Customer Segmentation Using Frequency Domain Analysis of Multivariate Time Series

Shan Hu

Joint work with Prof. Nalini Ravishanker,Univ. of Connecticut and Prof. Raj Venkatesan, Univ of Virginia,

Department of Statistics, University of Connecticut

June 5th, 2009

Outline

- Market Segmentation
- Frequency Based Clustering Methodology
- Example: Customers of Organic Products
- Results for Simulated Data
- Segmentation of a Grocery Store Customer Base
- Research in Progress

Market Segmentation

- Definition of Market Segmentation: Classify customers into homogeneous groups based on their demands.
- "Goods can no longer be produced and sold without considering customer needs and recognizing the heterogeneity of those needs"

Smith (1956): The Journal of Marketing

Wedel & Kamakur (2000):ISQM

Segmentation Methods:

	A priori	Post hoc
Descriptive	Contingency tables,Log-linear models	Clustering methods
Predictive	Cross-tabulation, regression, logit and discriminant analysis	AID, CART, Clusterwise regression, ANN, mixture models.

Spectral methods help us to extend the classical cluster analysis to incorporate both multivariate and time dependence in the data.

Wedel & Kamakur (2000):ISQM

Frequency Based Clustering Methodology

Let $\mathbf{X}_t = \left(X_t^{(1)}, ..., X_t^{(m)}\right), t = 1, 2, ..., T$ be an m-dimensional real-valued stationary time series

- The $m \times m$ covariance function: $\mathbf{R}(s-t)$.
- The spectral density: $\mathbf{f}(\lambda)$, $\lambda \in (-\pi, \pi] \backslash \{0\}$
  - $\mathbf{f}(\lambda)$ is an $m \times m$ complex Hermitian matrix;
  - Consistent estimate of $\mathbf{f}(\lambda)$ is the smoothed periodogram $\mathbf{f}_T(\lambda_j) = \frac{1}{2L+1} \sum_{|k| \leq L} \mathbf{I}(\lambda_{j+k})$;
  - where $\{\mathbf{I}(\lambda)\}_{ij} = \frac{1}{2\pi T} \left(\sum_{t=1}^{T} x_{ti} \exp(-it\lambda) \right) \left(\sum_{t=1}^{T} x_{tj} \exp(it\lambda) \right)$

Measure of Disparity: Consider Two Different Hypotheses

- Kullback-Leibler (KL) discrimination information:

$$I(p;q) = E_p\left\{\log \frac{p(\mathbf{x})}{q(\mathbf{x})}\right\} = \frac{1}{2}\left[tr(\mathbf{R}_p\mathbf{R}_q^{-1}) - \log \frac{|\mathbf{R}_p|}{|\mathbf{R}_q|} - mT\right]$$

with limit form:

$$I(\mathbf{f};\mathbf{g}) = \lim_{T\to\infty} T^{-1} I(p;q) = \frac{1}{2}\int_{-\pi}^{\pi} (tr(\mathbf{f}\mathbf{g}^{-1}) - \log \frac{|\mathbf{f}|}{|\mathbf{g}|} - m)\frac{d\lambda}{2\pi}$$

Kakizawa, Shumway, and Taniguchi (1998): JASA

Measure of Disparity

- Chernoff (CH) Information Measure: $0 \leq \alpha \leq 1$

$$B_\alpha(p;q) = -\log E_p\left\{\left(\frac{q(\mathbf{x})}{p(\mathbf{x})}\right)^\alpha\right\}$$
$$= \frac{1}{2}\left[\log\frac{|\alpha\mathbf{R}_p + (1-\alpha)\mathbf{R}_q|}{|\mathbf{R}_q|} - \alpha\log\frac{|\mathbf{R}_p|}{|\mathbf{R}_q|}\right]$$

  - Antisymmetry property: $B_\alpha(p;q) = B_{1-\alpha}(q;p)$
  - $B_\alpha(p;q)$, scaled by $\alpha(1-\alpha)$, converges to $I(p;q)$ for $\alpha \to 0$ and to $I(q;p)$ for $\alpha \to 1$.

- The limit form is:

$$B_\alpha(\mathbf{f};\mathbf{g}) = \lim_{T\to\infty} T^{-1}B_\alpha(p;q)$$
$$= \frac{1}{2}\int_{-\pi}^{\pi}\left(\log\frac{|\alpha\mathbf{f} + (1-\alpha)\mathbf{g}|}{|\mathbf{g}|} - \alpha\log\frac{|\mathbf{f}|}{|\mathbf{g}|}\right)\frac{d\lambda}{2\pi}$$

Frequency Based Clustering Methodology

- Approximation for the Quasi-distance measure based on estimated smoothed spectra:

$$JB_\alpha(\mathbf{f}_T;\mathbf{g}_T) = B_\alpha(\mathbf{f}_T;\mathbf{g}_T) + B_\alpha(\mathbf{g}_T;\mathbf{f}_T)$$
$$= \frac{1}{2}T^{-1}\sum_s[\log\frac{|\alpha\mathbf{f}_T+(1-\alpha)\mathbf{g}_T|}{|\mathbf{g}_T|} + \log\frac{|\alpha\mathbf{g}_T+(1-\alpha)\mathbf{f}_T|}{|\mathbf{f}_T|}]$$

- The measures of disparity can be used to cluster multivariate time series. In general, we may consider hierarchical methods using the quasi-distance matrix as an input.

Johnson & Wichern (2001): Pearson.

Example: Customers of Organic Products

Market Segmentation of Customers of Organic Products

- A grocery firm would like to
  - Design market campains towards a particular group of customers, who bought organic products;
  - Decide what kind of organic products to shelf in its stores;
  - Decide what other related products to co-promote.
- To achieve those goals, we first explore if heterogeneity exists in this customer population based on four aspects:
  - Calorie content
  - Price
  - Quality
  - Uniqueness: mainstream vs non-mainstream

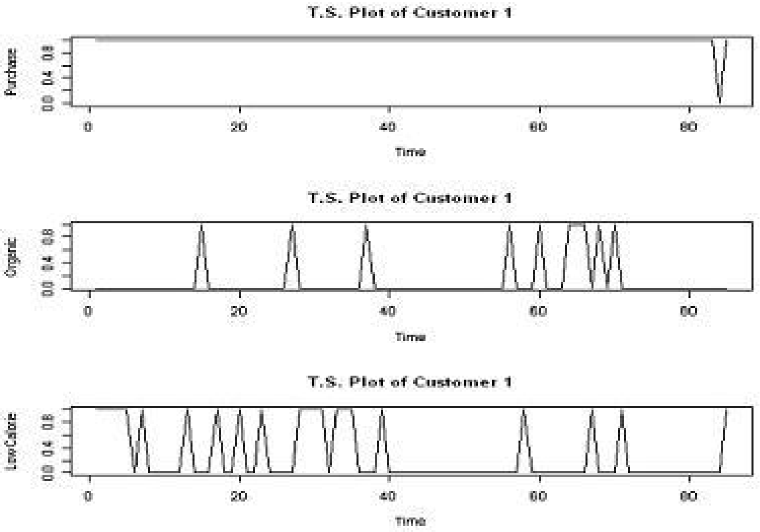
Example: Customers of Organic Products

Brief Description of the Grocery Store Database

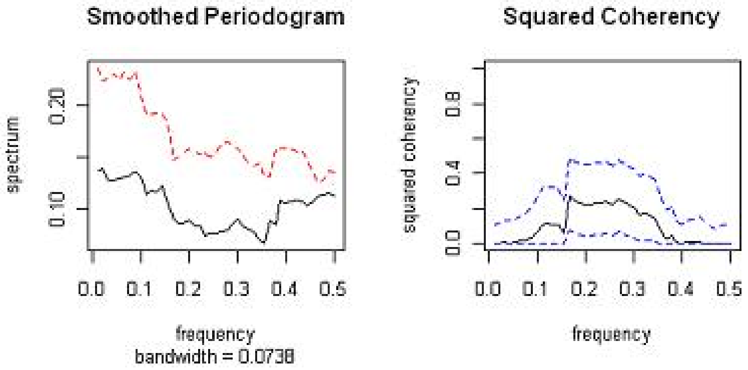
- Grocery store database:
  - 2500 customers
  - 102 weeks
  - 2,595,732 purchasing records
- Content:
  - Household ID;
  - ID numbers of the products they had purchased;
  - In which day/week they had made the purchase; etc.
  - Characteristics of all the products

Example: Customers of Organic Products

Observed Multivariate Time Series



Market Segmentation Example



Results for Simulated Data

Simulated Data

- Bivariate binary time series $\mathbf{X}_t = \left(X_t^{(1)}, X_t^{(2)}\right)$, $t = 1, ..., 200$
- $X_t^{(i)}$ follows a Bernoulli distribution: $E[X_t^{(i)}] = p_i \in [0,1]$ for $i = 1, 2$.
- Each $X_t^{(i)}$ is an $AR(1)$ process: $Corr(\ X_t^{(i)},\ X_{t-1}^{(i)}) = \phi_i$ for $i = 1, 2$.
- $Corr(\ X_t^{(1)},\ X_t^{(2)}) = \delta$ and its range:
  - Lower bound: $\max\left\{-(p_1 p_2 / q_1 q_2)^{1/2}, -(q_1 q_2 / p_1 p_2)^{1/2}\right\}$
  - Upper bound: $\min\left\{(p_1 q_2 / p_2 q_1)^{1/2}, -(p_2 q_1 / p_1 q_2)^{1/2}\right\}$

Note:The simulated data can be used to check the misclassification rate of the frequency based clustering method.

Emrich & Piedmonte (1991): TAS

Results for Simulated Data

We fixed the means of $X_t^{(1)}$ and $X_t^{(2)}$ under three scenarios, but the correlation between them is increasing from Scenario 1 to Scenario 3.

			Scenario 1	Scenario 2	Scenario 3
	p_1	p_2	δ	δ	δ
Population 1	0.2	0.1	0	0.2	0.6
Population 2	0.3	0.4	0	0.1	0.8
Population 3	0.2	0.4	0	0.2	0.6
Population 4	0.3	0.1	0	0.1	0.4
MisRate MV			*24.83%*	*23.78%*	*15.49%*
MisRate UV1			34.10%	33.39%	30.53%
MisRate UV2			25.11%	26.22%	31.09%

Note: Each population has 25 objects.

Results for Simulated Data

Then we fixed the correlation between $X_t^{(1)}$ and $X_t^{(2)}$ under the following scenarios, but the marginal means of the four populations are more scattered in Scenario 6 than in Scenario 3.

	δ		Scenario 3	Scenario 6
Population 1	0.6	p_1	0.2	0.1
		p_2	0.1	0.2
Population 2	0.8	p_1	0.3	0.4
		p_2	0.4	0.3
Population 3	0.6	p_1	0.2	0.7
		p_2	0.4	0.5
Population 4	0.4	p_1	0.3	0.9
		p_2	0.1	0.6
MisRate MV			*15.49%*	*12.63%*

Segmentation of a Grocery Store Customer Base

Result from using frequency based clustering method on the organic product buyers:

	Number of Customers in each group			
Number of Groups	Group 1	Group 2	Group 3	Group 4
2	429	3	N/A	N/A
3	418	11	3	N/A
4	254	164	11	3

Research in Progress

- It will be very useful to profile customers in each group based on frequency of the purchase of organic and low calorie products, as well as the demographic characteristics (gender, marital status, etc.).
- Incorporate "Price", "Quality", and "Uniqueness" variables into analysis.
- Walsh-Fourier Analysis
  - Use Walsh-Fourier transformation to calculate the quasi-distance measure.
  - Walsh-Fourier spectral analysis rather than Fourier analysis may be more suited to the analysis of discrete-valued and categorical-valued time series, and of time series that contain sharp discontinuities.

Stoffer, D.S.(1991), "*Walsh-Fourier Analysis and Its Statistical Applications*", JASA, 86,461-485

Stoffer, D.S.(1988), "*Multivariate Walsh-Fourier Analysis*", Journal of Time Series Analysis, 11, 57-73


Thank You!