개인 맞춤형 화장품 추천을 위한 사용자 SNS 정보 기반의 스코어링 기법

하은지, 문지훈, 황인준
고려대학교 전기전자공학과
e-mail:{eunji1114, johnny89, ehwang04}@korea.ac.kr

User's SNS Data-Based Scoring Scheme For Personalized Cosmetics Recommendation

Eunji Ha, Jihoon Moon and Eenjun Hwang
School of Electrical Engineering, Korea University

요 약

최근 남녀노소를 불문하고 피부 관리에 대한 관심이 증가하면서, 피부 개선에 효과적인 화장품의 선택에 관심이 높아지고 있다. 하지만 다양한 화장품들을 대상으로 자동화된 고객 맞춤형 화장품 추천은 그 발전이 더디고, 이와 관련된 연구 또한 아직 미미한 실정이다. 또한, 다양한 특성을 가지는 고객 피부 데이터 셋의 확보가 어려운 상황에서, 소수의 데이터 표본만을 이용하여 화장품 추천이 진행되고 있어 추천의 정확도를 확보하지 못하고 있다. 본 논문은 스마트폰용 휴대용 카메라를 이용하여 고객의 피부 상태를 진단한 후, 고객의 피부 개선에 적합한 화장품을 자동으로 추천하는 기법을 제안한다. 먼저, 화장품 추천을 위해 사용자의 SNS 데이터와 피부 데이터를 수집 및 분석하여 추천 리스트를 생성한다. 이를 기반으로, 추천된 각 화장품의 스코어를 계산한다. 그 다음, 피부 개선 순위와 스코어 기반의 화장품 특성 순위 간의 상관계수를 이용하여 가장 높은 상관계수의 화장품을 우선 추천한다. 성능 평가를 위해 실제 화장품 회사에서 제시한 화장품 추천 리스트와 본 논문에서 제안한 방법을 적용한 화장품 추천 리스트를 비교함으로써 효용성과 타당성을 입증하였다.

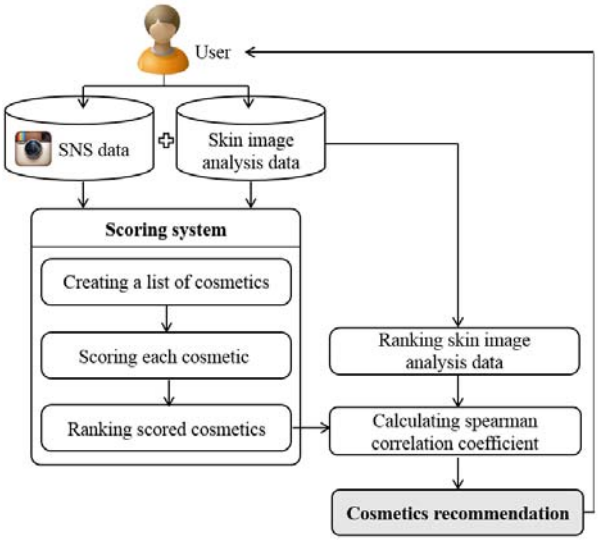
1. 서론

최근, 남녀노소를 불문하고 많은 사람들이 피부 관리에 관하여 지속적인 관심을 보이고 있다. 이러한 추세를 반영하여 많은 화장품 회사들이 고객의 피부 고민을 해결할 수 있는 다양한 유형의 화장품들을 출시하고 있다. 하지만 많은 고객들이 자신의 피부에 대해 어떠한 화장품을 이용해야 할 지 정확하게 모르기 때문에, 오프라인의 전문가를 통해서 추천을 받거나, 나와 비슷한 피부 고민을 가지는 사람의 후기나 경험을 참고하여 화장품을 구매하게 된다. 이러한 문제점을 해결하기 위한 고객의 피부 고민에 적합한 개인 맞춤형 화장품 추천에 대한 필요성은 그 어느 때보다 크지만 그러한 추천을 실시간으로, 언제, 어디서나, 쉽게, 제공해 주는 기법에 관한 연구는 아직 미미한 상태이다.

기존의 추천 시스템[1]은 고객 정보에서 속성 변수의 특징 벡터가 적용된 클러스터링 작업과 군집 내의 제품 카테고리 선호도 계산 작업의 전처리를 수행하여 제품을 추천하였다. 하지만 고객 정보와 제품 정보, 구매 이력 정보 등을 이용한 추천을 진행되기 때문에, 개인 정보의 사용으로 인한 사생활 침해, 구매 이력 정보의 부족으로 인한 추천의 낮은 정확도, 공공성 부족 등의 문제점이 발생한다.

본 논문은 스마트폰용 휴대용 카메라를 이용하여 사용자의 피부 상태를 진단하고, 사용자의 피부 개선에 적합한 개인 맞춤형 화장품 자동 추천 시스템을 제안한다. 사용자의 피부 상태를 진단하기 위하여, 50배율 카메라의 이미지에서 사용자의 피부 특징을 추출하고 분석한 기존 연구를 기반으로 진행하였다[2-4].

SNS(Social Network Service) 사용자들은 개인의 일상을 태그, 이미지 등을 이용하여 나타낸다. 본 연구에서는 개인 맞춤형 화장품을 추천하기 위해 사용자의 SNS 데이터를 활용하였다. 사용자가 SNS에 남긴 태그와 실제 사용자의 피부 이미지 분석 데이터 값들을 이용해 화장품 리스트를 생성하여 추천에 사용될 화장품 정보를 구축하였다. 시스템의 공공성을 갖추기 위해 각 연령대별로 화장품 스코어를 측정하였으며, 사용자의 피부 고민 순위에 따라 가중치를 적용하였다. 그 다음 스피어만 상관계수(Spearman correlation coefficient)를 사용하여 화장품과 사용자의 피부 고민 간의 연관성을 비교하여 적합한 화장품을 추천하였다. 최종적으로, 실제 전문가가 추천한 화장품 리스트와 본 실험에서 제안한 기법을 적용한 추천 화장품 리스트를 비교함으로 시스



(그림 1) 시스템 구성도

템의 성능을 평가하였다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 제안하는 기법과 이를 기반으로 구현한 시스템에 관해 설명한다. 4장에서는 시스템에 대한 실험을 통해 성능 평가를 진행하고 5장에서 본 논문의 결론과 향후 연구에 관하여 기술한다.

2. 개인맞춤형 화장품 자동 추천 시스템

뷰티나 미용 등에 관한 관심이 꾸준히 증가하면서 고객에게 관련 제품을 추천해주는 서비스들이 제공되고 있지만, 실제 서비스를 이용하는 사용자 개개인의 특성에 맞는 맞춤형 추천 서비스는 상당히 부족한 실정이다. 본 논문에서는 사용자의 피부 고민 개선에 적합한 개인 맞춤형 화장품 자동 추천 시스템을 제안한다. 선행연구를 기반으로, 실험에 사용되는 화장품의 스코어를 측정하고 스피어만 상관계수를 이용하여 사용자 피부 고민 개선에 적합한 화장품들을 추천하였다. (그림 1)은 본 논문에서 제안하는 시스템의 전체적인 구성도이다.

2.1 Dataset

제안하는 화장품 추천 시스템을 구축하기 위해서는, '다양한 피부 고민에 따른 추천'의 기반이 될 수 있는 전문적 정보를 가진 표본이 필요하다. 본 연구에서는 화장품 전문가를 통해, 선행 연구에서 제안한 실제 피부 이미지 분석 데이터의 수치 값인 다섯 가지 도메인을 이용하여 <표 1>의 '피부 고민별 제품 추천 리스트'를 작성하였다. 주름, 거칠기, 민감성, 미백, 유분은 0에서 1까지 나타내며, 1로 가까워질수록 해당 특징의 개선이 필요한 것을 의미하며 사용자의 피부 고민으로 간주한다.

2.2 화장품의 스코어 측정

추천에 사용되는 화장품의 스코어를 측정하는 방법은 다음과 같다. 먼저 앞서 작성된 '피부 고민별 제품 추천 리스트'를 이용하여 모든 화장품별로 추천된 사용자의 나이, 피부고민, 추천 순위로 리스트를 생성하고, 이를 연령대별로 분류한다. 제품 리스트의 구성은 다음과 같다.

$$List = \{나이, 피부\ 고민\ 1순위, 2순위, 3순위, 추천된\ 순위\}$$

사용자의 피부 고민 개선을 위한 제품을 추천하기 위해서 피부 고민의 상위 3개를 고려하였으며, 피부 고민에 대한 가중치를 식 (1)과 같이 지정하였다.

$$Weight_R = \frac{(N+1)-R}{N} \quad (1 \le R \le 3) \tag{1}$$

$$Score_D = \sum_{n=1}^{N} \{(Weight_n \times Number_D) + Tf{\cdot}idf_D\} \tag{2}$$

식 (1)에서 $Weight_R$은 고민 순위에 대한 가중치를 말하며, N은 도메인의 총 개수, R은 고민 순위를 말한다. 그 다음, 각 화장품의 도메인별 스코어 $Score_D$를 측정하기 위하여 식 (2)을 이용한다. 주름, 거칠기, 민감성, 미백, 유분 등 다섯 개의 도메인 각각의 스코어를 측정한다. 이 때, 각 도메인을 태그로 가정하여 사용자 SNS에서 관련 해시태그를 수집해 하나의 문서를 생성하고, 각 도메인 태그 문서의 tf·idf값을 측정한다. 순위에 따라 적용되는 가중치와 해당 도메인의 개수 $Number_D$, 각 태그로 구성된 도메인의 tf·idf 값 $Tf{\cdot}idf_D$를 곱하여 스코어를 구한다. <표 2>는 제품 리스트의 구성에 맞게 화장품 p1의 대한 사용자 피부 고민 순위 및 추천된 순위를 나타낸 것이다. 사용자 피부 고민 순위는 3순위까지 만을 나타내었다.

그 다음, 스코어가 측정된 도메인을 순위로 나타내어, 다섯 개의 도메인 모두 순위 값을 가질 수 있도록 한다. <표 3>는 화장품 p1에 대한 도메인별 모든 정보를 나타낸다.

<표 1> 피부 고민별 제품 추천 리스트

피험자 정보	피험자 번호		A101	A358	A688
	나이		25	40	57
실제 피부이미지 분석데이터	주름	노화방지필요	0.380327	0.501104	0.699049
	거칠기	수분필요	0.222222	0.234551	0.444752
	민감성	민감성제품필요	0.549057	0.409434	0.349057
	미백	톤개선필요	0.165217	0.482609	0.382609
	유분	유분컨트롤필요	0.437901	0.132462	0.183203
피부고민	1순위		민감성	주름	주름
	2순위		유분	미백	거칠기
	3순위		주름	민감성	미백
추천하는 화장품	1순위		p1	p15	p1
	2순위		p10	p16	p2
	3순위		p13	p1	p12
	4순위		p3	p13	p11
	5순위		p6	p12	p14

<표 2> 화장품 p1의 20대 피부 고민 순위 및 추천된 순위

화장품 p1				
순위 / 나이	사용자의 피부 고민			추천된 순위
	1순위	2순위	3순위	
20	미백	주름	민감성	2
20	미백	민감성	거칠기	2
21	미백	거칠기	민감성	5
26	미백	거칠기	주름	3
22	미백	민감성	거칠기	1
27	민감성	주름	미백	1
27	민감성	주름	미백	5
평균 추천 순위				2.714

<표 3> 화장품 p1의 도메인별 정보

나이	특징값	주름	거칠기	민감성	미백	유분
20대	스코어	3.6	3.4	4.6	6.6	0
	순위	3	4	2	1	5
30대	스코어	5.2	2.2	5.4	5.3	0.6
	순위	3	4	1	2	5
40대	스코어	6.6	3	3	7	0
	순위	2	3	3	1	5
50대	스코어	3.8	0	2.4	2.6	0.8
	순위	1	5	3	2	4

2.3 스피어만 상관계수를 이용한 화장품 추천

스피어만 상관계수는 데이터가 서열을 나타내는 척도일 때 데이터의 값 대신 순위를 이용한 상관계수이다. 두 변수 간의 연관 관계가 있는지 없는지를 밝혀주며 데이터에 이상점이 있거나 표본 크기가 작을 때 유용하다 [5, 6].

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = 1 - \frac{6\sum d_i}{n(n^2-1)} \quad (4)$$

$$(-1 \le r \le 1, d_i = x_i - y_i)$$

스피어만 상관계수는 -1과 1 사이의 값을 가지며, 두 변수 간의 상관계수가 높을수록 1에 가까워진다. 본 논문은 2.1에서 측정한 화장품 스코어와 실제 피부 이미지 분석 데이터 간의 상관계수가 높은 화장품을 추천한다. 스피어만 상관계수가 같은 경우, 화장품의 추천 평균 순위를 비교하여 순위가 앞선 화장품을 먼저 추천하였다.

2.4 우선 추천을 위한 예외사항

실험을 진행하기에 앞서, 화장품 추천 시스템은 기업의 목적에 맞는 추천이 진행되어야 하기에 우선 추천이 가능할 수 있도록 다음과 같은 예외 사항을 고려하였다.

1. 고객의 피부 고민 Top 3 안에 '유분'이 존재할 경우: 크림, 아이크림의 화장품 라인을 제거하고, 스피어만 상관계수의 차순위 화장품을 추천
2. 연령대에 적합하지 않은 제품 라인의 화장품 제거: 연령대별로 적합하지 않은 제품 라인을 사전에 제거한 후, 남은 제품 리스트를 통해 화장품을 추천
3. 기업 주력 화장품은 추천 리스트에 항상 포함 : 주력 화장품이 추천 리스트에 포함되어 있지 않은 경우, 추천 리스트의 마지막 화장품을 제거하고 주력 화장품을 대신 추천

<표 4>는 추천에 사용된 화장품 정보와 각 화장품의 특성을 나타낸다. p2, p5, p9, p11, p16은 유분기가 많은 '크림/아이크림' 타입이기에 피부 고민에 유분이 존재하는 사용자에게는 추천을 제외하였다. p14, p15, p16은 장년층을 위한 화장품이기에 사용자의 나이가 40세 이상일 경우에만 추천을 진행하였다. 35세 이상일 경우는 피부 고민이 평균 수치가 높을 때만 추천을 진행하였다. 또한, 기업의 주력상품이 p1이라고 가정하여, 항상 추천 리스트에 포함되도록 하였다.

<표 4> 추천에 사용된 화장품 리스트
(A: Aging(노화방지), M : Moist(수분공급), I : Irritate(저자극), W : Whitening(피부미백), O : Oil(피지모공컨트롤))

화장품 이름	제품의 기능					참고
	A	M	I	W	O	
p1	1	1	0	1	0	Always
p2	1	1	0	1	0	NOT to 'oil'
p3	1	0	0	0	0	
p4	1	0	0	0	0	
p5	1	0	0	0	0	NOT to 'oil'
p6	0	1	0	1	1	
p7	0	1	0	1	0	
p8	0	1	1	1	0	
p9	0	1	1	1	0	NOT to 'oil'
p10	0	1	0	0	0	
p11	0	1	0	0	0	NOT to 'oil'
p12	0	1	0	0	0	
p13	0	1	0	0	0	
p14	0	0	0	1	0	Above 35 or 40 years old (r16-NOT to 'oil')
p15	1	0	0	1	0	
p16	1	0	0	1	0	

<표 5> 사용자의 피부 이미지 분석 데이터 순위

피험자번호	나이	주름	거칠기	민감성	유분	주름
A358	40	1	4	3	2	5

<표 6> 사용자 정보와 화장품간의 스피어만 상관계수

번호	화장품 이름	**스피어만 상관계수**	상관계수 순위	평균 추천순위
1	p1	**1**	**1**	**3**
2	p2	**0.9**	4	3.13
3	p3	**-1**	12	3
4	p4	**-1**	12	4.333
5	p5	**0.75**	7	5
6	p6	**-1**	12	5
7	p7	**0.35**	10	2
8	p8	**-1**	12	3.111
9	p9	**0.65**	9	4.5
10	p10	**0.7**	8	2
11	p11	**0.9**	4	3.8
12	p12	**0.9**	4	2.75
13	p13	**0.95**	3	2.5
14	p14	**0.2**	11	2.5
15	p15	**1**	**1**	**1.85**
16	p16	**1**	**1**	**3.44**

3. 실험 및 결과

본 논문의 실험은 20대부터 50대까지 4개의 연령대별로 진행되었으며, 각 16명씩 총 64명에게 화장품 추천을 진행하였다. 사용자의 피부 고민에 따라 개인 맞춤형 화장품 추천이 진행될 수 있도록 각기 다른 피부 개선 효과를 가지는 p화장품 16개의 리스트를 작성하였다. 또한, 추후 시스템의 정확도 평가를 위해 a화장품 15개 리스트를 작성하였다.

<표 5>, <표 6>에는 피험자의 피부 이미지 분석 데이터 순위와 각 화장품 간의 스피어만 상관계수를 나타내었다. 상관계수가 1에 가까울수록 피부고민에 가장 적합한 화장품이라는 것을 의미하기 때문에, 상관계수의 순위에 따라 제품 추천을 진행하였다. p1, p15, p16 제품들은 모두 1의 상관계수를 가지기에, 화장품의 평균 추천 순위가 높은 제품을 먼저 추천하였다.

<표 7>은 화장품 전문가를 통한 기존의 화장품 추천 리스트와 본 연구에서 제안한 기법을 적용한 자동 화장품 추천 리스트를 비교한 것이다. 제안한 기법을 적용한 화장품 추천 리스트 결과가 5순위 제품만 제외하고 기존 전문가의 추천 리스트와 모두 일치하는 것을 알 수 있다. 또한, 시스템의 정확도는 기존 추천 리스트에 제안 기법을 적용한 추천 화장품이 있을 경우로 측정하였다. 각 연령별 추천의 정확도와 다른 제품군을 적용한 추천의 정확도를 평가하였다. <표 8>, <표 9>에서 나타나듯이, 연령별로 유사한 정확도로 추천된 것을 볼 수 있으며, p화장품과 a화장품 모두 매우 높은 정확도로 구현된 것을 확인할 수 있다.

4. 결론 및 향후계획

본 논문은 사용자의 피부 고민 개선에 적합한 개인맞춤형 화장품 자동 추천 시스템을 제안하였다. 추천을 위한 화장품 스코어 측정과 스피어만 상관계수를 이용하여 매우 높은 정확도의 화장품 추천 리스트를 구현할 수 있었다.

또한 본 논문에서 제안하는 화장품 추천 시스템의 경우, 화장품뿐만이 아닌 다양한 제품군에도 적용될 수 있으므로 다양한 서비스에도 적용될 수 있을 것으로 기대된다.

5. Acknowledgements

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임.
(NO.R0190-15-2012, 빅데이터 처리 고도화 핵심기술개발 사업 총괄 및 고성능 컴퓨팅 기술을 활용한 성능 가속화 기술 개발)

<표 7> 기존의 화장품 추천 리스트와 제안한 기법을 적용한 자동 화장품 추천 리스트의 비교 (p화장품)

추천하는 화장품 순위	기존의 화장품 추천	**제안한 기법을 적용한 화장품 추천**
1순위	p15	**p15**
2순위	p16	**p16**
3순위	p1	**p1**
4순위	p13	**p13**
5순위	p11	**p12**

<표 8> 연령별 추천 정확도

연령	정확도
20대	0.89
30대	0.94
40대	0.91
50대	0.98
평균	**0.93**

<표 9> 제품군에 따른 추천의 정확도

	정확도
p화장품	0.90
a화장품	0.96
p화장품 + a화장품	**0.93**

참고문헌

[1] Young Sung Cho, Mi Sug Gu, and Keun Ho Ryu, “Development of Personalized Recommendation System using RFM method and k-means Clustering,” *Journal of The Korea Society of Computer and Information*, Vol.17, No.6, pp. 163-172, 2012.

[2] K. Kim, Y.-H. Choi, and E. Hwang, “Wrinkle feature-based skin age estimation scheme,” *Proc. of International Conference on Multimedia and Expo,* pp. 1222-1225, 2009.

[3] Y.-H. Choi, Y.-S. Tak, S. Rho, and E. Hwang, “Skin feature extraction and processing model for statistical skin age estimation,” *Multimedia tools and applications,* Vol.64, No.2, pp. 227-247, 2013.

[4] Y.-H. Choi, D. Kim, E. Hwang, and B. J. Kim, “Skin texture aging trend analysis using dermoscopy images,” *Skin Research and Technology,* Vol.20, No.4, pp. 486-497, 2014.

[5] Jiang. M, Cui. P, Liu. R, Yang. Q, Wang. F, Zhu. W, and Yang. S, “Social contextual recommendation,” *Proc. of the 21st ACM international conference on Information and knowledge management,* pp. 45-54, 2012.

[6] Sorokowski. P, Sorokowska. A, Frackowiak. T, Karwowski. M, Rusicka. I, and Oleszkiewicz, A. “Sex differences in online selfie posting behaviors predict histrionic personality scores among men but not women,” *Computers in Human Behavior,* Vol.59, pp. 368-373, 2016.