

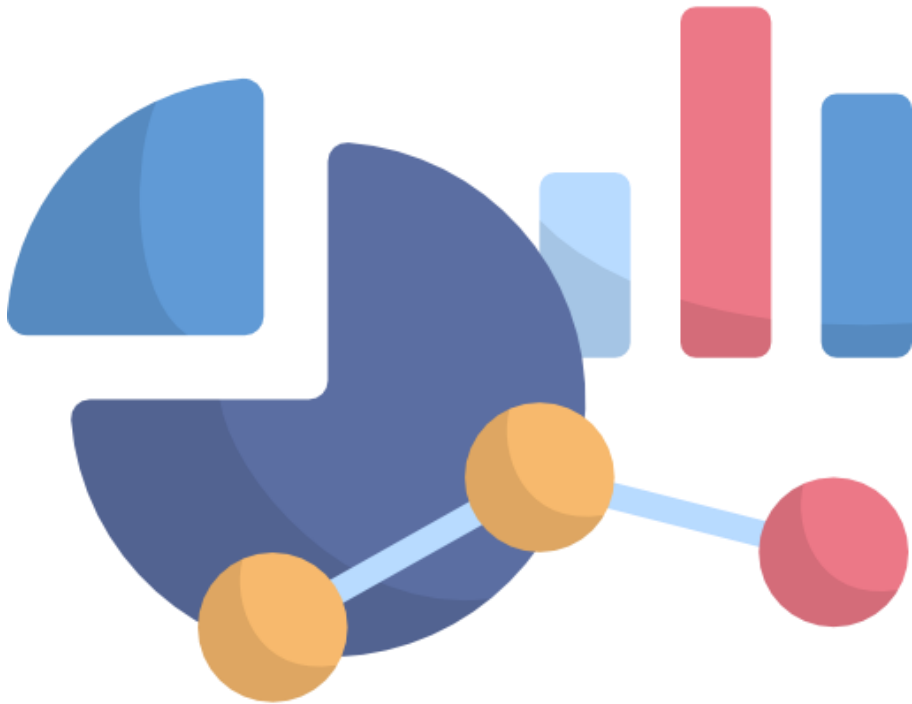


BIGTATO

: 연관분석을 활용한 Audience Targeting

(부제 : 잠재고객 맞춤 콘텐츠 추천)

안병우, 우영목, 한수호



연관분석을 활용한 Audience Targeting

(부제 : 잠재고객 맞춤 콘텐츠 추천)

01. 서론

- 아이디어 제안 배경
- 분석 목적 및 기법

02. 알고리즘 및 분석 결과

- 알고리즘 설명
- 알고리즘 적용
- 분석 결과

03. 비즈니스 모델 및 전략

부록. 변수 설명



PART.01

서론

- 아이디어 제안 배경
- 연구 목적

아이디어 제안 배경

온라인 쇼핑 이용객 증가

- 정보의 급격한 증가로 인해 소비자는 원하는 정보를 얻기는 점점 힘들어지고, 제공되는 정보들의 가치는 저하되는 문제 발생
- 이에 따라 ‘대중 맞춤(Mass Customization)’이 기업 경쟁력 강화에 중요한 마케팅 이슈로 작용

빅데이터 축적

- 기업들은 수집된 데이터를 통해 소비자들의 행동패턴을 분석하여, 비즈니스 인사이트를 적용할 수 있는 환경이 조성

분석 목적 및 기법

분석 목적

- 2019년 7월 ~ 2019년 9월 LPOINT 사용자 행동 데이터를 분석
- AI 기반 고객 맞춤 상품 추천 알고리즘을 도출하여 비즈니스 전략 수립

데이터 마이닝 기법

- 내용 기반(Content-Based) 필터링 방법:
 - 사용자가 과거에 선호했던 아이템과 유사한 성격을 가진 아이템을 추천하는 방식
 - 대표적 기법: 연관규칙 마이닝 (Association Rule Mining), 순차패턴 마이닝 (Sequential Pattern Mining)
- 사용 프로그램
 - Python

BIGTATO

PART.02

알고리즘 및 분석 결과

- 알고리즘 설명
- 알고리즘 적용
- 분석 결과

알고리즘 설명

연관규칙 마이닝 (Association Rule Mining)

- 대형 데이터베이스에서 변수 간의 흥미로운 관계를 발견하기 위한 규칙-기반 기계 학습 방법
- 동시에 구매될 가능성이 큰 상품들을 찾아냄으로써 장바구니 분석(Market Basket Analysis)에서 다루는 문제들에 적용 가능
- 소비자들의 구매이력을 토대로 “X 아이템을 구매하는 고객들은 Y 아이템 역시 구매할 가능성이 높다”는 식의 결론 가능
- Ex) 슈퍼마켓의 판매 데이터에서 발견된 “{양파, 감자} \Rightarrow {버거}” 규칙에서 고객이 양파와 감자를 함께 구매하면 햄버거 고기도 사기 쉽다는 것을 알 수 있다.

알고리즘 설명

연관규칙 알고리즘 (Using Apriori Algorithm)

○ 순서

1) 빈발항목집합을 찾는다.

- 미리 결정된 최소지지도 $supp_{min}$ 이상의 트랜잭션 지지도를 갖는 모든 빈발항목 생성

2) 생성된 빈발항목집합으로 연관규칙 생성

- $I = \{i_1, i_2, \dots, i_n\}$ 을 빈발항목들의 집합이라 할 때, 각 트랜잭션 T 는 $T \subseteq I$ 인 항목들의 집합
- 모든 빈발항목집합 I 에 대하여 I 의 모든 공집합이 아닌 부분집합들을 찾고 최소신뢰도 $conf_{min}$ 및 최소향상도 $lift_{min}$ 이상인 규칙을 생성

3) 연관규칙은 $X \Rightarrow Y$ 로 표시한다.

- Ex) 판매 제품 간의 연관규칙이 $\{\text{onion, potato}\} \Rightarrow \{\text{meat}\}$ 이면, $\{\text{onion, potato}\}$ 를 구매하면 $\{\text{meat}\}$ 도 구매하는 규칙

알고리즘 설명

사용 척도

- 의미 있는 연관규칙의 선택을 위해 다음의 척도가 유용하게 사용된다.

- 지지도(support): 전체 구매 건수 가운데 상품 X와 Y를 동시에 구매한 비율

$$supp(X) = \frac{|\{t \in T; X \subseteq Y\}|}{|T|}$$

- 신뢰도(confidence): 상품 X를 구매한 건수 가운데 Y도 같이 구매한 비율

$$conf(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)}$$

- 향상도(lift): 전체에서 상품 Y를 구매한 비율에 비해 X를 구매한 고객이 Y를 구매한 비율이 몇 배인지 알 수 있어 독립성 검증에 활용

$$lift(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X) \times supp(Y)}$$

알고리즘 설명



알고리즘 적용

| | clnt_id | new |
|----|---------|--|
| 0 | 2 | A03_Ramens |
| 1 | 2 | A03_Ramens |
| 2 | 2 | A03_Cream and Condensed milk |
| 3 | 2 | A03_Coffee Drinks |
| 4 | 2 | A03_Canned Vegetable Foods |
| 5 | 2 | A02_Men's T-shirts |
| 6 | 2 | A02_Infant / Toddlers' T-shirts / Tops |
| 7 | 2 | A03_Coffee Drinks |
| 8 | 2 | A03_Coffee Drinks |
| 9 | 2 | A03_Ramens |
| 10 | 2 | A03_Crab Sticks |
| 11 | 2 | A03_Fried Tofu |
| 12 | 9 | A03_Pasta Noodles |
| 13 | 9 | B01_Fresh Milk |
| 14 | 9 | B01_Tofu |
| 15 | 9 | A03_Pasta Sauces |

구매 품목과 biz_unit을 합쳐
새로운 column으로 만들

| | support | itemsets |
|------|----------|---|
| 0 | 0.034353 | (A02_Others) |
| 1 | 0.030598 | (A02_Women's T-shirts / Tops) |
| 2 | 0.033658 | (A02_Women's Underwear Sets) |
| 3 | 0.070515 | (A03_Apples) |
| 4 | 0.043533 | (A03_Australian Imported Beefs - Shoulders) |
| ... | ... | ... |
| 8034 | 0.030042 | (A03_General Snacks, A03_Ramens, A03_Chicken E... |
| 8035 | 0.030042 | (A03_General Snacks, A03_Ramens, A03_Chicken E... |
| 8036 | 0.030459 | (A03_General Snacks, A03_Ramens, A03_Chicken E... |
| 8037 | 0.030181 | (A03_General Snacks, A03_Ramens, A03_Chicken E... |
| 8038 | 0.030459 | (A03_General Snacks, A03_Ramens, A03_Chicken E... |

Clnt_id를 기준으로 묶은 뒤 연관 분석 실행

알고리즘 적용

| | antecedents | | consequents | antecedent support | consequent support | support | confidence | lift |
|-------|-----------------------|---|-------------|--------------------|--------------------|----------|------------|----------|
| 7634 | (B02_Trash Bags) | (A03_Bibim Ramens, A03_Ramens) | | 0.246592 | 0.144784 | 0.036439 | 0.147772 | 1.020636 |
| 18273 | (A03_Ramens) | (B02_General Snacks, B02_Chicken Eggs) | | 0.281085 | 0.105285 | 0.030459 | 0.108362 | 1.029226 |
| 18279 | (A03_Ramens) | (B02_Tofu, B02_Chicken Eggs) | | 0.281085 | 0.111544 | 0.032545 | 0.115784 | 1.038016 |
| 18260 | (B02_Trash Bags) | (A03_Tofu, A03_Ramens) | | 0.246592 | 0.139777 | 0.035883 | 0.145516 | 1.041055 |
| 18248 | (B02_General Snacks) | (A03_Tofu, A03_Ramens) | | 0.232545 | 0.139777 | 0.034771 | 0.149522 | 1.069711 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 20695 | (B02_Bar Ice Creams) | (B02_Trash Bags, B02_Tube Ice Creams) | | 0.103755 | 0.048540 | 0.034492 | 0.332440 | 6.848829 |
| 21997 | (B02_Cone Ice Creams) | (B02_General Snacks, B02_Tube Ice Creams) | | 0.088317 | 0.052156 | 0.031711 | 0.359055 | 6.884283 |
| 20653 | (B02_Bar Ice Creams) | (B02_General Snacks, B02_Tube Ice Creams) | | 0.103755 | 0.052156 | 0.037274 | 0.359249 | 6.888007 |
| 20655 | (B02_Tube Ice Creams) | (B02_Bar Ice Creams, B02_General Snacks) | | 0.073713 | 0.071627 | 0.037274 | 0.505660 | 7.059608 |
| 21999 | (B02_Tube Ice Creams) | (B02_Cone Ice Creams, B02_General Snacks) | | 0.073713 | 0.060918 | 0.031711 | 0.430189 | 7.061773 |

9915 rows × 11 columns

support : 0.03 이상, confidence : 0.001 이상
- 최대한 많은 패턴을 보고싶어서 confidence값을 낮게함

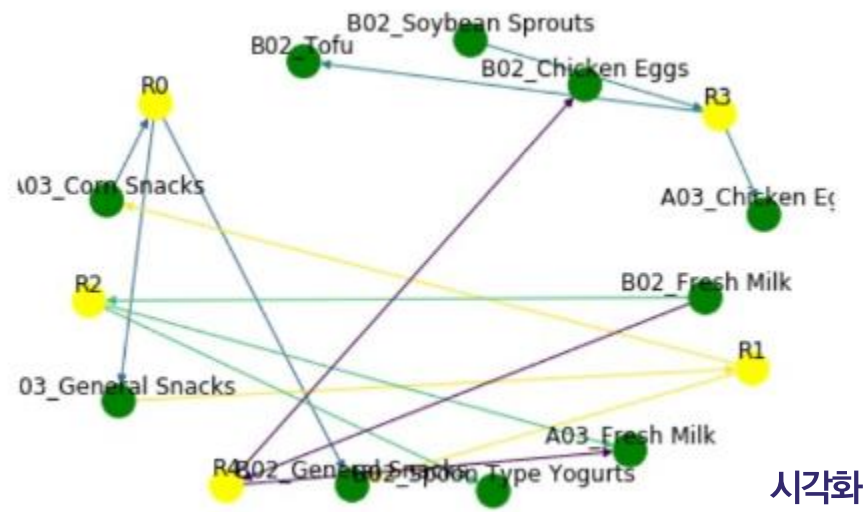
antecedents개수는 1개, consequents의 개수는 2개로 고정
- 전략 수립을 위해 그룹을 세분화함.

9915개의 패턴이 나왔으며 이 중에서 같은 biz_unit간 거래를 제외하면 396개의 패턴이 나옴.

분석 결과

| | antecedents_x | consequents_x |
|-------|-----------------------|--|
| 14433 | (A03_Corn Snacks) | (B02_General Snacks, A03_General Snacks) |
| 14432 | (A03_General Snacks) | (B02_General Snacks, A03_Corn Snacks) |
| 16328 | (B02_Fresh Milk) | (B02_Spoon Type Yogurts, A03_Fresh Milk) |
| 12386 | (B02_Soybean Sprouts) | (A03_Chicken Eggs, B02_Tofu) |
| 16310 | (B02_Fresh Milk) | (B02_Chicken Eggs, A03_Fresh Milk) |

- 연관 분석 결과 중 업종간 거래가 있는 패턴만 추출
- 특정 업종에서 거래를 한 고객이 다른 업종에서는 어느 상품을 구매하는지 알아냄



시각화

A speech bubble with a dashed outline and a tail pointing towards the bottom left. Inside the bubble, the word "BIGTATO" is written in a dark blue, hand-drawn, sans-serif font.

BIGTATO

PART.03

비즈니스 모델 및 전략

비즈니스 모델 및 전략

타업종 쿠폰 및 프로모션 제공

- 분석 결과, A03 업종에서 ComSnacks 상품을 구매한 고객이 B03 업종 GeneralSnacks 상품을 구매하는 형태의 패턴을 발견함
- A03 업종에서 ComSnacks 상품을 구매한 고객이 모두 B03 업종 GeneralSnacks 상품을 구매하는 것은 아니므로, A03 업종에서 ComSnacks 상품을 구매한 고객은 B03 업종 GeneralSnacks 상품을 구매하는 잠재고객이 될 수 있다.

비즈니스 모델 및 전략

| | antecedents_x | consequents_x |
|-------|-----------------------|--|
| 14433 | (A03_Corn Snacks) | (B02_General Snacks, A03_General Snacks) |
| 14432 | (A03_General Snacks) | (B02_General Snacks, A03_Corn Snacks) |
| 16328 | (B02_Fresh Milk) | (B02_Spoon Type Yogurts, A03_Fresh Milk) |
| 12386 | (B02_Soybean Sprouts) | (A03_Chicken Eggs, B02_Tofu) |
| 16310 | (B02_Fresh Milk) | (B02_Chicken Eggs, A03_Fresh Milk) |

A03 이용고객에게 B02 쿠폰 및 프로모션 정보 제공

B02 이용고객에게 A03 쿠폰 및 프로모션 정보 제공

기대 효과

A 업종에서 X 상품을 구매한 고객에게 B 업종의 Y 상품 쿠폰 및 프로모션을 제공한다면, 타업종으로의 유입이 예상 됨

연관규칙 마이닝의 결과로 나온 패턴을 바탕으로 타업종 간 쿠폰 및 프로모션 제공 가능



부록

변수 설명

| 변수명 | 설명 |
|-------------------|--|
| df_trans | 거래 정보 csv 데이터 프레임 |
| df_prodt | 상품 분류 정보 csv 데이터 프레임 |
| trans_Poten | biz_unit이 2개 이상인 고객만 나와있는 데이터 프레임 |
| Biz_nm3 | trans_poten데이터에서 pd_c기준으로 df_prodt와 merge하여 소분류명을 뽑아낸 데이터 프레임 |
| df_poten | biz_unit 열과 clac_nm3열을 합쳐서 새로 만든 데이터 프레임(ex. A03에서 onions를 구매한 고객: A03_onions) |
| lst | Biz_nm3 데이터를 clnt_id로 그룹핑하여 clnt_id별 구매 품목을 담은 리스트 |
| frequent_itemsets | apriori 연관분석 결과 |
| arules | frequent_itemsets의 패턴 분석 |
| result | antecedents의 개수가 1개 이상이고 consequents의 개수가 2개 이상인 패턴들의 규칙 |
| result_ | 같은 biz_unit에서 거래된 패턴 제외한 규칙 |
| final | result_ 상에서는 상품명이 제외된 상태이므로 index를 기준으로 merge하여 상품명이 명시된 최종 데이터 프레임 |