

트위터 이용자의 음악 청취 행태 분석 및 국내 음악 순위와의 비교 연구

유영석*, 손방용**

대전대학교 컴퓨터공학과 석사과정*, 대전대학교 컴퓨터공학과 교수**

Music Listening Behavior analysis of Twitter User and A Comparative Study of Domestic Music Ranking

Young-Seok Yoo*, Bang-Yong Sohn**

The master's course, Department of Computer Engineering, Daejin University*

Professor, Department of Computer Engineering, Daejin University**

요 약 음악 소비행태가 온라인으로 변화하면서 온라인 음악 플랫폼들이 등장하기 시작한지 오래다. 사람들이 인기 음악이나 인기 뮤지션의 음악 등 추천 정보 제공을 선호하게 되면서 온라인 음악 플랫폼의 차트를 이용하기도 하고, 소셜 네트워크 서비스(SNS)를 통해 정보를 공유하기도 한다. 온라인 음악 순위 차트는 조금씩 차이가 있으며, 그 이유 중 하나는 플랫폼마다 보유하고 있는 회원들의 특성이 다르기 때문이라고 볼 수 있다. 한편, SNS 이용자의 특성을 지닌 음악 차트 순위가 필요하고 이와 관련하여 각 온라인 음악 플랫폼의 차트들을 종합하려는 시도가 있어왔다. 더불어 음악적 특성을 SNS와 연결 지어 연구하는 관련 논문 역시 계속 등장하고 있다. 본 연구에서는 국내의 다양한 SNS중 트위터 이용자의 음악적 특성을 분석하고 기존의 음악 플랫폼의 순위차트와 트위터의 해시태그라는 기능을 통해 SNS 이용자들이 공통적으로 선호하는 음악 기반의 국내 인기 음악 차트 순위를 개발하고 그 순위의 특성을 분석하여 기존의 온라인 음악 순위 차트, 공영방송사 음악 순위 차트와의 결과를 비교 연구한다.

주제어 : 음악 플랫폼, 해시태그, 트위터, SNS, 차트

Abstract While consumption patterns have changed online music, online music platform began to emerge. While people prefer popular music recommendation, they use the online music platform chart or use the SNS Platform to share information. Online platform Ranking is different because of different properties held by members. Meanwhile, we need music charts characteristics of SNS users. So there were a lot of attempts to chart a comprehensive variety of platforms. And continue to emerge theses linking the musical characteristics and SNS. In this paper, We have developed a new chart using the behavior of Twitter Users who listen to music, and studies comparing the results with existing chart.

Key Words : Music Platform, Hashtag, Twitter, SNS, Chart

Received 1 April 2016, Revised 30 April 2016
Accepted 20 May 2016, Published 28 May 2016
Corresponding Author: Bang-Yong Sohn(Daejin University)
Email: bysohn@daejin.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

최근 온라인 음악 시장이 커지면서 음악이라는 디지털 콘텐츠에 대한 관심이 높아지고 이에 대한 연구가 지속되어 왔다[1]. 여기에 발 맞춰 각종 음악 관련 서비스를 제공하는 온라인 음악 플랫폼이 많이 생겨났으며 이러한 플랫폼들에서는 각자의 방법으로 음악 시장의 트렌드와 인기를 반영한 음악 순위 차트 정보를 제공하고 있다. 다만 온라인 음악 플랫폼들마다 보유 회원의 특성이 다르기 때문에 차트 정보들이 다를 수 있으며, 이는 플랫폼들만이 가지는 특성이 반영되었다고 할 수 있다. 수많은 음악들 사이에서 음악 청취자들에게 추천 서비스를 할 수 있는 유용한 음악 차트인 만큼 다양한 특성을 반영한 차트 데이터의 필요성이 있어 본 연구를 진행하게 되었다.

연구 초기에는 온라인 음악 플랫폼에서 제공하는 API[2] 등을 활용하여 종합적인 데이터를 수집해 이용자의 인기 차트를 파악하려고 하였으나, 일부의 플랫폼에서만 API를 제공하기 때문에 데이터 수집의 한계가 있어 음악 플랫폼이 아닌 다른 영역에서 음악 이용 행태 분석이 불가피하게 되었다. 이에 요즘 화두가 되고 있는 SNS를 이용하는 방법을 고안하게 되었다.

SNS 이용자들은 제한적 분량의 자신의 상태 혹은 상태 메시지를 위해 자신의 일상을 업로드하고, 이를 통해 주변 사람들의 관심을 유도하는 등의 특성을 지니고 있다[3]. 때문에 SNS를 이용해 기존의 유명인들의 팬 페이지, 특정 주제에 대해 같은 관심을 가지는 유저들 간의 새로운 커뮤니티의 형성을 주도하기도 하고, 새로운 광고수단으로 주목을 받고 있다[4]. 이러한 특성으로 인해 타인의 관심을 위해 대개 트렌디한 내용이나 타인에게 유용한 정보를 주로 업로드하게 되며, 이 중 음악이라는 요소는 대중들에게 깊숙이 자리 잡고 있는 요소임에도 불구하고 큰 대가를 지불하지 않으면서 음악 선곡 형태를 통해 자신이 트렌드에 발 맞춰 나간다는 것을 어필할 수 있는 수단이다.

본 연구에 활용된 트위터는 세계를 대상으로 어마어마한 이용자를 보유한 SNS이다. 여기에 더불어 자신이 업로드한 게시물의 연관된 키워드 등을 표현할 수 있는 해시태그[5]라는 기능을 제공하기 때문에 이는 데이터 수집을 위한 키워드 선정의 한 기준이 될 수 있다. 특히 음악, 영화 등의 미디어를 재생중이라는 #nowplaying,

#np 라는 해시태그의 경우 트위터에서 대표적으로 사용되는 해시태그이다. 자신이 음악, 영화 등의 특정 미디어를 감상중일 경우 #nowplaying, #np 라는 해시태그를 이용하는 방법이 처음부터 정해져 있지는 않았지만 마치 규칙처럼 해당 해시태그를 사용하는 것이 트렌드화 되었다. 이를 이용해 필요한 조건에 맞게 필터링이 가능하고, 효율적인 트위터 내의 데이터 마이닝을 위한 전처리가 가능하게 되었다[6]. 여기에 누구나에게 무료로 제공되고 있는 통계 분석 툴인 R[7]에서 최근 트위터 데이터 마이닝 관련 패키지를 제공함에 따라 편리하게 트위터 데이터를 수집이 가능하다.

본 연구에서는 R과 데이터 마이닝 패키지를 이용해 음악 트윗 차트를 작성하여 기존 존재하는 온라인 음악 순위 차트, 공영방송사 음악 순위 차트와의 비교를 하려고 한다.

2. 선행 연구

2.1 연구 배경

작은 새가 지저귀는 소리를 나타내는 트윗(tweet)의 의미를 가진 트위터는 2006년 3월에 서비스를 시작한 소셜 네트워크 서비스로 140자 단위의 트윗이라고 하는 메시지를 전송할 수 있다[8]. 이렇게 전송된 트윗은 이용자의 프로필 메시지에 표시되며 해당 이용자를 팔로우(follow)하는 다른 이용자에게 즉시 전달되는 실시간성을 가지고 있다. 이는 기존의 전화, SMS, 전자우편, 블로그 등 기존의 소통 형태와의 차별 점으로 작용하여 하나의 온라인 트렌드가 되었고, 단문의 글을 작성하기 쉽다는 특징 덕분에 다양한 디바이스에서 이용이 가능하다. 현재는 다양한 계층에서도 트위터를 이용하고 있으며, 단순 상태 표시 뿐 아니라 홍보, 고객 소리 접수 등의 다양한 용도로 활용하고 있다.

2.1.1 해시태그 #nowplaying 또는 #np

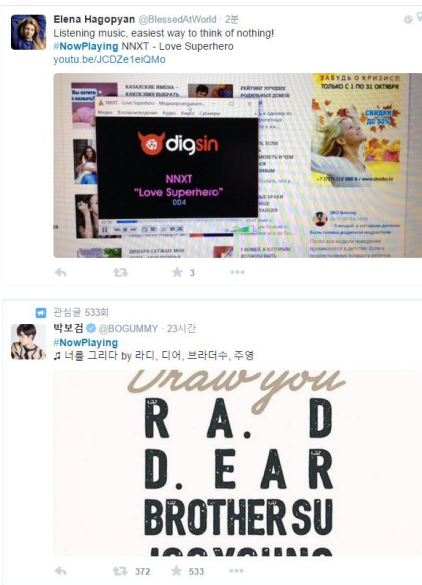
이처럼 다양한 용도로 사용되는 트위터지만 미미한 검색기능과 수많은 이용자의 이유로 인해 빅데이터 수집에 대한 어려움이 있어, 해시태그를 이용하여 원하는 정보를 필터링하여 얻을 수 있다.

#nowplaying 또는 #np는 [Fig. 1]과 같이 트윗 본문에

직접 기재하여 사용할 수 있다. #nowplaying 또는 #np 해시태그는 기존 해시태그의 특징인 단기 소멸성을 보이지 않고 지금까지 꾸준히 사용된다는 특징이 있으며 현재는 하루 450,000건 이상의 트윗이 생성되고 있다. 다만 음악 뿐 아니라 동영상 재생 시에도 같은 키워드가 사용되고, 해당 해시태그가 포함되어 있다고 해도 음악 제목이 함께 명시되지 않은 트윗이 존재한다는 특징 때문에 분석에 어려움이 존재한다.

그럼에도 현재까지 모색된 방법 중 이용자 청취 행태를 파악하기 위해 가장 적합한 수단으로 판단되어 2012년 이후 #nowplaying 또는 #np 해시태그와 이용자 청취 행태를 파악 후 이를 마케팅 영역으로 접목시켜 사용자를 유형별로 분류하고, 새로운 분석 모델을 만들어 내는 등의 연구가 계속 진행되고 있다[9].

본 연구에서는 트위터의 음악 청취 행태를 분석하고 음악 트렌드와의 관계성을 분석 및 SNS의 청취자 행태를 반영한 순위 차트를 통해 음악 트렌드를 보다 잘 설명할 수 있는 특성을 찾아내려고 한다.



[Fig. 1] Examples of use #nowplaying hashtag in Twitter

2.1.2 통계분석 패키지 twitterR

R은 현재 가장 유용하게 사용되는 통계 분석 도구중 하나이며, 그 기능이 매우 강력하고 누구나 무료로 이용

할 수 있다는 큰 장점을 가지고 있다. 이로 인해 현재 사회과학, 자연과학 등의 분야에서 통계 결과 도출 뿐 아니라 시각화된 결과 자료를 작성하는 방법으로 많이 사용되는 추세이다[10].

R은 통계 분석 도구로는 매우 유용하지만 특성상 데이터가 무수히 많아지면 데이터의 처리에 시간적인 제약을 받는다. 따라서 빅데이터의 처리에는 적합하지 않다고 판단되어 이를 빅데이터 수집도구와 병행하여 사용하려는 시도 아래 R HADOOP, RHIVE, RHIFE 등의 빅데이터 수집 및 분석 도구가 등장하기 시작했다[11]. 하지만 본 연구의 목적은 국내 사용자들의 추세를 분석하고, 주로 주 단위로 집계되는 음악 통계 특성상 만 개가 넘어가지 않는 데이터를 이용하기 때문에 R 한가지만을 사용하게 되었다.

여기에 twitterR[12]이라는 트위터 데이터 마이닝 관련 패키지를 통해 전체 사용자들의 트윗 정보들을 토대로 데이터를 수집할 수 있었다. twitterR은 R에서 트위터 API에 접근을 가능하게 하는 패키지로 twitterR의 정상적인 기능을 이용하기 위해서는 트위터 개발자 센터[13] 어플리케이션 개발 등록을 통해 접근 코드와 API키를 할당 받는 작업이 선행되어야 한다.

twitterR 패키지에서 제공하는 함수를 통해 할당받은 접근 코드와 API키를 이용해 트윗 데이터의 접근 인증을 받은 후 정상적인 트윗 수집이 가능하다.

2.2 기술적 배경

2.2.1 API(Application Programming Interface)

API는 응용 프로그램에서 사용하기 위해 특정 서비스에 요청을 보내고 응답을 받는 형식으로 접근을 가능하게 하는 인터페이스이다. 현재 API는 지도 API, 버스정보 API, 백과사전 API 뿐 아니라 다양한 기관, 기업에서 제공하는 서비스들이 API 형태로 제공되고 있다[14]. 특히 본 연구에서 이용하는 트위터 API는 R의 twitterR 패키지를 통해 트위터에서 제공하는 웹 API이며 메시지를 통해 요청을 보내고 응답을 받는 개념으로 사용된다.

보통 API 메시지 내용은 주로 JSON[15] 혹은 XML[16]이라는 메시지 교환 형태를 따르지만, 본 연구의 R 패키지인 twitterR에서 제공하는 메시지는 데이터 테이블 형식으로 제공되고 있다.

흔히 누구나 사용 가능하도록 공개되어 있는 API를

공개 API 혹은 오픈 API라고 하며, 제한 없이 다른 서비스 등에 맞물려 사용 가능하다.

2.2.2 통계분석 방법

본 연구에서 작성한 트윗 차트와 기존 존재하는 온라인 음악 순위 차트, 공영방송사 음악 순위 차트 간에 관계성을 확인하기 위해 분산분석과 상관분석이라는 통계분석 방법을 이용하였다.

분산분석은 두 개 이상의 집단을 비교하는 경우 각 집단의 분산, 총 평균과 각 집단 간 평균의 차이에 의해 생기는 집단 간의 분산을 비교하여 집단간 유사성을 확인하는 방법으로써, 분산의 비교를 통해 얻어지는 F분포를 이용한다[17]. 분산분석의 종류로는 일원분산분석, 이원분산분석, 다원분산분석, 공분산분석 등이 있는데, 본 연구의 경우 음악 플랫폼이라는 독립변인에 따라 순위라는 종속변인을 비교하는 것이므로 독립변인이 1개이다. 따라서 일원분산분석을 이용하며, 비교 플랫폼의 개수가 멜론, 엠넷, SBS, KBS, 트윗 차트로 3개 이상이기 때문에 집단 간 분산의 차이가 있다는 분석결과가 나왔다면 사후분석으로 이어져 각 집단의 수치에 따라 같은 분산을 가지는 그룹을 확인하게 된다. 결국 트윗 차트의 상위 순위권 음악이 온라인 음악 순위 차트, 공영방송사의 음악 순위 차트 상위권에 얼마나 위치하고 있는지, 하위 순위권 음악이 얼마나 위치하고 있는지를 판단하기 위해 사용한다.

상관분석은 두 변수 간에 어떠한 선형적 관계가 있는지 분석하는 방법으로써 변수간의 관계의 강도를 상관관계라고 한다. 상관관계의 정도를 파악하는 상관계수는 두 변수간의 연관된 정도를 나타내며[18], 함께 측정되는 유의확률을 통해 변수 간 서로 영향을 미치는지 여부를 알 수 있다. 통상적으로 유의확률이 0.05 이하의 값이 측정되는 경우 변수 간 서로 영향을 미친다고 판단할 수 있다.

상관계수는 피어슨 상관계수, 스피어만 상관계수, 크론바흐 알파 계수 등의 척도로 구할 수 있는데, 본 연구의 경우에는 음악 순위를 기준으로 상관관계를 구하기 때문에 자료의 값 대신 순위를 이용하는 스피어만 상관계수를 이용하였다. 스피어만 상관계수를 이용한 상관분석은 특정 매체에서 찾고자 하는 키워드의 발현 횟수 및 구매 만족도 조사 등 이외 다양한 분야에서의 요인 간

관계 분석을 위해 많이 쓰이는 기법이다[19]. 상관계수는 1에 가까울수록 한 변수가 증가하면 다른 변수도 함께 증가하는 양의 상관관계와 -1에 가까울수록 한 변수가 증가하면 다른 변수는 감소하는 음의 상관관계를 나타내며, 0에 가까울수록 변수 간의 상관관계가 약하다고 말할 수 있다.

3. 데이터 수집 및 정형화

3.1 트윗

트위터에서는 데이터를 수집하기 위한 다양한 API를 제공하는데, R에서 데이터를 제공받기 위해서 twitterR 패키지를 이용한다. 여기에서 등록된 트윗의 언어 설정을 통해 한국어가 포함된 트윗만을 수집하여 한국인 기준의 트윗 데이터를 획득하거나, 특정 기간을 조건으로 트윗을 수집하는 등의 조건으로 데이터 수집이 가능한데, twitterR을 통해 데이터를 수집하는 과정은 [Fig. 2]와 같으며, 제공받는 데이터 항목은 <Table 1>과 같다.

<Table 1> Information contained in the tweets collected from R

Item	
text	favorited
favoriteCount	replyToSN
created	truncated
replyToSID	id
replyToUID	statusSource
screenName	retweetCount
isRetweet	retweeted
longitude	latitude

```
> library(twitterR)
...
> setup_twitter_oauth(API_key, API_secret, acc_T, acc_T_Sec)
...
> Search <- c("#np")
> Result <- searchTwitter(Search, lang='ko', since='2015-09-28')
> Dataset <- twListToDF(Result)
...
```

[Fig. 2] Data collection code with R

항목 중 수집 기간 설정을 통해 트윗 생성일인 created를 기준으로 트윗 본문인 text내 #nowplaying, #np 해시태그가 포함되는 트윗들을 수집하였다.

3.2 각종 플랫폼 차트

대한민국 온라인 음악 산업에서 국내 온라인 음악 플랫폼들은 매우 다양하지만, 그 중 보유 회원, 제공 서비스, 운영 기간 등을 기준으로 가장 대표적인 유명 온라인 음악 플랫폼은 멜론, 엠넷, 벅스, 올레뮤직, 소리바다 등이 있으며[20], 각각의 점유율은 <Table 2>와 같다.

<Table 2> Online Music Platform Status

Platform	Visitor	Page View	Retention time
Melon	6,110,908	34,203	6,872
Mnet	1,508,820	6,872	2,294,684
Bugs	791,158	4,262	149,215
Ollehmusic	654,094	6,601	86,245
Soribada	301,838	1,781	23,517

<Table 2>의 방문자 수를 기준으로 온라인 음악 플랫폼은 엠넷, 멜론의 차트를 이용하였으며, 공영방송사의 경우, KBS, SBS를 이용 하였다. MBC의 경우 본 연구의 작성일을 기준으로 마지막 음악 차트 데이터 생성 날짜가 7월 이므로 포함시키지 않았으며, 각 차트 집계 기준은 <Table 3>과 같다.

<Table 3> Aggregate way of public broadcasters and Online Music Platform

Music Chart	Aggregate way
Melon	Streaming in Melon 40% + Download in Melon 60%
Mnet	Streaming in Mnet 50% + Download in Mnet 50%
SBS	Digital Sound Source 55% + SNS 35% + Viewers advance voting 5% + Album Score 5%
KBS	Digital Sound Source 65% + Music Sales 5% + Broadcast count 20% + Viewers rating 10%

국내 유명 음악 플랫폼과 공영 방송사의 음악차트에서는 동일한 집계기간을 가진 주 단위의 인기 차트 정보를 제공하기 때문에 이를 주 단위로 수집하여 종합적인 차트를 작성한다. 이 때, 주간 인기차트의 상한선은 SBS가 30위까지의 정보 제공하므로 각 음악 차트들의 기준을 같게 하기 위해 30위까지의 데이터를 수집하는 것으로 한다. 마찬가지로 플랫폼과 공영방송사별 30위 내의 음악 타이틀의 순위를 집계한 뒤 중복된 음악 타이틀을 제거하여 총 58개의 타이틀이 나열되었다.

3.3 트윗 데이터의 정형화

데이터 수집은 R에서 제공하는 twitterR 패키지 내의 searchTwitter 함수를 통해 각종 음악 순위 플랫폼의 차트 집계기간을 고려하여 월요일 00:00부터 차주 월요일 00:00 직전까지의 데이터 중 #nowplaying 해시태그와 #np 해시태그를 포함하는 트윗을 모두 추출한다. 이 다음 추출된 데이터 중 #np와 #nowplaying 해시태그를 함께 사용한 중복 트윗을 제거한 뒤 #np 해시태그를 포함하는 트윗과 #nowplaying 해시태그를 포함하는 트윗을 결합하여 하나의 데이터 셋으로 만든다.

9월 마지막 주 데이터(2015. 09. 28 ~ 2015. 10. 04)를 기준으로 총 4,004개의 트윗 데이터가 수집되었으며, 이 데이터 셋 내에서 원활한 필터링을 위해 데이터 전처리 과정으로 이어진다.

전처리 과정에는 R의 gsub()함수를 이용하여 트윗이 포함하고 있는 정보 중 text 항목으로 분류되어 있는 트윗 본문의 불필요한 문자를 제거한다. 정해진 불필요한 문자의 기준은 \n, \t, RT, http, #np, #nowplaying, '-' 이다. 근래의 일부 음악 타이틀의 경우 '\$', '#'과 같은 일부 특수문자의 사용이 빈번하게 발생하는 것을 고려하여 특수문자를 제거하는 과정은 배제하였다.

불필요 문자가 제거된 데이터 셋 내에서 다양한 트윗 게시자의 특성을 고려, 음악 타이틀의 검색을 원활히 하기 위해 띄어쓰기를 없앤 타이틀로 검색이 이루어지기 때문에 수집한 데이터의 공백을 모두 제거한다.

이렇게 완성된 데이터 셋을 R의 twitterR 패키지의 twListToDF 함수를 통해 리스트화 하여, 트윗 내용과 중복 횟수로만 이루어진 데이터 셋으로 간략화 한다.

마지막으로 집계된 데이터 셋을 아티스트 명으로 다시 한 번 필터링해 최종적인 데이터를 얻는다. 이렇게 얻어진 데이터 셋을 통해 중복된 트윗들의 중복 횟수를 산출하고, 트윗을 이용한 차트 집계 시 중복 횟수를 반영하여 트윗 차트를 작성한다.

4. 결과 비교 및 분석

4.1 차트별 동일성

정형화 과정을 거친 트윗 데이터 내에서 플랫폼 차트 종합을 통해 집계된 58개의 타이틀을 순차적으로 검색해

타이틀 별 트윗에 등장된 횟수를 기록한다. 타이틀 검색에는 R 내에서 데이터 수집시의 한국어 인식의 원활함을 위해 R의 Unicode 패키지라는 한국어 인코딩 함수를 이용하였다. 타이틀로 검색된 데이터 셋 내에서 마찬가지로 방법으로 해당 곡의 아티스트 명을 한국어, 영어 등으로 다시 검색을 실시하여 동명의 타이틀이라 할지라도 음악 플랫폼 차트에 등장한 타이틀과 아티스트가 매치된 결과가 도출될 수 있도록 한다. 이렇게 작성된 데이터 셋은 <Table 4>와 같다.

<Table 4> Tweet chart according to the number of tweets

Ranking in Twitter	Title	Artist	Tweet Count
1	울지말자	Homme	306
2	보통연애	Park Kyung	175
3	니가 하면	GOT7	163
4	Okey Dokey	Song Min-Ho ZICO	119
5	리듬 타	iKON	108
6	겹	Song Min-Ho	92
7	너나 잘해	Ailee	53
8	어깨	So-you Kwon Jeong-Yeol	49
9	거북선	Ja Mezz Andup Song Min-Ho	34
10	Ah-Choo	Lovelyz	27

집계된 결과를 온라인 음악 순위 차트, 공영방송사 음악 순위 차트를 통해 30위까지의 순위를 비교한 결과는 <Table 5>와 같다. 여기에 10위까지의 순위는 굵게 표시하였다.

<Table 5> Tweet chart and Platform-specific chart comparison

Ranking in Twitter	Title	Ranking in Platform			
		Melon	Mnet	SBS	KBS
1	울지말자		8		
2	보통연애	5	15	8	8
3	니가 하면			2	4
4	Okey Dokey	12	28		
5	리듬 타	16	11	4	13
6	겹	7	21		
7	너나 잘해	20	1	9	6
8	어깨	3	2	6	2
9	거북선	21			
10	Ah-Choo		22	14	24

<Table 5>를 살펴보면 각 플랫폼 및 공영방송사 내의 10위 권 내 음악이 트윗 순위 10위 내에 포진되어 있는 것을 확인하였지만 그리 많은 수의 음악이 포함되었다고 볼 수 있는 수치적인 증거가 없기 때문에 이를 통계분석틀을 이용하여 9월 마지막 주, 10월 첫째 주, 10월 둘째 주의 분산분석을 실시한 결과는 <Table 6>와 같다.

<Table 6> Analysis of variation results of the weekly Tweeter chart and platform chart(10 to top)

Week	Degrees of Freedom	F-distribution	p-value
Last week in September	33	0.981	0.433
First week of October	36	1.959	0.125
Second week of October	35	0.505	0.733

그룹 간의 분산을 분석한 결과 3개 주치의 유의확률은 일반적으로 사용되는 유의수준인 0.05보다 훨씬 크기 때문에 트윗 차트 10위 까지의 결과를 타 플랫폼과의 비교를 실시한 결과는 그룹 간 평균의 차이가 없다고 볼 수 있어 트윗 차트의 상위권 내에 위치한 음악은 타 플랫폼의 음악 차트 상에서도 상위에 위치하고, 트윗 차트의 하위권 내에 위치한 음악 또한 타 플랫폼의 음악 차트 상에서도 하위권에 위치하는 등의 유사한 결과를 갖는다고 결론지을 수 있다.

여기에 그치지 않고 온라인 음악 순위 차트, 공영방송사 음악 순위 차트 30위까지의 데이터를 수집한 것을 감안하여 트윗 차트와 각 음악 차트간의 9월 마지막 주, 10월 첫째 주, 10월 둘째 주의 분산분석을 실시한 결과는 <Table 7>과 같다.

<Table 7> Analysis of variation results of the weekly Tweeter chart and platform chart(30 to top)

Week	Degrees of Freedom	F-distribution	p-value
Last week in September	99	0.227	0.923
First week of October	99	0.707	0.589
Second week of October	111	0.232	0.920

30위까지의 각 주차별 분산 분석을 실시한 결과 10위까지 데이터와 마찬가지로 유의확률이 0.05보다 높았으며, 30위까지의 데이터 역시 그룹 간 차이가 없다고 볼 수 있다. 여기에 더해, 관찰된 유의확률은 10위까지의 분산분석 결과보다 훨씬 높은 수치를 보여 30위까지의 분산분석 결과가 더 명확하다고 볼 수 있다.

4.2 순위 예상

같은 주차별 트윗 차트와 플랫폼 차트의 10위와 30위까지의 분산분석 결과 <Table 6>와 <Table 7>의 결과를 바탕으로 트윗 차트와 온라인 음악 순위 차트, 공영방송사 음악 순위 차트 간에는 차이가 없는 것으로 나타났다. 여기에 더해, SNS를 이용하는 목적 중 사용자 간 정보교류가 크다는 측면을 염두하여[21] 매니아 혹은 얼리어답터와 같은 정보 공유자를 통해 다수의 사용자의 의사가 결정될 수 있어 트윗 데이터가 온라인 음악 순위 차트, 공영방송사 음악 순위 차트를 예측한 데이터라고 볼 수 있다. 트윗 차트와 타 플랫폼 순위간의 연관성을 확인하기 위하여 온라인 음악 순위, 공영방송사 음악 순위 30위까지의 차트 데이터를 한 주 전의 트윗 차트와 상관분석을 실시하였다. 사용한 데이터는 9월 마지막 주부터 10월 첫째 주까지의 트윗 차트와 10월 첫째 주부터 10월 셋째 주까지의 온라인 음악 플랫폼, 공영방송사의 데이터이며, 그 결과는 <Table 8>과 같다.

<Table 8> Correlation Analysis of weekly Tweet chart and next week platform chart (30 to top)

		Melon	Mnet	SBS	KBS
Last week in September	Correlation coefficient	0.262	-0.382	0.510	0.341
	p-value	0.294	0.276	0.044	0.166
First week of October	Correlation coefficient	0.389	-0.05	0.576	0.492
	p-value	0.123	0.872	0.012	0.045
Second week of October	Correlation coefficient	-0.1	0.475	0.577	0.374
	p-value	0.712	0.140	0.015	0.154

결과적으로 일부 플랫폼에서 연관성이 있다고 볼 수 있는 매우 유의한 확률이 관측되었는데, 특히 SBS는 모든 주차에서 트윗 차트와 연관성이 있다는 유의확률을 확인할 수 있었다. 이는 <Table 3>에 의해 온라인 음악

플랫폼의 경우 보유 회원의 자사 플랫폼 내의 청취 형태를 기준으로 순위를 책정하는 반면, SBS의 경우 SNS라고 할 수 있는 유튜브의 선호도를 포함하여 집계한 특성을 반영한 차트라고 볼 수 있기 때문이다.

5. 결론

트윗 정보를 이용하여 트윗 차트 순위를 개발하였고 <Table 4>, 인지도를 기준으로 한 유명 음악 플랫폼 2곳과 공영방송사 2곳의 음악차트와 개발된 트윗 차트 상위 순위를 비교하였다<Table 5>. 또한 기존 음악차트순위와 트윗 차트 순위의 동질성을 확인하기 위하여 분산분석을 행하였다. 1단계는 가장 인기 있는 10위까지의 음악을 이용하여 분석한 결과 집단 간의 차이가 없어 통계적으로 동일한 차트라고 말 할 수 있다고 볼 수 있으며 <Table 6>. 2단계에서 30위까지의 음악을 이용하여 분석한 결과 역시도 기존 차트와 트윗 차트 간의 차이가 없어 좀 더 통계적으로 1단계보다 더 가까워진 것을 <Table 7>에서 확인할 수 있다. 이를 통해 개발된 트윗 차트는 기존 존재하는 온라인 음악 순위 차트, 공영방송사의 음악 순위 차트와 동일한 그룹이라고 말 할 수 있으며 유사한 순위를 갖는 특성을 가진다고 볼 수 있다.

한편 트윗 차트가 온라인 음악 순위 차트, 공영 방송사 음악 순위 차트를 예측하는 선행지수 특성을 확인하기 위해 실시한 해당 주차의 트윗 차트와 차주의 온라인 음악 플랫폼, 공영 방송사 음악 차트와의 상관분석 결과, 트윗 차트가 공영방송사 2곳과의 공영방송사의 차트와는 강한 상관성을 나타내었다<Table 8>. 따라서 트윗 차트의 선행지수적 특성을 통해 트윗 차트는 일부 음악 차트 순위를 예상하는데 유용하다고 볼 수 있으며, 이는 플랫폼별 음악차트 순위를 선정하는 방식에서 공영방송사의 경우 음악차트순위를 결정하는 자료수집 대상 영역에 SNS인 유튜브의 점수가 포함되기 때문인 것으로 분석된다.

R상에서 제공되는 트위터 API가 2주 전까지의 데이터만을 제공하기 때문에 관련 연구를 지속하기 위해서 주기적인 데이터의 수집이 필요하므로 자동으로 트윗 데이터를 수집하도록 하는 프로그래밍이 병행되어야 한다.

REFERENCES

[1] H. G. Kang, "Music Pattern Analysis of K-POP", Journal of the Korea Convergence Society, Vol. 11, No. 3, pp. 95-100, 2013.

[2] SK Planet Developers : <https://developers.skplanetx.com/apidoc/kor/melon/chart>, September 1, 2015.

[3] H. Y. Chan, "Analysis of Opinion Social Data on the SNS by analyzing of Collective Damage Reply", Journal of the Korea Convergence Society, Vol. 11, No. 5, pp. 41-51, 2013.

[4] D. I. Tag, "A Study on The Influence of Convergence Benefit of Facebook Fan Page in Brand Attachment and Brand Commitment", Journal of the Korea Convergence Society, Vol. 6, No. 5, pp. 199-206, 2015.

[5] Twitter Information Center-Hashtag : <http://support.twitter.com/articles/402995>, September 1, 2015.

[6] J. S. Lee, "A Study on the Data mining preprocessing Tool For Efficient database marketing", Journal of the Korea Convergence Society, Vol. 12, No. 11, pp. 257-264, 2014.

[7] The R Project for Statistical Computing : <https://www.r-project.org>, September 1, 2015.

[8] H. W. Kwak, "Why Twitter is a social network", 2011.

[9] Y. K. Kim, "Crowds vs. Broadcasters: Predicting Billboard Rank Using Characteristic of Music Listening Behavior in Twitter", 2015.

[10] K. H. Choi, "A reviews on the social network analysis using R", Journal of the Korea Convergence Society, Vol. 6, No. 1, pp. 77-83, 2015.

[11] D. H. Byun, "Method for Selecting a Big Data Package", Journal of the Korea Convergence Society, Vol. 11, No. 10, pp. 47-57, 2013.

[12] twitterR : <http://cran.nexr.com/web/packages/twitterR/index.html>, September 1, 2015.

[13] Twitter developers : <https://dev.twitter.com>, September 1, 2015.

[14] Y. M. Kim, "A Study on plan for promoting innovation and utilization of information sharing", Journal of the Korea Convergence Society, Vol. 12,

No. 4, pp. 43-49, 2014.

[15] The JSON Data Interchange Format : <http://www.json.org>, September 1, 2015.

[16] Extensible Markup Language(XML) : <http://www.w3.org/XML>, September 1, 2015.

[17] IBM Knowledgecenter-One way ANOVA : http://www-01.ibm.com/support/knowledgecenter/SSLVMB_20.0.0/com.ibm.spss.statistics.help/idh_owew.htm?lang=ko, September 1, 2015.

[18] IBM Knowledgecenter-Bivariate Correlations : http://www-01.ibm.com/support/knowledgecenter/SSLVMB_21.0.0/com.ibm.spss.statistics.help/idh_corr.htm?lang=ko, September 1, 2015.

[19] J. H. Choi, "A Study on the Correlation of Agendas between Politicians' Twitters and traditional News Media", 2011.

[20] Korean Click, "October to December 2013 Total PC/Mobile(Web + APP) service status using online sources", 2014.

[21] H. Y. Kim, "Domestic and international SNS status and sociological implications", pp. 19-42, KISDI, 2012.

유 영 석(Yoo, Young Seok)



- 2015년 2월 : 대전대학교 컴퓨터공학과(이학사)
- 2015년 2월 ~ 현재 : 대전대학교 컴퓨터공학과 석사과정
- 관심분야 : 통계자료분석, 빅데이터
- E-Mail : ce_young@daejin.ac.kr

손 방 용(Sohn, Bang Yong)



- 1983년 2월 : 전북대학교 전산통계학과 (이학사)
- 1985년 2월 : 고려대학교 통계학과 (경제학석사)
- 1994년 2월 : 고려대학교 통계학과 (이학박사)
- 1995년 4월 ~ 현재 : 대전대학교 컴퓨터공학과 교수
- 관심분야 : 컴퓨터알고리즘, 통계자료분석, 빅데이터
- E-Mail : bysohn@daejin.ac.kr