

마케팅자료에서 특성점들을 이용한 군집방법

문숙경[†], 김우성

목원대학교 정보통계학과

Clustering Method Using Characteristic Points with Marketing Data

Soog-Kyung Moon, Woo-Sung Kim

Dept. of Information & Statistics, Mokwon University

Key Words : growth velocity curve, characteristic points, characteristic function

Abstract

We got the growth distance curve by spline smoothing method with observed marketing data and the growth velocity curve by the derivation of the growth distance curve. Using this growth velocity curve, we defined the several characteristic points which describe the variation of marketing data. In this paper, to specify several patterns of marketing data, we suggested characteristic function by using these characteristic points. In addition, we applied characteristic function to the seventeen brands of electric home products data.

1. 서론

이윤을 최대의 목표로 하는 기업에서는 다른 경쟁업체보다 앞서기 위해서는 여러 가지 시장 자료들을 취합하여 정보를 얻고 생산량 조절, 광고의 필요성 여부 및 시기 결정, 기존 제품 폐기 및 신제품 출시 등에 대한 결정을 신속하고 정확하게 내려야 할 것이다. 이런 시장 자료 중 대표적인 것이 특정 제품에 대한 소비자들의 구매량 즉, 판매량에 대한 자료일 것이다. 본 고에서는

특정 제품에 대하여 여러 경쟁 업체들이 각각의 고유의 상표를 가지고 생산하는 제품들의 판매량 자료를 바탕으로 비슷한 패턴을 가지는 기업들을 그룹화하는 군집분석을 실시하는 방법을 제시하려 한다.

비슷한 판매유형의 경쟁사들에 대한 광고 전략 및 판매 전략 등을 참고하여 시장에서의 여러 경영 의사결정을 신속하고 용이하게 내릴 수 있을 것이다. 이러한 군집분석 연구에서는 비선형모형 보다는 오히려 평활법(Smoothing Method)이나, 비선형함수의 혼합함수를 종종 이용한다. 여기에 주로 사용되어지는 기저 함수를 성장거리곡선

[†] 교신저자 skmoon@mwus.mokwon.ac.kr

(Growth Distance Curve)이라하며, 1차 혹은 2차 미분에 의하여 도출된 함수를 성장속도곡선(Growth Velocity Curve)이라 한다. 성장속도곡선을 통해, 성장기, 쇠퇴기 등과 같은 변곡점(Inflexion Point)들을 발견할 수 있는데, 이를 동물이나 사람의 성장에 관련된 여러 특징을 결정 지우는 특성점(Characteristic Points)이라 불리운다.

평활법은 오래 전부터 다수의 연구가 진행되어 왔고, 이의 응용예가 최근, 동물연구를 중심으로 한 성장곡선분석에도 많이 적용되고 있다. 이와 같이 성장곡선분석에 적용하기 위한 평활법은, 그 특징에 따라 여러 가지 방법들이 응용되고 있다. 특히, Spline법, 커널에 의한 함수추정법, LOWESE(Locally Weighted Smoothing Estimation), 게다가 최근에는 붓스트랩에 의한 방법 등이 응용되고 있다. 이에 대한 연구로는 Ashizawa(1995), Hamada(1994), Leigh(1996), Koppe *et al.*(1999), Lee *et al.*(1999) 등이 있다. 참고로, 대부분의 시장조사 자료의 경우 매우 고가이며, 조사의 어려움 때문에 장기간의 데이터를 입수하기가 곤란한 것이 대부분이다.

본 고에서는 관찰된 판매량 자료로부터 Spline 평활법을 이용하여 성장거리곡선을 구하고, 성장거리곡선에서 간단한 도함수에 의하여 성장속도곡선을 구할 것이다. 얻어진 이 속도곡선을 통해, 다양한 특성점(Characteristic Points)들을 2장에서 정의하였고 3장에서는 이들 특성점들을 이용하여 특성 함수를 제시할 것이다. 이 특성 함수의 결과 값에 따라 몇 가지 유형으로 분류할 것이다. 본 고에서 제시하는 특성 함수의 특징은 최근 자료가 수개월 전의 자료보

다 더 비중 있게 처리되도록 가중치를 부여한 것이라 할 수 있다. 실제로 특정 국가에서 17개 상표별로 3년간 조사된 한 가전제품의 판매대수 자료를 4장에서 예제로 적용하였고, SAS의 클러스트 분석(Average연결법)에 의한 기존의 군집 방법과 비교함으로써 기업의 생존과 번영을 결정짓는 전략적 특성점의 도출이 기대된다.

2. 성장 곡선 및 특성점 정의

2.1 성장 곡선 도출

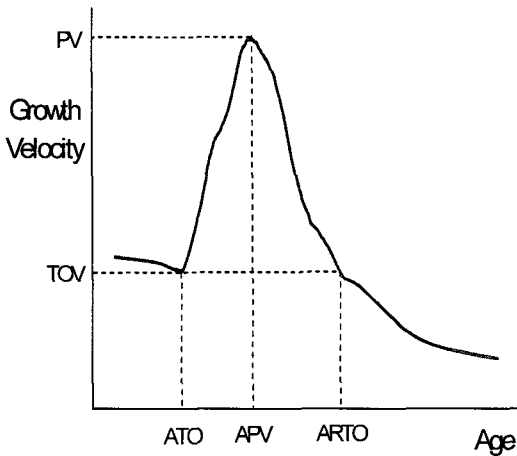
특정 시장에서의 A 브랜드에 대해 조사된 판매량 y_A 는 $y_{A1}, y_{A2}, \dots, y_{At}$ 와 같이 쓸 수 있다. 이 때, t 는 시점을 나타내는 것으로, 대체로 작은 숫자이며, 이는 단기간임을 나타낸다고 볼 수 있다.

성장거리곡선(Growth Distance Curve)은 상기 언급한 것처럼 판매량 자료로부터 Spline 평활법을 적용하여 구하였는데, SAS macro를 작성하여 프로그램을 실행시킨 결과물이다. 성장거리곡선(Growth Distance Curve)을 미분하면 성장속도곡선(Growth Velocity Curve)을 도출할 수 있다. 본 고에서는 다음과 같이 수치미분을 적용하여 성장속도 곡선을 구하였다. 시점 t 가 동일한 간격으로 변동하고 있으므로 t 시점에서의 수치미분은 t 시점에서의 단위 성장율(r_t)을 계산한 것과 동일해진다. 단, 아래 (1)식에 사용되어진 판매량 y_{At} 들은 Spline 평활법이 적용된 성장거리곡선에서 도출된 값들이다.

$$r_t = \frac{\delta_t}{y_{At-1}} = \frac{(y_{At} - y_{At-1})}{y_{At-1}} \dots\dots\dots(1)$$

2.2 특성점 정의

일반적으로 동물의 성장패턴에서 그 특성을 대표하는 특성점들로 APV(Age at Peak ; 성장이 가장 활발한 지점), ATO(Age at Takeoff Velocity ; 성장이 가장 둔한 지점), ATRO(Age at Return to Takeoff Velocity ; 성장의 재감소 지점) 등이 있다(Leigh, 1996). <그림 1>은 대표적인 속도함수에서의 특성점들을 보여 준다.



<그림 1> 성장속도곡선의 특성점 (Leigh, 1996)

위에서 구한 성장 속도 곡선에서 다음과 같이 여러 개의 특성점들을 정의한다.

동물의 성장패턴의 경우는 상기와 같은 3개의 특성점들로 충분한 설명이 이루어진다. 그러나 시장데이터의 경우는 경영정책, 마케팅전략, 경영환경 등에 따라 동물 성장과 같은 일관성이 결여되는 경우가 비일비

재하다. 따라서, 동물성장과 같이 3개의 성장 특성점만으로는 개체의 성장패턴을 파악하는 것은 불가능하다.

제품의 측면에서 고려해 보면 제품의 판매주기에 따라 PV, TOV 등이 존재할 수 있다. 경쟁시장의 측면에서는 경쟁브랜드의 침투에 따라 PV, TOV가 여러 개 발생할 수 있다. 또한, 의사결정자의 입장에서는 시장 점유 목표를 유지하고 있는지, 최대 피크, 최저 피크, 전반적인 추세 등에 따라 시장 진입에 대한 여러 경영 판단을 내릴 수 있다. 특히, 이러한 시장특성들이 언제 발생했는지 파악한다는 것은 향후 마케팅 활동의 매우 중요한 요인으로 작용한다. 본 고에서는 <표 1>과 같은 성장 특성점들을 고려한다

- ① PV, PPV(Period at Peak Velocity): 성장속도 값들의 증가 후 감소하는 것들 중에서 최대값과 그 시점
- ② TOV, PTOV(Period at Takeoff Velocity): 성장속도 값들의 감소 후 증가하는 것들 중에서 최소값과 그 시점
- ③ Median, PMedian: 성장속도 값의 크기 순서에서 가운데 위치한 값과 그 시점
- ④ Mean, Mean times: 성장속도 값들의 평균값과 평균을 통과하는 시점
- ⑤ Min, PMin: 성장속도 값들 중에서 최소값과 최소값이 있는 시점
- ⑥ Max, PMax: 성장속도 값들 중에서 최대값과 최대값이 있는 시점

<표 1>은 앞서 정의한 여러 특성점들을 제품의 측면에서 시장에서의 판매/광고 및 신제품 투입 여부 등에 대한 결정을 내려할 전략 시점을 파악하는데 도움이 될 수 있음을 보여준다.

〈표 1〉 특성점들을 이용한 시장에서의 전략

| Characteristic Point | Usage | Strategy |
|----------------------|-------------------|-------------|
| PV, PPV | 매출 최대 성장률 및 시기 | 시장 경쟁 전략 |
| TOV, PTOV | 매출 쇠퇴 시작 성장률 및 시기 | 신제품 투입 전략 |
| Median, PMedian | 평균 매출 성장률 추세 | 향후 추세 판단 |
| Mean, Mean times | 평균 매출 성장률 추세 | 향후 추세 판단 |
| Min, PMin | 최소 매출 성장률 및 시기 | 마케팅전략(광고 등) |
| Max, PMax | 최대 매출 성장률 및 시기 | 이미지 메이킹 등 |

3. 특성함수 도출

3.1 성장 속도 곡선의 유형

대체로 성장 속도 곡선은 곡선 모양이므로 PV와 TOV의 존재 여부에 따라 다음과 같이 몇가지 유형으로 분류할 수 있다. PV만 존재하는 경우는 위로 볼록한 유형이므로 위로 볼록한 지점(PPV)과 최소치를 파악하는 것으로, TOV만 존재하는 경우는 아래로 볼록한 유형이므로 아래로 볼록한 지점(PTOV)과 최대치를 파악하는 것으로, 그리고 PV와 TOV가 모두 존재하지 않는 경우는 선형을 띠는 유형이므로 기울기와 절편을 파악하는 것으로 비교적 용이하게 판매에 관한 유형을 파악할 수 있을 것이다. 그러나 PV와 TOV가 모두 존재하는 경우에는 각 회사들의 판매유형들을 분류하기가 매우 복잡한 다양한 형태를 취하는 것이 보통이다. 흔히 이 경우에는, PV의 크기, TOV의 크기, PPV, PTOV, 최소성장, 최대성장, 최소성장 시점, 최대성장 시점 등을 동시에 고려하여야만 할 것이다. 본 고에서

는 PV와 TOV가 모두 존재하는 경우 여러 가지 특성점들을 동시에 고려하는 특성 함수를 개발하여 이를 통해 몇 가지 유형으로 분류할 수 있는 기준을 제시하고자 한다.

3.2 특성 함수 도출

얻어진 특성점들에 대하여 전체 성장률의 범위 대비 성장피크와 쇠퇴성장의 높이의 비율을 고려한 아래 식을 우선 제시 하고, 이 식이 가진 의미를 해석하여 본 고에서 제시하는 특성 함수를 개발 할 수 있었다.

$$i_B = \frac{\text{growth spread}}{\text{growth range}} = \frac{(PV - TOV)}{(Max - Min)} \quad (2)$$

식(2)에서의 비율은 전체 성장의 변화를 나타낸다고 볼 수 있다. 즉, 이 비율이 1에 가까우면 가까울수록 성장변화가 매우 크게 되고, 0에 가까울수록 성장변화가 매우 적음을 의미한다. 그러나, 이 비율은 시점을 고려하지 못하였다. 즉, 전반기의 성장피크와 후반기의 성장피크를 동일한 비율로 나타낸다. 따라서, 성장피크 시기 및 쇠퇴 후

재성장의 시기에 대하여 동시에 고려되어야 할 것이다.

시장에서의 매출전략이 최근 시점의 데이터에 주안점을 두는 것을 근거로 하여 각 시점들을 전체 시점수로 나누어 준 값을 가중치로 고려할 수 있다. 즉, 최근시점일수록 가중치는 1에 가까워지고, 과거시점일수록 0에 가까워지도록 가중치가 결정된다. 이와 같은 가중치를 위식에 적용하면 다음과 같은 식을 도출할 수 있다.

$$w_{iB} = \frac{\text{weighted growth spread}}{\text{weighted growth range}} \dots (3)$$

여기서,

분자(weighted growth spread)

$$= \frac{(PV \times PPV - TOV \times PTOV)}{\text{period 수}}$$

이고,

분모(weighted growth range)

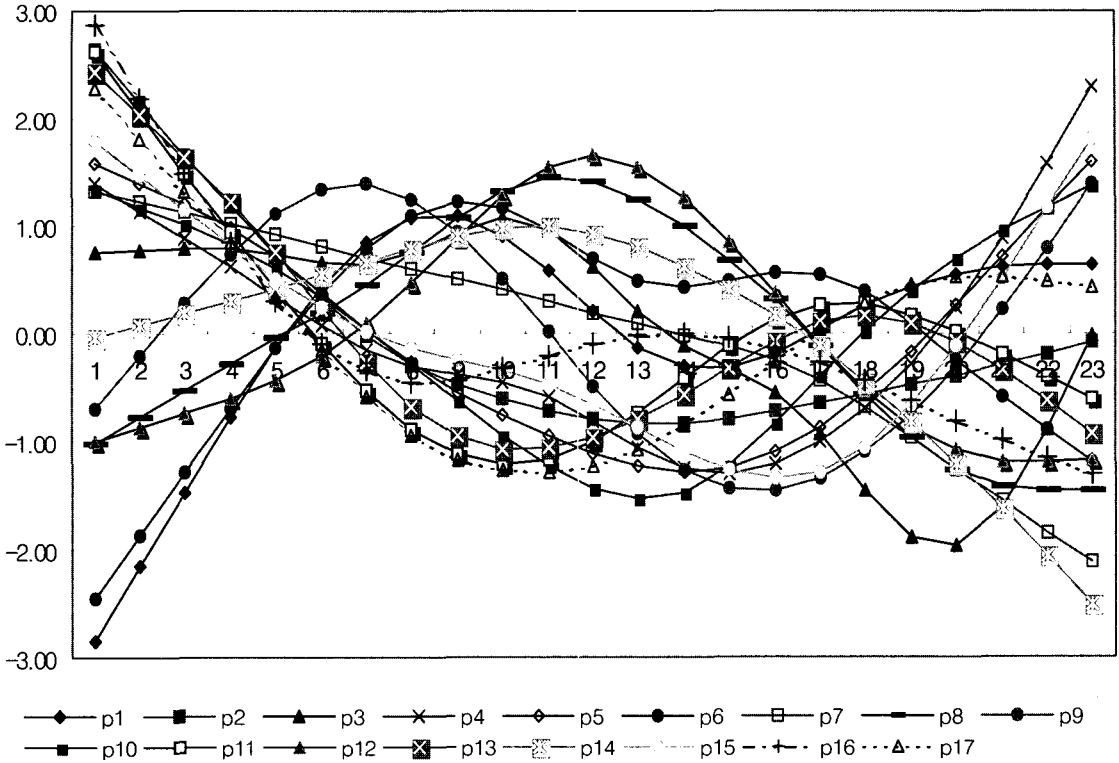
$$= \frac{(Max \times PMax - Min \times PMin)}{\text{period 수}}$$

이며, 여기서의 period수는 PPV와 PTOV 사이의 시점 차이를 의미한다.

식(3)의 결과 값은 상대적인 수치이지만, 그 값의 크기(경험적인 측정 기준)에 따라 <표 2>와 같이 성장패턴의 특성을 해석할 수 있다.

<표 2> 결과값에 따라 해석된 유형 분석

| w_{iB} | 특 징 |
|-------------------|---|
| 큰 음수 (-100 미만) | 대개 W자 모양의 성장패턴을 나타낸다. growth spread가 매우 작으며 전반, 후반에 최대값을 가지는 특징이 있다. (매출변화가 크며 증가나 감소 패턴이 없음) |
| 작은 음수 | 대개 M자 모양의 성장패턴을 나타낸다. growth spread가 매우 작으며 전반, 후반에 최소값을 가지는 특징이 있다. (매출변화가 크며 증가나 감소 패턴이 없음) |
| 0 | 성장 특성점(PV, TOV)이 존재하지 않거나 growth spread가 매우 작으며 지수 증가형태의 성장패턴을 나타낸다. (전반적으로 증가추세이나 중반기에 커다란 매출변화를 나타냄) |
| 작은 양수 (100 미만) | 성장 특성점(PV, TOV)이 존재하지 않거나 growth spread가 매우 작으며 지수 증가형태의 성장패턴을 나타낸다. (전반적으로 증가추세이나 중반기에 커다란 매출변화를 나타냄) |
| 양수 | 전반기에 매우 높은 성장을 보이다가 중반기에 PV와 TOV가 공존하며, 이후 후반기는 감소하는 성장패턴을 나타냄 (전반, 중반, 후반에 걸쳐 불안정한 매출변화를 나타냄) |



<그림 2> 17개 브랜드의 성장 속도 곡선

4. 적용 예

이 절에서는 특정 국가에서 판매되고 있는 주요 17개 브랜드들의 특정 가전제품에 대한 판매량 자료에 기초한 사례를 제시하려 한다. 기업 사정상 자세히 밝힐 수는 없지만 1998년 2월부터 2001년 11월까지 매월 판매량을 기록해놓은 자료를 적용하였다. 우선 각 브랜드에 대하여 성장 속도 곡선을 구하고 이를 그래프로 나타내면 <그림2>와 같다. 이로부터 도출된 17개 상표에 대한 특성점을 정리하면 <표 3>과 같다. <표 3>에서

①그룹(TOV만 존재하는 경우)은 특성 함수와 무관하게 동일한 성장패턴을 한다고 생각할 수 있다. 2, 4, 5, 10, 15번 브랜드는 TOV만 존재하기 때문에 아래로 볼록한 성장패턴을 나타내고 있다. 공교롭게도 이들 브랜드의 특성 함수의 범위는 45로 전체 특성화 함수의 결과 값이 5%내이다. 그러므로 이들 5개의 브랜드는 특정 국가에서의 시장 매출의 성장패턴이 동일하다고 말할 수 있다.

<표 3>의 결과를 요약하면 17개 브랜드의 매출 성장패턴을 모두 5개의 그룹으로 분류할 수 있다. 이 결과를 <표 4>에 정리하였다.

<표 3> 성장 특성점 (단위: %, 개체번호)

| ID | PV | PPV | TOV | PTOV | Median | PMedian | Min | PMin | Max | PMax | Mean | Mean times | RANGE | DUMMY | wi |
|----|------|-----|-------|------|--------|---------|-------|------|------|------|------|------------|-------|-------|-------|
| 1 | 10.5 | 9 | 5.6 | 15 | 7.6 | 18 | -3.1 | 1 | 10.5 | 9 | 6.7 | 3 | 4.2 | ④ | 10.7 |
| 2 | | | 0.1 | 13 | 2.6 | 20 | 0.1 | 13 | 19.3 | 1 | 4.7 | 1 | 0.8 | ① | -7.6 |
| 3 | 3.6 | 10 | 2.2 | 20 | 3.4 | 12 | 2.2 | 20 | 3.6 | 10 | 3.1 | 1 | -0.3 | ④ | 100.0 |
| 4 | | | -0.2 | 15 | 1.7 | 19 | -0.2 | 15 | 6.4 | 23 | 2.2 | 2 | 6.5 | ① | 1.5 |
| 5 | | | -2.6 | 14 | 1.1 | 19 | -2.6 | 14 | 6.9 | 23 | 1.1 | 2 | 8.4 | ① | 18.6 |
| 6 | 7.8 | 9 | 6.0 | 14 | 5.9 | 18 | -0.8 | 1 | 7.8 | 9 | 5.0 | 2 | 3.1 | ④ | -18.6 |
| 7 | | | | | 1.9 | 13 | -13.9 | 23 | 9.6 | 1 | 0.5 | 1 | 14.3 | ③ | 0.0 |
| 8 | 3.5 | 11 | -3.8 | 22 | -0.3 | 5 | -3.8 | 22 | 3.5 | 11 | -0.1 | 2 | 5.3 | ④ | 100.0 |
| 9 | 7.3 | 7 | 0.5 | 16 | 4.0 | 11 | 0.5 | 16 | 7.3 | 23 | 4.0 | 3 | 7.0 | ④ | 27.0 |
| 10 | | | -6.6 | 13 | 0.3 | 7 | -6.6 | 13 | 6.3 | 23 | 0.2 | 2 | 10.0 | ① | 37.4 |
| 11 | 10.3 | 18 | -1.3 | 10 | 7.3 | 15 | -1.3 | 10 | 28.5 | 1 | 8.1 | 3 | 1.8 | ⑤ | 474.8 |
| 12 | 9.1 | 12 | -3.3 | 22 | 1.0 | 6 | -3.3 | 22 | 9.1 | 12 | 1.9 | 2 | 7.9 | ④ | 100.0 |
| 13 | 4.5 | 18 | -1.4 | 10 | 2.4 | 7 | -1.4 | 10 | 15.2 | 1 | 3.7 | 3 | 1.3 | ⑤ | 322.7 |
| 14 | 3.7 | 11 | | | -3.0 | 4 | -30.1 | 23 | 3.7 | 11 | -5.9 | 2 | 31.9 | ② | 5.6 |
| 15 | | | -17.4 | 16 | -1.7 | 8 | -17.4 | 16 | 23.7 | 23 | 0.1 | 2 | 35.7 | ① | 33.8 |
| 16 | 10.5 | 14 | 5.0 | 8 | 8.1 | 11 | -5.3 | 23 | 45.5 | 1 | 10.5 | 1 | 7.3 | ⑤ | 63.5 |
| 17 | 10.5 | 21 | -0.4 | 11 | 7.6 | 17 | -0.4 | 11 | 20.9 | 1 | 7.3 | 2 | 1.1 | ⑤ | 891.2 |

* Dummy는 PV와 PTOV의 관계*

- * ① PV가 없는 경우, ② TOV가 없는 경우, ③ PV/TOV가 없는 경우,
- ④ PPV<PTOV의 경우, ⑤ PPV >PTOV의 경우

<표 4> 그룹 분류 결과

| 그룹 | 브랜드 번호 | 특성 |
|------|-----------------|-----------------|
| 그룹 A | 2, 4, 5, 10, 15 | 아래로 볼록한 성장 패턴 |
| 그룹 B | 14 | 위로 볼록한 성장패턴 |
| 그룹 C | 7 | 선형적으로 감소하는 성장패턴 |
| 그룹 D | 6 | -18.6 |
| | 1 | 10.7 |
| | 9 | 27.0 |
| | 3 | 100.0 |
| | 8 | 100.0 |
| 그룹 E | 12 | 100.0 |
| | 16 | 63.5 |
| | 13 | 322.7 |
| | 11 | 474.8 |
| | 17 | 891.2 |

M자 모양의 성장패턴
양의 3차 함수모양
종모양에서 왼쪽으로 치우친 성장패턴

음의 3차 함수 모양
음의 3차 함수 형태이나 전반의 성장률이 매우 큼

특히, 그룹D와 그룹E는 특성 함수의 결과 값에 따라 몇 개의 소그룹으로 분류할 수 있다. <표 4>에서 보면 브랜드 번호 14와 브랜드 번호 7 각각 한 개의 브랜드로 각각 한개씩의 그룹을 생성하여야하는 사실이 다소 아쉬움으로 남는다.

끝으로, 본 고에서 제시한 방법과 기존의 방법을 사용하여 구한 결과를 비교 분석하기 위하여, 군집분석으로 많이 활용되어져 온 방법들 중 SAS에서 제공하는 클러스터 분석(Average연결법)에 의한 결과를 <표 5>에 요약하였다.

<표 5> 클러스터분석에 의한 군집분류 결과

| 그룹 | 브랜드 번호 |
|------|---------------------------------|
| 그룹 ① | 1, 6, 9, 10, 11, 12, 13, 14, 15 |
| 그룹 ② | 16, 17 |
| 그룹 ③ | 2, 5 |
| 그룹 ④ | 7, 8 |
| 그룹 ⑤ | 3, 4 |

비교 분석 결과 두 방법 간에는 다소 차이가 존재하는 것을 볼 수 있으나, 최종적으로 어느 것이 나은지는 더 많은 실험으로 판정될 것이다. 다만, 현실적으로 매출액 자료만 존재하는 상황 하에서 클러스터 분석으로는 그룹화는 가능하지만 그룹별 특징에 관한 정보를 얻기에는 어려움이 존재한다는 것을 알 수 있고, 본 고에서 제시하는 방법으로는 각 그룹에 속하는 브랜드의 판매량에 관한 성장 패턴을 파악할 수 있었다는 것이 이 연구의 소결론이다.

5. 결론

본 논문은 인류학, 생명과학 등의 분야에서 일부 소개된 성장(growth) 곡선과 관련된 기법을 marketing 분야의 clustering의 문제에 적용해본 연구 결과입니다. 성장곡선분석에서 계산되는 성장 특성점들 중 PV와 TOV가 모두 존재하는 경우에 유형이 복잡함으로 특성함수를 제시하여 성장패턴을 쉽게 분류할 수 있는 기준을 제공하고 있다. 일반적인 경우, 클러스터분석이나 성장 특성점만으로 해석이 가능하지만, 성장패턴이 복잡한 마케팅자료에서는 해석의 기준이 되는 특성 함수를 개발 적용할 필요가 있다. 또한, 특성 함수는 동일한 개체를 분류하거나, 단기간의 데이터에 대하여 분류할 때에도 적용될 수 있다. 또한 기업에서는 이러한 전략적 특성점을 통하여 경쟁기업과의 경쟁에서 신속한 전략도출이 용이하게 될 것으로 기대된다. 그러나 본고의 특성상 자료수집의 제한으로 인하여 소량의 데이터를 분석대상으로 하였으나 차후 이러한 단점을 보완하기 위하여 시뮬레이션에서의 데이터 생성을 통하여 결과의 확인 및 검토를 통하여 수정 보완될 것으로 생각된다.

참고문헌

- [1] Ashizawa, K., Kato, S. and Eto, M.(1995), Individual adolescent growth of stature, body weight, and chest circumference of girls in Tokyo, Journal of Anthropological Science, Vol. 102, pp.421-426.
- [2] Hamada, Y.(1994), Standard Growth Patterns and Variation in Growth Patterns of the Japanese Monkeys

- (*Macaca fuscata*) Based on an Analysis by the Spline Function Method, *Anthropological Science* 102, pp.57-76.
- [3] Han, J. and Kamber, M.(2000), *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers
- [4] Koppe, T., Swindler, D.R. and Lee, S.H.(1999), A longitudinal study of the growth pattern of the maxillary sinus in the pig-tailed macaque (*Macaca nemestrina*), *Folia Primotologica*, Vol. 70 No. 6, pp.301-312.
- [5] Lee, S.H., Otake, M. and W. J. Schull(1999), Changes in the pattern of growth in stature related to prenatal exposure to ionizing radiation, *International Journal of Radiation Biology*, Vol. 75 No. 11, pp.1149-1158.
- [6] Leigh, S.R.(1996), Evolution of human growth spurts, *Journal of American Physics Anthropology*, Vol. 101, pp.455-474.
-