

# 고객 Clustering 을 위한 Blog 감성 추출에 대한 연구

배상근\*, 강재우\*\*

\*고려대학교 컴퓨터정보통신대학원

\*\*고려대학교 컴퓨터학과

e-mail : [sam90@korea.ac.kr](mailto:sam90@korea.ac.kr)

## A Study on Extracting Customer Emotions from Blog and Clustering for Target Marketing

Sangkeun Bae\*, Jaewoo Kang\*\*

\*Graduate School of Computer Information & Communication, Korea University

\*\*Dept. of Computer Science & Engineering, Korea University

### 요 약

Blog 는 개인의 여러 미묘한 감정과 감성들을 표현하고, 이를 소통하는 Communication Channel 역할을 하고 있으며, 또한 누구나 접근할 수 있게 되었다. 이는 각 기업에게, 기존의 비효율적인 Mass Marketing 방법에서 벗어나, 소비자의 감성을 자연스럽게 추출하여 세련된 Target Marketing 을 할 수 있는 훌륭한 기회를 제공하게 된다. 하지만, 고객의 Blog 로부터 미묘한 감성지수를 추출하고, 이를 마케팅 방법에 접목시키는 것은 쉽지 않은 일이다. 이러한 문제를 해결하기 위해서 본 논문에서는 고객 회원정보에 등록된 Blog 를 이용하여, Target Marketing 에 활용할 수 있는, 고객 Clustering 을 위한 Blog 감성 추출에 대한 연구를 수행하였다. Blog 의 Main Skin Image 를 통해 지배적인 채도와 명도를 추출하여 수치화하고, 이를 바탕으로 고객 Blog 를 테스트 스케일법(\*일본감성연구소 개발방법)의 실증된 감성 Group 별로 Clustering 하였다. Clustering 된 각 Blog 사용자를 대상으로 연관 배색에 대한 감성 설문조사를 실시한 결과, 유의한 실험결과가 도출되어 향후 고객 감성을 기반으로 한 Target Marketing 에 활용할 수 있는 가능성을 볼 수 있었다.

### 1. 서론

고객을 기업 최고의 가치로 생각하는 시대에 성공적인 기업 활동을 위해서는 고객의 Needs 에 대한 조사는 정확하고 과학적이어야 하며, 그 결과 또한 정확하게 마케팅 활동에 반영되어야 한다. 현대의 고객 Needs 란 단순한 Function 충족 중심의 욕구가 아닌, 고객의 깊은 내면에 자리잡고 있는 기호와 감성적 욕구를 포함하는 것이다. 이러한 고객의 Needs 는 구매와 매우 밀접하게 연결되어 있지만, 인간이 지니고 있는 감성과 기호란 매우 복잡하고 미묘한 것이어서 그것을 알아내기란 매우 어렵다.

하지만, 사용자 중심의 Web 2.0 시대에 고객들은 On-line 상에서 끊임없이 Communication 하고 자신을 표현하고 있다. 이에 지속적 변경시키고, 내면의 감정을 표현하는 고객 Blog 로부터, 감성과 기호를 자동적으로 추출할 수 있다면, 기업 입장에서는 매우 효과적이고, 세련된 감성마케팅 속성으로 활용할 수 있다.

최근에 와서 '감성마케팅(Taste marketing)'기법이 도입되어 선행 기업에서는 활용하기 시작하였지만, 사실상 결음마 단계에 머물러 있는 실정이다. 각 기업에서는 인간 내면에 자리 잡고 있는 감성적 요구를 계량화하고 마케팅에 적용해 보려는 노력들은 지속적으로 이루지고 있으며, 특히 일본감성연구소에서 개발

한 테스트 스케일법(오감의 가치관을 분석하고 체계화한 기법)은 감성마케팅의 기반을 마련했다.

이에 본 논문에서는 감성 Target Marketing 을 위하여, Blog 와 테스트 스케일법을 접목한 고객감성 Clustering 방법을 제안한다. 고객 회원 정보에 등록된 Blog 만을 이용하여, 주기적으로 고객의 감성 기호를 자동적으로 추출하고, 감성적으로 유사한 Group 별로 Clustering 하는 방법을 제공한다.

본 논문은 2 장에서 Blog Image 와 감성지수에 대한 이론적 배경들, 3 장에서는 감성 Clustering 에 대한 구현방법, 4 장에서는 연구에 대한 감성 Clustering 결과와 검증분석, 그리고, 마지막으로 5 장에서는 결론과 향후 과제로 구성되어 있다.

### 2. 이론적 배경

#### 2.1 Blog 구조와 Image 에 대한 이해

Blog 의 디자인과 기능이 점차 발달되고 다양화됨에 따라, 고객의 Blog 에 나타난 Image 들을 분석하고, 그 중 지배적인 Image 를 추출하는 것은 어려운 일임에 틀림이 없다. 하지만 Web 이라는 기술적 구조의 특성상, 네트워크의 부하를 감안하여, Web 페이지에 나타난 Image 숫자를 대폭 줄이거나, 하나의 Image 크기를 작게 하고 이를 반복시키는 Option 으로 구성하는

것이 일반적인 것도 사실이다.

Web 페이지를 보았을 때, 거기로부터 얻어지는 전체 색감은 보통 Body 부분에 나타난 Main Image 나 색상으로 인하여 지배적이게 되는 경우가 대부분이다. Body에 사용된 주요 Image가 없는 경우라도, Title 부분에 사용된 Image 나 색상이 전체 색감을 결정하는 경우가 많다. 본 연구에서는 이를 근거로, Blog에 사용된 전체 Image 나 색상을 추출하여 분석하지 않고, Repeat Option으로 Body나 Title의 Background에 사용된 Main Image 만을 이용하여 고객의 감성지수를 추출하였다.

**2.2 Image 파일에 대한 컬러 분석**

컬러 Image 검색에 있어서 주로 2 개의 컬러 모델이 사용된다. 하나는 Red, Green, Blue의 조합으로 이루어진 RGB 컬러모델인데, 컬러 CRT 모니터나 대부분의 computer에서는 이 컬러 모델을 사용한다. 다른 나머지 하나는 색상(Hue), 채도(Saturation), 명도(Intensity)로 이루어진 HSI 컬러 모델이다[1]. 사람은 단지 H, S, I 세가지 형태의 조합에 대한 반응으로만 컬러를 인식할 수 있기 때문에, computer로부터 획득되는 Image 파일의 RGB 값을 인지가 가능한 HSI 모델의 색상, 채도, 명도로 변환해야 할 필요가 있다. RGB와 HSI 사이의 변환은 다소 복잡하지만, 아래와 같은 방법으로 획득될 수 있다.

$$I = \frac{1}{3}(R + G + B)$$

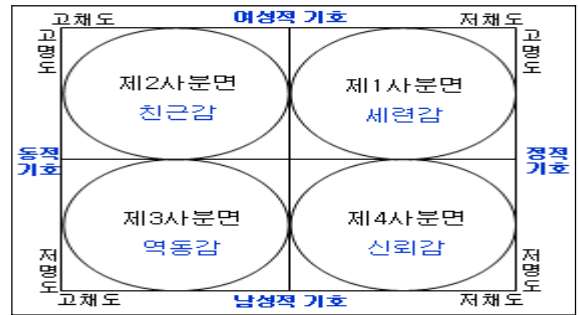
$$S = 1 - \frac{3}{(R + G + B)}[\min(R, G, B)]$$

**2.3 테스트 스케일법과 색채에 의한 감성분류**

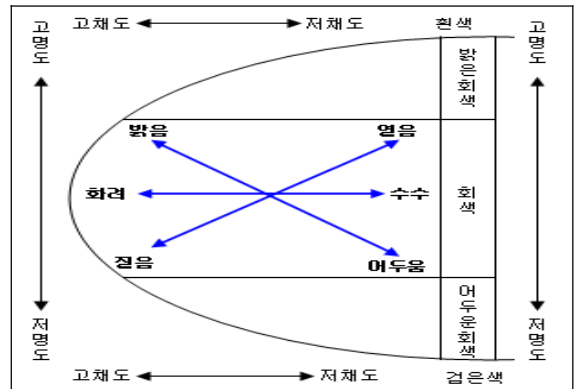
정보과학적 감성마케팅 방법을 현실에 구현시키기 위하여 객관적인 척도기준을 마련한 '테이스트 스케일법(기호감성 좌표축법)'은 일본 감성연구소가 약 7만명의 소비자를 대상으로 각 기호 감성을 조사하고 그 데이터를 통계학적인 다변량 해석법으로 재해석하여 개발해낸 이론이다. 이 방법으로 실물이나 또는 그에 가까운 시각적 모델을 대상으로 획득된 정직한 '감성'으로 사람들의 기호 정도를 분석해 보면 '좋고 싫음'에는 보편적인 심리 척도가 있다는 사실을 알게 된다[2]. 테스트 스케일법에서는 기호 감성을 여는 열쇠로 색채를 최우선으로 삼아, 조사를 시작한다. 많은 감각자극 중 색채가 가장 정량화 표현이 가능하고, 정성적 반응까지 조사하기 쉽기 때문이다. 더욱이 예로부터 색채는 인간의 관심을 끄는 중요한 인자(CDTPM: Color, Design, Touch, Price, Material)중에서도 으뜸으로 여겨져 왔으며 제품의 선호도와 가장 밀접한 관련이 있다[3]. 이러한 이론을 바탕으로 몇 만 명의 색채 기호 조사를 실시하고, 여기서 얻어진 데이터를 해석한 결과, 몇 가지 의미 있는 결과를 발견하게 된다. 무엇보다도 좋고 싫음을 결정하는 요소가 색상자극 보다는 톤 자극에 의해 결정되는 경우가 대부분이라는 사실(예컨대 화려한 톤의 빨강을 싫어하면 화려한 파랑도 싫어한다는 것)과, 좋고 싫어하는 톤은 색채학의 분류상 완전 대립되는 특성으로 나타

난다는 것이다. (예를 들면, 고명도/고채도를 좋아하는 사람은 저명도/저채도를 싫어한다는 것)

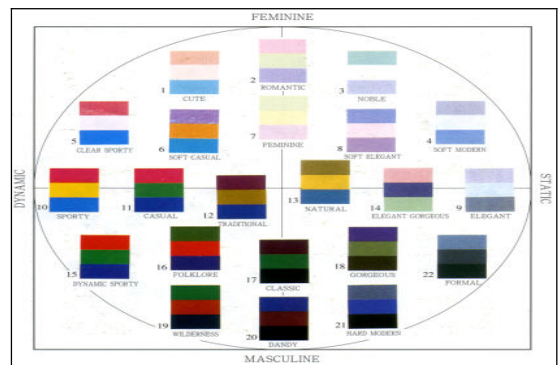
그림 1, 2, 3은 위에서 언급한 테이스트 스케일법의 특성과 그 결과를 보여주는 도표들로서, 그림 1은 스케일법을 통해 색채 자극으로 얻어진 감성 기호를 나타낸 좌표도 그림이다. 예를 들면, 제1사분면인 고명도/저채도 Group의 감성 기호는 여성적이면서 정적인 면이 많고, 세련된 느낌이 난다는 것이다. 그림 2는 색 입체를 가로로 자른 단면도로서, 가로에는 채도를, 세로에는 명도를 나타내어 그림 1에 대응되는 색상의 톤과 해당되는 위치의 형용사적 성질을 보여준다. 그림 3도 그림 1에 대응하는 대표적인 배색 관계를 나타내는 그림이다.



(그림 1) 테이스트 스케일 좌표도[2]



(그림 2) 자극의 톤 체계도[2]



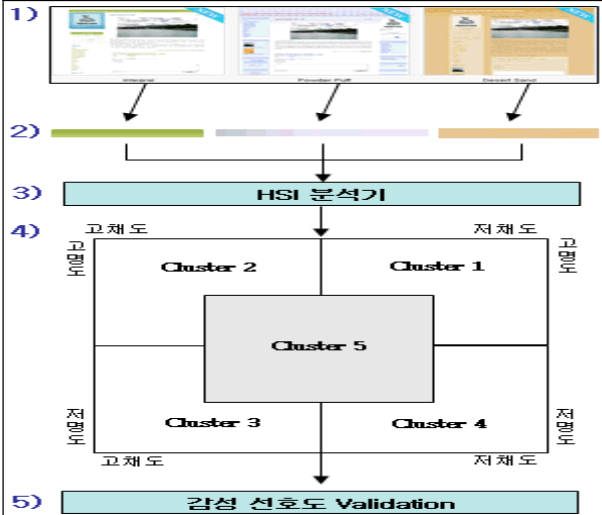
(그림 3) 배색 테이스트 스케일[2]

**3. 구현 방법**

**3.1 개요**

전체적인 구현 방법을 살펴보면, Blog로부터 각 대상 Skin의 Main Image를 추출하는 단계(그림 4의 1,

2), 추출된 Image 로부터 자체적으로 구현한 HSI 색상 분석기를 통해 Blog 의 색채 감성기호를 판단하는 단계 (그림 4 의 3), 이것을 그림 1 의 테스트 스케일 법 좌표도에 적용하여 Blog 를 Clustering 하는 단계(그림 4 의 4), 마지막으로 Clustering 한 결과를 연관 배색에 대한 Blog 사용자 선호도 조사를 통해 검증하는 단계로 이루어진다 (그림 4 의 5). 연구 실험에 대한 전체적인 진행 순서는 아래 그림 4 와 같다.



(그림 4) 실험 전체 흐름도

**3.2 Main Image 추출과 색채 감성**

Blog 로부터 Main Image 를 추출하기 위해, 한국의 대표적인 Blog 사이트인 다음(Daum), 네이버(Naver), 그리고 이글루스(Egloos)의 HTML 과 Javascript 소스코드를 분석하였다. 그림 5 에서 보는 소스코드 예와 같이 Blog 의 색채 감성을 지배하는 Image 를 추출할 수 있다.

```
[다음-Daum]
background-image:url(http://pimg.hanmail.net/blog3/skin/070903/bg.jpg?ver=1.27); background-repeat:repeat;
[네이버-Naver]
background-color:transparent;background-image:url(http://blogimgs.naver.com/nblog/skins/background/0011_body.gif);
[이글루스-Egloos]
background:#000000
url("http://md.egloos.com/skn/img/a/01/04/9/thunder_img1.gif")
repeat-x;
```

(그림 5) Blog 소스코드 예제

본 연구에서는 이글루스에서 기본적으로 제공하는 Blog Skin 들을 대상으로 Main Image 를 추출하였다. 추출된 Image 파일들은 자체적으로 구현한 HSI 색상 분석기를 통해 모든 Pixel 의 명도와 채도 값을 산정하게 된다. 전체 Image 파일을 Pixel 단위로 하나씩 R, G, B 값을 계산하고, 이를 다시 HSI 변환 공식에 대입한다. 이렇게 얻어진 모든 Pixel 의 명도와 채도 값에 대한 각 평균치가 해당 Image 를 대표하는 명도와 채도 대표값으로 정해진다.

$$I_{avg} = \frac{1}{n} \left( \sum_{n=1}^n \frac{1}{3} (R_{ij} + G_{ij} + B_{ij}) \right)$$

$$S_{avg} = \frac{1}{n} \left( \sum_{n=1}^n 1 - \frac{3}{(R_{ij} + G_{ij} + B_{ij})} [Min(R_{ij}, G_{ij}, B_{ij})] \right)$$

i 는 추출된 Image 의 넓이이며, j 는 높이를 말한다. 0 부터 i-1 과 j-1 까지 각각 1 씩 증가하여 Pixel 의 가로와 세로 좌표로 대응된다. n 은 (i-1)\*(j-1)로 계산된 전체 Pixel 의 수를 나타내게 된다. 상기와 같은 공식으로 산출된 채도와 명도 대표값은 0~1 사이의 범위를 가지게 되며, 소수점 둘째 자리까지 계산된다.

**3.3 감성 Clustering 과 Validation**

HSI 분석기를 통해 얻어진 채도와 명도 대표값을 가지고, Blog 를 그림 1 의 테스트 스케일 좌표도에 할당한다. 우선 명도는 세로축, 채도는 가로축으로 각 0.5 를 기준으로 하여 4 사분면인 4 개의 Cluster 에 나누어 할당한다. 다음엔 색채에 대한 뚜렷한 감성적 특징을 얻기 어려운 중채도/중명도(각각 0.25~0.75 범위)에 해당하는 Blog 를, Cluster 5 로 정하여 재 할당한다. 상기와 같이 Clustering 된 Blog 사용자들의 감성기호에 대한 검증은 그림 3 의 배색 테스트 스케일의 22 개 색상에 대한 선호도 설문 조사를 통하여 실시한다. 설문 대상자들에게 각 색상마다 0 부터 5 까지 선호도 점수를 주게 한 후, 색상번호 3, 4, 8 번의 선호도 점수 평균치를 Cluster1 의 색상선호도로, 1, 5, 6 번 평균치를 Cluster2, 15, 16, 19 번 평균치를 Cluster3, 마지막으로 18, 21, 22 번 평균치를 Cluster4 의 색상선호도로 계산한다. 나머지 중채도/중명도에 해당되는 색상들과 그림 3 좌표도의 사분면 경계에 위치한 색상들은 감성적 기호 특성을 유추하기 모호하여 색상선호도 검증대상에서 제외한다.

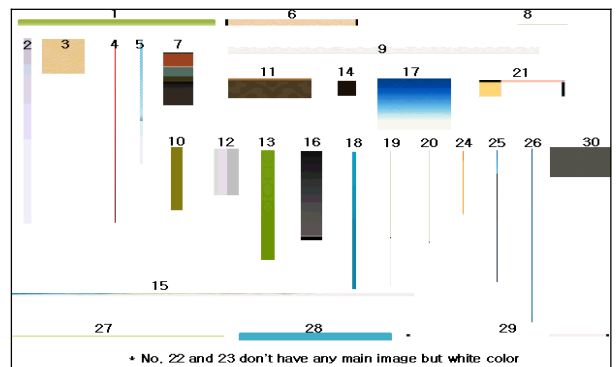
**4. 결과**

**4.1 Main Image 추출 결과**

본 논문에서는 이글루스 사이트에서 많이 사용되고 있는 30 개의 Skin(그림 6)을 실험 대상으로 선정하였으며, 그림 5 와 같은 코드 분석을 통하여 그림 7 과 같은 Main Image 들을 추출하였다.



(그림 6) 선정된 Blog Skins



(그림 7) 추출된 Main Images

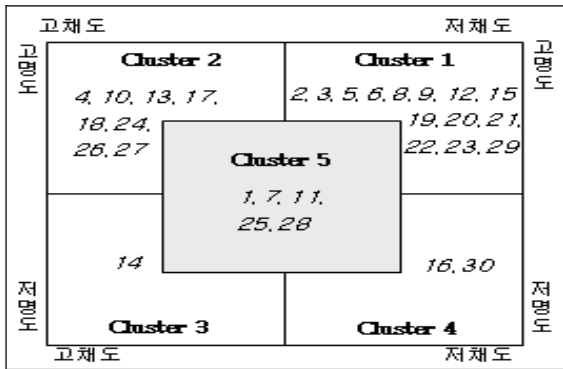
### 4.2 Blog 감성 Cluster 할당 결과

HSI 분석기를 통해 산정된 채도와 명도 대표 값을 기준으로, 표 1 과 같이 실험 대상 Blog 30 개를 5 개의 감성 Cluster 로 할당하였다.

<표 1> HSI 분석기를 통한 감성 Cluster 결정 도표

Image 번호	색상	채도	명도	감성 Cluster
1	71.46	0.54	0.74	5
2	255.6	0.09	0.98	1
3	38.31	0.37	0.91	1

그림 1 의 테스트 스케일 좌표도를 바탕으로 하여, 5 개 감성 Cluster 로 Blog 들을 매핑하면, 그림 8 과 같이 Cluster1 은 14 개, Cluster 2 는 8 개, Cluster3 은 1 개, Cluster4 는 2 개, Cluster5 는 5 개의 할당 결과를 얻을 수 있다



(그림 8) Blog 의 감성 Cluster 할당 결과

### 4.3 감성설문 조사결과

본 연구에 대한 검증은 Cluster5 를 제외한 4 개의 Cluster 로 할당된 Blog 사용자 30 명에게 그림 3 의 배색 테스트 색상 22 개에 대한 선호도 설문조사로 실시하였다. Cluster1 에 속한 Blog 사용자 14 명, Cluster2 는 10 명, Cluster3 과 4 는 각 3 명씩 설문을 조사하였다.

설문조사 결과, 전체 30 명의 대상자 중 본인이 속한 Cluster 의 색상 선호도 평균이 가장 높은 사용자는 19 명으로 나타났으며, 전체적으로는 63% 정도의 정확도를 보였다.

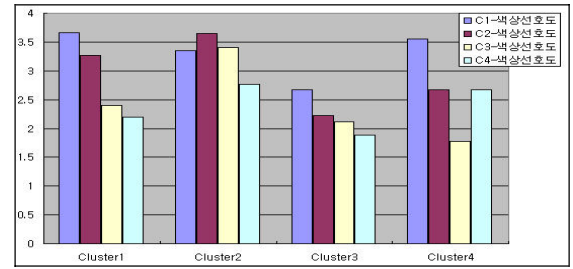
$$\text{Precision (63\%)} = \frac{\text{해당 색상 선호도 일치 수 (19)}}{\text{전체 Blog 사용자 수 (30)}} \times 100$$

결과를 세부적으로 보면, 그림 9 에서와 같이, Cluster1 저채도/고명도 그룹은 전체 14 명 중 11 명의 선호도가 일치하여 약 78%의 정확도를 보였으며, Cluster 2 는 10 명 중 6 명(60%), Cluster 3 과 4 는 똑같이 3 명 중 1 명이 일치하여 약 33%의 정확도를 보였다. 그림 10 은 감성 Cluster 에 속한 Blog 사용자들의 각 Cluster 에 해당되는 색상 선호도를 보여준다. Cluster1 과 2 의 경우, 다른 감성 Cluster 의 사용자 보다 자신이 속해있는 Cluster 의 색상에 대한 선호도가 가장 높았지만, Cluster3,4 의 저명도 Cluster 들 경우에는, 월등한 색상 선호도를 보이지 못했다. 마지막으로, 22 개 전체 색상에 대한 색상 선호도를 분석해 보면, 전체적으로 고명도 색상들(8, 9, 2 등)에 대한 높은 선

호도를 보이는 데, 이것은 우리나라 국민성 및 지역 성과도 밀접한 연관이 있다고 할 수 있다.



(그림 9) 감성 Cluster 별 소속 색상선호도 일치 분포



(그림 10) 감성 Cluster 별 색상선호도 분포

## 5. 결론 및 향후 과제

본 논문에서는 고객의 등록된 Blog 로 부터 감성지수를 추출하고, 이를 감성 Cluster 와 Mapping 함으로써 Targeting 된 감성기반의 마케팅 속성을 기업에게 제시하는 방안 에 관하여 논의하였다.

제안된 방법을 이용하여 Clustering 된 Blog 의 사용자 들은 배색 테스트 스케일 좌표상의 해당 Cluster 에 속한 색상들에게 높은 선호도를 보여주고 있다. 특히 감성 Cluster1 과 2 에서는 상당히 높은 정확도를 나타냈다. 이것은 고객의 Blog 로부터 얻어진 색채에 대한 정보를 통하여 고객의 감성이나 기호를 유추해 낼 수 있음을 실증하였고, 이를 기반으로, 고도화된 감성 마케팅의 고객 속성으로 활용될 수 있음을 보여주었다. 향후에는 고객 Blog 의 Main Image 에 대한 색상과 질감을 조합한 감성지수 추출이나, text mining 을 이용하여 Blog Content 로부터 고객의 관심상품을 유추해 낼 수 있는, 더욱 정교한 감성기반의 Target Marketing 방법 에 대하여 지속적으로 연구하고 실증해 나갈 것이다.

### 참고문헌

- [1] Youngkwan Choi, "Color Component Analysis for Image Retrieval", Ph.D. Thesis, Konkuk Univ., pp. 7-12, 2005.
- [2] Sato Kunio, "Taste Marketing-emphasizes the importance of designing five senses", Greenbee Press, 2001.
- [3] Duksang Lee, JungHee Cho, "Application to Emotional Engineering in Color Coordination", 인문논집. 2 권, pp. 322, 2000.
- [4] Sangoh Jeong, "Histogram-Based Color Image Retrieval", Psych221/EE362 Project Report, Mar.15, 2001.