Market Segmentation with Latent Class Regression

Applications of the package “FlexMix”

November 14, 2010 @ Shanghai University of Finance and Economics
Segment and Segmentation

- A **segment** is a group of end-users that share a unique set of wants/needs and/or purchase behaviors.
- **Segmentation** is the process that companies use to divide large heterogeneous markets into small markets that can be reached more efficiently and effectively with products and services that match their unique needs.
## Segmentation Bases

<table>
<thead>
<tr>
<th></th>
<th>General</th>
<th>Product-specific</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observable</strong></td>
<td>Cultural, geographic, demographic and socio-economic variables</td>
<td>User status, usage frequency, store loyalty and patronage, situations</td>
</tr>
<tr>
<td><strong>Unobservable</strong></td>
<td>Psychographics, values, personality and life-style</td>
<td>Psychographics, benefits, perceptions, <strong>elasticities</strong>, attributes, preferences, intention</td>
</tr>
</tbody>
</table>

Pricing Segments

Brand Loyals

Deal Prone
Advertising Segments

Response to Ad 1
Response to Ad 2

Segment A
Segment B
## Segmentation Methods

<table>
<thead>
<tr>
<th></th>
<th>a priori</th>
<th>Post hoc</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive</strong></td>
<td>Contingency tables, Log-linear models</td>
<td>Clustering methods: Nonoverlapping, overlapping, Fuzzy techniques, ANN, mixture models</td>
</tr>
<tr>
<td><strong>Predictive</strong></td>
<td>Cross-tabulation, Regression, logit and Discriminant analysis</td>
<td>AID, CART, Clusterwise regression, ANN, mixture models</td>
</tr>
</tbody>
</table>
Segmentation: Distance?

- Segmentation is the process of clustering consumers on basis of distances between them?
Segmentation with Cause-Effect Relation

Aggregate (100%): $Loyal = -2.46 + 1.06 \times Satisfaction$
Segment A (25.22%): $Loyal = 1.65 + 0.77 \times Satisfaction$
Segment B (48.38%): $Loyal = -4.25 + 1.33 \times Satisfaction$
Segment C (26.40%): $Loyal = 0.17 + 0.33 \times Satisfaction$

王霞, 赵平, 王高, 刘佳 (2005)
Multiple Correspondence Analysis

Segment A
- Senior high school
- Modest income
- 40 < Age < 50

Segment B
- Primary school or junior high school
- Lowest income
- Age > 50

Segment C
- College or above
- Average or high income
- Age < 30

王霞, 赵平, 王高, 刘佳 (2005)
Relationship for Segment Targeting

Descriptor Variables

Identifying the Particular Members of a Segment

Behavioral Variables

The Aspects of a Segment that Define Marketers’ Efforts

The Key to Targeting a Segment
Latent Class Regression

- Clusterwise / (finite) mixture regression
  - Consider finite mixture models with $K$ components of form
    \[
    h(y \mid x, \psi) = \sum_{k=1}^{K} \pi_k f(y \mid x, \theta_k)
    \] (1)
  - $\pi_k \geq 0$, $\sum_{k=1}^{K} \pi_k = 1$

  - where $y$ is a (possibly multivariate) dependent variable with conditional density $h$, $x$ is a vector of independent variables, $\pi_k$ is the prior probability of component $k$, $\theta_k$ is the component specific parameter vector for the density function $f$, and $\psi = (\pi_1, \ldots, \pi_K, \theta'_1, \ldots, \theta'_K)'$ is the vector of all parameters
Latent Class Regression

- If $f$ is a univariate normal density with component-specific mean $\beta'k x$ and variance $\sigma_k^2$, we have $\theta_k = (\beta'_k, \sigma_k^2)$ and Equation (1) describes a mixture of standard linear regression models.
- If $f$ is a member of the exponential family, we get a mixture of generalized linear models.
Posterior Probability

- The posterior probability that observation \((x, y)\) belongs to class \(j\) is given by

\[
P(j | x, y, \psi) = \frac{\pi_i f(y | x, \theta_j)}{\sum_k \pi_k f(y | x, \theta_k)}
\]

- The posterior probabilities can be used to segment data by assigning each observation to the class with maximum posterior probability

- Individual-level predictions of finite mixture models are a weighted combination of the segment-level regression functions, weighted with the posterior membership probabilities (DeSarbo, Kamakura, and Wedel 2006)
Parameter Estimation

- The log-likelihood of a sample of \( N \) observations \( \{(x_1, y_1), \ldots, (x_N, y_N)\} \) is given by

\[
\log L = \sum_{n=1}^{N} \log h(y_n | x_n, \psi) = \sum_{k=1}^{K} \log(\sum_{k=1}^{K} \pi_k f(y_n | x_n, \theta_k))
\]

- The most popular method for maximum likelihood estimation of the parameter vector \( \psi \) is the iterative EM algorithm (Leisch 2004)
Using FlexMix

• As a simple example we use artificial data with two latent classes of size 100 each:
  - Class 1: $y = 5x + \epsilon$
  - Class 2: $y = 15 + 10x - x^2 + \epsilon$
  - with $\epsilon \sim N(0, 9)$ and prior class probabilities $\pi_1 = \pi_2 = 0.5$

• We can fit this model in R using the commands

```r
> library(flexmix)
> data(NPreg)
> m1 = flexmix(yn ~ x + I(x^2), data = NPreg, k = 2)
> m1
```

Leisch (2004)
Call:
\texttt{flexmix(formula = yn \sim x + I(x^2), data = NPreg, k = 2)}
Cluster sizes:
\begin{tabular}{ll}
1 & 2 \\
100 & 100
\end{tabular}
convergence after 15 iterations
\texttt{> parameters(m1, component = 1)}
\texttt{$\text{coef}$}
\begin{tabular}{cccc}
(Intercept) & x & I(x^2) \\
-0.20989331 & 4.81782414 & 0.03615728
\end{tabular}
\texttt{$\text{sigma}$}
\texttt{[1] 3.47636}
\texttt{> parameters(m1, component = 2)}
\texttt{$\text{coef}$}
\begin{tabular}{cccc}
(Intercept) & x & I(x^2) \\
14.7168295 & 9.8466698 & -0.9683534
\end{tabular}
\texttt{$\text{sigma}$}
\texttt{[1] 3.479809}
\texttt{Leisch (2004)
Using FlexMix

> summary(m1)
Call:
flexmix(formula = yn ~ x + I(x^2), data = NPreg, k = 2)

prior size post>0 ratio
Comp.1  0.494 100  145    0.690
Comp.2  0.506 100  141    0.709
`log Lik.' -642.5453 (df=9)
AIC: 1303.091 BIC: 1332.775
> table(NPreg$class, m1@cluster)
   1 2
1 95 5
2  5 95

Leisch (2004)
## Significance Test

```r
> rml1 = refit(m1)
> summary(rml1)
```

**Call:**

```r
refit(m1)
```

**Component 1:**

|                  | Estimate | Std. Error | t value | Pr(>|t|) |
|------------------|----------|------------|---------|----------|
| (Intercept)      | -0.208996| 0.673900   | -0.3101 | 0.7568   |
| x                | 4.817015 | 0.327447   | 14.7108 | <2e-16   |
| I(x^2)           | 0.036233 | 0.032545   | 1.1133  | 0.2669   |

**Component 2:**

|                  | Estimate | Std. Error | t value     | Pr(>|t|)     |
|------------------|----------|------------|-------------|--------------|
| (Intercept)      | 14.717541| 0.890843   | 16.521      | < 2.2e-16    |
| x                | 9.846148 | 0.390385   | 25.222      | < 2.2e-16    |
| I(x^2)           | -0.968304| 0.036951   | -26.205     | < 2.2e-16    |
```

Leisch (2004)
Automated Model Search

- In real applications the number of components is unknown and has to be estimated
- Fit models with an increasing number of components and compare them using AIC or BIC

```r
> m7 = stepFlexmix(yp ~ x + I(x^2), data = NPreg, 
control = list(verbose = 0), K = 1:5, nrep = 5)
> sapply(m7, BIC)
     1      2      3      4      5
946.7477 925.9972 942.1553 960.0626 960.9347
```

- Choose the number of components minimizing the BIC

Leisch (2004)
Finite Mixtures with Concomitant Variables

- If the weights depend on further variables, these are referred to as concomitant variables.
- The model class is given by

$$ h(y \mid x, \omega, \psi) = \sum_{k=1}^{K} \pi_k (\omega, \alpha) f_k (y \mid x, \theta_k) $$

  - Where $w$ denotes the concomitant variables, $\alpha$ are the parameters of the concomitant variable model.

$$ \sum_{k=1}^{K} \pi_k (\omega, \alpha) = 1 \quad \pi_k (\omega, \alpha) > 0, \forall k $$

Grün and Leisch (2008)
## Segmenting Newspaper Readers

<table>
<thead>
<tr>
<th>变量类型</th>
<th>变量名称</th>
<th>细分市场1</th>
<th></th>
<th>细分市场2</th>
<th></th>
<th>细分市场3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>系数</td>
<td>T值</td>
<td>系数</td>
<td>T值</td>
<td>系数</td>
<td>T值</td>
</tr>
<tr>
<td>感知变量</td>
<td>截距</td>
<td>5.022*</td>
<td>5.454</td>
<td>0.621*</td>
<td>2.136</td>
<td>0.355</td>
<td>1.004</td>
</tr>
<tr>
<td></td>
<td>新闻栏目评价</td>
<td>-0.268</td>
<td>-1.5544</td>
<td>0.125*</td>
<td>2.051</td>
<td>0.418*</td>
<td>6.280</td>
</tr>
<tr>
<td></td>
<td>经济栏目评价</td>
<td>0.059</td>
<td>0.46541</td>
<td>0.074</td>
<td>1.471</td>
<td>0.052</td>
<td>1.117</td>
</tr>
<tr>
<td></td>
<td>娱乐栏目评价</td>
<td>0.069</td>
<td>0.468</td>
<td>0.074</td>
<td>1.542</td>
<td>0.040</td>
<td>0.717</td>
</tr>
<tr>
<td></td>
<td>北京栏目评价</td>
<td>0.270</td>
<td>1.845</td>
<td>-0.032</td>
<td>-0.703</td>
<td>0.137*</td>
<td>2.580</td>
</tr>
<tr>
<td></td>
<td>版面设计评价</td>
<td>0.360*</td>
<td>2.071</td>
<td>0.060</td>
<td>0.913</td>
<td>0.161*</td>
<td>2.556</td>
</tr>
<tr>
<td></td>
<td>印刷质量评价</td>
<td>-0.025</td>
<td>-0.178</td>
<td>0.002</td>
<td>0.043</td>
<td>0.188*</td>
<td>4.022</td>
</tr>
<tr>
<td></td>
<td>广告评价</td>
<td>0.153</td>
<td>1.385</td>
<td>0.077*</td>
<td>1.968</td>
<td>0.071</td>
<td>1.694</td>
</tr>
<tr>
<td></td>
<td>购买便利性评价</td>
<td>-0.402*</td>
<td>-3.103</td>
<td>0.198*</td>
<td>4.978</td>
<td>-0.109*</td>
<td>-2.842</td>
</tr>
<tr>
<td></td>
<td>感知价格</td>
<td>0.119</td>
<td>1.337</td>
<td>0.280*</td>
<td>9.170</td>
<td>-0.032</td>
<td>-0.991</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>样本量</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>119</td>
<td></td>
<td>495</td>
<td></td>
<td>290</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>市场份额 (%)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>13.16</td>
<td></td>
<td>54.76</td>
<td></td>
<td>32.08</td>
</tr>
</tbody>
</table>

*Note:* The dependent variable is “Customer Satisfaction” (N= 904), *: p<0.05

王燕, 赵平 (2009)
<table>
<thead>
<tr>
<th>变量类型</th>
<th>变量名称</th>
<th>细分市场1</th>
<th>细分市场2</th>
<th>细分市场3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>系数</td>
<td>T值</td>
<td>系数</td>
</tr>
<tr>
<td>个人特征变量</td>
<td>截距</td>
<td>0.081</td>
<td>0.075</td>
<td>-0.578</td>
</tr>
<tr>
<td>阅读频率a</td>
<td>每天阅读</td>
<td>-0.069</td>
<td>-0.162</td>
<td>0.188</td>
</tr>
<tr>
<td></td>
<td>每次读报用时b</td>
<td>半小时以下</td>
<td>0.644</td>
<td>1.572</td>
</tr>
<tr>
<td>阅读地点c</td>
<td>家中</td>
<td>0.492</td>
<td>0.705</td>
<td>-0.966</td>
</tr>
<tr>
<td></td>
<td>上班</td>
<td>0.125</td>
<td>0.193</td>
<td>-0.001</td>
</tr>
<tr>
<td>性别d</td>
<td>男</td>
<td>1.138*</td>
<td>2.796</td>
<td>0.643</td>
</tr>
<tr>
<td>教育程度e</td>
<td>高中及以下</td>
<td>1.016*</td>
<td>2.091</td>
<td>1.319*</td>
</tr>
<tr>
<td>年龄f</td>
<td>25岁以下</td>
<td>0.163</td>
<td>0.292</td>
<td>-0.772</td>
</tr>
<tr>
<td></td>
<td>25-35岁</td>
<td>-0.235</td>
<td>-0.344</td>
<td>0.542</td>
</tr>
<tr>
<td>家庭月收入g</td>
<td>2000-4000元</td>
<td>-1.689*</td>
<td>-3.055</td>
<td>-0.746</td>
</tr>
<tr>
<td></td>
<td>4000元以上</td>
<td>0.725</td>
<td>1.036</td>
<td>0.948</td>
</tr>
</tbody>
</table>
Recap

• The underlying basis of customer heterogeneity (i.e., discrete market segments) is unknown a priori
• The objective is to simultaneously estimate the number of market segments, their size and composition, and the segment specific regression coefficients
• Concomitant variable mixtures allow for demographic variables to explain segment membership simultaneously
• This class of methods enables marketers to engage in response-based segmentation, i.e., from descriptive to predictive segmentation
References (I)

References (II)

Q & A

- Your comments are appreciated