to learn whether these models are able to group attributes correctly by comparing with the preset group labels, and whether the factor analysis model is able to provide information about panelists and products within a group and among the groups by factor plots. We are also interested in comparing the clustering performance between our proposed models and Ward’s hierarchical clustering method.

3.4.1 Data

We randomly generate two different groups of attribute observations where each group contains 30 attributes. Each attribute is a $6 \times 10$ matrix that represents panelists and products separately. For each model, we iterate 1000 times, and their performance is evaluated by the misclassification rate and the average value of the Rand index (Rand, 1971). The specific steps for generating attribute matrices are as follows:

1. Randomly generate two matrices (i.e., $X_1$ and $X_2$), based on uniform distribution (i.e., $U(7,9)$ and $U(1,3)$ separately.)

2. Randomly generate 30 matrices using $X_1$+error. The error term follows the $N(0,1)$.

3. Randomly generate 30 matrices using $X_2$+error. The error term follows the $N(0,1)$. 
3.4.2 Simulation Results

With 1000 iterations, the misclassification rate and the average value of the Rand index for the matrix normal mixture model (Full model), four parsimonious models (P.M.), the mixture factor analysis model (Fa.M.), and Ward’s hierarchical clustering method (H.C.) are given in Table 3.1.

Table 3.1: Misclassification rate and average value of the Rand index for all models.

<table>
<thead>
<tr>
<th></th>
<th>Full Model</th>
<th>P.M.1</th>
<th>P.M.2</th>
<th>P.M.3</th>
<th>P.M.4</th>
<th>Fa.M.</th>
<th>H.C.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misclassification rate</td>
<td>0.01</td>
<td>0.01</td>
<td>0.007</td>
<td>0.015</td>
<td>0.015</td>
<td>0.002</td>
<td>0</td>
</tr>
<tr>
<td>Avg. Rand Index</td>
<td>0.9982</td>
<td>0.9987</td>
<td>0.9991</td>
<td>0.998</td>
<td>0.9974</td>
<td>0.9994</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.1 shows that all models perform well. Using Ward’s hierarchical clustering method, we found that there are no misclassifications. However, this method only concerns partitions based on the Euclidean distance between the observations while neglecting the variation among the rows and columns within each observation. The factor analysis model is slightly better than other models, with only two classification errors and an average Rand index of 0.9994. However, the rest of models also have quite low misclassification rates of around 1-1.5% and quite high average values of the Rand index. Therefore, all models can be used to cluster the attribute data.

As far as the factor analysis model is concerned, from equation (3.9) the factor plots about panelists and products within each group can be produced using \( \hat{\lambda}_1 \hat{U}_{1ig} \) and \( \hat{U}_{1ig} \hat{\lambda}'_2 \), which are given in Figure 3.1 and Figure 3.2, respectively.
In group 1, panelists 3, 4, and 5 have roughly equal scores for factors 1 and 2. Panelists 2 and 6 have slightly higher scores for factor 1 than for factor 2. Panelist 1 has a far higher score for factor 1 than for factor 2. In group 2, panelist 1 had equal scores for factors 1 and 2, while the rest of the panelists have slightly higher scores for factor 1 than for factor 2. We cannot detect any clear pattern across the panelists in groups 2 as they gave dissimilar evaluations with respect to the attributes.
Figure 3.2: Factor plots about products for the simulation data set, when $G = 2$.

Figure 3.2 shows the patterns of 10 products for each group of attributes. In group 1, product 9 had a positive score for factor 2 and a negative score for factor 1. All other products had negative scores for both factors; however, all but product 5 had higher scores for factor 1 than for factor 2. We also observe that products 1, 2, 3, 4, 6, and 10 are similar because they are located in close proximity on the plot. In group 2, all products have higher scores for factor 1. Products 1 and 10 are similar and products 2, 3, 4, 5, 7, and 8 are similar, while product 9 is different from all others.

Therefore, Figure 3.1 and Figure 3.2 clearly show differences between panelists and products for two groups of attributes. We are also able to learn which
products have similar characteristics and which panelists disagree on the evaluation for a group of attributes.

3.5 Application

3.5.1 Data

In this section, we apply all models to white bread data collected by Com- pusense, Inc. in November, 2011. There are 42 sensory attributes shown in Table 3.2 to measure the characters of the white bread. A total of nine sensory panelists evaluated 12 different brands of bread, with three replications. According to the convergence technique described above, we start deterministic annealing at $\nu = 0.05$, and after every 5 iterations we replace $\nu$ by $\nu \times 1.05$ until the most recent $\nu$ is approximately equal to 1 and ECM convergence occurs.
Table 3.2: Forty-two attributes of white bread as clustered by the full model. The attributes marked with a number
are the clustering results from parsimonious Model 2.

<table>
<thead>
<tr>
<th>Group</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Colour Intensity of Crumb, Overall Aroma, Grain Aroma, White Flour Aroma, Overall Flavour, Yeasty/Fermented Aroma, Sweet Aromatics Aroma, Dairy, Grain Flavour, White Flour Flavour, Yeasty/Fermented Flavour, Sour Aromatics Flavour, Sweet Aromatics Flavour, Saltiness, Sweetness, Surface Roughness, Firmness (Crumb), Density (Crumb), Aroma⁴, Nutty Aroma⁴, Toasted Aroma⁴, Dairy Flavour⁴, Nutty Flavour⁴, Toasted Flavour⁴, Bitterness⁴, Particles (Crust)⁴</td>
</tr>
<tr>
<td>Group 2</td>
<td>Visual Roughness of Crust (Whole Loaf), Springiness of Crumb, Adhesiveness (Crumb), Cohesiveness (Crumb)</td>
</tr>
<tr>
<td>Group 3</td>
<td>Cell Uniformity (Crumb), Cell Size, Firmness (Crust), Moistness (Crust), Chewiness (Crust), Moistness (Crumb), Chewiness (Crumb), Sour Aromatics Aroma¹, Sourness¹</td>
</tr>
<tr>
<td>Group 4</td>
<td>Colour Intensity of Crust (Whole Loaf), Glossiness of Crust (Whole Loaf), Particles (Crumb), Crust/Crumb Difference³</td>
</tr>
</tbody>
</table>
3.5.2 Clustering Results

We fit the matrix normal mixture full model and four parsimonious models to the white bread data and cluster them into $G = 2, 3, 4$ groups. According to the Bayesian information criterion (BIC) criterion, dividing these attributes into four groups is the best choice for both the full model and parsimonious Model 2. This result is also shown in Table 3.2. The attribute marked with a group id shows different clustering results using parsimonious model 2. Generally speaking, Group 1 describes basic physical flavour and aroma, Group 2 represents viscosity of crumb, Group 3 measures size, moistness, and chewiness of the crust and crumb, while Group 4 shows colour and glossiness of crust and particles of crumb. Based on the results, the full model and Model 2 highly agree with respect to grouping these attributes. Therefore, using parsimonious Model 2 is highly recommended.

3.5.3 Visualization

Clustering results according to the factor analysis model are given in Table 3.3:
Table 3.3: Forty-two attributes of white bread as clustered by the factor analysis model.

<table>
<thead>
<tr>
<th>Group</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Glossiness of Crust (Whole Loaf), Grain Aroma, Dairy Aroma, Nutty Aroma, Toasted Aroma, Grain Flavour, Sour Aromatics Flavour, Dairy Flavour, Nutty Flavour, Toasted Flavour, Particles (Crust), Particles (Crumb)</td>
</tr>
<tr>
<td>Group 2</td>
<td>Visual Roughness of Crust (Whole Loaf), Colour Intensity of Crumb, Cell Size, White Flour Aroma, Sweetness, Yeasty/Fermented Aroma, Sour Aromatics Aroma, Sweet Aromatics Aroma, Bitterness, Springiness of Crumb, White Flour Flavour, Yeasty/Fermented Flavour, Sweet Aromatics Flavour, Sourness, Surface Roughness, Firmness (Crumb), Density (Crumb)</td>
</tr>
<tr>
<td>Group 3</td>
<td>Overall Aroma, Overall Flavour, Saltiness, Crust/Crumb Difference, Firmness (Crust), Moistness (Crust), Chewiness (Crust), Chewiness (Crumb)</td>
</tr>
<tr>
<td>Group 4</td>
<td>Colour Intensity of Crust (Whole Loaf), Cell Uniformity (Crumb), Moistness (Crumb), Adhesiveness (Crumb), Cohesiveness (Crumb)</td>
</tr>
</tbody>
</table>
The clustering results in Table 3.3 are different than those in Table 3.2. Clustering is unsupervised learning, and therefore researchers can choose a set of results according to their research objectives.

Now, from equation (3.9), the group structures affected by panelists and products can be observed by plotting $\hat{\lambda}_1 \hat{U}_{1ig}$ and $\hat{U}_{1ig} \hat{\lambda}_2'$, respectively.

Figure 3.3 illustrates the different distributions of panelists on the factor plots for four groups.

Figure 3.3: Factor plots for panelists, with $G = 4$. 
Based on the graphs, there is no obvious difference among these four groups, which implies all panelists remain consistent in their evaluation for all attributes. We also notice that panelists 5094 and 2652 give an opposite evaluation for all groups, while panelists 5399 and 3694 give similar evaluations for Groups 1, 2 and 3.

Figure 3.4 illustrates the different distributions of products on the factor plots for these four groups.

![Factor plots for products, with $G = 4$.](image-url)
The plots show that product 10 is different than all other products for all groups, whereas products 3 and 6 are almost the same in all groups. Other products are distributed differently across the groups. For example, product 11 is far from other products in group 1, but is very similar to products 3 and 6 in group 2. Therefore, product developers must only focus on a specific group of attributes with respect to improving their product.

3.5.4 Hierarchical Clustering Results

To compare with the results obtained from factor analysis model, we perform Ward’s hierarchical clustering method to cluster the attributes using the bread data set, and explore the pattern of the panelists and products within each group thereafter. Clustering results according to the Ward’s hierarchical clustering method are given in Table 3.4.
Table 3.4: Forty-two attributes of white bread as clustered by the Hierarchical model.

<table>
<thead>
<tr>
<th>Group</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Colour Intensity of Crust (Whole Loaf), Cell Uniformity (Crumb), Moistness (Crumb), Adhesiveness (Crumb), Cohesiveness (Crumb)</td>
</tr>
<tr>
<td>Group 2</td>
<td>Glossiness of Crust (Whole Loaf), Colour Intensity (Crumb), Cell Size, Grain Aroma, White Flour Aroma, Yeasty.Fermented Aroma, Sour Aromatics Aroma, Sweet, Aromatics Aroma, Grain Flavour, White Flour Flavour, Yeasty.Fermented Flavour, Sour Aromatics Flavour, Sweet Aromatics Flavour, Saltiness, Sourness, Sweetness, Bitterness, Surface Roughness, Firmness (Crumb.), Density (Crumb.)</td>
</tr>
<tr>
<td>Group 3</td>
<td>Visual Roughness of Crust (Whole Loaf), Overall Aroma, Springiness (Crumb) Overall Flavour, Crust/Crumb Difference, Firmness (Crust.), Moistness (Crust.), Chewiness (Crust.), Chewiness (Crumb.)</td>
</tr>
<tr>
<td>Group 4</td>
<td>Dairy Aroma, Nutty Aroma, Toasted Aroma, Dairy Flavour, Nutty Flavour, Toasted Flavour, Particles (Crust.), Particles (Crumb.)</td>
</tr>
</tbody>
</table>
As a result, Table 3.4 roughly agrees with Table 3.3. Group 1 in Table 3.4 is the same as the group 4 in Table 3.3. Group 2 of Table 3.4 and group 3 of Table 3.3 both contain Overall Flavour, Crust/Crumb Difference, Firmness (Crust), moistness (Crust), Chewiness (Crust), and Chewiness (Crumb). Moreover, the attributes of group 4 in Table 3.4 is a subset of group 1 in Table 3.3.

Figures 3.5 and 3.6 show the patterns of panelists and products for each group of Table 3.4 by PCA, respectively.
Figure 3.5: PCA plots for panelists, with \( G = 4 \). Note: 1: 2652, 2: 3248, 3: 3694, 4: 3820, 5: 5049, 6: 5399, 7: 6469, 8: 7193, 9: 9771
In Figure 3.5, as a result of PCA on each panelist matrix (i.e., 12 products as rows and attributes in groups of Table 3.4 as columns), we observe that each panelist is shown 12 times in each group. It is difficult for us to learn if panelists are
in agreement with each other. For example, Figure 3.3 clearly shows that panelists 5094 and 2652 gave an opposite evaluation for all groups, but we cannot obtain this information from Figure 3.5.

Figure 3.6 has the same problem as Figure 3.4. Each product is represented nine times (i.e., nine panelists as rows, attributes as columns) in each group. From Figure 3.4, we concluded that product 10 was different than all other products for all groups, but Figure 3.6 does not show this pattern in groups 2 and 4.

Overall, because the observation matrices are not able to reduce dimensions from rows and columns simultaneously by Ward’s hierarchical clustering method, the patterns of figures 3.5 and 3.6 are complicated and we cannot get precise information from them.

3.6 Summary

In this chapter, we extend the mixture model-based clustering technique from multivariate data to matrix data, which is required due to the features of sensory data. To reduce estimation of the matrix parameters, four parsimonious models are introduced and we used the BIC to select the models. A simulation study was used to investigate the performance of the models, and the results showed all models performed very well. By applying the models to white bread data, we found that parsimonious Model 2, which assumes a mean matrix, results in identical values within each column but varying values among the columns with respect to those produced by the full model according to the BIC. To visualize the results of clustering,
mixtures of factor analyzers were plotted and used to identify the differences within a group and among groups. The benefit of this technique is learning the motivation behind the clustering while also considering both variation in the products and sensory panelists. Moreover, factor plots allow the intensity of products within a small group of attributes to become self-evident. Based on this information, product developers can target a group of attributes when formulating their product development strategy.
Chapter 4

A Connection Between Sensory Data and Consumer Liking Score

4.1 Conditional Multivariate Normal Distribution of Liking Scores

To construct a bridge between sensory data and consumer liking scores, we introduce a conditional multivariate normal distribution of liking scores given the sensory data for all products. Consider a sensory matrix containing $k$ products and $m$ attributes. We let a liking score $Y_i$ for consumer $i$ given sensory matrix $X$ follow a multivariate normal distribution, that is

$$Y_i|X \sim MVN[\beta_0\mathbf{1} + (X - \mu^*)\beta_1, D], \quad (4.1)$$

where $i = 1, 2, \ldots, n$. In this distribution, vector $\mathbf{1}$ is of length $k$ and $D$ is a diagonal covariance matrix that assumes the products are independent. The coefficient $\beta_0$ is a constant and $\beta_1$ is an $m \times 1$ matrix that gives a different weight to each attribute within one product; however, $\beta_1$ has the same contribution for each product. To avoid this mean function being meaningless, i.e., $\beta_0 = 0$, we add a $k \times m$ matrix...
\( \mu^* \) to the function with all entries being the same. The sensory matrix \( X \) contains column vectors \( x_1, x_2, \ldots, x_k \), and each \( x_j \), where \( j = 1, 2, \ldots, k \), follows a multivariate normal distribution with mean \( \mu_j \) and covariance matrix \( \Sigma_j \). To estimate the coefficients \( \beta_0, \beta_1, \) and \( D \) when \( \mu_j \) and \( \Sigma_j \) are known, we use the ECM algorithm, which is an iterative method for finding maximum likelihood estimates of parameters where there are unobserved or missing data. In this chapter, the liking scores \( Y_i \) are observed data, whereas the sensory data \( X \) given by consumers are unobserved.

Based on (4.1), we have the complete-data log-likelihood function

\[
L = \sum_{i=1}^{n} [\log f(y_i|x_1, x_2, \ldots, x_k) + \log f(x_1, x_2, \ldots, x_k)]
\]

\[
= \sum_{i=1}^{n} [\log f(y_i|x_1, x_2, \ldots, x_k) + \log f(x_1) + \log f(x_2) + \cdots + \log f(x_k)].
\]  

(4.2)

Hence, the expected complete-data log-likelihood function at the \( t \)th iteration is

\[
Q(\theta) = \sum_{i=1}^{n} W_{1i} + \sum_{i=1}^{n} \sum_{j=1}^{k} W_{2ij}, \tag{4.3}
\]

where

\[
W_{1i} = E_{g(\epsilon)}[\log f(x_1, x_2, \ldots, x_k|\theta)|y_i],
\]

\[
W_{2ij} = E_{g(\epsilon)}[\log f(x_j|\theta)|y_i],
\]

and

\[
\theta = \{ \beta_0, \beta_1, D, \mu_j, \Sigma_j \}.
\]

Also, we can obtain the mean and variance of \( Y_i \) using

\[
E(Y_i) = E[E(Y_i|x_1, x_2, \ldots, x_k)] = \beta_0 1 + (M - \mu^*)\beta_1 \tag{4.4}
\]
and

\[ \text{Var}(Y_i) = \text{E}[\text{Var}(Y_i|x_1, x_2, \ldots, x_k)] + \text{Var}[\text{E}(Y_i|x_1, x_2, \ldots, x_k)] \]

\[ = \text{Var} [\beta_0 1 + (X - \mu^*) \beta_1] + D \]

\[ = \text{Var}(X \beta_1) + D, \tag{4.5} \]

where \( M \) is a \( k \times m \) matrix containing column vectors \( \mu_1, \mu_2, \ldots, \mu_k \) and \( \text{Var}(X \beta_1) \) is a diagonal matrix with elements \( l = \{ \beta_1^T \Sigma_1 \beta_1, \ldots, \beta_1^T \Sigma_k \beta_1 \} \).

The covariance of \( X \) and \( Y_i \) is given by

\[ \text{Cov}(X, Y_i) = \text{E}[\text{Cov}(X, Y_i|X)] + \text{Cov}[\text{E}(X|X), \text{E}(Y_i|X)] \]

\[ = \text{Cov}[X, \beta_0 1 + (X - \mu^*) \beta_1] \]

\[ = \text{Cov}(X, X \beta_1), \tag{4.6} \]

where \( \text{Cov}(X, X \beta_1) \) is a diagonal matrix with elements \( l' = \{ \Sigma_1 \beta_1, \ldots, \Sigma_k \beta_1 \} \).

Therefore, using equations (4.4), (4.5), and (4.6) at each E-step, we have

\[ U = \text{E}(X|Y_i) \]

\[ = M + \text{Cov}(X, Y_i)[D + \text{Var}(X \beta_1)]^{-1}[Y_i - \beta_0 1 - (M - \mu^*) \beta_1]. \tag{4.7} \]

Moreover, the covariance matrix of \( X \) given \( Y_i \) is obtained from

\[ \text{Cov}(X|Y_i) = \Sigma - \text{Cov}(X, Y_i)[D + \text{Var}(X \beta_1)]^{-1}\text{Cov}(X, Y_i)', \tag{4.8} \]

where \( \Sigma \) is an \( mk \times mk \) matrix. Therefore, we have

\[ \text{E}(XX'|Y_i) = \text{Cov}(X|Y_i) + UU'. \tag{4.9} \]

This will be used in \( \sum_{i=1}^{n} \sum_{j=1}^{k} W_{2ij} \).
Now, we expand $W_{ii}$ in equation (4.3) to get
\[
\sum_{i=1}^{n} W_{ii} = \sum_{i=1}^{n} \mathbb{E} \left\{ -\frac{k}{2} \log 2\pi - \frac{1}{2} \log |D| - \frac{1}{2} \text{tr}[D^{-1}(y_i - \beta_0 1 - (X - \mu^*)\beta_1)(y_i - \beta_0 1 - (X - \mu^*)\beta_1) \mid y_i] \right\} \\
= -\frac{n}{2} \log |D| - \frac{1}{2} \sum_{i=1}^{n} \left\{ \text{tr}[D^{-1}(y_i 1' - y_i 1' \beta_0 - y_i \beta_1' U' + y_i \beta_1' \mu^* - \beta_0 1 y_i') + \beta_0 11' \beta_0 + \beta_0 1 \beta_1' U' - \beta_0 1 \beta_1' \mu^* - U \beta_1 y_i' + \mu^* \beta_1 y_i' + U \beta_1 1' \beta_0 - \mu^* \beta_1 1' \beta_0 - U \beta_1 \beta_1' \mu^* - \mu^* \beta_1 \beta_1' \mu^*] + \text{tr}[D^{-1} E(X \beta_1 \beta_1' X')] \right\}.
\]

In the M-step, we take the partial derivative of $\sum_{i=1}^{n} W_{ii}$ with respect to $\beta_0$, $\beta_1$, and $D$ respectively to obtain
\[
\hat{\beta}_0 = \frac{1}{n} \mathbb{E} \left\{ \sum_{i=1}^{n} [y_i - (U - \mu^*) \beta_1] \right\} \\
\hat{\beta}_1 = \left\{ \sum_{i=1}^{n} \left[ \text{tr}(D^{-1}) \text{tr}(\Sigma) + (U - \mu^*)' D^{-1} (U - \mu^*) \right] \right\}^{-1} \sum_{i=1}^{n} (U'D^{-1} y_i - \mu^* D^{-1} y_i) \\
- U'D^{-1} \beta_0' 1 - \mu^* \beta_0' 1\right\}.
\]
and
\[
\hat{D} = \sum_{i=1}^{n} [y_i 1' y_i' - y_i 1' \beta_0 - y_i \beta_1' U' + y_i \beta_1' \mu^* - \beta_0 1 y_i' + \beta_0 11' \beta_0 + \beta_0 1 \beta_1' U' - \beta_0 1 \beta_1' \mu^* - U \beta_1 y_i' + \mu^* \beta_1 y_i' + U \beta_1 1' \beta_0 - \mu^* \beta_1 1' \beta_0 - U \beta_1 \beta_1' \mu^* - \mu^* \beta_1 \beta_1' \mu^*] \\
+ \text{tr}(\Sigma x \beta_1) + U \beta_1 \beta_1' U],
\]
where $\Sigma x \beta_1$ is a diagonal matrix with elements $l'' = \{ \beta_1' \Sigma_1 \beta_1, \beta_1' \Sigma_2 \beta_1, \ldots, \beta_1' \Sigma_k \beta_1 \}$. 

We iterate the E-step and M-step, and the maximum likelihood estimates for these parameters are obtained when the difference between $L^{(t)}$ and $L^{(t+1)}$ becomes sufficiently small.
4.2 Multinomial Logistic Regression

Multinomial logistic regression is a simple extension of binary logistic regression that allows for more than two categories of the outcome variable. Using this logistic regression, we can compare each category of an unordered outcome variable to the reference category, providing a number of logistic regression models.

In this section, we consider the liking score from 1 to 9 as categorical variable $S$ for $Y_{ij}$, and select score 5 as the reference level. Therefore, for consumer $i$, the multinomial logistic model can be written as

$$\log \frac{P(Y_{ij} = S)}{P(Y_{ij} = 5)} = \alpha_s + x_j \beta_s,$$

(4.10)

where $\alpha_s$ is a constant and $\beta_s$ are vectors of regression coefficients, for $s = 1, \ldots, 4, 6, \ldots, 9$. In this model, the eight multinomial logistic equations contrast each of the categories 1, $\ldots$, 4, 6, $\ldots$, 9 with reference category 5, whereas the binary logistic regression equation is a contrast between successes and failures. Therefore, in equation (4.10), the intercept $\alpha_s$ tells us the expected odds of falling into the given category $S$ vs. the reference category 5 when the variable $x_j$ is zero, while the slope $\beta_s$ tells us how the log-odds of falling into the given category $S$ vs. the reference change with one unit of $x_j$.

We can also express equation (4.10) in terms of the original probabilities $P(Y_{ij} = S)$ rather than the log-odds, resulting in

$$P(Y_{ij} = S) = \frac{\exp(\alpha_s + x_j \beta_s)}{\sum_{S=1}^{9} \exp(\alpha_s + x_j \beta_s)},$$

(4.11)

for $S = 1, 2, \ldots, 9$. 
4.3 Simulation

4.3.1 Data

To investigate the performance of the equation 4.1, we generate a sensory data set containing six products and 12 attributes and a consumer liking data set comprising 300 consumers and six products. The details with respect to the generation of the sensory and consumer data are given below.

1. For the sensory data set, we randomly generate 12 means based on the uniform distribution (i.e., $U(30, 70)$). We also generate a $12 \times 12$ positive defined covariance matrix for each product.

2. According to the means and covariance matrix in step 1, we randomly generate an observation for each product based on the multivariate normal distribution.

3. We arbitrary set an initial value $\beta_0 = 5$.

4. We randomly generate the initial value of $\beta_1$ from $U(0.01, 0.1)$.

5. We randomly generate the initial value of $D$ from $U(0, 0.01)$.

6. According to equation 4.4 and 4.5, we generate the mean and variance for each consumer’s liking scores.

7. Based on step 6, we finally generate the liking score data by multivariate normal distribution.
4.3.2 Simulation Results

Estimation of the coefficients for each attribute (att.) are given in table 4.1. All of the numerical results have been rounded to four decimal places.

Table 4.1: Estimation of $\beta$ coefficients for the simulation data.

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>att. 1</th>
<th>att. 2</th>
<th>att. 3</th>
<th>att. 4</th>
<th>att. 5</th>
<th>att. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>26.4685</td>
<td>0.1700</td>
<td>0.2382</td>
<td>-0.2610</td>
<td>-0.2293</td>
<td>-0.4646</td>
<td>-0.0460</td>
</tr>
</tbody>
</table>

Table 4.1 Continued:

<table>
<thead>
<tr>
<th></th>
<th>att. 7</th>
<th>att. 8</th>
<th>att. 9</th>
<th>att. 10</th>
<th>att. 11</th>
<th>att. 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.2967</td>
<td>-0.6022</td>
<td>0.0011</td>
<td>-0.3614</td>
<td>0.0697</td>
<td>0.1558</td>
</tr>
</tbody>
</table>

From Table 4.1 we know that attributes 1, 2, 7, 9, 11, and 12 have positive estimators for $\beta_1$ while the rest are negative. Attribute 7 has the highest positive value, which means that a change of one unit in intensity results in the mean liking score changing 0.2967 units, provided all other attributes are fixed. Attribute 2 also has a strong positive effect on the mean liking score. Attribute 8 has the strongest negative influence on the mean liking score. Attribute 9 has little effect on the mean liking score because the coefficient is only 0.0011.

Based on equation 4.7, we obtain the conditional estimation of sensory matrices given by each consumer liking score. Therefore, associating the consumer liking score with these estimated sensory matrices instead of the original sensory data collected from a few panelists, we perform a PCA on the correlation matrix and preference mapping technique using a circular model to detect the properties of an optimal product from the map. Figure 4.1 is similar to figures in Meullenet et al. (2007) and
Figure 4.2 is inspired by Schlich and McEwan (1992).

Figure 4.1: PCA performed on the correlation matrix for the simulation data.

Figure 4.1 shows 12 attribute loadings on the first two principal components for PCA performed on the correlation matrix. Overall, att.3, att.4, att.6, att.7, att.9, att.11, and att.12 have high loadings on PC1 while att.1, att.2, att.5,
att.8, and att.10 are represented by PC2 because their loadings are bigger on this dimension.

![Preference mapping](image)

Figure 4.2: Circular model-based preference mapping for simulation data.

Figure 4.2 shows circular model-based preference mapping by individual consumers. The two dimension scales are based on the loading scores of the first two
PCs, and the region for the ideal products is projected on the map. The scales lying in the contour lines indicate the percentage of likers in the regions of the sensory space and the scale decreases from inner to outer contours. Bearing in mind that the circular model is an extension of the quadratic model, this implies that the consumer liking is increasing as intensity increases until it reaches the maximum preference; thereafter, the liking decreases as intensity increases.

In this plot, the darkness of the red colour shows that more consumers agree with respect to their maximum liking according to the circular model. Conversely, the dark blue region indicates the least ideal location for the product. Therefore, this plot shows product 3 is closer to the ideal region (i.e, dark red), which implies more consumers will like it. Product 1, 4, and 5 are in the dark blue region, so these products still need to be improved to meet consumers’ preferences. Combining this information with that from Figure 4.1, we know that all attributes on the right side of the circle (i.e., att.2, att.3, att.4, att.6, att.9, and att.12) play an important role in deriving consumer liking towards an ideal product.

From another point of view, to observe the probability of each liking score vs. the reference score for a given product, we perform a multinomial logistic regression (i.e., equation (4.10)) on the simulation data by regressing the log-odds of falling into the given category $S$ vs. the reference category 5 on $U$ in equation (4.7). To simplify the plot, we group the consumers into three groups by Ward’s hierarchical cluster analysis using Euclidian distances. The resulting mean probability of each liking score for each consumer for each product is given in Figure 4.3.
In Figure 4.3, considering liking score 5 as the reference level, product 1 receives liking score 2 and 3 across all groups. Similarly, product 2 obtains liking score 6 and 7, and product 3 receives liking scores 7 and 8 from all consumers. Few products receive the highest score, i.e., 9. In group 1, product 5 largely receives liking score 5. Scores for product 4 are split: the mean probabilities of receiving liking scores 4 and 7 are each 0.5. In group 2, product 5 receives liking score 4 and product
4 receives score 5. In group 3, most of products 4 and 5 receive liking score 5 and product 4 sometimes obtains score 7.

The simulation study illustrates our method is able to link the sensory data with consumer liking and it allows us to learn information about optimal products by estimating sensory data based on consumer liking. In the next section, we will apply this model to real data sets to discover information about commercial products that is useful in practice.

4.4 Applications

In this section, we apply the conditional multivariate normal distribution to the liking scores and the multinomial logistic regression model to two real data sets (i.e., cupcake and meat) provided by Compusense Inc. For both, we aim to estimate $\beta$ coefficients to understand how attributes impact the mean liking scores given the sensory data. Due to the insufficient number of products in the meat data set, preference mapping to determine the location of the optimal product is only conducted for the cupcake data set using the SensoMineR package in R. Thereafter, the probability of each liking score is estimated vs. the reference score 5, which represents a neutral preference. The probability plots based on the products, liking scores, and grouped consumers are given for both data sets.
4.4.1 Cupcake Data

For the cupcake data, the sensory data set contains four products that are evaluated by nine panelists based on eight attributes, and each product is evaluated two times. The liking scores data set is created by 300 consumers who are asked to evaluate their overall liking for these four products. The scale of the sensory intensity value is from 0 to 10 to 2 decimal points, and the liking scores use the 9-point hedonic scale. We obtain the mean $\mu_j$ and covariance matrix $\Sigma_j$ for each product $j$ based on all panelists. The resulting $\hat{\beta}_0$ and $\hat{\beta}_1$s are given in Table 4.2.
Table 4.2: The estimation of $\beta$ coefficients for the cupcake data.

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Dairy includes Butter</th>
<th>Egg</th>
<th>Surface Shine</th>
<th>Porosity</th>
<th>Denseness</th>
<th>Firmness</th>
<th>Springiness</th>
<th>External Residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>28.8797</td>
<td>4.1338</td>
<td>2.7511</td>
<td>-0.0469</td>
<td>-0.1870</td>
<td>-0.7622</td>
<td>0.0417</td>
<td>-0.0445</td>
<td>1.1290</td>
</tr>
</tbody>
</table>
From Table 4.2, we know that the coefficients for **Surface Shine**, **Porosity**, **Denseness**, and **Springiness** are negative, whereas those for **Dairy includes Butter**, **Egg**, **Firmness**, and **External Residue** are positive. The **Dairy includes Butter** term has a strong positive value, which means that a change of one unit in intensity results in the mean liking score changing 4.1338 units, provided the other attributes are fixed. The attribute **Egg** also has a large influence on the mean liking score; for each change of one unit in intensity of **Egg**, the liking score changes 2.7511 units if the rest of the attributes are fixed. Compared to these two attributes, **Firmness**, and **External Residues** contribute moderately small positive effects to the liking scores. Conversely, **Denseness**, **Porosity**, **Springiness**, and **Surface Shine** will slightly decrease consumer liking if they change one unit. Overall, it appears that product developers should pay most attention to the **Dairy includes Butter** and **Egg** attributes to please consumers.

Based on the conditional estimation of sensory matrices given by each consumer liking score, we perform a PCA on the correlation matrix and preference mapping with a circular model to investigate the properties of an optimal cupcake product from the map. Using the same idea of Figures 4.1 and 4.2, the results are shown in Figures 4.4 and 4.5.
Variables factor map (PCA)

Figure 4.4: PCA performed on the correlation matrix for the cupcake data.

Figure 4.4 shows eight attribute loadings on the first two principal components for PCA performed on the correlation matrix. The attributes Egg, Surface Shine, External Residue, and Porosity have high loadings on PC1, whereas Denseness, Springiness, Firmness, and Dairy includes Butter have higher loadings on PC2.
Figure 4.5 shows circular model-based preference mapping for the cupcake data by individual consumers. From this plot, we see that product 3 is closer to the ideal region (i.e., dark red). Combining this information with that from Figure 4.4, we know that the Denseness, Egg, Surface Shine, and External Residue play an
important role in deriving consumer liking towards an ideal product. Moreover, these two plots show that product 4 is represented by Firmness and Porosity, product 2 is represented by Dairy includes butter, and none of attributes represent product 1.

Now, according to equation (4.10), we perform a multinomial logistic regression on the cupcake data by regressing the log-odds of falling into the given category S vs. the reference category 5 on U in equation (4.7). To simplify the plot, we group the consumers into five groups by Ward’s hierarchical cluster analysis using Euclidian distances. The resulting mean probability of each liking score for each consumer for each product is given in Figure 4.6.
In Figure 4.6, generally speaking, the liking scores for all of the products are above 6. Specifically, for group 1, products 1 and 2 typically obtain score 8 with mean probability 0.38, and products 3 and 4 receive scores of 7 and scores of 9, respectively, with mean probability 0.4. For group 2, product 4 consistently obtains scores of 9 and product 3 receives scores of 8 with higher mean probability. Products 1 and 2 receive scores both 9 and 7. For group 3, product 3 often scores 9 and product 1 and
4 receive scores of 9 and 8. Product 2 receives scores of 6 and 8. For group 4, all consumers give score 8 to products 3 and 4, and most consumers give a score of 8 to products 1 and 2. For group 5, no discrimination among the products is possible because consumers gave a score of 9 for all products.

4.4.2 Meat Data

The meat data contain eight attributes based on three different products given by 14 panelists. Liking scores were then provided by 211 consumers. The scale of the sensory intensity value is from 0 to 100 to 1 decimal point and the liking scores use a 9-point hedonic scale. The $\mu_j$ and $\Sigma_j$ are obtained before running the ECM algorithm. The results of the estimation of the $\beta$s and mean probability of each liking score are given in Table 4.3 and Figure 4.7, respectively.
Table 4.3: The estimation of $\beta$ coefficients for the meat data.

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Aroma 1</th>
<th>Aroma 2</th>
<th>Visual 3</th>
<th>Visual 5</th>
<th>Taste 1</th>
<th>Taste 2</th>
<th>Aftertaste 1</th>
<th>Aftertaste 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}$</td>
<td>14.3740</td>
<td>-0.0909</td>
<td>0.1527</td>
<td>-0.0575</td>
<td>0.0755</td>
<td>-0.2135</td>
<td>0.1355</td>
<td>-0.0544</td>
<td>0.2065</td>
</tr>
</tbody>
</table>
The attributes Aftertaste 2, Aroma 2, Taste 2, and Visual 5 have positive coefficients in decreasing order of magnitude. Therefore, they positively influence the mean of the liking score. However, Visual 3, Aftertaste 1, Aroma 1, and Taste 1 decrease the mean consumer liking score when they increase one unit. Among all of the attributes, Aftertaste 2 and Taste 1 are the most important with respect to consumer hedonic score although in opposite directions.

Figure 4.7: The mean probability of each liking score based on the four groups of consumers for the meat data.
Figure 4.7 displays the mean probability of each liking score for three products based on the five groups of consumers. We use the same clustering method, i.e., Ward’s hierarchical clustering, as used for the cupcake data to group the consumers and again consider score 5 as the reference level. Generally speaking, most liking scores are around 6 to 8, and only a few products in each group receive a score of 9. For group 1, product 2 receives scores of 8, whereas scores for product 1 are split: some consumers give a score of 8, while others give a score of 6 or lower. Product 3 also receives scores of 8; however, its mean probability is the lowest. For group 2, few products receive high score of either 8 or 9. Product 1 obtains a score of 7, and product 2 receives scores 4 and 6. Product 3 receives scores 5 and 6. For group 3, product 3 receives of score 8 and product 2 receives scores of 7. The highest score that product 1 can receive is 7. For group 4, product 2 obtains scores of 8, and both products 1 and 3 receive score 7.

4.5 Summary

In this chapter, we make a connection between the sensory data from panelists and the liking score data from a number of consumers by constructing a mean function of the conditional multivariate normal distribution given the sensory data for all products. The ECM algorithm is used to estimate the coefficients, which explains the influence on the mean liking score for each attribute. Based on the conditional estimation at the E-step, we perform preference mapping using a circular model and find the ideal region for both simulated and real data (i.e., cupcake) data sets. Such
results can be obtained due to the large sample size within the estimated sensory data set. We also implement multinomial logistic regression for all data sets to obtain plots that illustrate the mean probability for each liking score based on the products and grouped consumers. These plots also allow consumer segments to be detected.
Chapter 5

Conclusions and Future Work

In this thesis, we developed a mixture model-based clustering technique using multivariate data to matrix data. This technique concurrently considered the variation of the products and sensory panelists. Based on a simulation study, the matrix normal mixture model, four parsimonious models, and the mixture of factor analysis models are very able to cluster the attribute data. For the real data set (i.e., white bread data), we found that the parsimonious Model 2: \( \hat{\Omega} = 1\beta' \), which assumes mean matrix \( \Omega \) has identical values within each column but varying values among the columns, demonstrated clustering results similar to the full model. Therefore, we recommend product researchers use this model for the study of white bread data. Moreover, the mixture of factor analysis provides a way to observe the distributions of panelists and products after clustering the attributes. By creating factor plots for each group, product researchers will be able to not only observe the variation of panelists and products among the groups, but also learn about similarities and difference within a group. This valuable information can help product developers create strategies to meet consumer needs with respect to groups of attributes.

Moreover, to meet consumer needs and maximize their liking, we made a connection between the sensory data collected from panelists and the liking score
data rated by a number of consumers. By constructing a mean function of the conditional multivariate normal distribution given the sensory data for all products, we are able to explain the influence on the mean liking score according to estimation of the attributes’ coefficients by the ECM algorithm. Additionally, it is important for us to perform preference mapping using a circular model to find the ideal region of the product based on conditional sensory data given by each consumer’s liking. These results can be used as another reference for product developers who are accustomed to only using sensory data from trained panelists. The multinomial logistic regression method was able to illustrate the probability for each liking score based on the products and grouped consumers. These plots allow consumer segments to be detected. Application of the method to the cupcake and meat data sets showed that we were able to link the sensory data and liking score data and provide constructive suggestions for product researchers.

In future research on mixture model based clustering, I will consider developing parsimonious models by using eigenvalue decomposition of the group covariance matrices $\Sigma_1$ and $\Sigma_2$ to reduce the estimated parameters. To study the connection between sensory data and liking data, I will use a quadratic mean function of the conditional multivariate normal distribution to investigate the relationship between consumer liking and attribute intensity, and also consider the addition of correlations among the products, so the covariance of the conditional multivariate normal distribution $\mathbf{D}$ is not represented as a diagonal matrix. This can allow us to observe the effect on the mean liking scores from related products or alternative products by our connecting model.
Bibliography


deboned at two, four, six, and twenty-four hours postmortem. *Poultry Science* 83, 101–108.


Wang, Y., Y. Liu, Y. Zhang, and Z. Xu (2011). Quality evaluation of mixed


