Individualized Garment Recommendation System for Online Shopping

TERUJI SEKOZAWA
Department of Information Systems Creation, Kanagawa University,
Yokohama 221-8686, JAPAN
sekozawa@ie.kanagawa-u.ac.jp
http://www.is.kanagawa-u.ac.jp/

Abstract: This study proposes a system that supports online purchases of garments with functions to select and recommend garments on the basis of the customer’s tastes. The system has three basic capabilities: (1) It analyzes the customer’s tastes using the analytical hierarchy process (AHP) and selects and suggests clothes. (2) It designates a priority for suggesting the garments using correlations found by clustering. (3) It suggests second purchases of items that have been chosen by previous customers with similar tastes who bought a garment that the current online customer has just decided to buy. The second purchases are selected using market basket analysis.

Key Words: Analytical hierarchy process, Online shopping, One-to-one, Recommendation system, Coordination, Market basket analysis

1 Introduction

The only way the customer can judge an item on display is to read the descriptions on the computer monitor (for example: type of apparel; color; price; brand; quality label; size) and look at the image of the item. Customers who do not have full knowledge about fashion often find themselves unable to find any among the great number of possible purchases that match well with their own tastes. Others find it impossible to make up their minds among the choices they like and give up on the purchase. Others make purchases, but are still uncertain whether they have bought something that really suits them.

The aim of this study is to offer basic functions to resolve the above issues in online shopping. Specifically, the objectives are to (1) analyze a buyer’s tastes using AHP, and select and offer pertinent clothes items; (2) create a cluster of apparel based on their correlation with a buyer’s tastes, and prioritize the items in the cluster; and (3) suggest garments bought by other buyers with similar tastes, on the basis of market basket analysis.

2 Outline of One-to-one Recommendation System

Figure 1 is a diagram of the system. As already described, the essential procedures in the function of this system are: (1) AHP; (2) cluster analysis; and (3) market basket analysis. These were used in combination. Also, databases for (a) apparel attributes and (b) simultaneous purchases were used with the three procedures above. Databases (a) and (b) in Fig. 1 are described in Section 2, and examples of the application of procedures (1), (2) and (3) above are described in Sections 3, 4 and 5, respectively.

Fig. 1: Overall structure of online system

2.1 Databases required by the system
(a) Database of apparel (product) attributes
This is shown in (a) of Fig. 1. Each apparel type was entered with its digitized assessment scores before the experiment. One standard of these criteria, for example, was Design. The score was entered in absolute values as follows: 9, Very Good; 7, Good; 5, Average; 3, Bad; 1, Very Bad. Thus, the database was constructed with nine-level rankings of all the clothes according to their attributes.

(b) Simultaneous purchase database
A diagram of this database is presented in Fig. 1(b).
This database is a tabulated record of all apparel items that were bought together on any date, after analysis of the apparel items. The database can be used to calculate the incidence of purchase of items and the confidence that they will be bought with something else. Association rules can be constructed to anticipate what combinations customers are most likely to select.

2.2 Outline and procedure of analysis
(1) Grasping the customer’s tastes with AHP and selecting apparel
An apparel (product) database is created for digitized characterizations of clothing. The system surveys customers, analyzes their tastes with AHP procedures and searches for clothes popular with them.

(2) Cluster analysis for customer tastes
K-means clustering is applied in cluster analysis by grouping items in the database that have strong correlations with each other in order to classify popular apparel in clusters.

(3) Recommending apparel combinations
A simultaneous purchase database is created for use by the “fashion advisor.” Market basket analysis is used with the clustered apparel items to recommend apparel with a high degree of confidence that the customer will like, while the customer is buying a part of a combination.

3 Digitization of customer tastes with AHP
One feature of AHP is its capability to quantify the vague elements in human perceptions and assign numerical values for decisions. In this research, AHP was applied for digitization of human tastes and objective judgments.

3.1 Hierarchical structure for searching for popular attire
AHP allows the user to assess apparel from a wealth of different angles, and on this basis, to construct a system capable of searching for the optimal attire. Figure 2 shows the hierarchical structure that was created. Level 1 is the general “Apparel customers like” objective, and Levels 2 and 3 include assessment standards. The lowest, Level 4, has suggestions for alternative purchases, including Cut and Sewn, Shirts, Pants, and Jackets.

3.2 Analysis of priority of assessment standards by AHP
The priority of each element in the assessment standards is determined. First, the essential assessment standards on Level 2 of the hierarchical structure are compared with each other and analyzed. Customers are surveyed for their own priorities among the four assessment standards on Level 2, and the results are analyzed.

The paired-comparison matrix in Table 1 shows the results of the survey after multiplication by weighting vector W. For this customer, the most important factor was Factor 1 System (0.558), followed by Sensitivity (0.263), Design (0.122) and Silhouette (0.057).

The highest-valued eigenvalue of the paired comparison matrix for the standards provided by the survey was $\lambda_{\text{max}} = 4.118$, and the consistency index CI was 0.039 < 0.1. This value was significant. CI is defined in Eq.(1).

$$\text{C.I.} = \frac{\lambda_{\text{max}} - n}{n - 1}$$ (1)
Table 2 shows customers’ evaluations when there were three assessment standards for System. The reader can see that this customer’s rating of the factors was Mode, Casual, Individuality, in diminishing order of importance. This evaluation had a CI under 0.1, so it can be considered valid.

Let us analyze a survey of customers for the importance of the third-level factors in Sensitivity. The factors in Design and Silhouette were also prioritized.

Table 2: Paired analysis of three factors in System (evaluating priority)

<table>
<thead>
<tr>
<th></th>
<th>Mode</th>
<th>Casual</th>
<th>Individuality</th>
<th>(Weight)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>0.633</td>
</tr>
<tr>
<td>Casual</td>
<td>1/3</td>
<td>1</td>
<td>3</td>
<td>0.26</td>
</tr>
<tr>
<td>Individuality</td>
<td>1/5</td>
<td>1/3</td>
<td>1</td>
<td>0.106</td>
</tr>
<tr>
<td>$\lambda_{\text{max}}$</td>
<td>3.039</td>
<td>0.019</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 3.3 Procedure for absolute evaluations

Let us turn to the set of articles from which alternative purchases could be suggested, the fourth level of hierarchy. When the CI exceeded 0.1, the values in the paired comparison matrix had to be re-confirmed. However, it is difficult to identify which pair violates consistency if the set of clothes (to put it another way, the set of possible recommendations) contains 160 articles. In addition, when another item is added to the database, the paired comparison must be re-run. Thus, there are problems in the conventional method for comparative assessment.

Absolute assessments were employed instead of comparative assessments. This method compares standards, rather than comparing specific clothing articles according to the standards.

**Procedure 1: Setting absolute assessment standards**

The absolute standards are as follows: Very good; good; average; bad; and very bad. Table 3 shows these standards.

### Table 3: Absolute assessment standard

<table>
<thead>
<tr>
<th></th>
<th>Very good</th>
<th>Good</th>
<th>Normal</th>
<th>Bad</th>
<th>Very bad</th>
<th>(Weight)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very good</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>0.503</td>
</tr>
<tr>
<td>Good</td>
<td>1/3</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>0.26</td>
</tr>
<tr>
<td>Normal</td>
<td>1/5</td>
<td>1/3</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>0.134</td>
</tr>
<tr>
<td>Bad</td>
<td>1/7</td>
<td>1/5</td>
<td>1/3</td>
<td>1</td>
<td>3</td>
<td>0.068</td>
</tr>
<tr>
<td>Very bad</td>
<td>1/9</td>
<td>1/7</td>
<td>1/5</td>
<td>1/3</td>
<td>1</td>
<td>0.035</td>
</tr>
<tr>
<td>$\lambda_{\text{max}}$</td>
<td>5.543</td>
<td>0.061</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The following evaluation vector was normalized to the maximum eigenvalues of the assessment standard comparison matrix:

$$ u^T = (0.503, 0.260, 0.134, 0.068, 0.035) \quad (2) $$

**Procedure 2: Assessment matrix**

$S_{ij}$ shows assessed value for alternate garment $j$ with respect to standard $i$. $S_{ij}$ is defined as the assessed value $a_{ij}$ for alternate garment $j$ with respect to standard $i$ is divided by the maximum assessed value $a_{\text{max}}$.

$$ S_{ij} = \frac{a_{ij}(\text{garment } j \text{ value by standard } i)}{a_{\text{max}}(\text{maximum value under standard } i)} \quad (3) $$

**Procedure 3: Overall assessed value**

The survey results shown in <3.2> were processed with AHP and indicated the following weighting vectors for the assessment standards (four elements), System (three elements), Sensitivity (three elements), Design (three elements) and Silhouette (three elements).

$$ w_1^T = (0.558, 0.263, 0.122, 0.057) $$

$$ w_2^T = (0.633, 0.260, 0.106) $$

$$ w_3^T = (0.633, 0.260, 0.106) $$

$$ w_4^T = (0.633, 0.260, 0.106) $$

$$ w_5^T = (0.057, 0.295, 0.649) $$

$w_1^T$ corresponds to the second level of the hierarchical structure, and $w_2^T - w_5^T$ correspond to the third level.

The above shows how the relative weights were assigned between assessment standards and to the standards themselves. The assessment matrix was found in Procedure 2, so the overall evaluation scores for the clothing articles to be suggested are calculated as follows:

$$ E_i = S_{ij} \cdot W \quad (5) $$

In absolute comparisons, the evaluation vector $u^T$ is identified, the attire attributes database is established, and the assessment matrix is found with Eq.(3) above. This is calculated using the assessment matrix instead of the relative comparison weighting matrix.

Figure 3 provides an example of the results of this kind of analysis. It is graphs of three clothing articles selected by weighting for cut and sewn.

The M-4070 in the graph means that it is item #4070 from “Cut and sewn.” The reader can see from this figure that all three samples have high weights in System, a standard important to this customer. These
have high weights in Mode, Avant-garde, Elegant Design, and Slim.

4 Classification of tastes and recommending attire on the basis of cluster analysis

Clothes items were grouped into clusters using the clothing attributes database. The intent of this clustering was to assemble groups of clothes that are likely to appeal to a wide variety of people’s tastes. Clothes suiting customers’ likes (clothes with high overall scores) are classified into clusters on the basis of the items suggested by AHP. Cluster analysis also makes it easier to apply market basket analysis effectively, as described below. This generates suggestions for garments coordinated with other purchases that also meet the customer’s tastes.

4.1 Clustering using K-means

This study employed K-means clustering, a partitioned optimization procedure within nonconfigurational classifications. In K-means clustering\(^9\), the cluster centroid \(c_i\) is considered the representative point and the set of data is divided into \(k\) clusters by minimizing the following evaluation equation:

\[
J = \sum_{i=1}^{k} \sum_{x \in c_i} (D(x, c_i))^2 \tag{6}
\]

\(x\): Article of clothing classified into cluster \(D(x, c_i)\): Distance between centroid \(c_i\) and attire \(x\)

4.2 Clustering results

Clustering analysis is performed with the clothes attributes database in order to group articles of clothing that can correlate with each other in any of many different ways. The examples shown here were obtained using clustering about the System criterion and setting the number of clusters at 12. The cluster numbers are set automatically by the system in the order of increasing sample number, and the system can be set to create clusters of uniform size.

The number in the table shows to which cluster the item belongs. Each cluster is displayed in a simple manner, and it is easy to display what characteristic a cluster has in common.

Table 4: Combinations of similar clothing items

<table>
<thead>
<tr>
<th>Cluster No.</th>
<th>Number of samples</th>
<th>Mode</th>
<th>Casual</th>
<th>Individuality</th>
<th>Avant-garde</th>
<th>Elegant</th>
<th>Conservative</th>
<th>Simple</th>
<th>Plump build</th>
<th>Normal</th>
<th>Slim build</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-40</td>
<td>3</td>
<td>2.30</td>
<td>9</td>
<td>7</td>
<td>3.30</td>
<td>3</td>
<td>2.30</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>C-200</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>C-4-0</td>
<td>3</td>
<td>1</td>
<td>3.88</td>
<td>8.33</td>
<td>8.33</td>
<td>3.30</td>
<td>2.33</td>
<td>4.30</td>
<td>7.86</td>
<td>1.86</td>
<td>4.32</td>
</tr>
<tr>
<td>C-0-0</td>
<td>3</td>
<td>1</td>
<td>3.87</td>
<td>8.42</td>
<td>8.42</td>
<td>3.31</td>
<td>2.42</td>
<td>4.42</td>
<td>7.37</td>
<td>1.85</td>
<td>4.42</td>
</tr>
</tbody>
</table>

Fig. 3 Examples of overall scoring of three clothing articles
4.3 Prioritizing merchandise for recommendation to the customer

Clothes items bearing certain similarities were grouped into clusters, as described in the previous section, as a way of organizing them in accordance with a wide variety of customers’ tastes. The items in each cluster were then prioritized to plan the appropriate order to present them to the customer. Ordinarily, however, the first natural groups for selecting clothes are Cut and Sewn, Shirts, Pants and Jackets. Here, in order to avoid the bother of selecting names to apply to new clusters and keep things easy for the customers to understand, the clothes are displayed using the above classifications. The system can recommend articles selected to meet the customer’s tastes in order of priority, as shown in Table 5.

The priority is assigned, beginning with the item having the highest weighting. In the above example, the recommended items were the M-3070 cut and sewn, the S-1050 Pants, the V-3100 Jacket and the G-2020 Shirt.

Table 5: Priorities set for clothes items

<table>
<thead>
<tr>
<th>Customer's priority</th>
<th>1st place</th>
<th>2nd place</th>
<th>3rd place</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cut and sewn</td>
<td>M-3070</td>
<td>M-2020</td>
<td>M-1040</td>
</tr>
<tr>
<td>Pants</td>
<td>S-1050</td>
<td>S-3010</td>
<td>S-4100</td>
</tr>
<tr>
<td>Jacket</td>
<td>V-3100</td>
<td>V-3040</td>
<td>V-2020</td>
</tr>
<tr>
<td>Shirt</td>
<td>G-2020</td>
<td>G-3040</td>
<td>G-3100</td>
</tr>
</tbody>
</table>

5 Coordination of attire with market basket analysis

In market basket analysis, the “fashion advisor” recommends items to the current customer that other customers with similar tastes have bought. This form of recommendation is an essential role of the human sales staff in a brick-and-mortar store. A virtual sales girl on an online shopping site can perform the same task, using market basket analysis.

Here, to attain the goal of coordinating attire, the clothes are previously sorted into the above four clusters of Cut and Sewn, Shirts, Pants, and Jackets. It is investigated how to suggest clothes in combinations that will suit the client.

5.1 Issues and Policies in the creation of purchase databases

An attire attributes database is used to create a database that allows calculation of the probability that two items will be bought at the same time, and this is essential when applying market basket analysis. Performing market basket analysis with the simultaneous purchase database allows the seller to suggest attire besides the item that was just purchased that has a high potential of sale.

Association rules are constructed to calculate the rate of appearance of an item in the purchase database and the confidence that it will be bought along with something else.

5.2 Construction of association rules

Association rules are constructed using clothes items that have been clustered by many applicable characteristics and gathered into a simultaneous purchase table in order to calculate the probability of simultaneous purchases. An intermediate product such as a simultaneous purchase table can provide information about which combinations of garments appear most often in transactions.

The following shows how a simultaneous purchase table and association rules are created.

“Confidence” in Table 6 describes how high the probability is that the customer will buy any of the pants (S-1010 – 1030) after buying V-1010 and G-1010.

Table 6: Simultaneous purchase table (partial) and confidence levels

<table>
<thead>
<tr>
<th>Combination</th>
<th>Incidence of preceding garment and following garment</th>
<th>Incidence of preceding garment</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>V-1010</td>
<td>G-1010 S-1010</td>
<td>0.00966667</td>
<td>0.0366667</td>
</tr>
<tr>
<td>V-1010</td>
<td>G-1010 S-1020</td>
<td>0.00966667</td>
<td>0.0366667</td>
</tr>
<tr>
<td>V-1010</td>
<td>G-1010 S-1030</td>
<td>0.01</td>
<td>0.0366667</td>
</tr>
</tbody>
</table>

“Incidence” is the quantity obtained by dividing the total number of purchases of the given item simultaneously with some other item by the total number of items purchased at the same time as any other garment. “Incidence of sales of the preceding garment” means the incidence of sales of the combination of V-1010 and G-1010. “Incidence of simultaneous sales of the preceding and following garment” means the incidence of sales of all three items, V-1010, G-1010 and S-1010–1030.

The confidence is found in the following calculation:
The calculated confidences for all combinations of preceding garment purchases were used to establish the following example of the association rules:

(i) 「V-1010」 & 「G-1010」 → 「S-1030」 「S-1010」
(ii) 「V-1010」 & 「S-1010」 → 「M-1010」 「G-1020」
(iii) 「V-1020」 & 「S-1020」 → 「M-1010」 「G-1030」

It is possible to put the confidence values in numerical order as well, so a rule was written to recommend the most likely (highest confidence value) and, if necessary, the second most likely item for additional sale.

6 Conclusions

The authors wrote system to select and recommend garments on the basis of the customer’s tastes, and proposed a solution to the problems on the conventional online system. Currently, a customer wishing to purchase a garment online must read the descriptions and look at the images of clothes and rely completely on her own subjective judgment to find and select the garment. This system aids the customer in selecting garments and also makes recommendations.

1) Digitization of tastes using AHP

People’s tastes and perceptions of clothes must be expressed in the form of numbers. AHP was used to accomplish this. An AHP-based data search system was used to create a database of the attributes of garments, and this was used to perceive individual customers’ tastes and to suggest garments they were inclined to like.

2) Classification of tastes and recommendations of clothes on the basis of cluster analysis

Clustering of the data in the clothes attributes database allows the system to respond appropriately to customers of many different preferences. Clothes fitting customers’ tastes can also be classified into clusters consisting of the items recommended by AHP. By using K-means clustering, the number of items to be examined can be reduced to a manageable level. Alternative garments can be suggested in order of priority, within the cluster of garments meeting the customer’s preferences.

3) Coordination of garments by market basket analysis

Market basket analysis allows the system to suggest clothes to a customer on the basis of what has been bought by other customers who have similar tastes. A complete database must be prepared of garments clustered on the basis of their characteristics. This is used to calculate the probability of simultaneous purchase of pairs of garments, and this calculation is applied to suggest potential second sales to customers.

References: