Dynamic Visual Analytics on Point-of-Sales Data

Yuho Chung
Lingnan University
Hong Kong

Kin-nam Lau
The Chinese University of Hong Kong
Hong Kong
Dynamic Visual Analytics on Point-of-Sales Data

ABSTRACT

We propose a dynamic visual model to visualize POS transaction data on a map. This map has two features. First, we represent products and customer segments from a high-dimensional space as points on a 2D space without errors. Second, we can conduct customer segmentation and product association simultaneously, thereby allowing marketers to visualize complete results on a single map and apply these results to marketing campaigns. We apply the model to the temporal POS data set (28,092 customers and 25 products) of a supermarket over a period of four months. Thereafter, we illustrate the different applications and report the results.

Keywords: visual analytics; segmentation; market basket analysis; temporal visual model
1. INTRODUCTION

Temporal data are common in marketing. A retail Point-of-Sales (POS) system in a supermarket records every piece of information about a transaction (e.g., membership ID, transaction date and time, and price of purchased products) and keeps track of sales performance over time. It also identifies the products purchased by shoppers and allows managers to evaluate the effectiveness of marketing campaigns and special promotions offered at some point in time. The marketing applications of POS data are extensively discussed in literature (Cachon & Kök, 2007; Garber, Goldenberg, Libai, & Muller, 2004). Cachon and Kök (2007) used the sales data of basket-shopping customers to predict basket profits and then deduce optimal assortment planning decisions. Garber et al. (2004) applied the spatial divergence approach to POS data to predict the success of a new product during the early period of its launch.

Visualization has become an important component of communication in both business and society. A visual image can immediately create a picture of a particular knowledge, as well as communicate more information than a table and text within a considerably limited space. These features make visual display more efficient and effective than table and text in presenting data. A visual aid (e.g., a map) can also improve comprehension (Lurie & Mason, 2007) and allow users to discover interesting patterns that will otherwise be difficult to observe using standard statistical methods (Kolata, 1982). Visual analytics has recently emerged as among the popular research areas in computer science (Simoff, Böhlen, & Mazeika, 2008) and statistics (Rao, Wegman, & Solka, 2005). Numerous data visualization

---

1 Visualization can be classified into data visualization and artistic visualization. The objective of data visualization is to convey abstract information in an intuitive manner to improve comprehension and communication. Meanwhile, the objective of artistic visualization is to present data in a manner that appeals to the audience, thereby increasing interest in the data being presented. In this paper, visualization refers to any method in which a massive amount of data is graphically represented into meaningful images to enhance understanding and dissemination.
models such as principal component analysis (PCA), multidimensional scaling (MDS), correspondence analysis (CA), and multidimensional unfolding (MDU) have been proposed in the literature. The visualization of cross-sectional data is challenging in itself. Dynamic visualization is of particular interest because marketing-mix data (e.g., promotion and price) always evolve or change over time. As the popular saying goes, a picture paints a thousand words; thus, a map of products and customers saves millions of words regarding their interaction (Kamakura & Moon, 2014; Kolata, 1982). Therefore, the first research question is as follows:

*How can we transform dynamic POS data into a display of products and customers on a map for managers to visualize their interactions cognitively?*

This question involves two challenges. First, when sample size is large, simultaneously computing the coordinates of voluminous customer and product points can be difficult. Second, when we represent high-dimensional POS data on a 2D space, the map can be distorted due to serious dimension reduction errors. The map in the reduced space may not accurately reflect the actual data. Therefore, our second research question is as follows:

*How can we solve the dimensionality problem to represent a massive number of products and customers on a map without dimension reduction error?*

To address these two research questions, we propose a dynamic visual analytics model with two assumptions. We use a point to represent a “customer segment” instead of a “customer” to solve the data scalability problem. We also allow overlapping segments to solve the dimensionality problem; thus, a customer may belong to multiple segments. The model allows marketers to conduct customer segmentation and product association simultaneously. We present a new approach for customer segmentation in POS data, and apply matrix
factorization to estimate the number of segments, determine the features of each segment, and select the segment membership of each customer simultaneously. We can use nodes with different colors to indicate the sizes of various segments. We also use directed or undirected links to represent different types of product relationships.

2. LITERATURE REVIEW

2.1. Static Visual Analytics

MDU is a popular data visualization technique to simultaneously display products and customers on a map (Bijmolt & Wedel, 1999; DeSarbo, Grewal, & Scott, 2008). To illustrate the MDU model, a case of \( I \) customers and \( J \) products is considered. Let \( Z = [z_{ij}] \) be the \( I \times J \) purchase matrix, where \( z_{ij} \) is the choice indicator that is defined as follows:

\[
z_{ij} = \begin{cases} 
1 & \text{if customer } i \text{ chooses product } j \\
0 & \text{otherwise}
\end{cases}.
\]  

(1)

DeSarbo and Hoffman (1987) assumed that the probability that customer \( i \) would choose product \( j \) (denoted as \( \pi_{ij} \)) was negatively related to their distance as follows:

\[
\pi_{ij} = \frac{1}{\Pr(z_{ij} = 1)} = \frac{1}{1 + \exp[-(r - d_{ij})]} \]

(2)

where \( r \) is a scaling parameter and \( d_{ij} \) is the distance between customer \( i \) and product \( j \) on the map. The coordinates for products and customers are estimated by maximizing the joint likelihood function of \( z_{ij} \) as follows:

\[
\max_{r, z_{ij}, \pi_{ij}} L = \prod_{i=1}^{I} \prod_{j=1}^{J} \left[ (\pi_{ij})^{z_{ij}} (1 - \pi_{ij})^{1-z_{ij}} \right].
\]

(3)

The joint display is attractive, but the MDU solution is problematic when sample size is large. We encounter serious computational difficulties because POS data is large with many customers and products. Therefore, most MDU applications are limited to a relatively small sample (Borg & Groenen, 2005). In addition, the purchase vector \( (z_1, \ldots, z_J) \) is defined in \( J\)-
dimensional space. MDU attempts to represent all \( I \) customers from \( J \)-dimensional space into a 2D space. Accurate 2D results that can reflect \( J \)-dimensional information is not guaranteed, particularly when \( J \) is large, because of the curse of dimensionality reduction problem. Moreover, even if MDU results are accurate, the map is crowded with customer points and not customer insights.

2.2. Dynamic Visual Analytics

In temporal data visualization literature, researchers have proposed several methods to handle temporal data.

1. Temporal-aggregation Approach

The simplest method to handle time-series data is to aggregate the temporal component over time. The resulting data contain a customer-by-product matrix, and the time component is removed in the analysis. We can then apply MDU to the aggregate data and study the relationship between customers and products. However, the aggregation approach has several drawbacks. Wei (1994) conducted an extensive survey on the effect of aggregation and found that a substantial loss of information resulted from aggregation in parameter estimation. With the help of a series of Monte Carlo simulations, he also showed that as the order of temporal aggregation and the number of parameters in the model increase, information loss becomes increasingly serious in the estimation process. That is, the resulting joint space map between customers and products may not reflect their actual relationship. Moreover, the map is a biased representation, which leads to misinformed decision making.

2. Multiple-snapshot Approach

Temporal data can be regarded as a collection of snapshots or a combination of multiple
cross-sectional data. We can split the time-series data into \( T \) equal time intervals and apply the traditional MDU method to each of these cross-sectional data. For example, if we observe the transactions records of \( I \) customers and \( J \) products for \( T \) consecutive periods, then we can divide the temporal data into \( T \) separate data sets, each with dimension \( I \times J \). We can then use a MDU technique to create \( T \) independent visualization maps that individually represent the intermediate relationships between products and customers at each time interval. However, this approach may create \( T \) separate and unconnected maps without linkage between them, which leads to serious interpretation problems.

3. Scale-free Graphical Approach

Many graph-drawing approaches, such as the force-directed method (Kobourov, 2014), and visualization methods, such as spectral layouts (Koren, 2003), are solely developed to create a 2D aesthetic representation of graphs without a natural representation for Minkowski distance (e.g., Euclidean distance) in a map (Herman, Melançon, & Marshall, 2000). The visualization graph is scale-free, and the positions of objects in the graph are meaningless. For example, Brandes and Corman (2003) created a pseudo-3D graph that regarded the 2D layouts of each period as layers in the stack. However, the resulting visualization graph is difficult to read and interpret. A better approach is to present an animated 2D spectral layout sequence that evolves over time to represent changes in different snapshots (Leydesdorff & Schank, 2008). The challenge in this approach is to preserve the mental map between snapshots to enable users to visualize and interpret the animation easily. We compare the three aforementioned approaches for temporal data visualization in Figure 1.
Figure 1 Dynamic visualization approaches

- **Temporal Aggregation Approach**
  - J
  - I
  - t = 1
  - t = 2
  - t = T

- **Multiple Snapshot Approach**
  - J
  - I
  - t = 1
  - t = T

- **Free-scale Graphical Approach**
  - J
  - I
  - t = 1
  - t = T

Aggregated map

3D layout

2D layout with animation

...
3. DYNAMIC VISUAL ANALYTICS MODEL

3.1. Model Framework

We consider a POS data set of \( I \) customers, \( J \) products, and \( T \) transaction periods. Customers can freely choose any products among \( J \) products at time \( t \), where \( t = 1, \ldots, T \). The transactions of a customer can be regarded as a customer sequence, in which each transaction consists of a list of purchased products. The list of transactions is ordered by transaction time from time 1 to time \( T \). For example, customer \( i \) can be described as a series of vectors \( \{Z_{i1}, Z_{i2}, \ldots, Z_{iT}\} \), where \( Z_{it} \) represents the purchase set of customer \( i \) at time \( t \). Formally, let the transactions of customers be \( Z_1, Z_2, \ldots, Z_T \). We observe the matrix of customer purchases \( Z_t = [z_{ijt}] \) at transaction time \( t \), where \( z_{ijt} = 1 \) if customer \( i \) purchases product \( j \) at time \( t \) and 0 otherwise. We also define \( S_t \) and \( C_t \) as the matrix of segment features and the matrix of segment memberships at time \( t \), respectively. The objective is to unfold segment features (i.e., \( S_t \)) and cluster customers into overlapping segments (i.e., \( C_t \)) simultaneously by factorizing POS data at time \( t \) (i.e., \( Z_t \)) into a product of binary matrices \( C_t \) and \( S_t \), as follows:

\[
Z_t = C_t S_t. \tag{4}
\]

\( C_t = [c_{ikt}] \) is the matrix of segment membership at time \( t \), and \( S_t = [s_{kjt}] \) is the \( t \)th period segment feature matrix that consists of \( K \) rows (i.e., \( K \geq J \)), with each row corresponding to the characteristics of the segment. The columns consist of \( J \) products, and thus, feature indicator \( s_{kjt} = 1 \) if all members in segment \( k \) purchase product \( j \) and 0 otherwise. The factorization in Equation (4) is error-free. Therefore, matrix \( Z_t \) can be perfectly explained by \( C_t \) and \( S_t \). However, \( C_t \) and \( S_t \) have multiple solutions that satisfy Equation (4). Therefore, we impose condition to derive a meaningful solution.
1. Minimum Segment Membership for each customer

Although we allow overlapping segments, we intend to minimize unnecessary overlaps. For each customer \( i \), we prefer to use the minimum number of segments to explain \( Z_t \) perfectly as follows:

\[
\min n_t = \sum_{i=1}^{I} \sum_{k=1}^{T} c_{ikt} \quad \text{for } t = 1, \ldots, T
\]

subject to \( Z_t = C_t S_t \),

where \( n_t \) is the number of segment memberships at time \( t \), and the membership indicator \( c_{ikt} = 1 \) if customer \( i \) is assigned to segment \( k \) at time \( t \) and 0 otherwise.

2. Non-trivial Solution

However, Equation (5) may lead to a trivial solution that the segment-feature matrix is equal to the customer-purchase matrix (i.e., \( S_t = Z_t \)) and that each customer belongs to only one segment. In this case, the objective function is \( n_t = I = K_t \), where \( K_t \) is the number of segments at time \( t \). To rule out this trivial solution, we restrict the number of segments to less than the number of customers (i.e., \( K_t < I \)). Therefore, we rewrite Equation (5) as follows:

\[
\min n_t = \sum_{i=1}^{I} \sum_{k=1}^{T} c_{ikt} \quad \text{for } t = 1, \ldots, T
\]

subject to \( Z_t = C_t S_t \) and \( K_t < I \).

3.2. Model Formulation

The objective of dynamic visual analytic model is to jointly visualize customers and products on the 2D map. We split the estimation process into two stages. In stage 1, we determine the product positions for each time period using the autologistic regression. Based on product positions, we simultaneously determine the number of
segments \((K_t)\), segment memberships \((C_t)\), features of segments \((S_t)\), and the corresponding segment positions on the 2D map in stage 2. This approach is called as External Unfolding in literature (Borg & Groenen, 2005).

### 3.2.1. Dynamic Product-positioning Map

We consider a case with \(I\) customers, \(J\) products, and \(T\) periods. For each time transaction \(t\), customer \(i\) can freely choose any products among \(J\) products. Let \(z_{it}(k_t) = (z_{i1}(k_t),\ldots, z_{iJ}(k_t))\) be the \(k_t\)th pattern that customer \(i\) chose at time \(t\). The product choice indicator \(z_{ij}(k_t) = 1\) if pattern \(k_t\) contains product \(j\) and 0 otherwise. Let \(U(z_{it}(k_t))\) be the total utility associated with \(z_{it}(k_t)\) as follows:

\[
U(z_{it}(k_t)) = V(z_{it}(k_t)) + e_{ikt} \quad \text{for } t = 1,\ldots, T; \tag{7}
\]

\[
V(z_{it}(k_t)) = \sum_{j=1}^{J} \alpha_{jt} z_{ij}(k_t) + \sum_{j'=1}^{J} \sum_{j'=j}^{J} \beta_{j'j} z_{ij}(k_t)z_{ij'}(k_t) + \sum_{j=1}^{J} \sum_{j'=1}^{J} \sum_{j'=j}^{J} \gamma_{j'j-1,j} z_{ij}(k_t)z_{ij'}(k_t) \tag{8}
\]

where \(V(z_{it}(k_t))\) is the deterministic utility, \(\alpha_{jt}\) is the average utility per unit of product \(j\) at time \(t\), and \(e_{ikt}\) is the error term. The deterministic utility is further divided into two parts: \(\beta_{j'j}\), which represents the co-utility from the joint purchases of one unit of product \(j\) and one unit of product \(j'\) at time \(t\); and \(\gamma_{j'j-1,j}\), which denotes the lagged co-utility from the joint purchases of one unit of product \(j'\) at time \(t-1\) and one unit of product \(j\) at time \(t\). The former measures the association of the joint purchases of two products at the same period. It is concerned with intra-product association. The latter identifies the effect of introducing lagged purchased products. It is concerned with inter-product association. Therefore, the probability of customer \(i\) choosing pattern \(k_t\) is

\[
\Pr(z_{it}(k_t)) = \exp(V(z_{it}(k_t)))/\sum_{l=1}^{2^J} \exp(V(z_{it}(l))). \tag{9}
\]

According to Besag (1974), the multivariate logistic distribution can be approximated
by the full conditional distribution of the single binary variable \( z_{ijt} \).

\[
\Pr(z_{ijt} \mid z_{i<jt}, z_{j<i-1}) = \left[ 1 + \exp \left[ -(\alpha_{jt} + \sum_{j'=1}^{J} \beta_{j'jt} z_{j'jt} + \sum_{j'=1}^{J} \gamma_{j'jt-1,jt} z_{j'jt-1}) \right] \right]^{-1},
\]

(10)

where \( z_{i<jt} = (z_{it}, \ldots, z_{jt}) \) and \( z_{j<i-1} = (z_{jt}, \ldots, z_{j(i-1)}) \). The parameters of Equation (10) can then be determined using a pseudo-likelihood estimation. We maximize the pseudo-likelihood estimation using the autologistic regression framework and link co-utility to product positions.

\[
d^2_{jt,j't} = \delta_s - \beta_{j't},
\]

(11a)

\[
d^2_{j't-1,jt} = \delta_L - \gamma_{j't-1,jt},
\]

(11b)

\[
d^2_{j,p,j't} = (x_{pjtt} - x_{pj't})^2 + (y_{pjtt} - y_{pj't})^2;
\]

(11c)

\[
d^2_{j't-1,jt} = (x_{pjtt} - x_{pj't-1})^2 + (y_{pjtt} - y_{pj't-1})^2;
\]

(11d)

\[
\sum_{j=1}^{J} x_{pjtt} = \sum_{j=1}^{J} y_{pjtt} = 0 \quad \text{for } t = 1, \ldots, T;
\]

(11e)

where \((x_{pjtt}, y_{pjtt})\) and \((x_{pj't}, y_{pj't})\) are the coordinates of products \( j \) and \( j' \), respectively, at time \( t \); and \( d_{jt,j't} \) is the distance between products \( j \) and \( j' \). Equations (11a) and (11b) transform co-utility into distance, where \( \delta_s \) and \( \delta_L \) are the transformation constants. Equations (11c) and (11d) define the Euclidean distance, and Equation (11e) normalizes the coordinates to ensure uniqueness.

### 3.2.2. Dynamic Customer Segment Map

In stage 2, we simultaneously determine the number of segments \( K_t \) and segment features \( S_t \) at each period. In addition, we derive segment memberships \( C_t \) and represent customer segments as segment points in the customer segment map. Given a common radius \( r \) and a product model configuration for each period, we construct product circles using product positions as the center. The region of intersection is the possible segment point positions. Suppose that \( K_t \) regions of intersection exist, with
each region having a unique purchase pattern at time \( t \). The center of the intersection region is the segment points, and thus, the distance between segment point \( k \) and product \( j \) on the map is defined as follows:

\[
d_{kj} = \sqrt{(x_{sk} - x_{pj})^2 + (y_{sk} - y_{pj})^2},
\]

where \((x_{sk}, y_{sk})\) and \((x_{pj}, y_{pj})\) are the coordinates of segment point \( k \) and product \( j \), respectively. Two types of restrictions are imposed on \( d_{kj} \) and \( r \) to interpret the visualization results as follows:

\[
d_{kj} \leq r \quad \text{if} \quad s_{kj} = 1 \quad \text{for} \quad k = 1, \ldots, K_t, j = 1, \ldots, J; \quad (13)
\]
\[
d_{kj} > r \quad \text{if} \quad s_{kj} = 0 \quad \text{for} \quad k = 1, \ldots, K_t, j = 1, \ldots, J. \quad (14)
\]

If product \( j \) is a feature of segment point \( k \) (i.e., \( s_{kj} = 1 \)), then segment point \( k \) should be close to product \( j \). Consequently, we use Equation (13) to restrict their distance to less than or equal to \( r \). By contrast, if product \( j \) is not part of segment point \( k \), then Equation (14) is used to restrict the distance to greater than \( r \). For each customer \( i \), we can construct \( n_i \) customer circles with radius \( r \) and segment points as centers. Customer \( i \) purchases product \( j \) if customer \( i \) belongs to a segment with a corresponding segment point that is located within the circle of product \( j \). By contrast, customer \( i \) does not purchase product \( j \) if customer \( i \) belongs to a segment with a corresponding segment point that falls outside the circle of product \( j \). The estimation of the aforementioned process can be estimated using iterative integer program\(^2\).

\(^2\) The detailed estimation procedure of dynamic segment map is available upon request.
4. MODEL APPLICATION

4.1. Data Description

We illustrate the dynamic visualization model using the transaction data of a supermarket in Taiwan (Lin, Hsu, & Huang, 2005). The data set was collected from November 2000 to February 2001. It consists of 254,150 transactions from 28,092 customers with 5,085 food and beverage stock keeping units (SKUs). The supermarket has a loyalty program, that is, each customer has a membership card with a unique membership ID. The membership ID can be regarded as an identifier to recover the complete purchase history of each customer. The SKUs sold in the supermarket have a wide range of categories. The customers of the supermarket can purchase SKUs of more than one category. The SKUs of one category may always be purchased when SKUs from another category are purchased. Simultaneously, they may be rarely purchased when SKUs from a different category are purchased. We aggregate SKUs into 25 product categories (hereafter referred to as products) listed in Table 1 and apply monthly customer purchase transaction data ($T = 4$) to the model on a monthly basis. Each product category is a binary variable that indicates whether an SKU in the category is purchased. SKU aggregation has at least three benefits. First, the unit sales of many SKU items are minimal. Consequently, the analysis of SKU items for few customers may not be interesting. The prevalence of low-support items in the association analysis may make deriving useful association rules difficult. Second, the analysis of 100 SKU items already involves $2^{100} = 1.26 \times 10^{30}$ combinations. Computational complexity and the curse of dimensionality problem come to play if the number of items is not limited to a manageable size. Third, even if we can compute and determine the positions of many SKU items on a visual map, visual analytics will be difficult because the map will be crowded with product points.
but will not provide insights. In addition, many supermarkets periodically run
category-based discounts on all SKUs within the same category. They sell products
through special offers to increase category purchases. We also observe the same
pattern in supermarket transaction data, that is, SKUs within the same category are
always promoted together. Monthly category shares and percentage changes in price
are reported in

Table 1.

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Category Share Percentage Changes in Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canned Vegetable</td>
<td>A</td>
</tr>
<tr>
<td>Canned Soup</td>
<td>B</td>
</tr>
<tr>
<td>Canned Meat</td>
<td>C</td>
</tr>
<tr>
<td>Noodle</td>
<td>D</td>
</tr>
<tr>
<td>Frozen Food</td>
<td>E</td>
</tr>
<tr>
<td>Condiment</td>
<td>F</td>
</tr>
<tr>
<td>Oil</td>
<td>G</td>
</tr>
<tr>
<td>Rice</td>
<td>H</td>
</tr>
<tr>
<td>Egg</td>
<td>I</td>
</tr>
<tr>
<td>Salt</td>
<td>J</td>
</tr>
<tr>
<td>Dessert</td>
<td>K</td>
</tr>
<tr>
<td>Dried Grapes</td>
<td>L</td>
</tr>
<tr>
<td>Biscuit</td>
<td>M</td>
</tr>
<tr>
<td>Peanuts and Nuts</td>
<td>N</td>
</tr>
<tr>
<td>Potato Chips</td>
<td>O</td>
</tr>
<tr>
<td>Chocolate</td>
<td>P</td>
</tr>
<tr>
<td>Confectionery</td>
<td>Q</td>
</tr>
<tr>
<td>Coffee</td>
<td>R</td>
</tr>
<tr>
<td>Herbal Tea</td>
<td>S</td>
</tr>
<tr>
<td>Juice</td>
<td>T</td>
</tr>
<tr>
<td>Soft Drink</td>
<td>U</td>
</tr>
<tr>
<td>Beer</td>
<td>V</td>
</tr>
<tr>
<td>Energy Drink</td>
<td>W</td>
</tr>
<tr>
<td>Regular Milk</td>
<td>X</td>
</tr>
<tr>
<td>Skimmed Milk</td>
<td>Y</td>
</tr>
</tbody>
</table>
4.2. Results of the Dynamic Product-positioning Map

We apply the model to the four-month ($T = 4$) pick-any purchase data set and display the dynamic product-positioning map in Figure 2 with the corresponding figures and statistics in Table 2. The transformation constants $\delta_i$ and $\delta_L$ are 1.41 and 1.62, respectively. The main utilities $\alpha_{jt}$ are proportional to the number of customers purchasing the products at each period. For example, the average price of the egg category edged up 24.4% in December, whereas sales dropped 64.3% from November to December. In February, the egg category was on sale, with the price dropping by 34.7% on average. The promotion stimulated the demand and increased sales by 218%. The main utilities reflect the sales of the egg category during this observation period from $-1.70$ in period 1 (November 2000), $-3.48$ in period 2 (December 2000), $-3.59$ in period 3 (January 2001), and $-1.36$ in period 4 (February 2001). In addition, the distance between any two products is negatively correlated with the co-occurrence count between two products in the data set. Numerous customers purchased canned vegetables, canned soups, and canned meat together. Snacks, such as potato chips, chocolates, and confectioneries, are also complements that share similar locations on the map. Five distinct groups of product categories are identified: snacks and biscuits (biscuits, peanuts and nuts, potato chips, chocolates, confectioneries, and dried grapes), canned products (canned vegetables, canned meat, and canned soups), groceries (rice, oil, desserts, and condiments), beverages (soft drinks, juices, coffee, and herbal tea), and milk (skimmed milk and regular milk).
In addition, to analyze intra-product association within each period, the dynamic product-positioning map may be employed in inter-product association analysis. This map exhibits the following features.
Table 2 Parameter estimates of the dynamic product map

<table>
<thead>
<tr>
<th>Product</th>
<th>$a_{j1}$</th>
<th>$a_{j2}$</th>
<th>$a_{j3}$</th>
<th>$a_{j4}$</th>
<th>$T = 1$</th>
<th>$T = 2$</th>
<th>$T = 3$</th>
<th>$T = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canned Vegetable</td>
<td>-4.05</td>
<td>-4.31</td>
<td>-3.70</td>
<td>-4.09</td>
<td>-0.14</td>
<td>0.69</td>
<td>-0.06</td>
<td>0.36</td>
</tr>
<tr>
<td>Canned Soup</td>
<td>-3.87</td>
<td>-4.16</td>
<td>-3.98</td>
<td>-4.01</td>
<td>0.14</td>
<td>0.64</td>
<td>0.00</td>
<td>0.65</td>
</tr>
<tr>
<td>Canned Meat</td>
<td>-2.84</td>
<td>-2.77</td>
<td>-2.64</td>
<td>-2.85</td>
<td>-0.21</td>
<td>0.59</td>
<td>-0.14</td>
<td>0.51</td>
</tr>
<tr>
<td>Noodle</td>
<td>-1.61</td>
<td>-1.75</td>
<td>-1.65</td>
<td>-1.63</td>
<td>0.02</td>
<td>0.42</td>
<td>0.17</td>
<td>0.53</td>
</tr>
<tr>
<td>Frozen Food</td>
<td>-2.17</td>
<td>-2.04</td>
<td>-1.96</td>
<td>-2.32</td>
<td>-0.54</td>
<td>0.08</td>
<td>-0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>Condiment</td>
<td>-1.81</td>
<td>-1.46</td>
<td>-1.78</td>
<td>-1.34</td>
<td>-0.32</td>
<td>0.48</td>
<td>-0.54</td>
<td>0.16</td>
</tr>
<tr>
<td>Oil</td>
<td>-2.85</td>
<td>-2.43</td>
<td>-2.57</td>
<td>-2.88</td>
<td>-0.58</td>
<td>0.49</td>
<td>-0.75</td>
<td>-0.05</td>
</tr>
<tr>
<td>Rice</td>
<td>-2.61</td>
<td>-3.03</td>
<td>-1.60</td>
<td>-2.57</td>
<td>-0.71</td>
<td>0.43</td>
<td>-0.67</td>
<td>-0.10</td>
</tr>
<tr>
<td>Egg</td>
<td>-1.70</td>
<td>-3.48</td>
<td>-3.59</td>
<td>-1.36</td>
<td>-0.90</td>
<td>-0.08</td>
<td>-0.44</td>
<td>0.55</td>
</tr>
<tr>
<td>Salt</td>
<td>-2.10</td>
<td>-1.59</td>
<td>-4.36</td>
<td>-3.77</td>
<td>-0.71</td>
<td>-0.56</td>
<td>-0.65</td>
<td>-0.55</td>
</tr>
<tr>
<td>Dessert</td>
<td>-2.35</td>
<td>-2.29</td>
<td>-2.18</td>
<td>-2.16</td>
<td>-0.58</td>
<td>-0.10</td>
<td>-0.38</td>
<td>0.33</td>
</tr>
<tr>
<td>Dried Grapes</td>
<td>-3.29</td>
<td>-2.98</td>
<td>-2.78</td>
<td>-2.90</td>
<td>0.63</td>
<td>0.52</td>
<td>0.52</td>
<td>0.43</td>
</tr>
<tr>
<td>Biscuit</td>
<td>-1.60</td>
<td>-1.83</td>
<td>-2.10</td>
<td>-2.03</td>
<td>0.51</td>
<td>0.10</td>
<td>0.52</td>
<td>0.08</td>
</tr>
<tr>
<td>Peanuts and Nuts</td>
<td>-2.33</td>
<td>-2.30</td>
<td>-1.79</td>
<td>-2.47</td>
<td>0.64</td>
<td>0.41</td>
<td>0.60</td>
<td>0.35</td>
</tr>
<tr>
<td>Potato Chips</td>
<td>-2.07</td>
<td>-2.05</td>
<td>-2.26</td>
<td>-1.95</td>
<td>0.53</td>
<td>-0.01</td>
<td>0.54</td>
<td>-0.05</td>
</tr>
<tr>
<td>Chocolate</td>
<td>-3.32</td>
<td>-3.34</td>
<td>-3.34</td>
<td>-2.75</td>
<td>0.80</td>
<td>0.00</td>
<td>0.74</td>
<td>0.05</td>
</tr>
<tr>
<td>Confectionery</td>
<td>-2.23</td>
<td>-2.24</td>
<td>-1.98</td>
<td>-2.44</td>
<td>0.76</td>
<td>0.19</td>
<td>0.71</td>
<td>-0.06</td>
</tr>
<tr>
<td>Coffee</td>
<td>-2.85</td>
<td>-2.96</td>
<td>-3.00</td>
<td>-2.75</td>
<td>0.60</td>
<td>-0.44</td>
<td>0.48</td>
<td>-0.53</td>
</tr>
<tr>
<td>Herbal Tea</td>
<td>-2.95</td>
<td>-2.92</td>
<td>-2.95</td>
<td>-3.27</td>
<td>0.37</td>
<td>-0.49</td>
<td>0.25</td>
<td>-0.51</td>
</tr>
<tr>
<td>Juice</td>
<td>-2.32</td>
<td>-2.33</td>
<td>-2.51</td>
<td>-2.49</td>
<td>0.14</td>
<td>-0.56</td>
<td>0.15</td>
<td>-0.35</td>
</tr>
<tr>
<td>Soft Drink</td>
<td>-3.45</td>
<td>-3.30</td>
<td>-3.05</td>
<td>-3.41</td>
<td>0.05</td>
<td>-0.69</td>
<td>0.07</td>
<td>-0.61</td>
</tr>
<tr>
<td>Beer</td>
<td>-4.49</td>
<td>-3.97</td>
<td>-3.84</td>
<td>-4.22</td>
<td>0.28</td>
<td>-0.76</td>
<td>-0.23</td>
<td>-0.64</td>
</tr>
<tr>
<td>Energy Drink</td>
<td>-3.22</td>
<td>-3.68</td>
<td>-4.49</td>
<td>-4.50</td>
<td>-0.28</td>
<td>-0.82</td>
<td>-0.06</td>
<td>-0.76</td>
</tr>
<tr>
<td>Regular Milk</td>
<td>-1.83</td>
<td>-1.75</td>
<td>-1.92</td>
<td>-1.75</td>
<td>-0.12</td>
<td>-0.21</td>
<td>-0.17</td>
<td>-0.20</td>
</tr>
<tr>
<td>Skimmed Milk</td>
<td>-2.77</td>
<td>-2.78</td>
<td>-2.75</td>
<td>-2.71</td>
<td>-0.39</td>
<td>-0.32</td>
<td>-0.20</td>
<td>-0.05</td>
</tr>
</tbody>
</table>
1. Product Path

Figure 3 shows the path of products over time. The blue trace shows the movement of products from period 1 to period 2, period 2 to period 3, and period 3 to period 4. For example, skimmed milk is on the lower left quadrant of the map in period 1. It is close to regular milk, desserts, and frozen foods. In period 2, it moves up with regular milk, desserts, and frozen foods toward canned products. Then, they reach the upper quadrant of the map in period 3. In period 4, skimmed milk moves downward with regular milk to the lower left quadrant (to its original position in period 1) and leaves desserts and frozen foods in the upper quadrant. The size of dots is proportional to the size of customers purchasing the products. The consumption of regular milk and skimmed milk changed during the observation period. An increasing number of people purchased regular milk and skimmed milk in periods 1 and 4 than in periods 2 and 3. Similarly, energy drink is at the bottom part of the map and moves to the right from periods 1 to 4.
2. Dynamic Product Association

In a static product-positioning map, we can visualize product complements and product substitutes based on the distance between the members of a pair of products. Products close to each other are product complements, whereas products far from each other are product substitutes. We extend this concept temporally in the dynamic product-association map. Regular milk and skimmed milk are product complements.
temporally because they move together during each time stamp. Similarly, dried grapes and peanuts are product complements because they are close to each other during each period. However, beer and energy drink are product substitutes temporally. While beer moves from the right to the left during the observation period, energy drink moves in the opposite direction, that is, from the left in period 1 to the right in period 4.

3. Product Variation

While (a) regular milk and skimmed milk and (b) dried grapes and peanuts are both product complements temporally, they are different in nature. The first pair (i.e., regular milk and skimmed milk) moves up and down in the map and is highly associated with other products. When other products are on sale or subject to panic buying, the sale of regular milk and skimmed milk is affected by the sale of these products. Therefore, promoting these products will influence the consumption of regular milk and skimmed milk. By contrast, dried grapes and peanuts are invariant to the promotion of other products. They remain in the same position and are unaffected by other products. Therefore, the promotion of other products may not drive the additional sale of dried grapes and peanuts.

4.3. Results of the Dynamic Customer Segment Map

1. Dynamic Model Configuration

The model uses 322 segment points to create a dynamic joint space map for products and segments. Each segment point is time specific and represents the common purchase behavior of a group of consumers. We label each segment point with the product composition in square brackets in Figure 4. For example, 75 segment points
appear in period 1. The segment point [DF] corresponds to 1,357 customers who purchase products D and F in period 1. The size of the segment point is proportional to the number of people associated with the corresponding segment. The use of 75 segment points and 25 product points completely explain the relationships between customer segments and product purchases without error. The estimated radius is 0.20. Figure 4 is the basis of the dynamic visual analytic model, which allows viewing of both intra-segment and inter-segment relationships between customer segments and product purchases.

Figure 4 Dynamic segment map

---

**Period = 1, Number of segments = 75**

---

**Period = 2, Number of segments = 81**
Period = 3, Number of segments = 77  
Period = 4, Number of segments = 89

Note: Pairs of segment points are connected if over 150 customers belong to both segments.

2. Customer Path

The transactions of a customer can be regarded as a customer purchase sequence over time. Customer path, which is ordered by transaction time, can be visualized on the map shown in Figure 5. For example, customer 1 purchases one product category (i.e., rice) during each period and, therefore, belongs to the rice segment over the four consecutive periods (\{[H]\}; \{[H]\}; \{[H]\}; \{[H]\}). By contrast, another customer (i.e., customer 2) purchases several product categories during each period. During period 1, customer 2 only purchases soft drink. Then, he/she adds potato chips to his/her purchase category during period 2. During period 3, customer 2 seeks variety and purchases other snacks such as chocolate and confectionery. Simultaneously, he/she is attracted by the promotion of canned soups and buys canned soup and noodle, which is strongly associated with canned soups. Customer 2 continues to purchase soft drink, along with juice. During period 4, customer 2 purchases the same snacks bought during the previous two periods (i.e., potato chips, chocolate, and confectionery).
He/she also continues to purchase soft drink and juice and extends his/her purchase in the canned product category to include canned meat and canned vegetables. This lengthy description can be summarized and regarded as a customer path. Customer 2 expands his/her purchase categories and seeks variety across related categories in Figure 5. The purchase sequence of customer 2 is (\{[U]\}; \{[U], [O]\}; \{[TU], [PQ], [BD]\}; \{[TU], [OPQ], [ACD]\}). Unlike customer 1 who belongs to the same segment across the four periods, customer 2 uses one, two, three, and three segment points during periods 1, 2, 3, and 4, respectively.

Figure 5 Customer path

Notes:
Path of customer 1:
(\{[H]\}; \{[H]\}; \{[H]\}; \{[H]\})

Path of customer 2:
(\{[U]\}; \{[U], [O]\}; \{[TU], [PQ], [BD]\}; \{[TU], [OPQ], [ACD]\})
3. Intra-segment Association

We have 75, 81, 77, and 89 segments during periods 1, 2, 3, and 4, respectively. Each customer belongs to multiple segments in each period. A maximum of \( \binom{K}{2} \) segments pairs can exist for \( K \) number of segment points at time \( t \). For example, there are 2,775 segment pairs possible in period 1 and 3,916 segment pairs exist in period 4. The graph will be cluttered with lines and numbers if we display all pairs of segments on a map. Instead, we draw a link between pairs of segment points in which the number of customers associated with the segments exceeds a certain threshold value. In Figure 4, we identify 104 important links, with at least 150 customers each. These 104 links correspond to 104 pairs of relationships and represent the major overlapping segments in period 1. The same process is conducted for the other periods, and 70, 126, and 92 segment pairs are identified in periods 2, 3, and 4, respectively. Customers tend to purchase more than one group of product clusters. For example, they purchase groceries (left quadrant) and snacks and biscuits (right quadrant) together, as indicated by the numerous cross-connected links between these two categories.

4. Inter-segment Association

Each customer belongs to multiple segments, and the complete purchase history of a customer can be represented by a customer path. This visual analytics can be applied to longitudinal data by adding the temporal component of such data, which makes visual analytics useful and applicable in identifying cross-selling opportunities in time-series data. Temporal association analysis is not new in literature, and in fact, has captured the interest of managers and researchers because of its importance. Researchers have attempted to modify existing association rule algorithms to incorporate the temporal components of longitudinal data (Agrawal & Srikant, 1995;
Kumar, Alrabea, & Sekhar, 2010). They have presented the temporal association rule in the following form:

\[ P_{it} \rightarrow P_{jt+1}, \]  

(15)

where \( P_{it} \) and \( P_{jt+1} \) are products \( i \) and \( j \) at time \( t \) and time \( t + 1 \), respectively. The temporal association rule can be determined using support, lift, and confidence. Previous researchers have presented a single-product association rule and have examined temporal product association. We extend temporal product association to temporal segment association, and then visualize segment transitions in a group of consumers who share a common purchase behavior over time. Formally, we define the temporal segment association rule for segment \( i \) at time \( t \) and segment \( j \) at time \( t + 1 \) as follows:

\[ S_{it} \rightarrow S_{jt+1}. \]  

(16)

The rule is accurate if confidence is high. Temporal confidence measures how much segment \( j \) at time \( t + 1 \) depends on segment \( i \) at time \( t \).

\[ \text{confidence}(S_{it} \rightarrow S_{jt+1}) = \Pr(S_{jt+1} | S_{it}) \geq \sigma_c, \]  

(17)

where \( \sigma_c \) is the confidence threshold. We draw a directed link from segment point \( S_{it} \) at time \( t \) to segment point \( S_{jt+1} \) at time \( t + 1 \). Similarly, we define temporal lift as follows:

\[ \text{lift}(S_{it} \rightarrow S_{jt+1}) = \frac{\text{confidence}(S_{it} \rightarrow S_{jt+1})}{\Pr(S_{jt+1})} \geq \sigma_L, \]  

(18)

where \( \sigma_L \) is the lift threshold. Temporal lift measures how much a rule is better than chance. As an example, we set \( \sigma_c = 5\% \) and \( \sigma_L = 1.5 \) as the minimum confidence and lift, respectively, of the temporal association rules, along with a minimum of 150 customers (i.e., \( \sigma_S \)) for each segment transition. We use different colors, thickness values, and line styles to represent the strength of the relationships between segment
pairs on a visual dynamic map. In particular, we draw

(1) an undirected solid link if support is satisfied but lift and confidence are not,

(2) a directed dashed link if support and confidence are satisfied but lift is not,

(3) a directed dotted link if support and lift are satisfied but confidence is not, and

(4) a directed solid link if all the criteria are satisfied.

To create the dynamic knowledge map shown in Figure 6, we color the links in different shades. A dark link indicates a large lift, whereas a light link represents a small lift. Similarly, we use a thick (thin) link to represent a relationship with high (low) confidence. The size of each segment point is proportional to the number of customers in the segment. We use green, red, purple, and brown nodes to represent segment points in periods 1, 2, 3, and 4, respectively. The findings are summarized as follows.

Figure 6 Dynamic knowledge map

From period 1 to period 2
From period 2 to period 3

From period 3 to period 4

Note:
The green, red, purple, and brown dots represent the segment points at periods 1, 2, 3, and 4, respectively.
1. Dynamic Knowledge Map

We identify 21 undirected solid links, 216 directed dashed links, 2 directed dotted links, and 17 directed solid links in the knowledge map shown in Figure 6. The high proportion of segment pairs with high support and high confidence but low lift values suggests that many segments are substitutes of one another. In particular, segment pairs with low lift values (light-colored link) are far from each other. By contrast, segment pairs with high lift values (dark-colored link) are close to each other.

2. Repeat Purchase Behavior

We observe strong repeat purchase behavior in each successive period. In the transitional map from period 1 to period 2, segments with the same product features tend to be close to one another. For example, the salt category (i.e., [J]) is positioned on the left in both periods 1 and 2, whereas the coffee category (i.e., [R]) is in the right quadrant during these two periods. The association links of \([XY] \rightarrow [XY]\) are thick and dark, which suggests that customers who have purchased regular milk and skimmed milk in period 1 are more likely to purchase the same products in period 2. Thick and dark lines also appear in segment pairs \([R] \rightarrow [R], [T] \rightarrow [T]\), and \([N] \rightarrow [N]\), which suggests that segment self-transition is exceptionally higher than expected and supports the findings of strong repeat purchase behavior.

3. Price-sensitive Behavior

Although repeat purchase behavior is prevalent among customer segments, this behavior can be weakened by a price change. We observe a strong price-sensitive behavior across customer segments, that is, an increase (decrease) in price may lead to a decrease (increase) in product consumption. In the transition map from period 2 to
period 3, the rice category (i.e., [H]) is on sale at a 10% discount, which has prompted many people to buy rice during this period, and consequently, leads to a dramatic increase in sales, as reflected by the sudden increase in the segment size of [H]. Another example is the 35% reduction in the price of the egg category from period 3 to period 4. Although a decrease in price may lead to an increase in consumer surplus, the reverse also applies. For example, the price of the salt category (i.e., [J]) increases by 50% during the same transition period; the segment size of [J] in period 1 is large but disappears in period 2. The same pattern can also be observed in the rice category, in which price returns to normal when the promotion ends by the end of period 3. The segment size of the rice category in period 4 decreases to a quarter of its original size in period 3.

4. Relationships between Segment Association and Price Change

Intuitively, a decrease (increase) in price leads to an increase (decrease) in consumer surplus. From the observation of 25 products over a four-month period, a decrease in price has led to an increase in demand 19 times, and an increase in price has led to a decrease in product consumption 36 times. Surprisingly, the price-up/demand-increase pattern has occurred 13 times, and the price-down/demand-decrease pattern has occurred 7 times. Through a careful investigation of price-up/demand-increase products, we determine that these products are highly associated with other products that are on sale. For example, the price of oil increases by 4.5% from period 2 to period 3, but the demand for it also increases by 15.2%. This pattern may be driven by a customer segment [GH] (a segment of rice and oil), which has doubled in size from period 2 to period 3. During that time, rice is on sale at a 10% discount. Thus, the promoted products increase the sales of other associated products. We also observe a
similar behavior in which the sale on canned vegetables and canned soups has driven the sales of the associated price-up canned meat category from period 2 to period 3.

5. CONTRIBUTIONS, LIMITATIONS, AND FUTURE RESEARCH

POS data are the most valuable asset of retailers. These data record the purchase history, buying behavior, and product preferences of each customer. Thus, the research challenge is to identify ways to help marketers fully utilize POS data and knowledge on customers to identify and execute appropriate marketing actions.

5.1. Contributions

We present a novel dynamic visual analytic model to analyze massive temporal POS transaction data in marketing. The proposed model has two distinct features. First, it allows marketers to conduct customer segmentation and product association simultaneously. Second, it summarizes the segmentation and both intra-association and inter-association results in a map, which helps marketers cognitively visualize customer insights. We apply the model in a supermarket transaction data set comprising 28,092 customers and 25 product categories over a four-month period to illustrate their different applications, and report the empirical results and discuss the key features. The formulation of the proposed model is different from those of traditional statistical and data mining techniques. Such differences may be examined from three aspects.

1. Clustering perspective

Segmentation (also referred to as clustering) is a popular and important research topic in marketing and statistics (Allon & Federgruen, 2009; Wedel & Kamakura, 2000).
Numerous methods have been proposed (e.g., agglomerative algorithms, such as hierarchical clustering, and partitioning algorithms, such as $k$-means), and many applications have demonstrated the importance and usefulness of clustering in various domains. Clustering groups customers who are similar to one another into subgroups. We propose a new approach for clustering customers and products simultaneously in customer segmentation. It relies on the matrix factorization method to estimate the number of segments, determine the features of each segment, and choose the segment membership for each customer simultaneously. The proposed model allows overlapping of segments and can be solved using binary integer programming iteratively.

2. Dimensionality Reduction and Data Scalability Perspective

Marketing scientists have studied perceptual mapping for several decades. MDU is the most popular model for two-mode data (e.g., customers and products). However, this technique suffers from two problems, namely, dimensionality and scalability. First, if we cannot solve the dimensionality reduction problem, then the representation of high-dimensional objects in low-dimensional space will be inaccurate. Second, the number of parameters in MDU increases with sample size. Consequently, traditional MDU application is limited to survey data with a small sample size (Borg & Groenen, 2005). To address these challenges, we use the overlapping segments assumption to solve the dimensionality reduction problem by representing high-dimensional objects in 2D space without visualization error. To address the scalability issue, we group customers into customer segments and represent them as segment points instead of customer points on the map. Therefore, the proposed model is not only error-free but is also more scalable than MDU because the number of segments is significantly less
than the number of customers.

3. Visualization Perspective

Online analytical processing (OLAP) is frequently used to answer specific business questions raised by marketers (Thomsen, 2002). An example of an OLAP question is “How many customers have purchased product A this month?” Although OLAP is efficient in providing query-based answers, it is question-specific. Consequently, OLAP cannot inspire marketers to find the right questions during the data exploration stage. Moreover, OLAP cannot consolidate results from different queries into a unified picture. Therefore, marketers may have numerous OLAP results, typically as tables, but they do not gain insight from the data. In contrast to OLAP, the proposed model integrates clustering and association results into a single unified analysis to enable marketers to explore visually and interact with association and prediction results to improve understanding. In addition, the visual map presents the information in such a manner that the hidden structure of POS data is systematically unfolded using points, directed and undirected links through the visual map. Hence, important findings can be efficiently and effectively communicated to others.

Time is one of the natural components in numerous marketing data. For example, supermarket transaction data record products bought by customers at certain time stamps. We introduce the concepts of dynamic product path and dynamic product association in a dynamic product map, and present customer path and inter-segment association analyses in a dynamic segment map with interesting customer insights (e.g., repeat purchase behavior and price-sensitive behavior).
5.2. Managerial Implications

Dynamic visual analytic model assist marketers in two ways. First, it produces a joint space representation of customers and products in a single map without error. Managers can gain insight into the overall structure of customer–product relationship and obtain meaningful information on separate segments at the individual level. In Figure 2, milk and skimmed milk are clustered because of customer needs. Customers always purchased these items together, and these items are far from other product categories. Second, we can visualize product movement as a product path and customer purchase sequence over time as a customer path. This feature promises to be effective in identifying the factors that drive the movement of product and customer positions. For example, the price of product often influences product consumption. We can visualize how changes in prices affect product consumption. In addition, we can see how price-up or on-sale products affect the sales of other products. Figure 6 shows that promoted products can increase the sales of non-promoted products. For example, the demand for oil (a price-up product) increases by approximately 15% because of the stimulation of associated on-sale products (e.g., rice). It can help managers understand the underlying mechanism of an observed market phenomenon and execute actions faster.

5.3. Limitations and Future Research

The proposed dynamic visual analytics model is also subject to several limitations. First, it is applied to a temporal supermarket transaction data set comprising 28,092 customers and 25 product categories. Although the sample sizes is larger than many reported visual analytical research (e.g., MDU) in marketing, they are still small compared with the massive number of customers that companies frequently have in
their databases in the big data era. For example, the department store giant Walmart handles millions of customer transactions every day. Another question is how to enhance the algorithm to enable it to handle even larger samples within a reasonable time frame. Second, the required error-free visual map inevitably creates a massive number of customer memberships and segment points to represent customers. A trade-off exists between the number of parameters and the acceptable visualization error. Future research may allow users to input acceptable visualization errors to reduce computational complexity and create a joint space map with less points. In addition, only two-mode POS data (i.e., products and customers) are considered in the proposed framework. Although we examine the influence of price on product positioning and customer segmentation, such influence is not explicitly modeled in the visual framework. Other types of marketing information such as customer profile (e.g., age and gender) and marketing-mix information (e.g., price, place, and promotion) may provide interesting and useful insights to aid in decision making in marketing. Future research may incorporate such information into utility function and examine the effects of changes in these variables (e.g., price) on product association and customer segmentation.
6. REFERENCES


destiny: Using spatial dimension of sales data for early prediction of new

navigation in information visualization: A survey. *Visualization and Computer

ebape.fgv.br/sites/ebape.fgv.br/files/REVISION%20A%20picture%20is%20worth%20a%20thousand%20words.pdf

(Ed.), *Handbook of Graph Drawing and Visualization* (pp. 383-408). Boca
Raton, FL: CRC Press.

919-920.

(pp. 496-508): Springer.

Mining in Large Databases. *Knowledge Discovery Practices and Emerging
Applications of Data Mining: Trends and New Domains: Trends and New
Domains*, 48-65.

Indicators of structural changes and interdisciplinary developments. *Journal of
the American Society for Information Science and Technology*, 59 (11), 1810-
1818.


