

Development of big data based Skin Care Information System SCIS for skin condition diagnosis and management

Hyung-Hoon Kim*, Jeong-Ran Cho**

*Professor, Dept. of Cosmetic Science, Kwangju Womens University, Gwangju, Korea

**Professor, Dept. of Health Administration, Kwangju Womens University, Gwangju, Korea

[Abstract]

Diagnosis and management of skin condition is a very basic and important function in performing its role for workers in the beauty industry and cosmetics industry. For accurate skin condition diagnosis and management, it is necessary to understand the skin condition and needs of customers. In this paper, we developed SCIS, a big data-based skin care information system that supports skin condition diagnosis and management using social media big data for skin condition diagnosis and management. By using the developed system, it is possible to analyze and extract core information for skin condition diagnosis and management based on text information. The skin care information system SCIS developed in this paper consists of big data collection stage, text preprocessing stage, image preprocessing stage, and text word analysis stage. SCIS collected big data necessary for skin diagnosis and management, and extracted key words and topics from text information through simple frequency analysis, relative frequency analysis, co-occurrence analysis, and correlation analysis of key words. In addition, by analyzing the extracted key words and information and performing various visualization processes such as scatter plot, NetworkX, t-SNE, and clustering, it can be used efficiently in diagnosing and managing skin conditions.

▶ **Key words:** Skin information management system, Skin diagnosis and management, Big data, Natural language processing, Text mining

[요 약]

피부상태의 진단과 관리는 뷰티산업종사자와 화장품산업종사자에게 그 역할을 수행함에 있어서 매우 기초적이며 중요한 기능이다. 정확한 피부상태 진단과 관리를 위해서는 고객의 피부상태와 요구사항을 잘 파악하는 것이 필요하다. 본 논문에서는 피부상태 진단 및 관리를 위해 소셜미디어의 빅데이터를 사용하여 피부상태 진단 및 관리를 지원하는 빅데이터기반 피부관리정보시스템 SCIS를 개발하였다. 개발된 시스템을 사용하여 텍스트 정보 중심의 피부상태 진단과 관리를 위한 핵심 정보를 분석하고 추출할 수 있다. 본 논문에서 개발된 피부관리정보시스템 SCIS는 빅데이터 수집단계, 텍스트전처리단계, 이미지전처리단계, 텍스트단어분석단계로 구성되어 있다. SCIS는 피부진단 및 관리에 필요한 빅데이터를 수집하고, 텍스트 정보를 대상으로 핵심단어의 단순빈도분석, 상대빈도분석, 동시출현분석, 상관성분석을 통해 핵심단어 및 주제를 추출하였다. 또한 추출된 핵심단어 및 정보를 분석하고 산포도, NetworkX, t-SNE 및 클러스터링 등의 다양한 시각화 처리를 함으로써 피부상태 진단 및 관리에 있어 이를 효율적으로 사용할 수 있도록 하였다.

▶ **주제어:** 피부정보관리시스템, 피부진단 및 관리, 빅데이터, 자연어처리, 텍스트마이닝

- First Author: Hyung-Hoon Kim, Corresponding Author: Jeong-Ran Cho
- *Hyung-Hoon Kim (hhkim@kwu.ac.kr), Dept. of Cosmetic Science, Kwangju Womens University
- **Jeong-Ran Cho (jrcho@kwu.ac.kr), Dept. of Health Administration, Kwangju Womens University
- Received: 2022. 02. 03, Revised: 2022. 02. 03, Accepted: 2022. 03. 04.

I. Introduction

4차 산업혁명시대의 도래와 함께 화장품기술 환경과 뷰티산업 전반의 환경에 큰 변혁이 일어나고 있다. 최근 화장품산업과 뷰티산업의 동향은 기능과 성분을 중시하며 기능성 화장품과 개인 맞춤형 뷰티산업 방향으로 발전되어 가고 있다. 화장품과 뷰티산업에서 개인 맞춤형 기능성 화장품과 미용 서비스를 제공하기 위해서는 무엇보다 화장품 및 미용 소비자의 니즈를 파악하는 것이 가장 중요하다. 뷰티산업 분야에서 소비자의 니즈 파악 과정에서 가장 기본적인 내용은 소비자의 피부상태 파악과 요구사항 파악이라 할 수 있다. 이러한 관점에서 화장품과학 및 미용과학 분야의 뷰티산업 종사자와 연구개발자에게 피부상태의 파악과 진단은 중요한 관심사이다.

일반적으로 뷰티산업 종사자들이 피부 미용을 목적으로 피부상태 진단을 하는 경우에 대부분 육안 관찰 또는 스마트폰 등의 보편적으로 보급된 관찰 장비를 사용하여 관찰하고 피부미용을 원하는 일반인들의 피부상태에 대한 의견을 통하여 피부 미용적 진단과 관리를 진행하고 있다. 피부 미용에 있어서 피부를 건강하게 유지하거나 관리하는 첫 번째 단계는 해당 고객의 피부 상태를 정확하게 진단하고 분석하는 것이다[1, 2]. 그러나 정확한 진단을 위해 사용되는 피부 상태 측정 장비는 현재까지는 고가의 장비 가격 및 측정기기의 크기 등으로 특별한 설치 운영 환경이 요구되어 대부분의 뷰티산업 종사자에게는 현실적으로 사용하기 쉽지 않고 일반 피부 미용 이용자에게도 피부상태 진단에 불편함과 비현실적 비용 부담으로 대중화되어 사용하기가 매우 어려운 실정이다. 피부미용 분야에서도 다양한 비침습적 측정기의 발달이 이루어지고 있어 피부상태 진단에 더욱 편리함을 주는 환경이 개선 되어지고 있으나 이러한 측정장비를 사용할 수 없는 일반 사용자 및 뷰티 산업종사자에게는 아직 비현실적 환경이라 할 수 있다.

뷰티산업종사자에 의한 고객의 피부상태에 대한 진단 과정에서 피부 미용 고객 자신의 피부상태에 대한 평가와 표현이 매우 주관적이기 때문에 전문가가 기대하는 전문적 용어와 표현이 상이하고 불일치하는 현상이 발생하고 있다. 이와 같은 피부상태 평가와 표현에 대한 피부미용 고객과 뷰티산업종사자 사이의 차이로 인하여 결과적으로 피부상태 진단과 분석이 객관적이지 못하거나 올바르게 못하게 되는 문제가 발생할 수 있다. 즉, 피부상태에 대한 상담을 희망하는 일반인과 뷰티산업종사자간의 이해와 표현의 불일치는 정확한 피부상태 진단과 관리를 어렵게 만든다. 또한 기존의 직접 대면을 중심으로 하고 있는 시대와 다르게

많은 활동이 온라인 비대면 환경으로 급속하게 발전하고 온라인을 기반으로한 새로운 세대가 나타나고 있다. 그러나 뷰티산업종사자가 기존의 방식만을 의존하고 새로운 소셜미디어에 나타난 고객의 니즈를 파악하지 못하는 경우 피부미용 진단과 관리에 있어서 더욱더 많은 어려움이 발생하게 된다. 따라서 뷰티산업종사자는 피부미용 관리 및 피부상태 진단에 있어서 피부미용 고객이 소셜미디어 커뮤니티와 같은 다양한 온라인 환경에서 자신의 피부미용 및 피부상태와 관련하여 요구하는 사용자의 요구 사항을 이해 및 파악하고 이를 참고하여 고객의 피부미용 관리 및 피부상태 진단의 상호작용을 할 필요가 있다. 피부미용과 화장품과학 분야의 기술 추세에 의하면 포괄적 고객을 대상으로한 보편적 기능을 갖는 화장품과 보편적 피부미용 및 관리 방식에서 개인맞춤형 화장품과 개인에게 최적화된 피부미용을 위한 방향으로 발전되어 가고 있다.

본 논문에서는 이러한 피부상태 진단시 사용되는 육안 관찰과 문진방법의 과정에서 보다 객관적이고 정확한 피부상태 진단과 분석을 위해 최신의 정보통신기술과 빅데이터 기술을 활용하여 피부상태 진단에 필요한 정보를 추출 및 분석하는 소셜미디어 기반 빅데이터 분석시스템을 개발하고자 한다.

II. Related Research

2.1 Skin care in cosmetics and beauty industry

일반적으로 뷰티산업종사자에 의해 진행되는 피부상태 평가시 피부 표피 최상부의 각질층에 대한 평가가 매우 중요하다. 뷰티산업종사자들이 피부 미용 고객들을 대상으로 현재까지 일반적으로 많이 사용하고 있는 방법은 시진, 즉 육안으로 관찰하는 방법이다. 피부 각질층의 표면상태는 피부 각질층 하부의 다양한 변화를 반영하고 있으므로 육안으로 피부상태를 관찰하여 피부상태 유형을 판단하는 것은 타당성있는 뷰티산업분야의 판별 방법이라 할 수 있다. 피부 각질층에 대한 육안 판단으로 판단되는 피부 유형의 종류는 크게 피부표면의 형태적 요소에 대한 분류 종류와 피부표면의 색채적 요소에 대한 분류 종류로 나누어진다[1].

2.2 Skin care using big data and artificial intelligence technology

4차 산업혁명시대로 진입하고, 경제적 문화적인 생활 수준이 향상됨에 따라 피부 관리에 대한 관심이 매우 높아지고 있다. 이에 따라 화장품 업계에서는 피부 측정 도구

를 이용하여 고객 개별적인 피부상태를 진단하고 피부상태에 대한 컨설팅과 필요 제품을 추천해주는 등 이에 대한 기술 연구가 요구되고 있다.

이와 같은 화장품 및 뷰티산업 분야의 기술 요구에 따라 기계학습 인공지능 기술을 활용하여 피부상태 진단을 위한 연구가 진행되어 왔다. 피부상태 진단에 대한 기존 연구에서는 준비된 피부상태 이미지를 활용하여 선형적 분류 알고리즘인 SVM(Support Vector Machine)으로 기계학습을 진행하고 이를 근거로 피부 이미지를 지성과 건성으로 분류하는 피부상태 진단 모델에 관한 연구를 진행하였다[4]. SVM은 기계학습 분류 알고리즘 가운데 하나로 선형 분리 모델에 대한 대표적인 기계학습 방법이다.

기존에 연구된 기계학습 인공지능 기술을 활용한 피부상태 유형의 진단은 훈련데이터의 준비에 대한 한계점과 피부상태 이미지로부터 기하학적 특징점을 추출하는데 많은 시간적 비용부담과 모든 특징을 추출하기 어렵다는 한계점이 있다. 또한 피부상태 유형은 간단한 구분 경계선에 의해 구분되지 않는 비선형적 특성을 가지고 있기 때문에 SVM과 같은 선형 분리 모델을 사용하는 것은 실질적으로 사용하는데 한계점이 있다.

기존의 화장품 및 뷰티산업 분야에 대한 빅데이터의 활용 연구는 화장품 관련 기업에서 요구되는 고객과 시장의 특성을 파악하거나 화장품 및 미용소비에 대한 트렌드를 파악하기 위하여 사용되었다. 기능성 화장품 마케팅의 소셜 빅데이터 분석 활용에 대한 연구에서는 소셜미디어로부터 임상부 화장품 관련 본문과 댓글을 수집, 분석하여 화장품 중소제조기업인 H사를 대상으로 신제품 마케팅 전략 도출에 활용하였다[18]. 빅데이터 분석을 통한 화장품 트렌드 변화 추이에 관한 연구에서는 코로나 19로 인해 변화된 화장품 시장과 소비자의 니즈를 파악하고자 소셜미디어로부터 “화장품” 키워드를 중심으로 빅데이터를 수집하고 분석함으로써 피부를 위한 기초화장품과 성분에 대한 관심이 높아지는 분석 결과를 도출하였다[19]. 화장품 및 뷰티산업 분야에서 대부분의 빅데이터의 활용이 고객의 일반적인 소비 트렌드와 시장의 추이를 파악하는 것으로 주로 연구되어 기업 경영과 관련된 거시적인 정보 분석으로 제한되어 사용된 한계점이 있다.

의 피부상태를 파악하는 것이다. 피부미용 대상자의 피부상태는 연령, 성별, 생활지역, 생활환경 등 다양한 요인에 의해 많은 영향을 받는다. 따라서 보다 정확하게 고객의 피부상태 진단 및 관리를 위해서는 이에 영향을 줄 수 있는 다양한 요인에 대한 정보를 활용하는 것이 필요하다.

피부미용 목적으로 뷰티산업종사자가 피부상태 진단을 위해 일반적으로 사용하고 있는 요소는 표피 최상부의 각질층에 대한 피부상태이다. 본 논문에서 제안한 SCIS시스템은 소셜미디어 기반의 빅데이터를 수집, 분석하여 피부상태 진단 및 평가에 사용할 수 있는 중요한 정보를 제공함으로써 뷰티산업종사자가 보다 정확한 진단 및 관리를 할 수 있도록 한다.

본 논문에서 제안한 SCIS은 그림 1과 같이 빅데이터수집단계, 텍스트전처리단계, 이미지전처리단계, 텍스트단어 분석단계로 구성된다.

첫 번째 단계인 빅데이터수집단계(crawlingInstagram)는 인스타그램을 대상으로 피부상태 진단 및 관리에 있어서 가장 중요하게 사용하고 있는 표피 최상부의 각질층에 대한 평가에 사용되는 피부표면의 형태적 요소와 피부표면의 색채적 요소를 검색 키워드로 사용한다. 빅데이터수집단계는 인스타그램의 해시태그 기반의 탐색 기능을 사용한다. 피부표면의 피부상태 진단에 대한 평가요소를 인스타그램의 검색 키워드로 사용하여 검색된 검색 목록에 대한 게시물에 대해 이미지와, 게시물 본문 및 해시태그, 댓글 본문 및 댓글 해시태그, 게시된 날짜, 좋아요 등에 대한 빅데이터 정보를 수집한다.

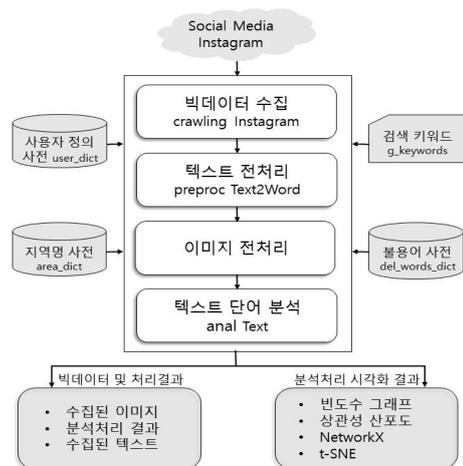


Fig. 1. Overview of SCIS

III. The Proposed System

3.1 Overview of SCIS

뷰티산업종사자와 화장품산업종사자에게 피부상태 진단 및 관리에 있어서 가장 기본적이며 필수적인 정보는 고객

두 번째 단계인 텍스트전처리단계(preprocText2Word)는 수집된 각 게시물의 데이터 가운데 텍스트 데이터인 본문 내용, 해시 태그, 댓글 내용, 댓글 해시 태그를 대상으

로 문장의 형태소 분석 과정을 진행하여 단어를 분리 추출하고 각 단어의 품사 유형을 구분한다. 본 논문에서 피부상태 진단과 분석을 위한 자료로 사용하기에 적합한 품사 유형을 중심으로 핵심 키워드에 해당하는 주요 단어를 추출하고 단어의 길이가 2미만인 것과 자체적으로 준비한 불용어 단어에 포함된 것을 제거한다. 각 게시물에 대하여 이와 같은 과정을 통해 단어를 추출하고 각 단어의 출현 빈도수를 계산한다. 이 결과는 파일로 저장되어 다음 단계에 이용될 수 있도록 한다.

세 번째 단계인 이미지전처리단계에서는 빅데이터수집 단계에서 수집된 이미지에 대해 피부이미지 및 얼굴이미지를 탐지하고 피부상태 분석과 진단에 사용될 피부이미지를 저장한다. 본 논문에서는 텍스트만을 대상으로 연구가 진행되어 피부이미지분석에 대한 처리과정은 생략한다.

네 번째 단계인 텍스트단어분석단계는 핵심단어 빈도수 및 상관성분석(analWordFrqRelation)과 핵심단어 동시출현 분석(analWordPairNetwork)과정으로 진행한다. 핵심단어 빈도수 및 상관성분석과정에서는 각 검색 키워드에 대한 텍스트전처리단계의 결과로 저장된 텍스트 단어 빅데이터 엑셀파일을 읽어들이어 단어의 출현 빈도수를 기준으로 검색 키워드별 핵심단어를 추출하고, 시기별 핵심단어의 상관성 분석, 지역별 핵심단어의 상관성 분석 등을 진행한다.

사용자정의사전은 텍스트전처리단계에서 게시물의 텍스트로부터 단어를 추출할 때 코엔엘파이 Komoran 모듈에 사용 가능한 옵션으로서 사용자가 필요에 따라 단어와 품사유형을 정의하고 단어 추출하는데 사용된다[3,4,5]. 피부상태 진단 및 관리에 필요한 단어를 사용자정의사전에 등록하여 관련 단어가 추출이 되도록 한다. 불용어사전은 사용자가 필요에 따라 수정가능하며 텍스트전처리단계의 코엔엘파이 Komoran 모듈에 의해서 추출된 단어 목록에서 불필요한 단어를 삭제하기 위해 사용된다. 지역명사전은 텍스트단어분석단계에서 지역과 관련된 정보를 분석하기 위하여 지역명을 등록하는데 사용된다.

3.2 Proposed system SCIS

본 논문에서 제안한 SCIS는 빅데이터수집단계, 텍스트전처리단계, 이미지전처리단계, 텍스트단어분석단계로 구성된다. SCIS는 소셜미디어 인스타그램을 통한 빅데이터를 활용한다. 빅데이터수집단계에서는 인스타그램으로부터 피부상태 분석에 필요한 빅데이터를 수집하고, 텍스트전처리단계에서는 수집된 빅데이터에 대한 한국어 문장분석과 불용어제거 등을 수행한다. 이미지전처리단계는 수집된 빅데이터의 피부이미지의 분석을 위한 전처리과정을

진행하고, 텍스트단어분석단계는 단어의 빈도수분석, 동시출현분석, 상관성분석 등을 수행한다. SCIS 각 단계에 대한 구체적인 내용은 다음과 같다.

3.2.1 Big data collection stage

SCIS의 첫 번째 단계인 빅데이터수집단계(crawlingInstagram)에서 피부상태 진단에 대한 빅데이터를 검색하기 위해 피부표면의 형태적 요소와 색채적 요소에 해당하는 키워드를 사용한다. 피부표면의 형태적 요소에는 피부결, 모공, 각질, 주름 등이 있고, 피부표면의 색채적 요소에는 피부색, 피부톤, 홍반, 홍조, 여드름, 흑색면포, 백색면포, 색소침착 등이 있다. 이는 뷰티산업종사자와 피부미용 관련 전문가들이 상담 및 관리할 때에 사용하는 육안 확인 방식의 분류요소이다[1, 2]. 빅데이터수집단계에서는 이러한 피부표면 분류요소를 빅데이터 수집에 대한 검색 키워드로 사용한다.

피부상태 진단 및 관리에 필요한 빅데이터를 수집하기 위해서 검색 키워드 목록 g_keywords를 정의한다. 그림 2는 검색을 위한 g_keywords와 이를 사용한 빅데이터 수집 과정을 보여주고 있다.

빅데이터수집단계는 소셜미디어 인스타그램에서 제공되는 기본 기능인 게시물의 해시 태그를 사용한 검색 기능으로 진행된다. 검색 키워드 목록 g_keywords에 포함된 각 검색 키워드를 사용하여 loginInstagram()함수에서는 인스타그램을 접속하여 로그인 과정을 진행하고, instagram_crawling_image_text()함수에서는 접속된 인스타그램에서 자동적 스크롤링 과정을 통하여 게시물을 검색하고 정보를 수집하는 과정을 진행한다.

```
# 검색 키워드
g_keywords = ["피부결", "피부모공", "피부각질", "피부주름", "피부톤", "피부홍반", "피부홍조", "피부여드름", "피부색소침착", "피부상태", "피부표면"]

def main_crawlingInstagram():
    for keyword in g_keywords:
        # 인스타그램 로그인 처리
        driver = loginInstagram()

        # 인스타그램 크롤링
        instagram_crawling_image_text(driver, keyword, g_yearStart, g_yearEnd)

        # 크롬 브라우저 종료
        driver.close()
        driver.quit()
```

Fig. 2. The process of the big data collection stage

그림 3은 빅데이터수집단계에 의해 자동적으로 인스타그램의 정보를 검색하는 과정을 보여주고 있다. 그림 3의

(a)에 보이는 것과 같이 특정 검색 키워드를 검색창에 자동 입력하면 관련 검색 해시 태그 목록이 표시되고 이 해시 태그 목록을 순차적으로 선택하여 검색이 진행된다. 그림 3의 (b)와 (c)에서는 검색 키워드에 대한 게시물을 자동적으로 스크롤링하면서 데이터를 수집한다. 본 논문에서는 프로그램 수행 날짜를 기준으로 게시물의 게시된 날짜를 특정 기간 g_yearStart=2000에서 g_yearEnd=2021로 제한하였고, 각 검색 키워드별 최대 검색 데이터량은 g_crawlMax=500으로 제한하여 빅데이터를 수집하였다.

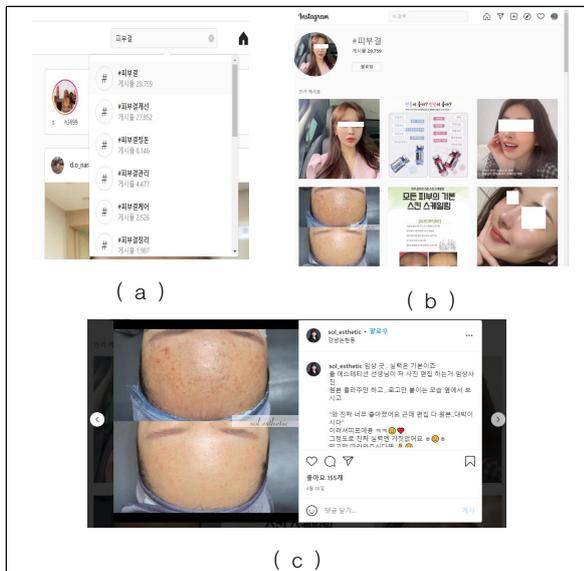


Fig. 3. An example of the search process in the big data collection stage

빅데이터수집단계는 인스타그램 게시물의 이미지, 게시물 본문, 해시태그, 게시날짜, 좋아요, 댓글 본문, 댓글 해시 태그를 수집한다. 수집된 데이터 가운데 이미지 파일들은 검색 키워드별 폴더 안에 일련번호가 부여된 JPG 형식 파일로 저장되고, 나머지 분석 대상의 텍스트 데이터는 검색 키워드별로 엑셀파일에 저장된다.

3.2.2 Text preprocessing step

SCIS의 두 번째 단계인 텍스트전처리단계는 피부상태 진단과 분석에 적합한 품사 유형 가운데 명사를 중심으로 한 핵심 단어를 추출한다. 한글 텍스트에서 원하는 품사 유형을 분리하여 추출하는 과정은 한글의 특성상 영어권 문장에 비하여 상대적으로 어려운 문제이다[6, 7]. 빅데이터 수집 및 분석 과정에서 텍스트에 대한 전처리과정인 자연어처리는 수집한 데이터를 효율적으로 가공 및 분석 처리하기 위한 필수적 과정이라 할 수 있다. 한국어 자연어 처리 과정은 코엔엘파이(KoNLPy) 오픈소스 프로젝트로

진행되고 있으며, 이 프로젝트에서 제공되는 한글 텍스트에 대한 문장 형태소 분석 모듈에는 Hannanum, Kkma, Komoran, Mecab, Okt 등이 있다. 본 논문의 텍스트전처리단계에서는 Komoran모듈을 사용하는데, 이는 효율성이 높고 사용자정의사전을 사용하여 한국어 자연어처리가 가능하다[8]. 사용자정의사전에 피부상태 진단 및 관리에 필요한 단어를 추가함으로써 한국어 자연어처리 과정에서 최대한의 이에 대한 핵심 단어와 주제가 상실되지 않고 추출될 수 있도록 한다.

텍스트전처리단계에서 추출한 형태소 유형은 Komoran에서 사용된 명사, 동사, 형용사, 보조용언, 명사추정법주, 분석불능법주이고 표 1과 같다. 본 논문에서는 Komoran의 기본적 기능을 통한 결과에 텍스트전처리단계의 사용자정의사전, 불용어사전, 지역명사전, 단어 처리기준을 적용함으로써 소셜미디어의 신조어와 새로운 용어도 중요한 분석 정보가 되도록 한다.

Table 1. Morpheme types used to extract words

Morpheme type	Tag	Morpheme type	Tag
일반명사	NNG	동사	VV
고유명사	NNP	형용사	VA
의존명사	NNB	보조용언	VX
명사추정법주	NF	분석불능법주	NA

Komoran 모듈의 옵션으로 사용되는 사용자정의사전에 피부진단과 관련된 용어를 추가하여 이를 반영한 형태소 유형 분석을 한다. 표 2는 본 논문에서 사용된 사용자정의사전의 일부 내용이다. 사용자정의사전은 빅데이터 수집 분야와 분석 목적에 따라 적절한 내용으로 구성할 필요가 있다.

Table 2. Example of User Defined Dictionary

Word	Tag	Word	Tag
피부상태	NNG	피지분비량	NNG
피부색	NNG	헤모글로빈	NNP
피부습	NNG	혈색소	NNP
피부톤	NNG	홍반	NNP
피부표면	NNG

텍스트전처리단계는 분석에서 불필요한 단어들을 제거하기 위하여 불용어사전에 포함된 단어를 제거하고 이후 텍스트 전처리과정을 진행한다. 불용어사전은 표 3과 같이 텍스트 분석 목적에 따라 사용자가 임의로 정의하고 수정할 수 있다.

Table 3. Example of a delete stopword dictionary

index	delWord	index	delWord
0	어제	3	모레
1	오늘	4	하루
2	내일

텍스트전처리단계에서는 분석 추출된 단어와 빈도수에 대한 결과 내용을 표 4와 같이 검색 키워드별로 저장한다.

Table 4. Example of text preprocessing result file for each search keyword

index	img File	date	like	reply Num	Keyword	Tag 1	Tag 1 Num	...
0	1.jpg	2021-09-09	29	1	피부표면	표면	2	...
1	2.jpg	2021-12-01	16	0	피부표면	오늘밤	1	...
2	3.jpg	2021-07-18	38	1	피부표면	콜라겐	4	...
3	4.jpg	2020-06-08	8	0	피부표면	체온	11	...
4	5.jpg	2020-04-24	29	0	피부표면	스킨세리티	2	...
5	6.jpg	2018-01-18	11	1	피부표면	모공	3	...
6	7.jpg	2021-10-27	0	1	피부표면	피부	2	...
...

3.2.3 Text word analysis stage

SCIS의 마지막 단계인 텍스트단어분석단계는 핵심단어 빈도수 및 상관성분석과정과 핵심단어 동시출현분석과정으로 구성된다. 빈도수 분석의 의미는 관심이 있거나 중요한 단어는 자주 언급이 될 것이고 빈도가 높은 단어가 중요한 주제라는 개념을 반영한 것이다. 본 논문에서는 피부상태 진단 및 관리와 관련된 주요한 단어를 인스타그램을 통하여 검색하고, 빈도수 분석에 기반하여 피부상태 진단 및 관리의 핵심단어를 추출한다.

핵심단어를 추출하기 위한 빈도수 분석은 각 단어에 대한 단순 빈도수(TF: Term Frequency)와 단어의 상대적 중요성을 고려한 상대적 빈도수(TFIDF: Term Frequency-Inverse Document Frequency)를 사용한다 [3]. 단어의 단순 빈도수와 상대적 빈도수를 근거로 분석에 필요한 단어와 정보를 추출하고, 이 단어의 상관성 및 근접성을 평가하는데 상관계수를 계산하여 사용한다. 동시출현분석은 사람들이 자신의 주장이나 의견을 표현할 때에 일반적으로 관련이 깊은 단어들을 자주 함께 사용하거나 더욱 근접하게 사용한다는 특성을 반영한 핵심정보 또는 주제를 파악하기 위한 기술이다. 본 논문에서는 핵심단어 동시출현 분석과정을 통하여 사용자들의 게시물로부터 핵심 정보 또는 주제를 추출하고 파악한다.

IV. Analysis on Proposed System SCIS

본 논문에서 제안하고 개발한 빅데이터분석시스템 SCIS는 소셜미디어 인스타그램(instagram)을 대상으로 빅데이터를 수집하고 분석하였다. 시스템 개발을 위해 파이썬(Python) v3.8과 한국어 자연어처리에 대한 공개 소스인 코엔엘파이 Komoran 등을 사용하였다. SCIS는 핵심단어 빈도수 및 상관성분석과 핵심단어 동시출현분석 과정의 수치적 결과와 함께 다양한 시각화 처리 결과를 제공하여 분석된 내용을 직관적으로 파악하고 이용할 수 있도록 하였다. SCIS에서는 분석 결과를 시각화하기 위하여 막대그래프, 산포도, 워드 클라우드, 히트맵, NetworkX, t-SNE, 클러스터링을 사용하였다.

4.1 Big data collection and Text preprocessing

빅데이터 수집에 사용한 검색 키워드 g.keywords는 "피부결", "피부모공", "피부각질", "피부주름", "피부톤", "피부홍반", "피부홍조", "피부여드름", "피부색소침착", "피부상태", "피부표면"이다. 빅데이터수집단계에서 검색된 게시물은 총 4,049개이고 각 키워드별 검색 수량은 표 5와 같다. 빅데이터 수집 결과는 각 키워드별 엑셀파일로 각각 저장되고, 그림 4는 "피부결" 키워드에 대한 엑셀파일의 일부이다.

텍스트전처리단계에서는 그림 4의 빅데이터수집결과로 얻어진 텍스트로부터 단어들을 추출하고 단어분석을 위한 전처리과정을 진행하였다. 또한 피부상태 진단 및 관리에 적합한 품사 유형의 단어를 선별하고 이 단어의 빈도수를 계산하여 이후 분석단계에 사용할 수 있도록 하였다.

A	B	C	D	E	F	G	H	I	J	K	L
1	content	tags	date	like	place	replyNum	plyConten	replyTags	imgFile	Keyword	
2	0	피부진정아 유효광고	2021-11-2	436		12	완전 순		/result_da#피부결게시물	29,720	
3	1	일상 속 스킨케어	2021-04-2	155		2	인진해		/result_da#피부결게시물	29,720	
4	2	NEW 필링 임산부가	2021-05-2	254		3	Hi plz		/result_da#피부결게시물	29,720	
5	3	구매하러가기	9시2021-07-2	148		4	오와	홈케어	/result_da#피부결게시물	29,720	
6	4	요즘 스킨: 유효광고	2021-06-0	706		6	역시 화산		/result_da#피부결게시물	29,720	
7	5	면니를 논리 리비양	2021-01-0	566		12	피부는 나	핀	/result_da#피부결게시물	29,720	

Fig. 4. Collected big data for "피부결"

Table 5. Summary of Big Data Search Results

Search keyword	Posts	Like	Date
피부결	515	68.2	2021-01-07~2021-11-29
피부모공	176	73.7	2018-01-18~2021-11-29
피부각질	539	65.3	2019-08-21~2021-11-29
피부주름	515	25.9	2020-11-12~2021-11-29
피부톤	500	64.8	2020-05-31~2021-12-01
피부홍반	44	38.1	2016-09-17~2021-11-04
피부홍조	500	49.0	2017-12-24~2021-11-28
피부여드름	264	14.3	2017-01-16~2021-11-29
피부색소침착	226	132.8	2019-11-13~2021-12-02
피부상태	521	53.4	2015-10-12~2021-12-01
피부표면	249	33.5	2015-07-15~2021-12-01
총계	4,049		

텍스트전처리결과는 각 키워드별 엑셀파일로 각각 저장되고, 그림 5는 “피부결” 키워드에 대한 엑셀파일의 일부이다.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	imgfile	date	like	replyNum	Keyword	Tag 1	TagNum 1	Tag 2	TagNum 2	Tag 3	TagNum 3	Tag 4	TagNum 4	Tag 5	TagNum 5	Tag
2	0	/result_da2021-11-2	436	12	피부결	피부	7	클레이	6	진정	3	마스크	3	제형	3	피부결
3	1	/result_da2021-04-2	155	2	피부결	여드름	4	피부결	2	실력	2	집	2	일상사	2	원본
4	2	/result_da2021-05-2	254	3	피부결	피부	6	스케일링	3	관리	3	여드름	2	재생	2	지극
5	3	/result_da2021-07-2	148	4	피부결	화장품	4	피부결	3	구매	3	볼트	3	발라	3	피부본
6	4	/result_da2021-06-0	706	6	피부결	모공	4	화산	4	중이	4	클레이	3	각질	2	피부
7	5	/result_da2021-01-0	566	12	피부결	피부	24	열물	20	모공	6	리버	6	블루	6	열드

Fig. 5. Result of the text preprocessing for “피부결”

텍스트전처리단계의 결과는 그림 5에 보인것과 같이 저장된 이미지 파일명, 게시된 날짜, 좋아요 수, 댓글수, 검색 키워드, 추출된 단어 개수 만큼의 단어와 단어 빈도수로 구성되어 각 검색 키워드별로 저장된다.

4.2 Text Word analysis

텍스트단어분석단계는 핵심단어빈도수 분석 및 상관성 분석과 핵심단어 동시출현분석의 두 개의 과정으로 진행된다. 핵심단어빈도수분석 및 상관성분석과정에서는 핵심단어와 중요한 주제를 파악하기 위하여 단어의 단순빈도수(tf)와 상대빈도수(tfidf)을 계산하고 이를 시각화하였다.

4.2.1 Key word frequency analysis and correlation analysis

1) Key word frequency analysis

그림 6은 각 검색 키워드에 대한 단순빈도수를 기준으로 추출된 상위 핵심단어를 막대그래프와 워드클라우드로 시각화한 결과의 일부이다. 이를 통하여 각 피부상태 진단 및 관리와 관련된 고객들의 주요 관심사와 중요 주제를 파악할 수 있었다. “피부결”과 관련되어서는 화장품, 수분, 여드름, 각질에 대한 내용이 고객들의 주요 관심사이고 중요한 주제였음을 알 수 있었다.

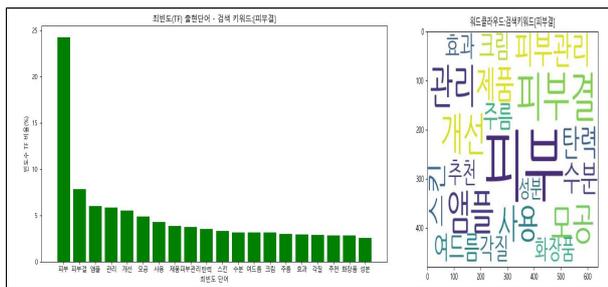


Fig. 6. Results of words frequency for “피부결”

또한 좋아요 항목에 대한 정보를 고려하여 단어의 빈도수를 평가하고 상위 빈도수 단어를 추출하였다. 그 결과 각 검색 키워드에 대한 상위 빈도수 단어의 대부분의 단어가 유사하게 추출 되었으나 일부 새로운 단어가 추가되거나 기존 단어가 상위 순위에서 제외되는 차이를 보였다.

그림 7은 전체 수집된 빅데이터로부터 추출된 단어의 단순 빈도수와 좋아요를 기준으로 상위 빈도수를 갖는 핵심 단어의 분포 상황을 막대그래프와 워드클라우드로 시각화 하여 보여주고 있다. 피부상태 진단 및 관리에 있어서 여드름, 각질, 주름, 모공 등의 주제에 많은 관심을 가지고 있으며, 이를 치료하고 개선하기 위한 화장품, 수분 등의 방법에 많은 관심을 보일 것으로 분석된다. 이와 같은 분석결과는 뷰티산업종사자와 화장품산업종사자 입장에서 피부상태 진단 및 관리 측면에 있어서 적절한 대응책을 수립하는데 사용될 수 있다.

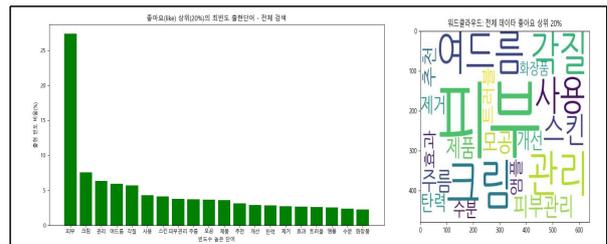


Fig. 7. Results of words frequency and likes

핵심단어빈도수분석과정에서는 단순빈도수와 상대빈도수(TF-IDF)를 사용하여 분석하였다. 상대빈도수는 단어 빈도수(TF)와 역문서 빈도수(IDF)의 곱이다. 단어빈도수 $tf(t, d)$ 는 문서 d 에서 단어 t 의 빈도수를 나타내고, 역문서 빈도수 $idf(t, D)$ 는 단어 t 가 포함된 문서수의 전체 문서수 D 에 대한 상대 빈도값이다. 여기에서 문서는 각 게시물을 의미한다. 단어빈도수와 역문서 빈도수의 두 값을 산출하는 방법에는 여러 가지가 있으며, 본 논문에서는 식 (1)의 수식을 사용하였다.

$$tf(t, d) = \begin{cases} 1, & \text{단어 } t \text{ 가 문서 } d \text{ 에 한 번이라도 나타난 경우} \\ 0, & \text{그외의 경우} \end{cases}$$

$$idf(t, D) = \log \left(\frac{|D|}{|\{d \in D : t \in d\}| + 1} \right) \quad (1)$$

$$tf.idf(t, d, D) = tf(t, d) \times idf(t, D)$$

그림 8은 각 검색 키워드에 대한 상대빈도수(TF-IDF)를 기준으로 추출된 상위 핵심단어를 막대 그래프와 워드클라우드로 시각화한 결과의 일부이다.

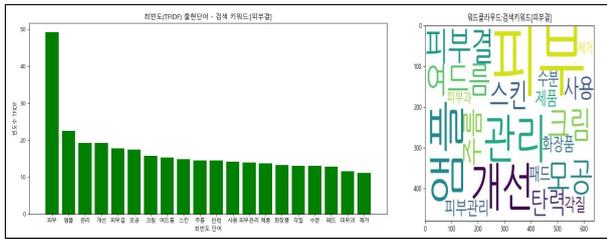


Fig. 8. Results of TFIDF for “피부결”

표 6은 각 검색 키워드에 대한 상대 빈도수를 기준으로 추출된 상위 핵심단어의 목록을 보여주고 있다.

Table 6. Top TFIDF Frequency Words by Search Keywords

Keyword	MostWords
피부각질	피부,각질,스킨,피부관리,관리,제거,사용,모공,플래닝,화장품,드리,예약,추천,시술,제품,트러블,인천,보습,크림,뷰티
피부결	피부,애플,관리,개선,피부결,모공,크림,여드름,스킨,주름,탄력,사용,피부관리,제품,화장품,각질,수분,패드,피부과,제거
피부모공	피부,모공,여드름,관리,탄력,스킨,화장품,플리,주름,피부과,피부관리,효과,뷰티,제주,치료,제품,개선,세안,닥터,버블
피부상태	피부,화장품,치료,크림,피부상태,스킨,여드름,뷰티,광주,관리,수분,피부관리,오가,피부과,사용,상담,소통,성분,제품,추천
피부색소침착	크림,여드름,강릉,피부,트러블,노스,색소침착,사용,관리,자국,기미,진정,들르,피부관리,테스트,완료,제품,페어,얼굴,바르
피부여드름	여드름,얼굴,피부,관리,트러블,피부관리,강남,오전,흉터,원인,오후,좁쌀,평일,없애,치료,영업시간,토요일,농성,전화,자국
피부주름	콜라겐,피부,공구,제품,조여정,만원,기능,건강,여성,피부관리,주름,관리,선물,자외선,식품,데려오,손상,가격,성분,구성
피부톤	피부,크림,순환,디톡스,탄력,아이오페,피부톤,포텐,고객,맞춤,라인,이지,중상,생기,개선,체험,관리,주름,바이오,사용
피부표면	스킨,플래닝,수강,광주,피부,전남,피부관리,전북,원데이,학원,수완,피부미용,프로그램,제거,개선,각질,더마플래닝,피부표면,관리,노폐물
피부홍반	피부,치료,피부과,홍조,아토피,관리,홍반,광주,의원,레이저,시크릿,화장품,효과,부스트,마티스,이지스,마스크,난치병,습진,사용
피부홍조	피부,쿠션,관리,크림,피부관리,커버,팩트,여드름,홍조,추천,로하,여름,재생,트러블,쉬드,뷰티,수분,자외선,진정,체온

상대빈도수분석을 통하여 단순빈도수분석에서 파악하지 못했던 핵심단어 및 중요 주제를 추가적으로 추출할 수 있었다. 예를 들면 표 6에서 피부각질에 대하여 단순 빈도수에 의한 정보에서는 파악되지 않았던 “크림”과 “인천”이라는 단어가 추가적으로 추출된 것을 볼 수 있으며, 피부결에 대하여 “패드”와 “피부과”라는 단어가 중요 주제로 추가 추출되었다.

2) Keyword correlation analysis

피부상태는 계절적 환경변화에 따라 영향을 받을 수 있다. 따라서 본 논문에서는 피부상태 진단 및 관리와 관련하여 계절별 상관성을 파악하기 위하여 단어와 시기의 상관계수를 계산하였고, 게시물에 대한 사용자들의 좋아요를 근거로 사용자들의 관심도와 지지도를 파악하였다. 그림 9에서는 분석된 각 핵심단어와 시기에 대한 유의미한 상관성 결과를 나타내고 있다.

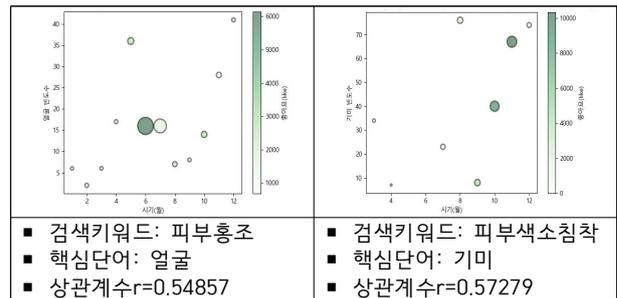


Fig. 9. Results of correlation analysis between keywords and times

그림 9에서 사용자들의 관심도 및 지지도를 좋아요수와 댓글수를 근거로 산점도의 색상의 농도와 크기로 표현하였다. 시기별로 피부홍조와 관련된 주제에서 핵심단어 얼굴과의 상관성은 상관계수 r=0.54857로 상관관계가 어느 정도 있는 것으로 평가되고 시기적으로 5~6월에 사용자들이 많은 관심을 보이고 있다는 것을 알 수 있다. 이와 같이 피부상태 진단 및 관리와 관련된 특정한 주제와 시기와의 관련성을 파악함으로써 효과적 피부상태 진단 및 관리에 요구되는 대응책을 수립할 수 있다.

본 논문에서는 피부상태 진단 및 관리와 관련하여 지역과의 상관성을 파악하기 위하여 월별 도시 출현 빈도수를 분석하였다. 그림 10은 월별 도시 출현 빈도수를 막대 그래프로 표현한 결과이다. 그림 10에서 피부주름이란 주제는 광주, 부산, 서울 등에서 6월과 8월에 집중적으로 언급되었음을 나타내고 있고, 피부각질이란 주제는 인천에서 6월에 집중적으로 언급되었음을 나타내고 있다.

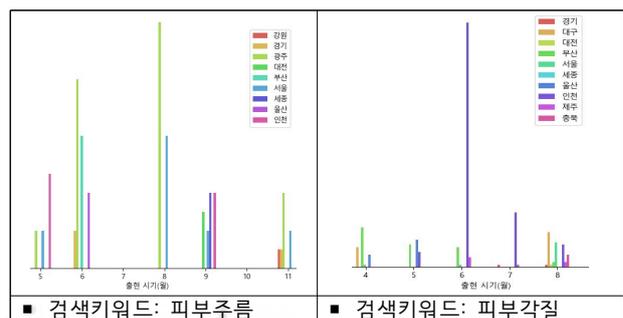


Fig. 10. Results of city name frequency by month

텍스트 정보 분석과정의 빈도수를 통해 검색된 단어는 매우 방대하여 분석된 내용을 쉽게 파악하기 어렵게 만든다. 이러한 문제점을 보완하기 위해 본 논문에서는 t-SNE(t-distributed Stochastic Neighbor Embedding)를 추가적으로 사용하였다. t-SNE는 추출된 방대한 분량의 단어로 된 높은 차원의 복잡한 정보를 2차원으로 축소하여 시각화하는 방법이다. 유사한 정보는 가깝게 배치하고, 유사하지 않은 정보는 멀리 떨어지게 배치하여 중요 정보 및 단어간의 관련성을 시각적으로 쉽게 파악할 수 있도록 한다[10].

그림 11은 추출된 단어를 t-SNE 방법으로 차원을 축소하였고 이 결과를 군집분석 방법으로 시각화한 결과이다. 군집 분석 방법으로는 K-평균(K-means) 방법을 사용하였으며 군집 개수 K=10으로 임의 지정하여 처리한 결과이다.

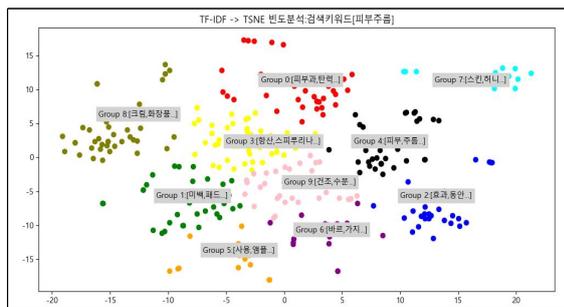


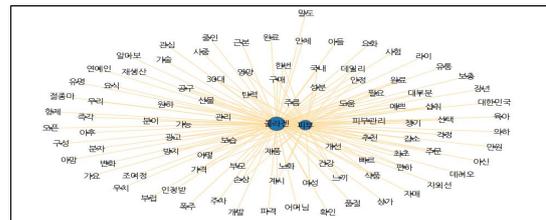
Fig. 11. Results of t-SNE and Clustering(k=10)

예를 들어 그림 11에서 검색 키워드 피부주름에 대하여, 그룹 0(Group 0)은 콜라겐, 탄력, 초음파 리프팅 등과 관련성이 있고, 그룹 1(Group 1)은 미백, 피부결, 히알루론산 등과 관련성이 있음을 확인할 수 있었다.

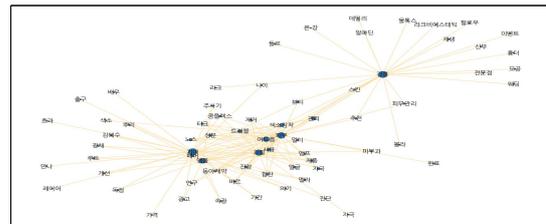
4.2.2 Keyword co-occurrence analysis

본 논문에서는 피부진단 및 관리에 대한 중요한 주제를 파악하기 위하여 단어 동시출현분석을 진행하였다. 여기에서 단어 동시출현분석이란 게시물 내용에 단어1과 단어2가 동시에 들어가 있을 때 두 단어는 연관성이 있다고 판단하는 기법이다.

그림 12는 SCIS에서 단어 동시출현분석을 통해 추출된 연관성 높은 핵심 단어의 NetworkX를 사용한 시각화된 결과이다.



(a) 검색키워드: 피부주름



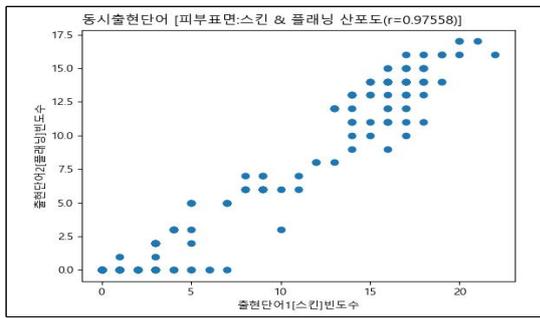
(b) 검색키워드: 피부색소침착

Fig. 12. Results of keywords co-occurrence frequency with NetworkX

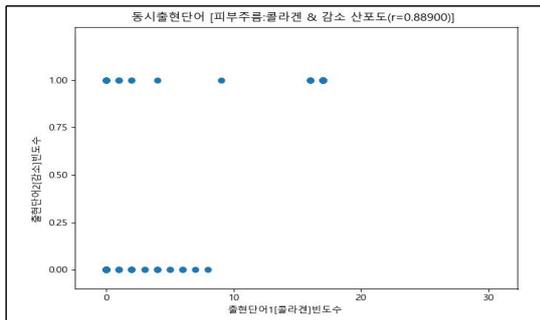
NetworkX로 표현된 그래프에서 노드(node)는 동시출현된 단어를 나타내고 있고, 노드 사이의 간선(edge)은 두 단어가 동시에 출현된 단어임을 나타낸다. 또한 노드의 크기는 해당 단어가 출현하는 빈도수의 크기를 상대적으로 나타내는 것으로 큰 크기는 더 많은 출현 빈도수를 의미한다[11].

그림 12의 (a)는 검색키워드 피부주름에 대한 동시출현 단어들의 NetworkX 그래프를 나타내고 있다. 여기에서 피부주름은 “콜라겐”과 “피부”를 중심으로 “주름”, “보습”, “노화”, “탄력”등의 핵심 주제와 밀접한 관련성이 있음을 알 수 있었다. 그림 12의 (b)는 검색 키워드 피부색소침착에 대한 동시출현 단어들의 NetworkX 그래프를 나타내고 있다.

그림 13은 단어 동시출현분석결과 획득된 단어들 간의 상관계수를 계산하고 이들 가운데 높은 상관성을 갖는 단어의 상관성계수와 산포도 결과를 나타내고 있다. 그림 13의 (a)는 검색 키워드 피부표면에서 동시출현된 단어 가운데 “스킨”과 “플래닝”의 높은 상관성(r=0.97558)을 보여주고 있다. 피부표면과 관련하여 “스킨”과 “플래닝”이 중요한 연관성을 갖는 주제로 다루어지고 있음을 알 수 있다. 그림 13의 (b)는 검색 키워드 피부주름에서 동시출현된 단어 가운데 “콜라겐”과 “감소”의 높은 상관성(r=0.889)을 보여주고 있다.



(a) “스킨” x “플래닝” (검색키워드:피부표면)



(b) “콜라겐” x “감소” (검색키워드:피부주름)

Fig. 13. Results of correlation coefficient for co-occurring words

V. Conclusions

컴퓨터 및 정보통신기술의 발전과 함께 4차 산업혁명기술이 다양한 분야와 실생활에 적용되어 가고 있으며, 화장품 및 피부미용 분야에서도 혁명과 같은 큰 변화와 발전이 일어나고 있다.

기존 연구에서는 피부 이미지로부터 정형적인 형태의 기하학적 특징점을 추출하는 제한점과 준비된 훈련데이터를 전제한 방법으로 실용화하는데 있어서 한계점을 갖고 있다. 또한 기존의 화장품 및 뷰티산업 분야에서 대부분의 빅데이터의 활용이 고객의 일반적인 소비 트렌드와 시장의 추이를 파악하는 것을 목적으로 주로 연구되어 기업 경영과 관련된 거시적인 정보 분석으로 제한되어 사용된 한계점이 있다. 따라서 변화되는 화장품 및 피부미용에 대한 고객들의 요구 사항을 지속적이면서 실시간적인 관점에서 파악하고 대응할 필요가 있고 이 과정을 체계적이며 효율적으로 진행할 수 있는 빅데이터분석시스템이 필요하다.

본 논문에서 개발한 SCIS에서는 인스타그램을 대상으로 피부상태 진단에 필요한 빅데이터를 수집 및 분석하였다. SCIS의 빅데이터수집단계에서는 피부상태 진단 및 관리와 관련된 주요 키워드를 검색 추출하였다. SCIS의 텍

스트전처리단계에서는 수집된 텍스트로부터 핵심 단어 및 중요한 주제를 추출하기 위해 문장분석, 불용어제거 등의 전처리과정을 진행하였다. SCIS의 텍스트분석단계에서는 텍스트 전처리 결과로부터 단어의 빈도수, 상대빈도수, 동시출현분석, 상관성을 분석하고 시각화하였다.

본 논문에서 개발된 시스템을 사용하여 피부상태 진단 및 관리와 관련된 빅데이터를 실험적으로 수집하고 분석하였다. 본 논문의 빅데이터분석 처리 결과 각 피부상태 진단 및 관리에 필요한 핵심단어 및 정보를 파악하고, 시기별, 지역별 상관성 등의 새로운 분석된 정보를 알 수 있었다. 본 논문의 시스템은 다양한 고객 니즈를 반영한 맞춤형 상품과 서비스 개발 및 개선에 활용될 수 있을 것이다. 화장품산업과 뷰티산업분야가 4차 산업혁명기술과 함께 급속하게 변화하고 발전해가는 상황에서 본 논문에서 제안한 빅데이터 활용 시스템은 매우 미래 지향적 시스템 체계라고 할 수 있을 것이다.

향후 빅데이터의 활용성을 높이기 위해서는 텍스트 분석뿐만 아니라 수집된 이미지에 대한 추가적인 이미지 분석 연구가 필요하다. 피부상태 이미지 분석 연구는 본 논문의 텍스트 기반 분석 결과와 함께 기계학습의 인공지능 기술과 같은 보다 발전된 피부상태 진단 시스템을 구축하는데 필요한 기초 연구자료가 될 것이다. 이러한 피부상태 이미지와 텍스트 정보를 활용하는 시스템은 더욱더 정확한 고객 니즈를 고려한 화장품산업 및 뷰티산업에서의 활용이 가능할 것으로 기대된다.

REFERENCES

- [1] Eui-Hyang Kim, Hyun-joo Kim, “A Study on the Analysis Method of Skin Condition through Visual Confirmation of Skin Surface”, Journal of Convergence for Information Technology, Vol. 11. No. 5, pp.267-275, May. 2021
- [2] J. S. Koh, “Observation of Sebum, Moisture content and Microtopography of Different aged Skin for classification of facial Skin Types”, Journal of the Society of Cosmetic Scientists of Korea, Vol. 15. No. 1, pp.63-70, 1989
- [3] Salton G. and McGill, M. J., “Introduction to modern information retrieval”, McGraw-Hill, ISBN 0-07-054484-0., 1983
- [4] Kim Ki Young, “A Study on Model of Skin Type Judgment Tool Using Machine Learning Technique”, The Treatise on The Plastic Media, Vo4. No. 21, pp.115-121, Nov. 2018
- [5] Zheng, Tong, Zha. Dongmiao, Wang, Xingtian, Gao, Weijun, Miao, Sheng, “Architecture Descriptions Analysis Based on Text Mining and Crawling Technology”, 2021 IEEE 3rd International

- Conference on Frontiers Technology of Information and Computer, pp.138-143, Nov. 2021
- [6] Jae-Been Ha, Do-Eun Lee, "A Trend Analysis and Policy proposal for the Work Permit System through Text Mining: Focusing on Text Mining and Social Network analysis", Journal of Convergence for Information Technology, Vol. 11. No. 9, pp.17-27, 2021
- [7] Yun-Mi Jun, Ki-Jung Kang, "Communitycare Key words and Social Network Analysis based on Big Data", Korean Journal of Family Welfare, Vol. 24. No. 2, pp.251-269, June. 2019
- [8] Shineware, KOMORAN, <https://komorandocs.readthedocs.io/>
- [9] Kim, Sun-ji, Chon, HaeJung, "Analysis of the Relationship between Knowledge and Praticce of Skin Moisture and Facial Parts WaterContent ", Journal of the Korean society for Wellness, Vol. 13. No. 4, pp81-89, Dec. 2018
- [10] Chatzimparmpas, A, Martins, R.M, Kerren, A, "t-viSNE: Interactive Assessment and Interpretation of t-SNE Projections", IEEE Transactions on Visualization and Computer Graphics, Vol. 26. No. 8, pp.2696-2714, Aug. 2020
- [11] NetworkX developers, NetworkX, <https://networkx.org/>
- [12] Hyung-Woo Lee, "Development of Supervised Machine Learning based Catalog Entry Classification and Recommendation System", Journal of Internet Computing and Services, Vol. 20. No. 1, pp.57-66, Feb. 2019
- [13] Dae-Chul Kim, Cheol_hee Lee, Myong-Hui Choi, Yeong-Ho Ha, "Skin Detection Method using Color Space based Methods and Focus Region", Journal of Korean Society for Imaging Science & Technology, Vol. 18. No. 4, pp.16-22, Dec. 2012
- [14] Jeonghoon Kang, Hojung Lim, Wonsik Ko, "Visualization of Text Mining Technology based on Big Data", The Korean Society of Visulalization, Dec. 2015
- [15] Won-Jo Lee, "A Study on Word Cloud Techniques for Analysis of Unstructured Text Data", The Journal of the Convergence on Culture Technology, Vol. 6. No. 4, pp.715-720, Nov.. 2020
- [16] M. Han, Y. Kim, C. Lee, "Analysis of News Regarding New southeastem Airport Using Text Mining Techniques", Smart Media Journal, Vol. 6. No. 1, 2017
- [17] Joon-Hwan Kim, Hyung-Jin Mun, Hang Lee, "A Study on Trend Analysis in Convergence Research Applying Word Cloud in Korea", Journal of Digital Convergence, Vol. 19. No. 2, pp.33-38, 2021
- [18] Sin-Hae Hwang, Dong-Young Ku, Jeoung-Kun Kim, "Application of Social Big Data Analysis for CosMedical Cosmetics Marketing : H Company Case Study", Jouranal of Digital Convergence, Vol. 17., No. 7, pp.35-41, 2019
- [19] Lim, Hee Suk, Shin, Jae Wook, "A Study on the Trends of Cosmetics through Big Data Analysis-Focusing on text mining and semantic network analysis", The Korean Society of Illustration Research, Vol. 66. pp.85-95, 2021

Authors



Hyung-Hoon Kim received the B.S. degrees in Computer Science from Chonnam National University, Korea, in 1986. He received the M.S. degrees in Computer Science from KAIST, Korea, in 1988.

He received the Ph.D. degrees in Computer Science from Hanyang University, Korea, in 2007. Dr. Kim joined the faculty of Kwangju Womens University, Gwangju, Korea, in 1994. He is currently a Professor in the Department of Cosmetic Science at Kwangju Womens University. He is interested in web programming, medical information system, information security, and AI.



Jeong-Ran Cho received the B.S., M.S. and Ph.D. degrees in Computer Science from Chonnam National University, Korea, in 1987, 1989 and 1999, respectively. Dr. Cho joined the faculty of Kwangju Womens

University, Gwangju, Korea, in 1994. She is currently a Professor in the Department of Health Administration at Kwangju Womens University. She is interested in database, parallel computing, internet and mobile computing, and multimedia contents service.