

ファジィ判別分析による日本人の飲酒態度の測定

朝野 熙彦

1. 伝統的なマーケット・セグメンテーション

1.1 日本人の酒嗜好に関する分析データ

日本人の飲酒態度を把握することを目的として、1991年9月に日本の首都圏でクォータ法で調査を行った。ふだん1週間に1回以上アルコール飲料を飲んでいる20~49才の人が調査対象者で、サンプル数は男性が400人、女性が300人の合計700人である。

アルコール飲料としては、ビール、日本酒、焼酎、ウィスキー、ブランデー、ワイン、リキュール、カクテル類を含む。

本分析で用いた変数は次の2群から成る。

第1群は、「酒嗜好変数」と略称する変数群である。表1に示すお酒の飲み方や態度に関する13項目について、どちらの方に近いかを強制選択させた。これらの変数を消費者のセグメンテーションに利用した。

表1 日本人の酒嗜好変数

外で飲むのが好き	家で飲むのが好き
1人で飲みに行ける	誰かと一緒にないとだめ
会社・仕事関係の人が多く	仕事以外の人が多い
2, 3軒とはしごする方	1軒だけで腰を落ち着ける方
大勢で飲むのが好き	2, 3人で飲むのが好き
賑やかな所がよい	静かな所がよい
新しい店を開拓する方	馴染みの店をひいきする方
強い酒が好き	弱い酒が好き
辛口が好き	甘口が好き
安い酒を沢山飲みたい	少量でも高級酒を飲みたい
新しい酒を試してみる方	飲みつけているものを守る
つまみなして飲める方	料理がないと飲みにくい
同じ酒を1本通して飲む方	チャンポンで飲む方

あさの ひろひこ (株)日本リサーチセンター
〒104 中央区八丁堀 4-5-4

第2の変数群は、酒嗜好のパターンを予測するための10個の説明変数である。その一覧を表2に示す。

表2 説明変数一覧

a) デモグラフィック変数

変数名	コーディングの方法
性別	男性: 1, 女性: 0
職業	就業者: 1, 無職: 0
未婚既婚	既婚者: 1, 未婚者: 0
年齢	実数でインプット

b) ライフスタイル変数

- ・タバコを1日3本以上吸う
 - ・おしゃれに関心がある方
 - ・甘いものが好きである
 - ・カラオケで歌うのが好き
 - ・健康に気を配っている方
 - ・カロリーに気をつけている
- 該当...1, 非該当...0
のダミー変数でデータを記述

1.2 伝統的なセグメンテーション

お酒に対する嗜好性の違いに着目して、全サンプルを6つのクラスターに分類した。まず、酒嗜好変数のデータを林(1965)のパターン分類で分析し、直交座標を2次元まで抽出した。固有値は、順に0.186, 0.148で、累積寄与率は35%であった。

パターン分類によって、2次元の空間に670人の消費者も同時に位置づけられた。ここでは30人がデリートされたが、その理由は13対の酒嗜好変数に対して、1つでも態度保留であった人を分析から除外したからである。

次に、2次元空間内での位置に近い消費者をグルーピングして、全サンプルを6つのクラスターにセグメンテーションした。

クラスター数の最適性基準としては、Milligan and Cooper (1985) のモンテカルロ法を用いた比較実験に従って、Calinski and Harabasz (1974) による疑似統

Cluster	Frequency	Cluster Means		Standard Deviations	
		Dim 1	Dim 2	Dim 1	Dim 2
1	91	0.722	0.288	0.442	0.464
2	161	-1.140	-0.451	0.441	0.488
3	73	1.650	-0.025	0.425	0.517
4	130	-0.628	0.937	0.365	0.422
5	103	0.481	-1.525	0.603	0.416
6	112	0.264	-0.067	0.376	0.362

<Statistics for Variables>
 Pseudo F Statistic=541.88
 Approximate Expected Over-All R-Squaered=0.836

図1 Cluster Summary

計量を用いた。

クラスター分析にはSASのFASTCLUS Procedureを用いた。クラスター分析のサマリーの結果を図1に示す。

各クラスターの類型を、次のように解釈した。

- C 1 : コンパ人間……多勢で賑やかに飲んで騒ぐ、人生を明るくエンジョイするタイプ (14%)
- C 2 : 孤独を愛する人間……誰にも邪魔されず1人静かに酒を飲みたいタイプ (24%)
- C 3 : アルコール大好き人間……アルコール分の強い安酒を沢山飲みたいタイプ (11%)
- C 4 : オバタリアン人間……お酒が大事なのではなく、むしろ食べながらお喋りするの好きなタイプ (19%)

表3 Canonical Discriminant Analysis

<Multivariate Statistics and F Approximation>					
Statistic	Value	F	Num DF	Pr> F	
Wilks' Lambda	0.624	6.52	50	0.0001	
<Test of Ho : The canonical correlations in the current row and all that follow are zero>					
Canonical Correlation	Likelihood Ratio	Approx F	Num DF	Pr > F	
1	0.477	0.624	6.52	50	0.0001
2	0.406	0.807	4.02	36	0.0001
3	0.130	0.967	0.93	24	0.5597
4	0.115	0.983	0.79	14	0.6760
5	0.059	0.997	0.39	6	0.8886

C 5 : 飲んだくれ人間……会社の同僚2, 3人と酒を飲みながらネチネチと愚痴を言っているうちに酔い潰れてしまうタイプ (15%)

C 6 : おつきあい人間……酒は別に好きではないが、つきあいでアルコールが出れば、仕方なく飲むというタイプ (17%)

1.3 グループ分けの構造を知る伝統的なアプローチ

日本人の酒の楽しみ方には、いろいろなタイプがあることがわかったが、次にどのような変数が影響して消費者がこのようなタイプに分れるのかを分析していこう。

この問題に対する伝統的なアプローチは、判別分析を用いることである。6つのクラスターを基準変数のグループ分けとし、表2の10変数を説明変数に用いてSASのCANDISC Procedureで分析した。

分析結果を表3に示す。

表3によれば、母集団において6つのクラスターの平均が等しいという帰無仮説は、危険率1%で棄却されている。

また、表3には、正準相関係数に対するF検定の結果も示されている。これから2番目までの正準相関係数が1%水準で有意であることが判る。

さらに、図2には標準判別係数を1, 2次元空間にプロットした。また、同図には判別スコアのクラスターの重心の位置も、C1~C6として示されている。

従来の消費者調査では、上に述べてきたようなセグメンテーション分析で消費者の類型化を行なうことが多かった。しかし、マーケットの構造自体がZadeh(1965)がいうところのファジィ集合から成っているかもしれない。したがって、グループが排反であるという前提は必ずしも自明なものとはいえないだろう。

同じ人間が「コンパ人間」であるのと同時に、「アルコール大好き人間」でもあるというオーバーラッピングが当

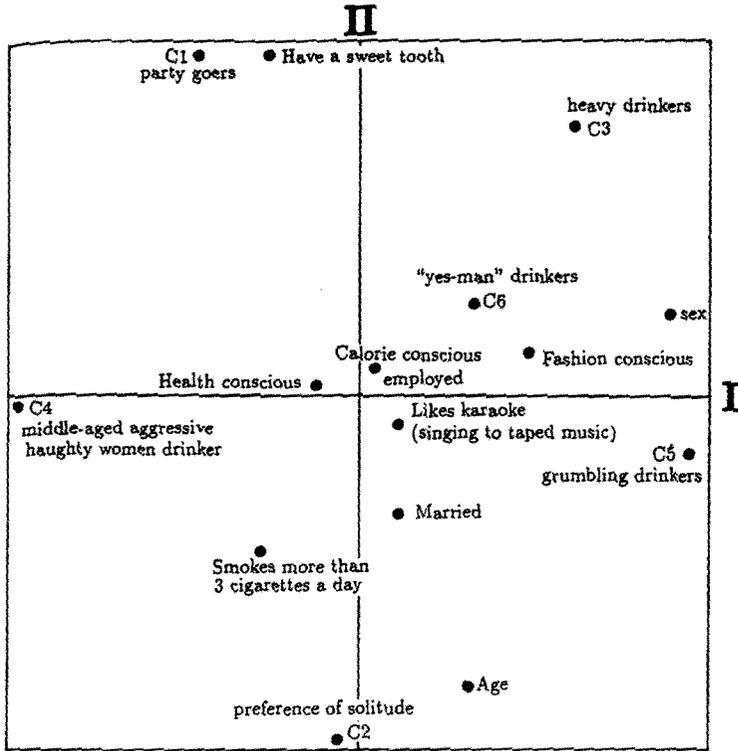


図2 判別空間

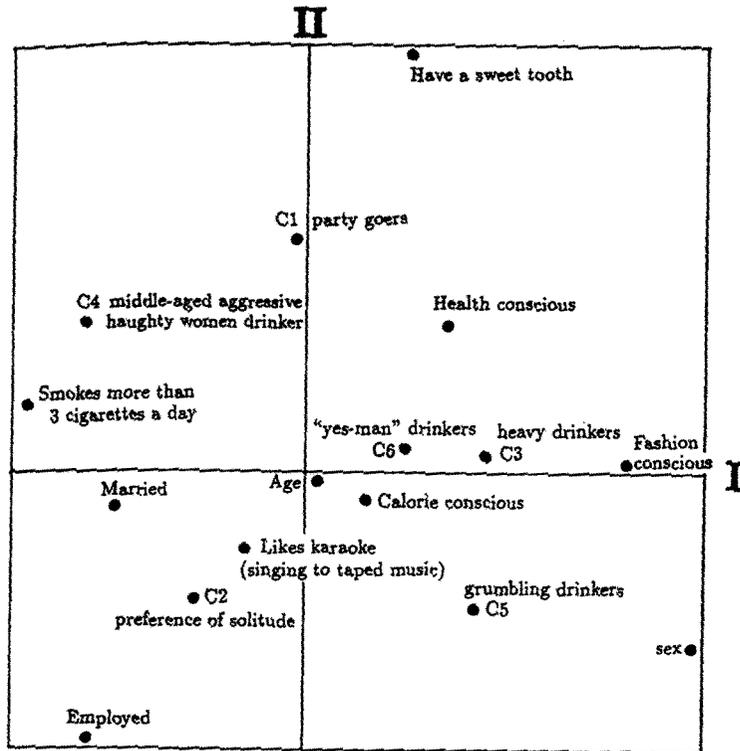


図3 ファジィ判別分析による判別空間

	ケース1	ケース2	ケース3
基準変数のメンバーシップ値行列	1 0 1 0 1 0 0 1 0 1	1 0 1 0 0 1 0 1 0 1	2 0 1 0 0 1 0 1 0 1
説明変数データ行列	4 2 0 4 2 0 1 5 3 0 2 6 6 0 4	4 2 0 1 5 3 0 2 6 6 0 4 6 0 4	4 2 0 1 5 3 0 2 6 6 0 4 6 0 4

図4 分析データ

然考えられよう。

オーバーラッピング・クラスター分析の方法は、すでにCorter and Tversky (1986) らが開発している。しかしオーバーラッピングの有無だけでなく、主として「オバタリアン人間」だが、いくぶんかは「おつきあい人間」でもある、というようにクラスター帰属のグレードまでも分析に加味できれば、なお望ましい。そこで、われわれは飲酒態度に関するデータに和多田ら (1982) のファジィ判別分析 (FDA) を適用した。FDAで分析するには、まず消費者が各クラスターに含まれる確からしさを、メンバーシップ関数で表わす。このメンバーシップ値によって、対象者とファジィ・クラスターとの関係が、クリस्पでない関係として記述できる。

2. ファジィ理論によるアプローチ

2.1 メンバーシップ関数

今回の調査では、メンバーシップ値を、調査時点であらかじめ調査対象者に聞いておくことはできなかった。そこで全く恣意的ではあるが、クラスターに帰属するメンバーシップ値を次の指数関数によって定義することにした。

メンバーシップ関数Mが、クラスターの重心からの距離dに対して単調減少する、と仮定することはブロージブルであろう。われわれは、ここで正規分布の関数を採用した。

平均 μ 、標準偏差 σ の変数 x の正規密度関数 $f(x, \mu, \sigma)$ を $f^*(\mu, \mu, \sigma)$ で割り、

距離を $d=x-\mu$ と変換すると、

$$M(d, \sigma) = f(x, \mu, \sigma) / f^*(\mu, \mu, \sigma)$$

$$= \frac{\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}{\left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{0}{\sigma}\right)^2\right]\right]}$$

$$= \exp\left[-\frac{1}{2}\left(\frac{d}{\sigma}\right)^2\right]$$

2.2 ファジィ判別分析

各クラスターへのメンバーシップ関数を基準変数とし表2の10個の変数を説明変数として、Appendix Aの解法にしたがってファジィ判別分析をかけた。

ファジィ判別分析の判別係数を空間に表現したのが、図3である。

図2と3を比べると、図3の方が、「喫煙」と「健康

	ケース1		ケース2		ケース3	
判別係数	(1)	(2)	(1)	(2)	(1)	(2)
	0.000	0.342	0.000	0.288	0.000	0.342
	-0.152	0.213	-0.167	0.464	-0.152	0.213
	0.152	0.107	0.167	0.290	0.152	0.107
判別スコア	(1)	(2)	(1)	(2)	(1)	(2)
	-0.365	0.022	-0.500	-0.698	-0.365	0.022
	-0.365	0.022	-0.500	0.698	-0.365	-0.044
	-0.365	-0.044	0.500	-0.111	0.548	-0.706
	0.548	-0.706	0.500	0.111	0.548	0.706
	0.548	0.706				

図5 FDAの実験結果

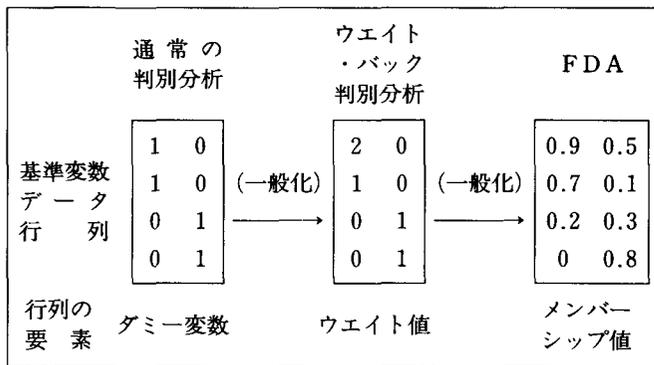


図6 判別分析の一般化

に気を配っている」というライフスタイル変数がセパレートしている点など、判別空間が理解しやすい。

ファジィ判別分析には、次の長所がある。それは、予測すべき基準変数がメンバーシップ関数であることである。その結果、1つのクラスターに属する確からしさだけではなく、それ以外のクラスターに属する確からしさも全く同等に取り扱われることになる。つまり、FDAは判別問題のごく自然な一般化をしていることになる。

3. FDAの理論的な位置づけ

FDAは、正準相関分析に応用上の特定の意味を与えたもの、とみなすことができる。その説明のためにまず標本構成と母集団構成が異なる場合のウエイト・バックの道具としても、FDAが利用できることを示そう。

3.1 ウエイト・バックの必要性

マーケティング・リサーチでは、クォータの設定次第で、調査サンプルが母集団における構成と食い違ってしまうことがよくある。

たとえば、大都市と地方の村で等確率抽出をすると、地方の村から抽出される標本数がきわめて少なくなり、村単位での統計的推論に耐えられなくなるのがふつうである。あらかじめ地方の村には、大きな抽出率を与えてサンプリングを行ない、集計段階で抽出率の逆数に比例した量をウエイトとして与えることが行なわれている。このような集計上の操作をウエイト・バックと呼んでいる。

多変量解析でも、このようなウエイト・バックができれば都合がよい。ウエイトづけに必要な公式をAppendix Bに示した。

3.2 数値例

図4に数値的な実験データを示す。

まず、ケース1を仮想的な母集団のデータだとしよう。基準変数および説明変数データ行列のそれぞれ1行目と2行目は本質的に同一の特性を持った対象者に対応している。同一地域に住み、同じ消費者特性を持ち、しかもデモグラフィックも全く等しいと仮定するので、各行が同等になっている。仮に、これをA地区と呼ぼう。第3行目以下は、B地区という識別の仕方を

する。

次のケース2は、A地区の標本抽出率が50%で、B地区では100%であった場合に相当する。データ数は、A地区が1サンプルとなり、B地区は3サンプルになる。これがクォータの設定によって、抽出率が意図的に変えられた場合を表わしている。ケース3は、基準変数行列の第1行に、ウエイトづけの係数を代入した場合である。ケース2と比べると、ケース2の1行目の要素が抽出率の逆数倍で、2倍になっているところだけが異なる。説明変数のデータ行列は、変わらない。

さて、これら3組のデータをそれぞれ判別分析した結果を、図5に示そう。

ケース1が母集団に対して、通常の判別分析を適用した場合に相当するので、このケースから導かれた判別係数と判別スコアが真のパラメータである。

ケース2は、抽出率にバイアスのかかった標本データを単純に判別分析した場合にあたる。マーケティング・リサーチでは、通常このタイプのデータ解析をしていることになる。パラメータの推定値は、判別係数、判別スコアとも真の値と乖離がある。ケース3がウエイト・バック情報をFDAで処理した場合にあたり、これは真のパラメータを正しく再現できている。

3.3 討論

伝統的な判別分析と、ウエイト・バック判別分析およびFDA三者の関係は、図6のように整理できる。

基準変数のデータ行列の要素がダミー変数の場合が判別分析であり、ウエイト値の場合が母集団ウエイト・バックであり、そして、メンバーシップ値の場合がファジィ判別分析になる。これら3通りの分析法は、原理的に等しい。ファジィ判別分析が最も一般的なケースで、他の2つの分析法はその特殊な場合にあたる。

分析方法を一般化するほど、分析データの制約条件がゆるめられていくことがわかる。FDAのデータ行列から明らかのように、FDAは2つの変数群の正準相関を最大化する正準相関分析と原理的に等しい。つまり、ノンメトリック・データのメトリック復帰という見方もできよう。

本研究の実施にあたっては、榎東急エージェンシーから分析データの提供を受けた。感謝の意を表す。

Appendix A Fuzzy Discriminant Analysis

〔記法〕

$y=(y_i)$: 合成変数ベクトル

$B=(b_{ig})$: 基準変数のメンバーシップ値行列

$Z=(Z_{ij})$: 説明変数のメンバーシップ値行列

$\omega=(\omega_j)$: 説明変数のウェイトベクトル

N_g : 行列 B の列和

N : 行列 B の要素の総和

対象者 $i=1,2,\dots,n$

マーケット・セグメント $g=1,2,\dots,G$

説明変数 $j=1,2,\dots,p$

y について次の構造モデルを考える。

$$y=Z\omega \quad (1)$$

y の平方和を T 、基準変数のグループ分けによる級内分散を C とおくと、相関比は次のように定義される。

$$\eta^2 = \frac{C}{T} \quad (2)$$

さらに y のグループ平均と総平均をそれぞれ \bar{y}_g, \bar{y} とする。

$$\bar{y}_g = \frac{1}{N_g} y' b_g \quad (3)$$

$$\bar{y} = \frac{1}{N} y' B1_c \quad (4)$$

ここで、 b_g は b の第 g 列ベクトルで1は1を要素とする定数ベクトルである。

さらに対角行列 N は次のように定める。ここで N_g は行列 N の主対角要素である。

$$N = \text{diag}(N_g) \quad (5)$$

$$M = (\mu_{ij}) = N^{-1} BZ \quad (6)$$

$$\mu = \frac{1}{N} Z' B1_c \quad (7)$$

また、

$$\text{diag}(b_i)_g = \begin{bmatrix} b_{1g} & & 0 \\ & \ddots & \\ 0 & & b_{ng} \end{bmatrix} \quad (8)$$

次に行列 M_n と M_c を(9)(10)式で定める。

$$\overline{M}_n = 1_n \mu' = \begin{bmatrix} \mu_1 & \dots & \mu_p \\ \mu_1 & \dots & \mu_p \end{bmatrix} \quad (9)$$

$$\overline{M}_c = 1_c \mu' = \begin{bmatrix} \mu_1 & \dots & \mu_p \\ \mu_1 & \dots & \mu_p \end{bmatrix} \quad (10)$$

〔問題の定式化と解法〕

通常の重判別分析と同様に、 η^2 を最大化するように ω を推定する。

$$\begin{aligned} T &= \sum_{g=1}^G \sum_{i=1}^n (y_i - \bar{y})^2 b_{ig} \\ &= \sum_{g=1}^G \left(y - 1_n \frac{1}{N} y' B1_c \right)' \text{diag}(b_i)_g \\ &= \sum_{g=1}^G (Z\omega - 1_n \omega' \mu)' \text{diag}(b_i)_g (Z\omega - 1_n \omega' \mu) \\ &= \sum_{g=1}^G ((Z - \overline{M}_n) \omega)' \text{diag}(b_i)_g (Z - \overline{M}_n) \omega \end{aligned} \quad (11)$$

$$\begin{aligned} C &= \sum_{g=1}^G N_g (\bar{y}_g - \bar{y})^2 = \sum_{g=1}^G N_g \left(\frac{1}{N_g} y' b_g - \frac{1}{N} y' B1_c \right) \\ &= \left(N^{-1} B' Z\omega - 1_c \frac{1}{N} \omega' Z' B1_c \right)' \\ &\quad N \left(N^{-1} B' Z\omega - 1_c \frac{1}{N} \omega' Z' B1_c \right) \\ &= (M\omega - 1_c \omega' \mu)' N (M\omega - 1_c \omega' \mu) \\ &= ((M - \overline{M}_c) \omega)' N (M - \overline{M}_c) \omega \end{aligned} \quad (12)$$

(2), (11), (12)式より $\eta^2 T = C$ を ω に関して偏微分し、

$$\frac{\partial \eta^2}{\partial \omega} = 0 \text{ とおけば} \quad (13)$$

$$\frac{\partial C}{\partial \omega} = \eta^2 \frac{\partial T}{\partial \omega} \quad (14)$$

$$\begin{aligned} \frac{\partial C}{\partial \omega} &= 2(M - \overline{M}_c)' N (M - \overline{M}_c) \omega = 2H\omega \\ \frac{\partial T}{\partial \omega} &= 2 \sum_{g=1}^G (Z - \overline{M}_n)' \text{diag}(b_i)_g (Z - \overline{M}_n) \omega \\ &= 2F\omega \end{aligned} \quad (15)$$

したがって、

$$H\omega = \eta^2 F\omega \quad (16)$$

固有値 η^2 と固有ベクトル ω は、(16)式の固有値関数を解いて推定される。

Appendix B 対象者にウェイトをつけるための内積統計量

〔記法〕

対象者 $i=1,2,\dots,n$

変数 $j=1,2,\dots,p$

原データ行列 $XY=(X_{ij})$

ウェイト行列 $W = \text{diag}(W_i)$ これは対角行列

ウェイトの和 $T = trW$

変数 j の加重平均 $\bar{x}_j = \frac{1}{T} \sum_i \omega_i x_{ij}$

平均偏差データ行列 $X = (x_{ij} - \bar{x}_j)$

X の第 j 列 x_j の分散は, $S_j^2 = \frac{1}{T} x_j' W x_j$

ここで, 標準偏差 S_j の逆数を主対角要素とする対角行列を

$D_c^{-1/2} = \text{diag}\left(\frac{1}{S_j}\right)$ とおけば,

基準化データ行列は, $Z = X D_c^{-1/2}$

以上の記法を用いれば, 多変量解析に必要な内積統計量は次のように表現できる.

分散共分散行列 $C_{xx} = \frac{1}{T} X' W X$

相関行列 $R = \frac{1}{T} Z' W Z$

参考文献

- [1] Calinski, R. B. and Harabasz, J. (1974) A dendrite method for cluster analysis. *Communications in Statistics*, 3, 1-27.
[2] Corter, J. E. and Tversky, A. (1986) Extended

similarity trees, *Psychometrika*, 51, 429-451.

- [3] Hayashi, C. (1965) On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematical statistical point of view. *Annals of the Institute of Statistical Mathematics*, 3, 69-98.
[4] Milligan, G. W. and Cooper, M. C. (1985) An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50, 159-179.
[5] Watada, J., Tanaka, H. and Asai, K. (1982) Fuzzy quantification theory type II. *The Japanese Journal of Behaviormetrics*, 9, No. 2, 24-32.
[6] Wedel, M. and Steenkamp J-B, E. M. (1991) A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. *Journal of Marketing Research*, 28, 385-396.
[7] Zadeh, L. A. (1965) Fuzzy sets. *Information & Control*, 8, 338-35.