

문제 설명

K-POP 기획사인 "JYH 엔터테인먼트"에서는 팬덤(특정 아이들에게 관심이 많은 집단)이 아이돌 성공에 중요한 역할을 하고 있다는 것을 파악하였다.

특히, 아이돌의 성공을 위해 특정 팬덤이 다른 팬덤들과 협력하는 경우가 많음을 알게 되었다. 예를 들어, 타 아이돌의 음반을 스트리밍 하거나 시상식 때 투표를 독려하며 도움을 준 후, 지지하는 아이돌이 음반을 발매하거나 시상식 후보로 오른다면 도움을 주었던 팬덤에 도움을 요청한다.

JYH 엔터테인먼트는 이러한 팬덤 활동의 영향력을 파악하기 위해서 팬덤 메타 데이터와 팬덤 간 협력 정보 데이터(11월, 12월)를 확보해 데이터 사이언티스트에게 분석작업을 의뢰하였다.

데이터 사이언티스트가 되어 아래의 문제를 해결하고, 보고서를 작성하여라.
JYH에 제출해야 할 결과물은 다음과 같다.

- 보고서(pdf 형식으로 제출, 파일이름: 학번), 소스코드(파일 이름: 학번.ipynb)
- 위의 결과물을 아이캠퍼스에 제출하라.

Dataset 설명

팬덤 도움 데이터는 다음의 다섯가지 데이터셋으로 구성되어 있다.

1. 팬덤 정보 데이터(fandom_nodes_2021.csv)

fandom_id	fandom_name (팬덤명)	nov_post (11월 총 게시물 수)	dec_post (12월 총 게시물 수)	nov_support (11월 총 supporting 수)	dec_support (12월 총 supporting 수)	type (boy or girl – if boy = 1, otherwise 0)
-----------	----------------------	------------------------------	------------------------------	--	--	---

2. 팬덤간 도움 기록 데이터(fandom_edges_2021.csv)

source (도움 준 fandom_id)	target (도움 받은 fandom_id)	nov_support (11월 도움 수)	dec_support (12월 도움 수)
----------------------------	-----------------------------	---------------------------	---------------------------

3. Weekly chart 정보(weekly_chart.json)

album	artist	랭킹	title	start	end
-------	--------	----	-------	-------	-----

4. Monthly chart 정보(monthly_chart.json)

album	artist	랭킹	title	start	end
-------	--------	----	-------	-------	-----

5. Fandom 메타 정보(fandom_meta_2021.json)

fandom 명	artist	agent	데뷔년도	팬 커뮤니티 내 전체 게시물 수
----------	--------	-------	------	-------------------

문제

* Annotation

- supporting: 도움 주는 / - supported: 도움 받은

[전체 기간 데이터 분석]

1. 팬덤 간 도움은 활발한가? 어떤 팬덤이 활발한 도움 활동을 하는가?

- 1) 팬덤 간 전체 supporting 수/ 팬덤 내 전체 활동 수 (2점)
- 2) 타입별 분석(전체, girl, boy) (3점)
 - Supporting Top 10 fandoms 들에 대한 정보
 - Supported Top 10 fandoms 들에 대한 정보
- 3) 소속사 별 분석(소속 가수들 정보 합계) (3점)
 - Supporting Top 10 소속사
 - Supported Top 10 소속사

2. 팬덤 활동이 많은 그룹이 도움을 주고 받는 데에도 활발한가?

1) Pearson correlation을 활용해서 측정하고 해석하라 (3점)

Hint!

- https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

- <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.pearsonr.html>

2) Spearman rank correlation을 활용해서 측정하고 해석하라 (3점)

Hint!

- https://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient

- <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.spearmanr.html>

3. 다음의 가설이 타당한지 independent t-test를 이용하여 분석하라. (3점/3점)

- 1) - H_0 : 걸그룹과 보이그룹 간 supported 활동 정도 차이가 없다.
 - H_1 : 걸그룹과 보이그룹 간 supported 활동 정도 차이가 있다.
- 2) - H_0 : 걸그룹과 보이그룹 간 supporting 활동 정도 차이가 없다.
 - H_1 : 걸그룹과 보이그룹 간 supporting 활동 정도 차이가 있다.

Hint!

- <https://ko.wikipedia.org/wiki/T-%ED%85%8C%EC%8A%A4%ED%8A%B8>

- https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html

[월별 데이터 분석]

4. 팬덤 도움 그래프를 생성하고 분석하라.

1) 전체 및 월별 팬덤 도움 (directed/weighted) 그래프 $G = (V, E, W)$ 를 생성하라. (G_0 : 전체 활동 그래프, G_1 : 11월 그래프, G_2 : 12월 그래프) (6점)

- node V : fandom
- node attributes: agent, debut year, gender type
- edge E : fandom \rightarrow fandom (supporting)

- weight W : 해당 월의 도움 수 (의미 있는 edge만 포함하기 위해서 상위 1% weight인 edge만 남기고 나머지 edge는 filtering out)

- 2) G0의 degree를 바탕으로 Cumulative Distribution Function (CDF)를 만들고 해석하라. (6점)
 - 3) G0에서 pagerank 및 degree centrality 기준 top 5 fandom의 정보를 기술하고 해석하라. (6점)
 - 4) G0에서 boy->boy, boy->girl, girl->boy, girl->girl 경우의 비율을 각각 구하여 기술하고 해석하라 (6점)
5. 각 월별로 생성된 두개의 팬덤 도움 네트워크 G1과 G2를 비교하라.
- 1) G1에서는 edge가 있었는데 G2에서 있거나 없어지는 경우, G1에서는 edge가 없었는데 G2에서는 있거나 없어지는 경우를 테이블에 기술하고 해석하라. 도움이 유지되는 경우가 팬덤 아이돌의 boy, girl 그룹 여부와 관련이 있는가? (7점)
 - 2) 네트워크 별 RolX algorithm 기반으로 Graph Role Extraction과 girvan_newman algorithm 기반 Community Detection 분석을 하고, 각각에 대해 의미하는 바를 해석한 후 월별로 그 역할과 그룹이 달라진 팬덤에 대해 분석하라. (7점)
- Hint!

-<https://github.com/dkaslovsky/GraphRole>

https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.community centrality.girvan_newman.html

https://networkx.org/documentation/stable/reference/classes/generated/networkx.DiGraph.to_undirected.html
- 3) G1과 G2의 pagerank 및 degree centrality를 구하고, G1에서 역할을 많이 하였던 팬덤들이 G2에서도 여전히 역할을 많이 하는지 분석하라. (7점)

[Implication 분석]

6. 팬덤간의 도움 현상이 음원 차트에 미치는 영향에 대해 분석하라. (15점)
7. (Optional) 팬덤간의 도움 현상의 본질을 이해할 수 있는 추가 분석할 내용이 있으면 분석 문제 정의 및 분석 결과를 기술하라. (5점)
8. 위의 결과를 바탕으로 분석 리포트를 작성하라. 분량은 테이블 등 그림 포함하여 A4용지 5장 이내로 작성하라. 문항 별 내용 정리 및 분석 결과 의미 해석, 분석 결과가 적용될 수 있는 가능성 등을 토의하라. 토의에 필요하면 추가적 분석을 수행해서 포함 할 수 있다. (15점)

- JYH 엔터테인먼트