

## Latent Dirichlet Allocation 토픽모델링을 이용한 한방 의료 서비스 분석에 관한 연구 : 의료 소비자의 온라인 리뷰를 중심으로

손채연<sup>1) #</sup> · 송연우<sup>1) #</sup> · 이승호<sup>2) \*</sup>

<sup>1)</sup> 우석대학교 한의과대학 학생

<sup>2)</sup> 우석대학교 한의과대학 병리학교실 조교수

### A Study on the Analysis of Korean Medical Services using Latent Dirichlet Allocation Topic Modeling : Focusing on online reviews by medical consumers

Chaeyeon Son<sup>1) #</sup> · Yeonwoo Song<sup>1) #</sup> · Seungho Lee<sup>2) \*</sup>

<sup>1)</sup> College of Korean Medicine, Woosuk University

<sup>2)</sup> Department of Pathology, College of Korean Medicine, Woosuk University

#### Abstract

**Objective :** This study aims to understand the consumer's needs for Korean medicine medical service using online review analysis of medical consumers.

**Methods :** We analyzed the purpose and satisfaction factors of medical service use using LDA (Latent Dirichlet Allocation) topic modeling. The data used in the study was 120,727 screened reviews written by medical consumers registered on Naver. The analyzed results were compared with the "2020 Korean Medicine Utilization Survey".

**Results :** From 2018 to 2021, the five most frequently used terms were "kindness", "treatment", "doctor", "Korean medicine", and "acupuncture". The main purpose of visiting Korean medicine medical clinic and hospital was to treat "traffic accidents" in 2018, "waist(back) pain" in 2019, "musculoskeletal pain" in 2020 & 2021. Based on the rating, reviewers were satisfied with "explanation of treatment" and "treatment attitude", and dissatisfied with "accessibility to the institution".

**Conclusion :** We concluded that the main purpose of use of Korean medicine institution was to treat musculoskeletal disorders. Based on the results of this study, it is expected that it will be used to improve Korean medicine medical service in the future.

---

**Key words :** Korean medicine, Hospital, LDA, Topic modeling, Online Review

---

---

• 접수 : 2022년 3월 21일    • 수정접수 : 2022년 4월 17일    • 채택 : 2022년 4월 23일

\*Corresponding author : Seungho Lee, Dept. of Pathology, College of Korean Medicine, Woosuk University, Jeonju-si 54986, Republic of Korea.

전화 : +82-63-290-9018, 전자우편 : eseungho@gmail.com

# Co-first author

## I. 서론

최근 우리 사회에 각종 온라인 플랫폼이 유입되어, 온라인 시장이 크게 성장하고 있다. 통계청에 따르면 온라인 쇼핑 시장은 2020년 159조에서 2021년 192조로 증가하여 20.7% 상승하였다<sup>1)</sup>. 온라인 플랫폼은 소비와 판매과정에 큰 변화를 주고 있다. 그 과정에서 작성되고 관리되는 온라인 리뷰는 다양한 분야에서 사용되고 있으며, 이는 소비자의 서비스 이용에 큰 영향을 미친다.

온라인 리뷰는 서비스를 경험한 다수의 소비자가 서비스에 대한 긍정 또는 부정적 의견을 표출하는 형태로<sup>2)</sup>, 리뷰를 통해 소비를 고민하는 이들의 불확실성을 낮추고 정보 수집에 도움을 줄 수 있어, 소비자들의 리뷰에 대한 의존도를 높인다<sup>3)</sup>. 특히나 소비자의 입장에서 공신력 있는 한 사람의 정보보다는 소비자와 유사한 다수의 일반인 정보가 더 신뢰성이 높기 때문에 리뷰를 활용하는 소비자가 점차 늘어나는 추세이다<sup>4)</sup>.

소비자들은 쇼핑, 관광, 호텔 등 다양한 분야에서 온라인 리뷰를 활용하고 있으며, 한방의료계에서도 의료소비자들이 병원을 선택할 때 온라인 리뷰를 활용하고 있다. 『2020년 한방의료이용 및 한약소비실태조사』<sup>5)</sup>에 따르면 한방 의료에 대한 정보/지식 획득 경로는 ‘방송매체(TV, 라디오 등)’, ‘종이 매체(신문, 책 등)’, ‘인터넷 사이트(포털, 한방 의료 관련 홈페이지 등)’, ‘인터넷SNS(인스타그램 등)’, ‘한방의료기관(의료기관 종사자와의 상담 등)’, ‘오프라인 홍보물(포스터, 현수막 등)’, ‘가족 등 주변 사람’, ‘기타’가 있다. 이 중 ‘주변 사람(38.6%)’을 통한 정보획득이 가장 높은 비율을 보였고, 그다음으로는 ‘방송매체(29.1%)’, ‘한방의료기관(22.9%)’, ‘인터넷 사이트(8.3%)’가 뒤를 이었다. 반면에 ‘종이매체(0.7%)’와 ‘오프라인 홍보물(1.2%)’이 가장 낮은 비율을 차지하였다. 2017년도와 비교하였을 때, ‘주변 사람’을 통한 정보 획득은 2.6%p (38.6% → 36%) 감소하였고, ‘한방의료기관’을 통한 정보 획득 또한 2%p (24.9% → 22.9%) 감소하였다. 반면에 ‘인터넷 사이트’에 의한 정보획득은 17년에 비해 3.9%p (4.4% → 8.3%) 올라 가장 높은 증가율을 보였다. 이상의 결과를 통해, 한방 의료계에서 온라인 리뷰를 포함한 ‘인터넷 사이트’가 환자들의 정보획득에 영향력이 있음을 확인할 수 있고, 이는 점차 커질 것으로 예상된다<sup>6)</sup>.

온라인 리뷰의 특성을 활용한 연구들이 쇼핑, 호텔, 음식 및 배달, 병원 등 여러 분야에서 다양하게 보고되고 있다. 특히 텍스트를 분석하는 대다수의 연구에서는 ‘LDA 토픽모델링(Latent Dirichlet Allocation Topic-Modeling)’기법을 활용하였다. LDA 토픽모델링이란, 텍스트를 대표하는 토픽들을 추출하여 의미 있는 정보를 찾아내는 확률 기반의 통계적 모델이다<sup>7)</sup>. 관련 연구로, 홍 등은<sup>8)</sup> 트립어드바이저(www.tripadvisor.com)에 등록된 전 세계 주요 관광도시 호텔에서 숙박한 고객이 영어로 작성한 리뷰와 평가점수 데이터를 LDA 토픽모델링 기법을 통해 분석하였다. 채 등은<sup>9)</sup> 구글 플레이에 등록된 국내 소셜커머스 업체와 오픈마켓 업체의 모바일 앱 리뷰를 LDA 토픽모델링 기법을 통해 감정분석과 동시출현단어분석을 수행하였다. 김 등은<sup>10)</sup> 온라인 배달플랫폼 요기요에 등록된 치킨 관련 리뷰 및 평점 데이터를 텍스트마이닝 기법을 통해 분석하고, 이를 통해 부정어 사전을 제작하였다. 의료와 관련하여 이 등<sup>11)</sup>은 중국의 Dianping (www.dianping.com)에 등록된 병원리뷰를 LDA 토픽모델링 기법을 통해 분석하여 의료서비스의 잠재된 만족요인을 추출하였다. 최 등<sup>12)</sup>은 Yelp의 공개된 특정 병원 리뷰를 텍스트마이닝 기법을 통해 분석하고, SERVQUAL 이론의 지표를 토대로 감성수준을 측정하였다.

이와 같이 다양한 분야에서 온라인 리뷰를 분석한 연구가 진행되고 있다. 그러나 현재까지 한방의료계에서 온라인 리뷰를 분석한 선행 연구사례는 드물었다. 따라서 본 연구에서는 2018년부터 2021년 10월 26일까지 국내 포털사이트 NAVER에 올라온 한방의료기관의 소비자 리뷰를 LDA (Latent Dirichlet Allocation) 토픽모델링을 하여 연도별 한방 의료 이용 동향을 파악하고, 더불어 평점 별로 긍부정 리뷰를 구분하여 리뷰에 잠재된 만족도와 관련된 요인을 파악해 의료서비스에 대한 소비자의 요구를 파악하고자 한다.

## II. 연구 설계

### 1. 연구 절차

본 연구는 한방 의료 이용 동향 파악 및 만족과 불만족의 요인을 파악하기 위해 LDA 토픽모델링을 통한 리뷰 텍스트 분석을 실시하였다. 분석 방법은 [Fig. 1]

과 같다.

## 2. 데이터 수집

데이터 분석 사이트인 AceCounter에 따르면 2021년 3분기 기준 국내 포털사이트 검색엔진 유입률<sup>13)</sup>은 네이버(https://www.naver.com) 63.99%, 구글(https://www.google.com) 28.17%, 다음(https://www.daum.net) 6.80% 순으로 네이버는 2위인 구글에 비해서도 2배 이상 많은 사용자를 확보하고 있다. 플랫폼 사용자 수와 데이터 수집 용이성을 고려하여 데이터 수집은 네이버를 대상으로 하였으며, 연구의 조사 대상은 국내 한방의료기관을 방문한 소비자가 작성한 리뷰이며, 조사 기간은 2018년부터 2021년 10월 26일까지로 한정하였다. 이는 2018년 이전의 리뷰 수가 200개로 다른 연도에 비해 적고, 네이버의 별점서비스(평점)가 2021년 10월 26일에 종료된 것을 고려하여 해당 기간 외의 리뷰는 제외하였다.

2022년 1월 11일 기준 건강보험심사평가원<sup>14)</sup>에 등록된 15,029개의 국내 한의원 및 한방병원 중 네이버 지도에서 검색되는 한방의료기관은 14,748개였고, 이 모든 한방의료기관의 방문자 리뷰를 수집하였다. 2022년 2월 2일 기준으로 총 375,258개의 리뷰를 추출하였고, 그중 2021년 10월 26일 이후에 등록된 별점이 없는 리뷰 101,090개와 2018년도 이전에 등록된 200개의 리뷰는 제외하였다. 또한 리뷰의 주제 파악이 연구의 주목적임을 고려하여 글자 수가 적은 리뷰는 유의미한 의미를 지니고 있지 않다고 판단해 글자 수가 20자 이하인 리뷰 153,241개도 제외하여 총 120,727개의 리뷰가 최종 연구자료로 사용되었다.

## 3. 자연어 전처리

LDA 분석에 앞서 추출한 리뷰는 비정형 데이터의 상태이기 때문에, 토픽모델링을 실시하기 전에 이를 구

조화하는 전처리 과정이 필요하다. 대부분의 선행 연구에서 한글 텍스트를 분석할 때는 Python의 한국어 전처리 패키지인 'KoNLPy'를 사용하였다. 그러나 'KoNLPy'는 개발자가 등록된 사전을 기반으로 텍스트를 분석하기 때문에<sup>15)</sup>, 사용하는 사전에 따라 결과가 달라지므로 일관성을 확보하기 어렵다. 따라서 본 연구에서는 Google cloud의 Natural Language API<sup>16)</sup>를 사용하여 추출한 120,727개의 리뷰를 품사별로 구분하고, 구분된 품사에서 명사만을 따로 추출하여 LDA 토픽모델링을 시행하였다.

## 4. 연도별 및 긍·부정 리뷰 분류

추출한 리뷰데이터에 대하여 연도별 환자들의 한방 의료 이용목적 및 치료 분야의 동향 파악을 위해 18년부터 21년까지 연도별로 리뷰를 구분하였다. 또한 한방 의료의 만족과 불만족요인을 파악하기 위해 각 리뷰 작성자가 등록한 별점(평점)에 따라 긍정·부정 리뷰로 분류하였다. 네이버 리뷰의 경우 평점이 0.5점부터 5점까지 0.5점 단위로 나뉘는데, 평점별 리뷰 개수를 분석한 결과 별점 5점이 104,065개로 전체 리뷰의 86.2%를 차지하였다. 이를 통해 네이버 리뷰가 우편향 되어있다고 판단하여 별점 0.5~3.5점은 부정 리뷰(4,052개), 4~4.5점은 중립(12,610개), 5점은 긍정 리뷰(104,065개)로 기준을 정하여 구분하였다. 연구 결과를 분석할 때는 부정 리뷰에서의 토픽은 불만족의 주요 원인으로, 긍정 리뷰에서의 토픽은 만족요인의 주요 원인으로 추정하였다.

그러나 리뷰 특성상 대부분의 리뷰에서 질환명에 대해 단순 기술하는 경우가 많으며, 긍정군에서 도출된 질환명 관련 토픽은 소비자가 특정 질환 치료에 만족한 것으로 해석할 수 있지만, 긍정 외에 중립, 부정 군의 경우 질환과 관련된 토픽이 도출되지 않아 상호비교가 불가능하였다. 또한, 부정 군의 '치료법' 토픽은 소비자 본인이 받은 치료법에 대한 단순 기술이므로 부정요인



Fig. 1. Research Process

으로 판단하기 어려워 고찰 대상에서 배제하였다.

### 5. LDA 토픽모델링(Latent Dirichlet Allocation Topic-Modeling)

‘토픽모델링(Topic Modeling)’은 텍스트를 대표하는 토픽들을 추출하는 통계적 모델이며, 텍스트에서 의미 있는 정보를 찾아내는 ‘텍스트 마이닝(Text mining)’의 한 분야이다. 잠재 디리클레 할당(LDA, Latent Dirichlet Allocation)은 토픽모델링의 한 기법으로, 문서의 집합체인 코퍼스(corpus)가 다수의 토픽으로 구성되어 있다고 가정하여 문서의 양, 문서 내 단어 개수, 코퍼스로 계산되는 파라미터 값 등을 Dirichlet 분포를 사용하여 잠재된 파라미터를 계산하고 이를 통해 각 주제에 맞는 단어를 할당하는 확률적 토픽 모델 기법이다<sup>7)</sup>. 즉, 의미적 관계가 있는 단어들은 하나의 주제로 묶이게 되며, 동일한 주제로 묶여 있는 단어들을 통해 각 문서에 어떤 토픽들이 존재하는지를 파악하는 기법이다<sup>7)</sup>. 최근 국내에서는 사회과학 등 다양한 분야에서 방대한 텍스트 자료를 주제별로 분류하고 자료에 잠재된 다양한 정보를 도출할 수 있는 LDA 토픽모델링을 활용한 연구가 활발하게 이루어지고 있다<sup>8)</sup>.

플레이트 표기법을 사용하여 LDA 모델<sup>7)</sup>을 도식화 하면 [Fig. 2]와 같다. 플레이트 표기법은 확률론적 그래픽 모델을 표현하는데 사용되는 방법으로 바깥쪽의 상자(plate)는 문서를 나타내며, 안쪽의 상자는 주어진 문

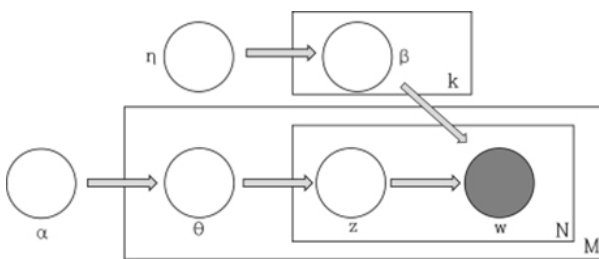


Fig. 2. Graphical model representation of Latent Dirichlet Allocation

M: The number of documents, N: The number of words in the document, k: The number of topics, w: Word, z: Topic of words,  $\theta$ : Topic distribution by document,  $\eta$ : Parameter of the Dirichlet prior on the per-topic word distribution,  $\alpha$ : Parameter of the Dirichlet prior on the per-document topic distributions,  $\beta$ : Word distribution for topic

서에서 반복되는 단어의 위치를 나타낸다. 또한 회색원(w)은 관찰 가능한 유일한 변수로 구체적 단어를 의미하며, 나머지 변수들의 경우 이를 통해 추론될 수 있는 값이다. M과 N은 각각 문서의 개수와 문서 내의 단어의 수를 의미하며, k는 토픽의 개수를 의미한다. w는 단어를 의미하며, z는 문서를 구성하는 특정 단어의 토픽이고,  $\theta$ 는 문서별 토픽의 분포를 나타낸다. 여기서  $\theta$ 는  $\alpha$ 에 의해 결정되는데,  $\alpha$ 는 토픽의 디리클레분포 파라미터가 된다.  $\beta$ 는 각 토픽의 단어 비중을 의미하며,  $\beta$ 는 단어의 디리클레분포 파라미터  $\eta$ 에 의해 결정된다. LDA기법의 원리는 관측값을 이용해 역으로 잠재된 토픽 구조를 추론하는 것이며, 토픽들의 확률 분포와 토픽에서 단어들의 확률 분포(디리클레 분포)들을 추정하여 토픽을 찾는 것이다. 수집한 리뷰데이터를 통해 리뷰 내에 잠재된 요인들을 비지도 학습(Unsupervised Learning)으로 찾아낼 수 있다. 관측변수(w)가 주어졌을 때 잠재 변수들의 조건부 확률을 계산하여 w를 제외한 나머지 잠재요소들을 구할 수 있다. 문서를 작성할 때는 이야기하고자 하는 주제를 중심으로 표현하게 되므로, 각 문서의 토픽 비율( $\theta$ )과 토픽에서 단어(z)를 알아낼 수 있다<sup>19)</sup>. 이러한 방식으로 진행된 LDA 분석을 통해 토픽 안에 단어가 등장할 확률이 Score로 표현되어, 해당 토픽 내 각 단어가 차지하는 비율을 파악할 수 있다. 본 연구에서는 LDA 분석에 활용되는 Python의 ‘gensim’ 모듈 4.1.2 버전을 사용하여 LDA를 실행하였다<sup>20)</sup>. 또한 결과값을 ‘pyLDAvis’ 3.3.1 및 ‘WordCloud’ 1.8.1 모듈을 사용하여 시각화하였다.

### 6. 최적 토픽 수 결정

LDA 토픽 모형을 활용하기 위해서는 사전에 토픽의 개수(k)를 입력하여야 한다. 토픽모델링에서 적당한 토픽 개수를 선정하는 데에는 여러 방법이 고안되어왔다. 선행연구에서는 연구자가 임의로 토픽 개수를 지정하기도 하였고<sup>21)</sup>, 일관성(coherence) 점수를 계산해 일관성이 가장 높은 지점의 토픽 개수를 지정하거나<sup>22)</sup>, 복잡도(perplexity)를 계산해 복잡도가 가장 낮은 지점의 토픽 개수로 선정하기도 하였다<sup>23)</sup>. 그러나 일관성만 보고 토픽의 수를 너무 많이 설정하면 불필요한 주제가 도출될 수 있고, 복잡도만 보고 너무 적게 설정하면 많은 키워드가 한 주제에 포함되어 주제설정이 어렵다<sup>24)</sup>. 따라서 일관성과 복잡도를 모두 고려하여 적정 토픽 개수를 선

정하기도 한다<sup>25)</sup>. 따라서 본 연구에서는 이러한 방법을 참고하여 Python의 ‘gensim’ 모듈을 활용해 일관성과 복잡도를 모두 계산하였다. 2개에서 14개까지 주제 개수로 토픽모델링을 실시한 결과, 높은 일관성과 낮은 복잡도를 보이는 최적 주제 개수인 8개(k=8)로 선택하였다.

### III. 연구 결과

#### 1. 표본의 특성

본 연구는 네이버에서 추출한 120,727개의 의료소비자 리뷰에 대하여 분석하였다. 한방 의료기관 별 리뷰 건수는 최대 826건이고, 최소는 1건, 중간값은 103이며, 평균적으로 약 145.53건의 리뷰가 작성되었다. 리뷰 별 글자 수는 최대 422자, 최소는 21자, 평균 76.91자, 중간값은 54자이다. 리뷰를 등록한 환자별 방문 횟수의 경우 최대 388회이고, 최소는 1회, 중간값은 1회, 평균 2.35회로 집계되었다. 결과는 [Table 1]과 같다.

Table 1. Characteristics of the sample

	Total Number	Maximum value	Minimum value	Median value	Mean value
The number of review by institution	14,748	826	1	103	145.53
The number of letters by review	120,727	422	21	54	76.91
The number of visits by user	59,110	388	1	1	2.35

리뷰 작성자의 방문 인증은 병원의 영수증 인증을 하거나 포털에서 제공하는 예약 서비스로 병원 방문 예약을 하는 방법이 있는데, 88,506건(73.31%)은 영수증으로 방문 인증을 하였고, 32,221건(26.69%)은 포털의 예약 서비스를 통해 방문 인증하였다. 리뷰 작성자의 방문 요일은 월요일 21,596건, 화요일 20,148건, 수요일 18,529건, 목요일 17,386건, 금요일 20,184건, 토요일 18,557건, 일요일 4,327건으로 월요일에 가장 많은 방문 횟수를 보였다.

연도별 리뷰 개수는 2018년 1,085건(0.9%), 2019년 5,821건(4.8%), 2020년 47,420건(39.3%), 2021년

66,401건(55%)이고 아래의 [Fig. 3]과 같다.

별점(평점)별 리뷰개수는 0.5점은 996건(0.82%), 1.0점은 487건(0.40%), 1.5점은 182건(0.15%), 2.0점은 367건(0.3%), 2.5점은 263건(0.22%), 3.0점은 978건(0.81%), 3.5점은 779건(0.65%), 4.0점은 6,240건(5.17%), 4.5점은 6,370건(5.28%), 5.0점은 104,065건(86.2%)이고 아래의 [Fig. 4]와 같다.

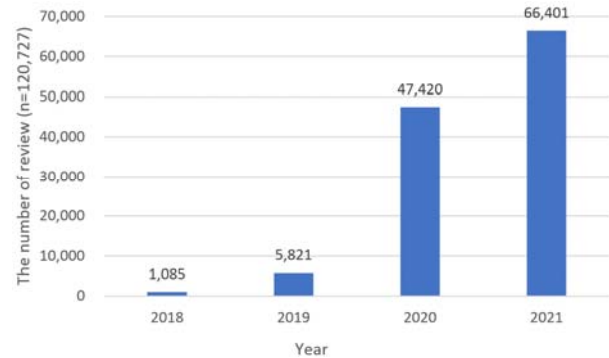


Fig. 3. The number of reviews by year

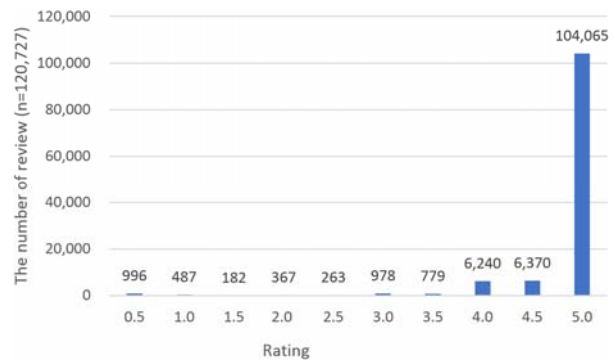


Fig. 4. The number of reviews by rating

#### 2. LDA 토픽모델링 최적 토픽 수 도출

2개부터 14개까지로 토픽 개수를 한정하여 토픽모델링을 실시한 결과 복잡도와 일관성은 [Fig. 5]와 같은 결과로 나왔다. Perplexity에서는 토픽이 증가할수록 복잡도는 점차 낮아지며, 토픽 개수가 8개 이상부터 급격한 감소세를 보였다. 또한 Coherence에서는 토픽 개수(k)가 8개일 때 점수 0.56으로 가장 높은 일치도를 보였다. 본 연구는 한 주제에 포함되는 단어들의 의미론적 일치성을 높이기 위해 낮은 복잡도를 갖는 토픽 개수 중에서 일치도가 가장 높은 토픽 개수를 최적 토픽 개수로 보았고, 따라서 토픽의 수를 8개로 선정하였다. 이러한 방법으로 도출된 최적 토픽 개수로 추출한 온라인

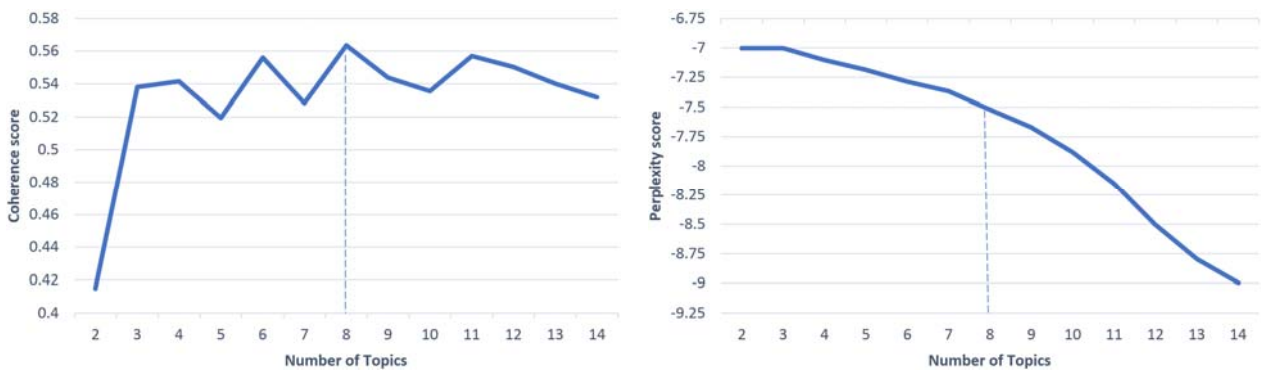


Fig. 5. Coherence score by the number of topics(Left) & Perplexity score by the number of topics(Right)

인 리뷰에 대해 LDA 분석을 실시하였다.

### 3. LDA 토픽모델링 결과

#### 1) 전체 리뷰 대상 LDA 토픽모델링 결과

본 연구에서는 리뷰 텍스트의 분석에 앞서 리뷰의 전반적인 토픽 분포와 리뷰에서 사용된 주요 단어를 파악하고자 하였다. 따라서 전체 리뷰에 대하여 적절한 주제

개수로 선정된 8개의 토픽에 맞추어 LDA 분석을 실시하였고, 결과는 [Fig. 6]과 같다. [Fig. 6] 왼쪽의 결과는 IDM(Intertopic Distance Map)으로, 8개 토픽의 구조적 관계 및 연관성을 시각화한 것이다. IDM은 원으로 표현된 토픽 간의 거리를 통해 토픽 간 유사도를 파악할 수 있으며, 원의 크기를 통해 각 토픽의 비중을 파악할 수 있다. 따라서 각 키워드 비중을 통해 각각의 토픽을 지정하였다. 1번 키워드는 [치료, 침, 허리, 추나]로 나왔고, 토픽은 ‘질환-근골격계’로 정하였다. 2번 키워

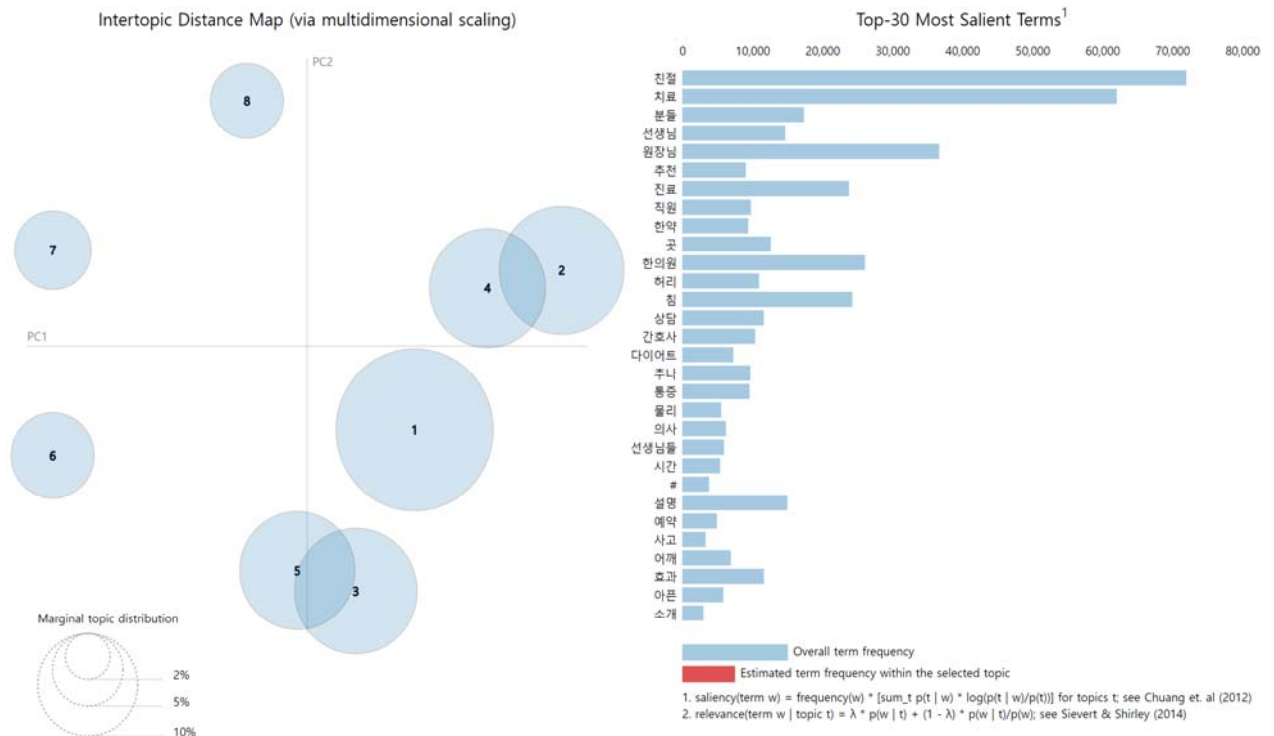


Fig. 6. Intertopic Distance Map and Top-30 most salient terms for all reviews

드는 [친절, 원장님, 진료, 설명]으로 나왔고, 토픽은 ‘치료법에 대한 설명’으로 정하였다. 3번 키워드는 [한약, 다이어트, 관리, 건강]으로 나왔고, 토픽은 ‘질환-다이어트’로 정하였다. 4번 키워드는 [친절, 선생님, 실력, 분위기]로 나왔고, 토픽은 ‘진료 태도’로 정하였다. 5번 키워드는 [진료, 시간, 예약, 대기]로 나왔고, 토픽은 ‘진료 대기시간’으로 정하였다. 6번 키워드는 [가격, 보험, 비용, 진료비]로 나왔고, 토픽은 ‘치료비’로 정하였다. 7번 키워드는 [추천, 소개, 지인, 주변]으로 나왔고, 토픽은 ‘의료기관 접근성’으로 정하였다. 8번 키워드는 [물리, 교통, 사고, 침]으로 나왔고, 토픽은 ‘질환-교통사고’로 구분하였다. [Fig. 6]의 오른쪽의 결과는 전체 리뷰의 상위 30개 키워드에 대한 단순 출현빈도 (frequency)를 나타낸 것이다. 단순 출현 빈도는 전체 문서 내에서 빈번하게 등장하는 키워드의 등장 횟수를 의미한다<sup>26)</sup>. 분석 결과, 전체 리뷰 대상으로 리뷰 내 가장 많이 사용된 키워드는 ‘친절’이고, 그 다음으로 ‘치료’, ‘원장님’, ‘한의원’ 등이 있다. 이를 바탕으로 ‘워드클라우드’를 통해 시각화한 것이 [Fig. 7]이다. 워드클라우드란 단어를 출현빈도에 따라서 글자의 크기와 굵기로 표현한 것이다.



Fig. 7. Wordcloud for all reviews

## 2) 연도별 LDA 토픽모델링 결과

한방 의료에 대한 전반적인 흐름과 트렌드를 파악하기 위해 소비자들의 방문 날짜(2018.1.1~2021.10.26)를 연도별로 분류하였고, 적절한 주제 개수로 선정된 8개의 토픽에 맞추어 LDA 토픽모델링을 실시하였다. 각 토픽 내 포함된 키워드를 등장 확률이 높은 순으로 정렬한 후 토픽을 대표할 수 있는 주요 키워드 4개씩 도출하여 정리하였고, 이때 ‘선생님-원장님’과 같이 비슷한 키워드는 하나만 택하였다. 8개의 토픽은 『2020년 한방의료이용실태조사』<sup>5)</sup>의 ‘이용목적’에 따른 ‘질환치

료’(비염, 어깨통증, 허리통증, 부인과, 근골격계), ‘건강 증진 및 미용’(다이어트, 피부-여드름, 두드러기), ‘교통사고’와 부문별 만족도의 만족 기준인 ‘진료 대기 시간’, ‘진료 태도’, ‘치료 효과’를 기준으로 명명하였다. WEIGHT는 특정 연도 내 특정 토픽의 비중을 의미하며, SCORE은 특정 토픽 내 특정 키워드의 비중을 의미한다. 결과는 아래 [Table 2]와 같다.

Table의 토픽 순서 배치는 서비스 부분별 토픽과 질환별 토픽으로 구분하였으며, 모든 연도에 공통적으로 나오는 토픽을 각 table의 같은 위치에 배치하여 연도별 토픽 비교에 용이하도록 하였다.

2018년 작성된 리뷰의 LDA 토픽모델링 결과, 1번 토픽의 주요 키워드는 [친절, 예약, 시간, 대기]로 나왔고, 토픽은 ‘진료 대기시간’으로 정하였다. 2번 토픽의 주요 키워드는 [친절, 만족, 실력, 편안]으로 나왔고, 토픽은 ‘진료 태도’로 정하였다. 3번 토픽의 주요 키워드는 [치료, 진료, 감사, 효과]로 나왔고, 토픽은 ‘치료 효과’로 정하였다. 4번 토픽의 주요 키워드는 [치료, 침, 사고, 추나]로 나왔고, 토픽은 ‘교통사고’로 정하였다. 5번 토픽의 주요 키워드는 [다이어트, 상담, 한약, 살]로 나왔고, 토픽은 ‘다이어트’로 정하였다. 6번 토픽의 주요 키워드는 [치료, 한약, 비염, 효과]로 나왔고, 토픽은 ‘비염’으로 정하였다. 7번 토픽의 주요 키워드는 [어깨, 치료, 침, 효과]로 나왔고, 토픽은 ‘어깨 통증’으로 정하였다. 8번 토픽의 주요 키워드는 [진료, 허리, 통증, 침]으로 나왔고, 토픽은 ‘허리 통증’으로 정하였다.

2019년 작성된 리뷰의 LDA 토픽모델링 결과, 1번 토픽의 주요 키워드는 [예약, 시간, 대기, 전화]로 나왔고, 토픽은 ‘진료 대기시간’으로 정하였다. 2번 토픽의 주요 키워드는 [친절, 진료, 설명, 감사]로 나왔고, 토픽은 ‘진료 태도’로 정하였다. 3번 토픽의 주요 키워드는 [다이어트, 한약, 관리, 감량]으로 나왔고, 토픽은 ‘다이어트’로 정하였다. 4번 토픽의 주요 키워드는 [보약, 산후, 출산, 케어]로 나왔고, 토픽은 ‘부인과’로 정하였다. 5번 토픽의 주요 키워드는 [침, 허리, 추나, 물리]로 나왔고, 토픽은 ‘허리통증’으로 정하였다. 6번 토픽의 주요 키워드는 [치료, 아이, 비염, 감사]로 나왔고, 토픽은 ‘비염’으로 정하였다. 7번 토픽의 주요 키워드는 [한약, 치료, 여드름, 추천]으로 나왔고, 토픽은 ‘피부-여드름’으로 정하였다. 8번 토픽의 주요 키워드는 [후유증, 두드러기, 얼굴, 부종]으로 나왔고, 토픽은 ‘두드러기’로 정하였다.

2020년 작성된 리뷰의 LDA 토픽모델링 결과, 1번 토픽의 주요 키워드는 [진료, 시간, 예약, 대기]로 나왔고, 토픽은 '진료 대기시간'으로 정하였다. 2번 토픽의 주요 키워드는 [친절, 원장님, 설명, 편안]으로 나왔고, 토픽은 '진료 태도'로 정하였다. 3번 토픽의 주요 키워드는 [치료, 감사, 추천, 효과]로 나왔고, 토픽은 '치료 효과'로 정하였다. 4번 토픽의 주요 키워드는 [코로나, 선물, 가격, 공진단]으로 나와서 예외적으로 '코로나'라는 토픽으로 정하였다. 5번 토픽의 주요 키워드는 [병원, 사고, 입원, 보험]으로 나왔고, 토픽은 '교통사고'로 정하였다. 6번 토픽의 주요 키워드는 [다이어트, 한약, 운동, 성공]으로 나왔고, 토픽은 '다이어트'로 정하였다. 7번 토픽의 주요 키워드는 [치료, 침, 추나, 통증]으로 나왔고, 토픽은 '근골격계'로 정하였다. 8번 토픽의 주요 키워드는 [관리, 피부, 시술, 얼굴]로 나왔고, 토픽은 '피부'로 정하였다.

2021년 작성된 리뷰의 LDA 토픽모델링 결과, 1번 토픽의 주요 키워드는 [시간, 예약, 대기, 주말]로 나왔고, 토픽은 '진료 대기시간'으로 정하였다. 2번 토픽의 주요 키워드는 [친절, 원장님, 설명, 감사]로 나왔고, 토픽은 '진료 태도'로 정하였다. 3번 토픽의 주요 키워드는 [친절, 침, 실력, 효과]로 나왔고, 토픽은 '치료 효과'로 정하였다. 4번 토픽의 주요 키워드는 [병원, 사고, 물리,

입원]으로 나왔고, 토픽은 '교통사고'로 정하였다. 5번 토픽의 주요 키워드는 [치료, 침, 통증, 추나]로 나왔고, 토픽은 '근골격계'로 정하였다. 6번 토픽의 주요 키워드는 [한약, 다이어트, 효과, 소개]로 나왔고, 토픽은 '다이어트'로 정하였다. 7번 토픽의 주요 키워드는 [치료, 아이, 비염, 감기]로 나왔고, 토픽은 '비염'으로 정하였다. 8번 토픽의 주요 키워드는 [몸, 관리, 피부, 회복]으로 나왔고, 토픽은 '피부'로 정하였다. 위의 결과는 [Table 2.]와 같다.

또한 2018년부터 2021년까지 연구 대상 기간 동안 각 연도별로 3회 이상 나타난 토픽에 대하여 연도별 순위 및 변화 추이를 살펴보면 [Fig 8.]과 같다. 공통적으로 '진료 대기시간'과 '진료 태도'가 높은 순위를 차지하고 있었으며 질환으로는 '다이어트'가 높은 수준을 일정하게 유지하였다.

### 3) 평점별 LDA 토픽모델링 결과

한방 의료에 대한 만족과 불만족요인을 파악하기 위해 평점을 기준으로 별점 0.5~3.5점은 부정, 4~4.5점은 중립, 5점은 긍정 리뷰로 구분하여 LDA 토픽모델링을 실시하였고, 결과는 아래 [Table 3]과 같이 나왔다. [Table 3]의 주제 배열은 [Table 2]에서 사용한 방법을 동일하게 적용하였다. 긍정 요인과 부정 요인에 대한 차이점을 비교하기 위해 각 결과에 대한 8개의 토픽은 『2020년 한방의료이용실태조사』<sup>5)</sup>의 '부문별 만족도'의 만족 기준인 '진료 대기시간', '진료 태도', '치료 효과', '치료비', '치료법에 대한 설명', '의료기관 접근성', '의료기관 시설환경', '치료법'을 기준으로 명명하였다.

부정 리뷰에 대한 LDA 토픽모델링 결과, 1번 토픽의 주요 키워드는 [시간, 대기, 오픈, 사람들]로 나왔고, 토픽은 '진료 대기시간'으로 정하였다. 2번 토픽의 주요 키워드는 [친절, 진료, 의사, 간호사]로 나왔고, 토픽은 '진료 태도'로 정하였다. 3번 토픽의 주요 키워드는 [치료, 효과, 느낌, 침]으로 나왔고, 토픽은 '치료 효과'로 정하였다. 4번 토픽의 주요 키워드는 [치료, 실비, 진료비, 과잉]으로 나왔고, 토픽은 '치료비'로 정하였다. 5번 토픽의 주요 키워드는 [친절, 원장님, 치료, 상담]으로 나왔고, 토픽은 '치료법에 대한 설명'으로 정하였다. 6번 토픽의 주요 키워드는 [예약, 시간, 전화, 진료]로 나왔고, 토픽은 '의료기관 접근성'으로 정하였다. 7번 토픽의 주요 키워드는 [불편, 주차장, 무료, 만족]으로 나왔

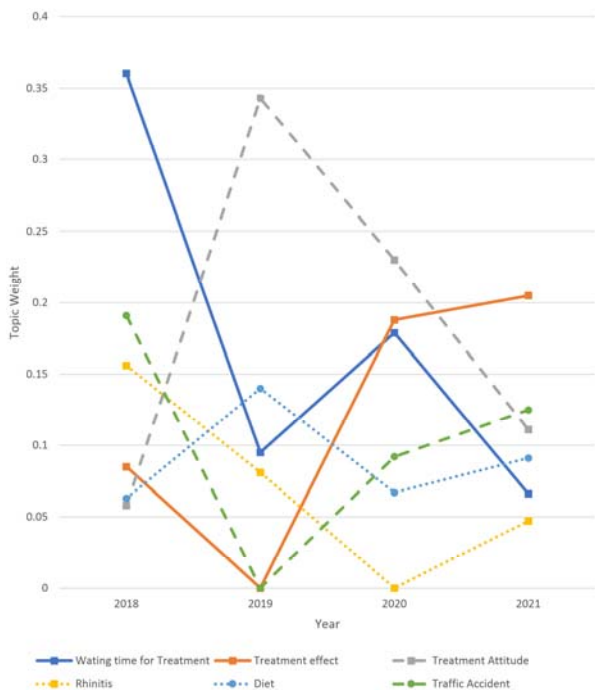


Fig 8. Trends in major topics by year



Table 2. Results of review analysis by year

TOPIC	WEIGHT	KEY WORD	SCORE	TOPIC	WEIGHT	KEY WORD	SCORE	TOPIC	WEIGHT	KEY WORD	SCORE	TOPIC	WEIGHT	KEY WORD	SCORE
Waiting time for Treatment	0.36	친절	0.074	Treatment Attitude	0.058	친절	0.027	Waiting time for Treatment	0.095	예약	0.083	Treatment Attitude	0.343	친절	0.107
		예약	0.031			만족	0.011			시간	0.05			진료	0.029
		시간	0.014			실력	0.008			대기	0.015			설명	0.026
		대기	0.005			편안	0.007			전화	0.008			감사	0.019
		치료	0.02			치료	0.057			다이얼트	0.047			보약	0.011
Treatment effect	0.085	진료	0.012	Traffic Accident	0.191	침	0.013	Diet	0.14	한약	0.022	The gynecology	0.048	산후	0.009
		감사	0.007			사고	0.012			관리	0.018			출산	0.009
		효과	0.005			추나	0.007			감광	0.01			케어	0.006
		다이얼트	0.054			치료	0.035			침	0.04			치료	0.029
		상담	0.033			한약	0.01			허리	0.026			아이	0.02
Diet	0.063	한약	0.013	Rhinitis	0.156	비염	0.007	Waist(back) pain	0.204	추나	0.023	Rhinitis	0.081	비염	0.008
		살	0.01			효과	0.007			물리	0.012			감사	0.006
		어깨	0.01			진료	0.02			한약	0.016			후유증	0.01
		치료	0.008			허리	0.009			치료	0.014			두드러기	0.008
		침	0.007			통증	0.007			여드름	0.007			얼굴	0.006
Shoulder pain	0.023	효과	0.004	Waist(back) pain	0.062	침	0.006	Skin-Acne	0.055	추진	0.006	Skin-Urticaria	0.033	부종	0.004
		진료	0.044			진료	0.166			시간	0.055			친절	0.166
		시간	0.024			원장님	0.063			예약	0.038			원장님	0.084
		예약	0.021			설명	0.037			대기	0.015			설명	0.065
		대기	0.007			편안	0.011			주말	0.012			감사	0.056
Treatment effect	0.188	치료	0.037	COVID-19	0.028	코로나	0.018	Treatment effect	0.205	친절	0.138	Traffic Accident	0.125	병원	0.039
		감사	0.017			신물	0.016			침	0.021			사고	0.015
		추천	0.013			가격	0.014			실력	0.008			물리	0.01
		효과	0.012			공진단	0.014			효과	0.007			입원	0.008
		병원	0.038			다이얼트	0.068			치료	0.108			한약	0.063
Traffic Accident	0.092	사고	0.029	Diet	0.067	한약	0.045	Musculoskeletal system	0.214	침	0.058	Diet	0.091	다이얼트	0.057
		입원	0.01			운동	0.016			통증	0.033			효과	0.041
		보험	0.006			성공	0.01			추나	0.033			소개	0.012
		치료	0.123			관리	0.043			치료	0.042			몸	0.028
		침	0.064			피부	0.041			아이	0.04			관리	0.011
Musculoskeletal system	0.177	추나	0.039	Skin	0.039	시술	0.024	Rhinitis	0.047	비염	0.023	Skin	0.141	피부	0.009
		통증	0.034			얼굴	0.018			감기	0.007			회복	0.007

Table 3. Results of review analysis by rating

N E G A T I V E				P O S I T I V E			
TOPIC	WEIGHT	KEY WORD	SCORE	TOPIC	WEIGHT	KEY WORD	SCORE
Waiting time for Treatment	0.026	시간	0.047	Explanation of treatment	0.162	진절	0.166
		대기	0.021			진료	0.061
		오른	0.015			원장님	0.023
		사람들	0.014			상담	0.016
Treatment effect	0.145	치료	0.055	Treatment effect	0.142	치료	0.07
		효과	0.025			침	0.063
		느낌	0.012			효과	0.023
		침	0.012			추천	0.022
Explanation of treatment	0.152	진절	0.084	The gynecology	0.041	진맥	0.023
		원장님	0.034			한약	0.022
		치료	0.022			산후	0.014
		상담	0.025			생리통	0.008
Facilities & Environment	0.069	불편	0.026	Musculoskeletal system	0.208	치료	0.124
		주차장	0.016			추나	0.04
		무로	0.01			침	0.036
		만족	0.008			통증	0.033
N E U T R A L							
TOPIC	WEIGHT	KEY WORD	SCORE	TOPIC	WEIGHT	KEY WORD	SCORE
Waiting time for Treatment	0.068	예약	0.084	Treatment Attitude	0.068	감사	0.011
		시간	0.081			고생	0.008
		대기	0.041			직원분	0.006
		진화	0.008			감동	0.006
Treatment effect	0.127	효과	0.023	Treatment expenses	0.045	가격	0.024
		진료	0.017			임원	0.015
		차방	0.011			이용	0.01
		체질	0.009			보약	0.007
Explanation of treatment	0.126	진절	0.049	Accessibility	0.058	한의원	0.023
		상담	0.031			동네	0.013
		설명	0.029			사람들	0.007
		원장님	0.026			주차장	0.006
Facilities & Environment	0.218	시설	0.011	Medical Treatment	0.29	치료	0.113
		편안	0.008			침	0.044
		만족	0.007			추나	0.018
		안내	0.005			물리	0.017

고, 토픽은 ‘의료기관 시설환경’으로 정하였다. 8번 토픽의 주요 키워드는 [치료, 침, 추나, 물리]로 나왔고, 토픽은 ‘치료법’으로 정하였다.

중립 리뷰에 대한 LDA 토픽모델링 결과, 1번 토픽의 주요 키워드는 [예약, 시간, 대기, 전화]로 나왔고, 토픽은 ‘진료 대기시간’으로 정하였다. 2번 토픽의 주요 키워드는 [감사, 고생, 직원분, 감동]으로 나왔고, 토픽은 ‘진료 태도’로 정하였다. 3번 토픽의 주요 키워드는 [효과, 진료, 처방, 체질]로 나왔고, 토픽은 ‘치료 효과’로 정하였다. 4번 토픽의 주요 키워드는 [가격, 입원, 이용, 보약]으로 나왔고, 토픽은 ‘치료비’로 정하였다. 5번 토픽의 주요 키워드는 [친절, 상담, 설명, 원장님]으로 나왔고, 토픽은 ‘치료법에 대한 설명’으로 정하였다. 6번 토픽의 주요 키워드는 [한의원, 동네, 사람들, 주차장]으로 나왔고, 토픽은 ‘의료기관 접근성’으로 정하였다. 7번 토픽의 주요 키워드는 [시설, 편안, 만족, 안내]로 나왔고, 토픽은 ‘의료기관 시설환경’으로 정하였다. 8번 토픽의 주요 키워드는 [치료, 침, 추나, 물리]로 나왔고, 토픽은 ‘치료법’으로 정하였다.

긍정 리뷰에 대한 LDA 토픽모델링 결과, 1번 토픽의 주요 키워드는 [친절, 진료, 설명, 상담]으로 나왔고, 토픽은 ‘치료법에 대한 설명’으로 정하였다. 2번 토픽의 주요 키워드는 [진료, 친절, 원장님, 감사]로 나왔고, 토픽은 ‘진료 태도’로 정하였다. 3번 토픽의 주요 키워드는 [치료, 침, 효과, 추천]으로 나왔고, 토픽은 ‘치료 효과’로 정하였다. 4번 토픽의 주요 키워드는 [한의원, 예약, 동네, 근처]로 나왔고, 토픽은 ‘의료기관 접근성’으로 정하였다. 5번 토픽의 주요 키워드는 [진맥, 한약, 산후, 생리통]으로 나왔고, 토픽은 ‘부인과’로 정하였다. 6번 토픽의 주요 키워드는 [한약, 다이어트, 건강, 몸]으로 나왔고, 토픽은 ‘다이어트’로 정하였다. 7번 토픽의 주요 키워드는 [치료, 추나, 침, 통증]으로 나왔고, 토픽은 ‘근골격계’로 정하였다. 8번 토픽의 주요 키워드는 [관리, 피부, 여드름, 두피]로 나왔고, 토픽은 ‘피부과’로 정하였다. 긍정 리뷰에서는 질환과 관련된 키워드의 빈도가 높아 4개의 토픽에 대해서는 이용목적에 따른 질환으로 명명하였다. 위의 결과는 [Table 3]과 같다.

#### IV. 고 찰

오늘날 온라인 플랫폼이 발달함에 따라 소비자들의

온라인 리뷰 활용도가 높아지고 있다. 서비스를 경험한 사람들은 자신의 경험을 바탕으로 리뷰를 작성해 정보를 제공하고, 잠재 소비자들은 이를 통해 정보를 획득한다. 뿐만 아니라 사업자 등 서비스 제공자의 경우 온라인 리뷰를 통해 소비자의 니즈와 의사결정 과정을 파악할 수 있다. 이러한 온라인 리뷰 데이터는 실제 사용자의 생각을 엿볼 수 있는 강력한 도구이며 호텔 서비스 품질 요인 분석, 사용자 선호도 예측, 영화 및 음악 큐레이션 및 추천 시스템, 스팸 탐지 등 다양한 산업 분야에 걸쳐 연구와 활용이 활발히 이루어지고 있다. 따라서 본 연구는 온라인 리뷰의 특성을 활용하여 네이버에 등록된 한방 의료소비자의 리뷰를 토픽모델링으로 분석해 한방 이용목적의 전반적인 흐름과 만족요인을 확인하였다.

데이터 분석을 통해 얻은 토픽별 weight를 통해 연도에 따른 토픽 점유율을 파악하여, 연도별로 weight가 가장 큰 토픽이 해당 연도를 대표한다고 해석할 수 있다. 연도별 토픽을 분석한 결과, ‘진료 대기시간’과 ‘진료 태도’는 모든 연도에 공통으로 나타났으며, ‘치료 효과’ 또한 2019년을 제외한 모든 연도에 공통으로 도출되었다. ‘진료 대기시간’, ‘진료 태도’, ‘치료 효과’는 병원을 평가하는 리뷰의 특성상 연도와 관계없이 자주 등장하므로 해당 토픽을 제외하였다. 따라서 질환 중심으로 연도별 주요 토픽을 분석하면, 2018년은 ‘교통사고’가 19.1%로 가장 큰 비중을 차지하였고, 2019년은 ‘허리통증(20.4%)’, 2020년은 ‘질환-근골격계(17.7%)’, 2021년도 ‘질환-근골격계(24.1%)’가 가장 큰 비중으로 각 연도를 대표하였다. 2019년의 ‘허리통증’ 토픽 또한 근골격계 질환으로 포함될 수 있으며, 이 결과를 통해 ‘근골격계’ 질환은 2018년을 제외한 모든 연도에서 가장 큰 비중을 차지하는 질환임을 파악할 수 있었다. 조사 결과를 토대로 실제 한방 의료 이용목적과 연계 고찰하기 위해 『2020년 한방의료이용실태조사』<sup>5)</sup> 결과와 비교하였다. 비교 결과, 실태조사에서도 근골격계 질환 치료 목적이 가장 큰 비중을 차지하였으며, 이를 토대로 한방 의료소비자의 주 이용목적은 ‘근골격계 질환 치료’임을 파악하고, 온라인 리뷰의 실효성을 입증할 수 있었다. 또한 2018년에 교통사고가 주요 토픽을 차지한 이유에 대해 조사한 결과, 자동차보험 한방 진료비가 2015년 3,576억원에서 2018년 7,139억원으로 약 2배 증가함을 파악하였으며<sup>27)</sup>, 이를 통해 2018년 전후 교통사고로 한방의료기관을 방문하는 환자 수가 증가한 것과 연결

지을 수 있었다.

이 외에도 2020년은 코로나 유행으로 인해 이와 관련하여 이슈가 많아 리뷰 분석에서도 특징적인 결과를 확인할 수 있다. 먼저 LDA모델링 결과에서 ‘피부미용’ 토픽에 대해 자세히 분석해보면, 피부미용 토픽의 weight가 전년 대비 꾸준히 상승한 것을 확인할 수 있으나, 2020년에 피부 관련 질환은 코로나로 전년 대비 4.9%p 감소(8.8% → 3.9%)하였다. 코로나의 영향이 줄어든 2021년에 다시 10.2%p 증가(3.9% → 14.1%)하였다. 더불어 ‘다이어트’ 토픽을 분석해보면, 다이어트는 2018부터 2021까지 모든 연도에 나타난 토픽이었으며, 이 또한 전체적으로 증가하는 추세로 나타났으나, 마찬가지로 코로나의 영향으로 2020은 전년 대비 7.3%p 감소(14.1% → 6.7%)하였다. 그러나 2021년에 다시 토픽의 비중도가 2.4%p 상승(6.7% → 9.1%)하며 관심이 높아짐을 확인할 수 있다. 통계청에서 조사한 「화장품산업현황」결과에 따르면 2006년부터 꾸준히 상승 폭을 보였던 화장품생산 금액이 전년대비 2020년에 -6.8% 하락하였다<sup>28)</sup>. 또한 통계청에서 조사한 「국민생활체육조사」에 따르면 체육활동 참여 이유 중 체중조절 및 체형관리가 전년 대비 2020년에 2%p 감소(15% → 13%)하였다<sup>29)</sup>. 이 결과로 2020년에는 화장품 및 다이어트 관련 산업이 전년 대비 감소하였음을 확인할 수 있었으며, 이는 온라인 리뷰에서 피부미용과 다이어트 관련 토픽의 비중이 줄어든 것과 비슷한 결과이다. 이를 통해 2020년 코로나로 인해 거리두기 및 비대면 문화가 형성되며 외출이 감소하게 되어 화장, 다이어트 등 전반적인 미용 분야의 필요성과 관심도가 낮아졌을 것이라 예측할 수 있다<sup>30)</sup>. 또한 2020년의 토픽 중 ‘코로나’의 키워드인 [코로나, 선물, 가격, 공진단]에 대해 분석하였고, 이는 2020년에 발생한 코로나19로 인해 소비자들이 한약에 대한 관심도가 늘어난 것으로 예상해 관련 추가 조사를 하였다. 그 결과 코로나로 면역력에 대한 관심이 급증하면서 면역 관련 대표 한약재인 홍삼의 판매량이 코로나19 이전인 2018년도에 비해 2019년에 증가한 것을 파악해 연관성을 확인할 수 있었다<sup>31)</sup>.

별점별 주요 토픽을 분석하면, 긍정 군은 ‘치료법에 대한 설명’과 ‘진료 태도’가 16.2%, 부정 군은 ‘의료기관 접근성’이 18.8%, 중립 군은 ‘시설환경’이 21.8%로 가장 큰 비중을 차지하였다. 조사 결과를 토대로 실제 한방 의료서비스 부문별 만족도와 연계 고찰하기 위해 「2020

년 한방의료이용실태조사」<sup>5)</sup> 결과와 비교하였고, 만족요인에 대한 실태조사에서도 ‘의료기관 종사자의 진료 태도’가 92.4%로 가장 높아 리뷰 분석 결과와 차이가 없었다. 불만족요인에 대한 실태조사는 ‘진료비’가 6.3%로 가장 높았고, 그 다음으로 ‘한방의료기관의 접근성(1.4%)’이었다. LDA 토픽모델링 결과로는 ‘의료기관 접근성(18.8%)’이 가장 큰 불만족요인이었으며, ‘치료비’의 비중은 4.5%로 크지 않지만, 다른 토픽과는 달리 ‘치료비’에 속한 키워드 중 [과잉]이라는 직접적인 부정 키워드가 등장한 것으로 보아 불만족요인에 큰 영향을 미친 것으로 해석할 수 있다.

추가로 주목할 점은 ‘진료 대기시간’, ‘치료비’, ‘의료기관 시설환경’이라는 토픽은 긍정 군에서는 나타나지 않고 부정 군에서만 등장한 것으로, 해당 토픽은 만족보다 불만을 느끼는 요인으로 작용한다고 판단하였다. 따라서 불만족요인에 대한 이유를 파악하고자 토픽과 해당 키워드의 연관성을 확인하였다. ‘대기시간’ 토픽의 [시간, 대기, 오픈, 사람들] 키워드를 통해 소비자들이 의료기관 내에 대기 인원이 많거나 의료기관의 오픈 시간, 진료 대기에 불만을 느끼는 경우가 많다고 해석할 수 있었다. 또한 ‘의료기관 시설환경’ 토픽의 [불편, 주차장, 무료, 만족] 키워드를 통해, 소비자들이 시설에 불편을 느끼고, 특히 주차장 시설에 불만을 느끼는 경우가 많다고 해석할 수 있었다.

본 연구는 ‘LDA 토픽 모델링 기법을 활용한 텍스트 분석’을 시행하였으며, 이를 활용해 한방 의료 현황과 소비자의 요구를 분석하는 새로운 방법을 제시하였다는 데에 의의가 있다. 그러나, 모든 리뷰를 분석함으로써, 구체적인 정보획득에 어려운 부분이 있었다. 또한, 본 연구에서 이용한 포털사이트의 특성상 리뷰 텍스트 외에 리뷰 작성자의 연령대, 성별 및 지역에 대한 정보를 추출하는 것에 한계가 있었으며, 홍보 등으로 인한 데이터의 편향이 있을 수 있어 결과의 해석에 신중을 기할 필요가 있다. 따라서 향후 연구에서는 작성자의 인구학적 특성 자료를 추가로 수집하고 편향을 제거하기 위한 방법론을 개발 및 적용하여 보완 분석할 필요가 있다. 도출된 연구 결과는 한방의료서비스의 질적 향상에 기여할 것으로 생각되며, 더불어 LDA 토픽모델링 기법을 활용하여 발전된 한방의료 서비스를 제공하고 관련된 연구가 향후 수행될 것을 기대하는 바이다.

## V. 결론

본 연구는 한방 의료의 이용목적 동향 조사와 만족 및 불만족요인을 파악하기 위해 2018년부터 2021년 10월 26일까지 한방 의료소비자가 작성한 온라인 리뷰를 수집하여 LDA 토픽모델링으로 텍스트 분석을 하였으며 다음의 결과를 도출하였다.

1. LDA 토픽모델링을 실시한 결과, 전체 리뷰에서 가장 많이 출현한 상위 5개 키워드는 ‘친절’, ‘치료’, ‘원장님’, ‘한의원’, ‘침’이고, 토픽은 8개로 분류되며 각각 ‘질환-근골격계’, ‘치료법에 대한 설명’, ‘질환-다이어트’, ‘진료 태도’, ‘진료 대기시간’, ‘치료비’, ‘의료기관 접근성’, ‘질환-교통사고’로 나타났다.
2. 연도별 한방 의료 이용목적에 대해 가장 큰 비중을 차지하는 토픽은 2018년은 ‘교통사고’, 2019년은 ‘허리통증’, 2020년은 ‘질환-근골격계’, 2021년은 ‘질환-근골격계’ 치료로 나타났다.
3. 평점을 기준으로 분석한 한방 의료 만족 및 불만족요인에 대해 만족은 ‘치료법에 대한 설명’과 ‘진료 태도’, 불만족은 ‘의료기관 접근성’이 가장 큰 요인이었다.

### 감사의 글

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2020R1G1A1100725).

## 참고문헌

1. KOSTAT. [Report] Online shopping trends in December 2021 and online direct sales and purchase trends in the fourth quarter. [cited in] 2022.02.20. Available from: [https://kostat.go.kr/portal/korea/kor\\_nw/1/1/index.board?bmode=read&aSeq=416587](https://kostat.go.kr/portal/korea/kor_nw/1/1/index.board?bmode=read&aSeq=416587)
2. Chung HN, Nam YJ. An eye-tracking study on factors relating the credibility of the review text. *txtlng*. 2017;43():193-219.
3. Jung H. The Impact of Others' Choices on the Affective and Cognitive Processing When Choosing a Credit Card. *The Journal of Digital Contents Society*. 2019;20(6):1181-1189.
4. Kim HY, Park JM. Advertising, Public Relations, and Marketing Practitioners, Perception on Media Richness and Marketing Communication Characteristics of Media and Marketing Public Relations Tactics. *Advertising Research*. 2013;97:217-246.
5. Ministry of Health and Welfare. [Report] 2020 Korean Medicine Utilization-Basic Report (User-Outpatient, Hospitalization). [cited in] 2022.02.20. Available from: [https://www.koms.or.kr/board/researchReport/view.do?post\\_no=162&menu\\_no=21](https://www.koms.or.kr/board/researchReport/view.do?post_no=162&menu_no=21)
6. Yoon LS, Lim BM. The Determinants and Behavioral Intentions of Korean Medicine Utilization in Youth Aged 19-39 Years : Based on the microdata of national survey on Korean Medicine Utilization and Herbal Medicine Consumption, 2017. *Journal of Society of Preventive Korean Medicine*. 2021;25(2):85-98.
7. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of machine Learning research*. 2003;3(Jan):993-1022.
8. Hong TH, Niu Hy, Ren G, Park JY. Multi-Topic Sentiment Analysis using LDA for Online Review. *The Journal of Information Systems*. 2018;27(1):89-110.
9. Chae SH, Lim JY, Kang JY. A Comparative Analysis of Social Commerce and Open Market Using User Reviews in Korean Mobile Commerce. *Journal of Intelligence and Information Systems*. 2015;21(4):53-77.
10. Kim JK, Choi ES, Yoon SH, Lee YB, Kim DH. BEHIND CHICKEN RATINGS: An Exploratory Analysis of Yogiyo Reviews Through Text Mining. *Journal of the Korea Contents Association*. 2021;21(11):30-40.
11. Lee SH, Jo AR, Lee HY. The Medical Service Customer's Satisfaction Factors Extracted from

- Online Hospital Review Data Using Latent Dirichlet Allocation Method. *Journal of Korea Service Management Society*. 2017;18(5):23-44.
12. Choi JE, Kim SD, Kim HW. A Study on Sentiment Score of Healthcare Service Quality on the Hospital Rating. *Information Systems Review*. 2018;20(2):111-137.
  13. ACE Counter. [Report] Analysis of inflow rate of search engines in the 3rd quarter of 2021. [cited in] 2022.02.20. Available from: [https://www.acecounter.com/www2/education/trendReportDetail.amz?rno=257&cpage=1&f\\_serach1=&i\\_serach=#self](https://www.acecounter.com/www2/education/trendReportDetail.amz?rno=257&cpage=1&f_serach1=&i_serach=#self). 2021
  14. Health Insurance Review & Assessment Service. [cited in] 2022.01.11. Available from: <https://www.hira.or.kr/rd/hosp/getHospList.do?pgmid=HIRAA030002020000#tab01>.
  15. Park EJ, Cho SZ. KoNLPy: Korean natural language processing in Python. *Annual Conference on Human and Language Technology*. 2014;26:133-136.
  16. Google cloud natural language. [cited in] 2022.02.20. Available from: [https://cloud.google.com/natural-language/docs/basics#natural\\_language\\_features](https://cloud.google.com/natural-language/docs/basics#natural_language_features)
  17. Nahm CH. An Illustrative Application of Topic Modeling Method to a Farmer's Diary. *Cross-Cultural Studies*. 2016;22(1):89-135.
  18. Choi SC, Park HW. A Study on the Trend of Topic Modeling in South Korea using KCI Journal Publications. *Journal of The Korean Data Analysis Society*. 2020;22(2):815-826.
  19. Blei DM. Probabilistic topic models. *Communications of the ACM*. 2012;55(4):77-84.
  20. Rehurek R, Sojka P. Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks 2010*.
  21. Lee JH, Lee IS, Jung KS, Chae BH, Lee JY. Patents and Papers Trends of Solar-Photovoltaic (PV) Technology using LDA Algorithm. *Journal of Digital Convergence*. 2017;15(9):231-9.
  22. Park YU, Chung KY. A study on the analysis of customer's sentiment using DMR(Dirichlet Multinomial Regression) topic modeling based on online review big data: Focusing on the foreign customer's reviews of domestic 5 star deluxe hotels. *Korean Journal of Hospitality & Tourism*. 2021;30(2):1-20.
  23. Kim JE, Baek SG. Analysis of Issues on the College and University Structural Reform Evaluation Using Text Big Data Analytics. *Asian Journal of Education*. 2016;17(3):409-436.
  24. Greene D, O'Callaghan D, Cunningham P. How many topics? stability analysis for topic models. In *Joint European conference on machine learning and knowledge discovery in databases*. 2014;14: 498-513. Springer, Berlin, Heidelberg.
  25. Lee BC, Kim DS. Deriving the Determinants of Hotel Service Quality using OTA Reviews - LDA Topic Modeling -. *Journal of Hotel & Resort*. 2020;19(4):41-58.
  26. Park JH, Chun MS, Bea SH, Kim HJ. Research Trends on Factors Influencing the Quality of Life of Cancer Survivors: Text Network Analysis and Topic Modeling Approach. *Asian Oncology Nursing*. 2021;21(4):231-40.
  27. National assembly research service. [Report] Current status and improvement tasks of auto insurance oriental medical treatment. [cited in] 2022.02.27. Available from: <https://www.nars.go.kr/report/view.do?page=48&cmsCode=CM0043&categoryId=&searchType=&searchKeyword=&brdSeq=30053>
  28. KOSIS. Current status of the cosmetics industry, number of cosmetics companies, number of items, and production performance. [cited in] 2022.02.27. Available from: [https://kosis.kr/statHtml/statHtml.do?orgId=145&tblId=DT\\_145011\\_A004&vw\\_cd=MT\\_ZTITLE&list\\_id=F\\_64\\_002&scrId=&seqNo=&lang\\_mode=ko&obj\\_var\\_id=&itm\\_id=&conn\\_path=MT\\_ZTITLE&path=%252FstatisticsList%252FstatisticsListInde](https://kosis.kr/statHtml/statHtml.do?orgId=145&tblId=DT_145011_A004&vw_cd=MT_ZTITLE&list_id=F_64_002&scrId=&seqNo=&lang_mode=ko&obj_var_id=&itm_id=&conn_path=MT_ZTITLE&path=%252FstatisticsList%252FstatisticsListInde)

- x.do
29. KOSIS. Year-to-year comparison of reasons for participating in sports activities. [cited in] 2022.02.27. Available from: [https://kosis.kr/statHtml/statHtml.do?orgId=113&tblId=DT\\_113\\_STBL\\_1027846&vw\\_cd=MT\\_ZTITLE&list\\_id=113\\_11303\\_003&seqNo=&lang\\_mode=ko&language=kor&obj\\_var\\_id=&itm\\_id=&conn\\_path=MT\\_ZTITLE](https://kosis.kr/statHtml/statHtml.do?orgId=113&tblId=DT_113_STBL_1027846&vw_cd=MT_ZTITLE&list_id=113_11303_003&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE)
  30. Park YM, Kim HY. A Study on Changes in Women's Makeup Interest and Use Patterns Before and After the Outbreak of COVID-19. 2021;19(1):129-138.
  31. Jo JM. "Hongsam," which takes care of health in daily life of COVID-19, is also being consumed non-face-to-face. fnnews. 2021. [cited in] 2022.02.27. Available from: <https://www.fnnews.com/news/202112011825508748>