문제해결과 기업성과 달성을 위한 정보시스템 활용

고객 가치 분석(Customer Value Analytics)

SNU Business School

고객가치분석(Customer Value Analytics)

고객가치분석이란?

· 고객가치분석(Customer value analytics): 고객에 관한 데이터를 활용하여, 시장 세분화나 예측 시스템 등을 통해 중요한 비즈니스 의사 결정을 지원하기 위한 과정
[source: https://en.wikipedia.org/wiki/Customer_analytics]

- 고객가치분석을 통해 얻어진 정보는 마케팅, 상권분석 및 입지선정(Site selection), 고객관계관리(Customer Relationship Management) 등을 위해 폭넓게 쓰임

- 고객가치분석의 핵심은 구매(purchase), 이탈(churn) 등 고객의 행동을 예측하는 데에 있으며 유통, 금융 등 다양한 분야에 수익성을 높이고 비용을 절감하기 위해 활용됨

고객가치분석 사례 – 이탈 예측

· 통신사 고객들의 이탈 예측(churn prediction) 사례
[source: https://www.experfy.com/blog/building-predictive-models-for-customer-churn-in-telecom]

- 배경: 통신사들이 제공하는 서비스는 크게 구별되지 않으며 번호이동도 쉽게 가능해 고객 충성도(customer loyalty)를 유지하는 것이 상당히 어려움. 이에 따라, 고객들의 이탈 여부를 예측하는 것이 마케팅에 있어 상당히 중요한 정보가 될 수 있음.

- 목적: 가까운 미래에 이탈할 가능성이 매우 높은 고객들을 찾아내는 것

- 활용한 데이터(변수): 인구통계학적 정보(나이, 성별, 결혼여부, 거주지 등), 통화 기록, 사용하는 서비스, 불만사항 접수 기록, 신용정보 등

- 의사결정나무 알고리즘(Decision Tree Algorithm)을 활용하여 분류한 결과

		예측	
		이탈	미이탈
실제	이탈	30,219	1,535
	미이탈	20,685	247,561

고객가치분석 사례 – 이탈 예측

· 통신사 고객들의 이탈 예측(churn prediction) 사례
[source: https://www.experfy.com/blog/building-predictive-models-for-customer-churn-in-telecom]

- 결과 해석

● 전반적인 정확도(accuracy)는 상당히 높은 92.6%임

● 이탈 예측 알고리즘이 이탈할 것이라고 예측한 고객 50,904명 중 30,219명이 실제로 이탈하였음

● 이탈 예측 알고리즘이 이탈하지 않을 것이라고 예측한 고객 249,096명 중 247,561명이 실제로 이탈하지 않았음

● 즉, 이탈할 것으로 예측했는데 실제로는 이탈하지 않은 제2종 오류가 상대적으로 높게 나타남

● 고객의 이탈에 중요하게 나타난 변수는 나이(age), 연봉(salary), 데이터 서비스 사용량(data usage)로 밝혀짐

- 이와 같이, 데이터를 통해 얻어낸 정보는 마케팅이나 고객관계관리 등에 유용하게 활용될 수 있음

● 이탈할 것으로 예상되는 고객들에게 프로모션을 제공

● 이탈하지 않을 것으로 예상되는 고객에게 업셀(upsell) 시도

고객가치분석 사례 – 이탈 예측

· 통신사 고객들의 이탈 예측(churn prediction) 사례

- 수업 때 학생들에게 아래 영상을 보여주시면 이해가 빠를 것 같습니다

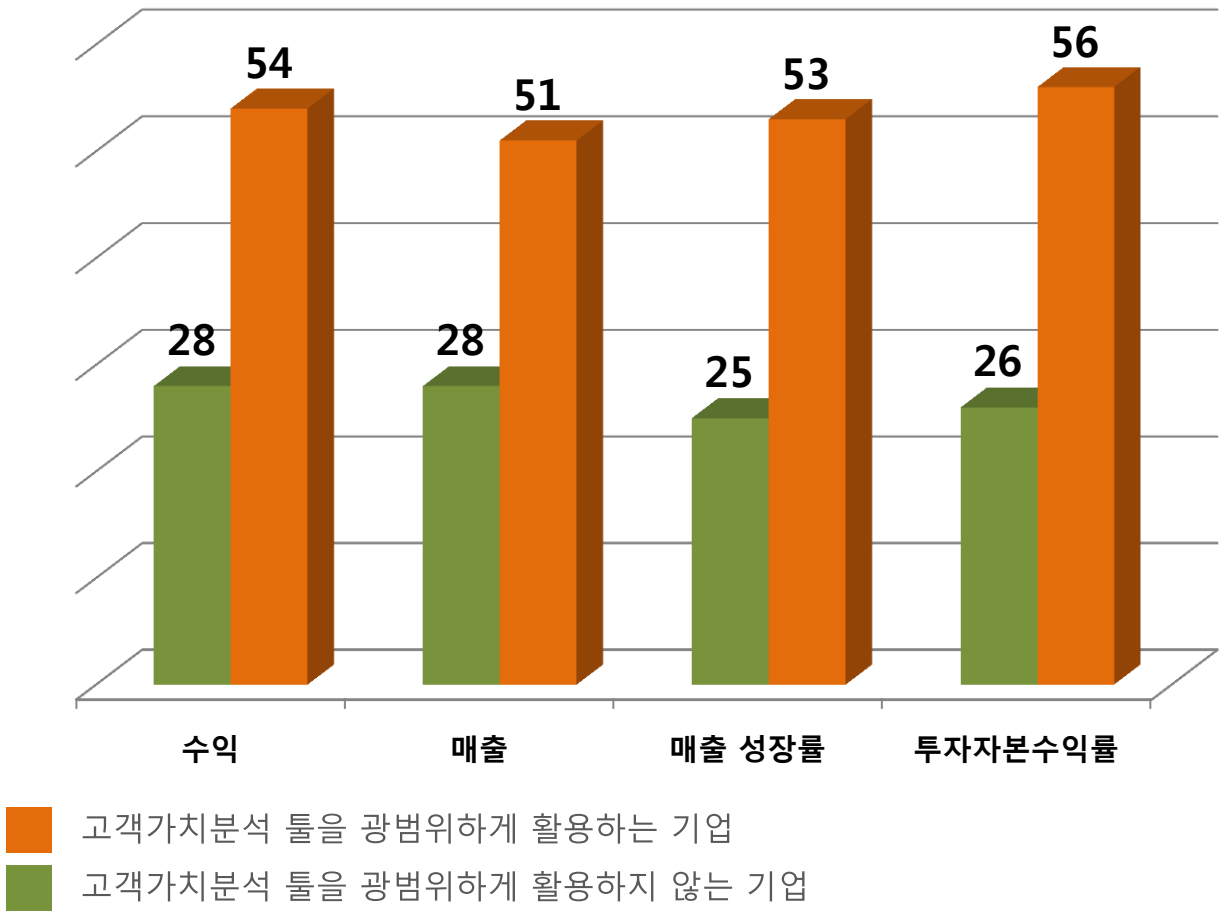
- VIDEO: Telecom Churn Prediction Model

- link: https://www.youtube.com/watch?v=LbXJHzy_7Qk

고객가치분석의 중요성

· 고객분석을 폭넓게 활용하는 기업들은 높은 성과를 자랑한다
[source: http://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/why-customer-analytics-matter]

- 고객가치분석 툴을 광범위하게 활용하지 않는 기업보다 광범위하게 활용하는 기업 중에 경쟁우위에 있는 기업의 비율이 유의미하게 높음



고객가치분석의 중요성

· IBM의 customer predictive analytics 서비스

- **수업 때 학생들에게 아래 영상을 보여주시면 이해가 빠를 것 같습니다**

- VIDEO: Use predictive analytics to anticipate your customers' life changes

- link: https://www.youtube.com/watch?v=hRySquvfeeM

추천 시스템(Recommender Systems)

추천 시스템이란?

· 추천 시스템(Recommender system) [source: https://en.wikipedia.org/wiki/Recommender_system]

- 정보 필터링 시스템(Information filering systems)의 일종으로, 특정 사용자가 특정 제품이나 서비스에 대해 가지는 선호도(preference)를 예측하기 위한 시스템

- 최근 10년 간 비약적으로 성장해 요식, 금융서비스, 보험, 온라인 데이팅, 소셜 미디어, 뉴스, 영화, 음악 등 다방면에서 추천 시스템이 개발되고 있음

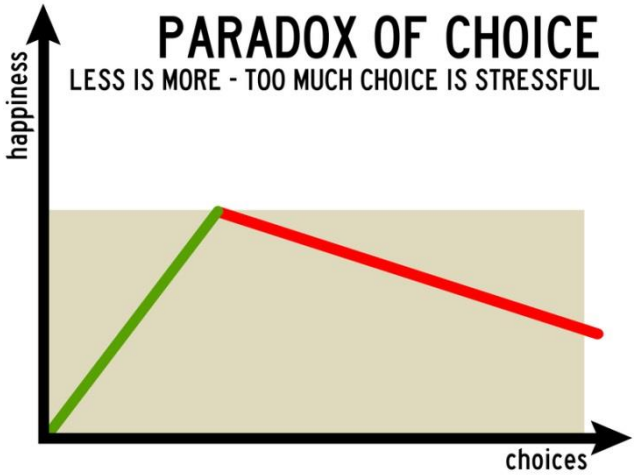
- 기계학습, 텍스트 분석 등 관련 기술이 발달하면서 추천 시스템도 고도로 다각화되고 정교화되고 있는 추세임

추천 시스템이란?

· 추천 시스템의 중요성

- 아마존의 CEO 제프 베조스는 *"우리는 제품을 파는 것이 아니라 고객의 구매 의사결정을 도와줘서 돈을 번다"*라고 할 정도로 추천시스템은 전자상거래(e-commerce)에서 핵심적인 역할을 하고 있음

- 선택의 가짓수가 너무 많아져 오히려 선택의 효용이 떨어지는 "선택의 역설(the paradox of choice)" 현상이 일어나는 현대 사회에 의사결정을 도와주는 핵심적인 도구가 되고 있음[Shcwartz 2004]



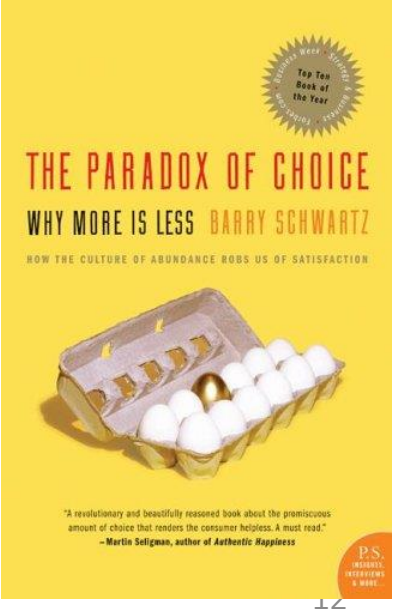
추천 시스템이란?

· 선택의 역설(Paradox of Choice): 현대 사회에서는 선택의 자유(freedom of choice)가 행복을 증진시킨다고 알려져 있지만, 선택의 가짓수가 어느 정도 이상 증가할 경우 오히려 사람들이 선택을 못하고 선택의 효용이 감소함 [Schwartz 04]

- 고객들에게 30가지 종류의 잼(jam)을 선택하게 했을 때보다 6가지 종류의 선택지를 주었을 때 더 구매율이 높았다

- 학생들에게 추가적인 점수(extra credit)을 주기 위한 에세이를 작성할 때 30개의 토픽을 준 경우보다 6개의 토픽을 준 경우 제출한 학생 수가 더 많았고 에세이의 질(quality)도 더 높았다

- 암 환자가 아닌 응답자들에게 치료 방식을 선택할 권리를 원하냐고 묻자 65%가 그렇다고 응답한 반면, 실제로 투병중인 암 환자들에게 물어본 결과 12%만 선택권을 원한다고 응답했다



추천 시스템이란?

· 추천 시스템의 중요성[Jannach and Friedrich 2013]

- 소비자 측면

● 흥미로운 제품을 손쉽게 찾아준다

● 선택의 가짓수를 줄임으로써 선택을 용이하게 한다

● 선택 사항들을 탐색하는데 도움을 준다

● 이전에는 찾을 수 없던 새로운 옵션을 발견하게 한다

추천 시스템이란?

· 추천 시스템의 중요성[Jannach and Friedrich 2013]

- 공급자 측면

● 소비자들에게 개인화된 서비스(personalization)를 제공한다

● 고객 충성도와 신뢰도를 제고한다

● 매출, 고객 유지율(retention) 등을 높인다

● 소비자들에 대한 이해를 제고한다

● 개인화된 마케팅 및 프로모션이 가능하다

추천 시스템의 분류

· 추천 시스템의 핵심은 소비자의 취향을 ‘예측’하는 메커니즘으로, 이러한 메커니즘에는 크게 세 가지 방식이 있음[Jafarkarimi1 et al 2012]

- 협업 필터링(Collaborative filtering) 방식

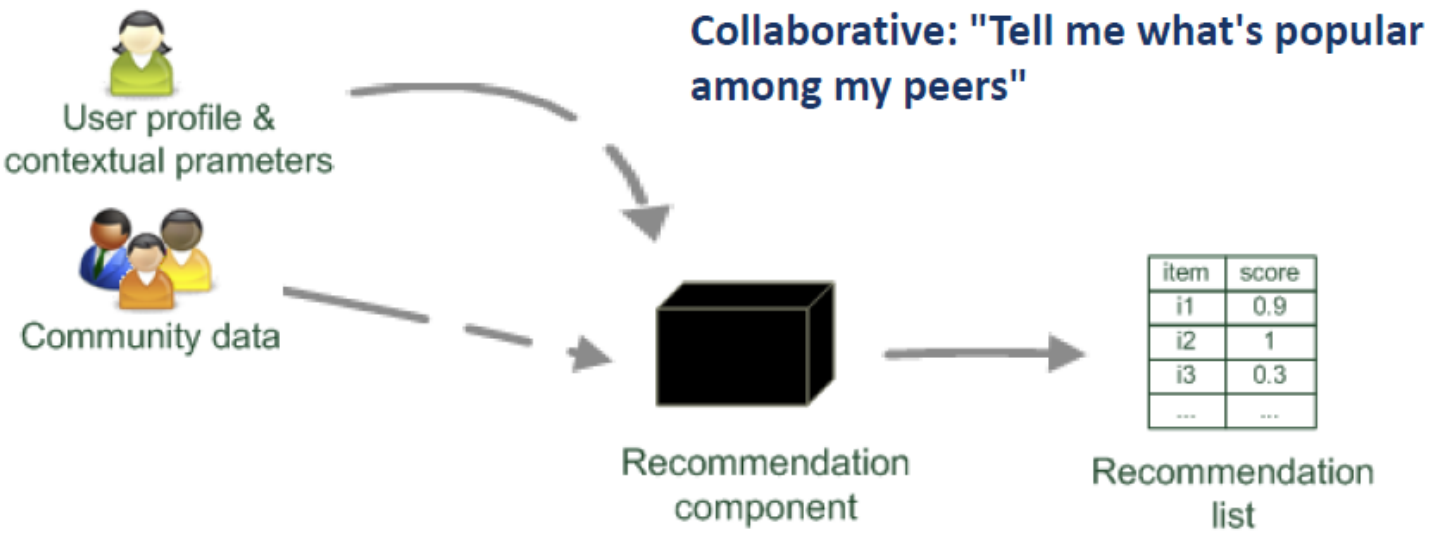
- 컨텐츠 기반 필터링(Contents-based filtering) 방식

- 하이브리드(Hybrid) 방식: 협업 필터랑과 컨텐츠 기반 필터링 방식의 혼합

추천 시스템의 분류

· 협업 필터링(Collaborative filtering): 특정 사용자의 구매 기록, 과거 행위 등을 바탕으로 유사한 사용자 층(user group)을 밝혀내고, 이를 바탕으로 그 사용자가 선호할 만한 제품이나 서비스를 추천
[source: https://en.wikipedia.org/wiki/Recommender_system, Jannach and Friedrich 2013]

- "Tell me what's popular among my peers"



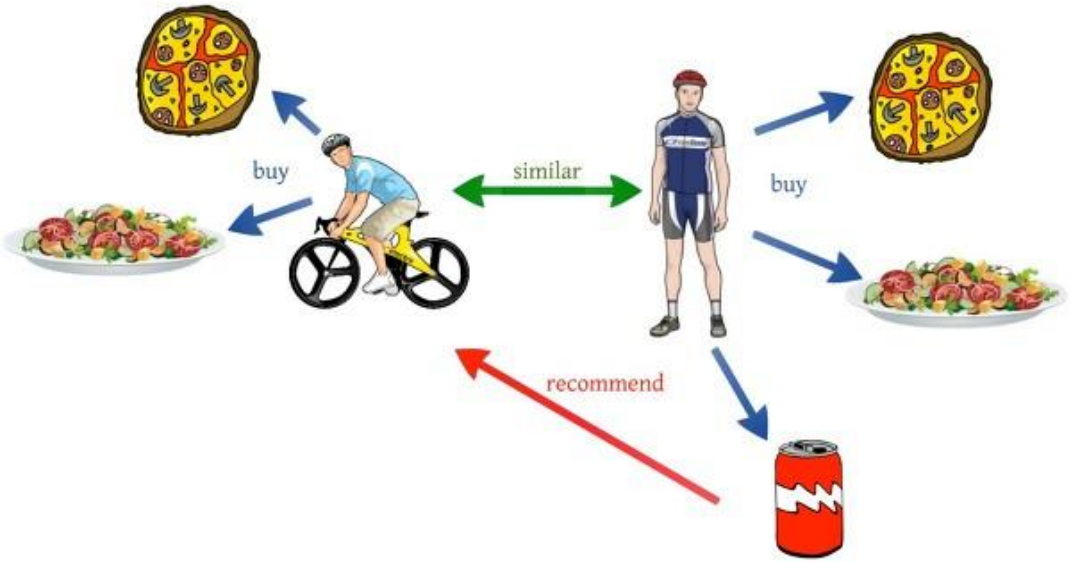
추천 시스템의 분류

· 협업 필터링(Collaborative filtering): 특정 사용자의 구매 기록, 과거 행위 등을 바탕으로 유사한 사용자 층(user group)을 밝혀내고, 이를 바탕으로 그 사용자가 선호할 만한 제품이나 서비스를 추천
[source: https://en.wikipedia.org/wiki/Recommender_system, Janach and Friedrich 2013]

- **"Tell me what's popular among my peers"**

- 협업 필터링 방식의 가장 큰 장점은 "유사성(similarity)"를 바탕으로 추천해주므로 특정 제품이나 서비스의 컨텐츠에 대한 심도 있는 이해가 필요 없다는 것임

- 예를 들어, 아래 두 사용자가 피자와 샐러드를 구매하며 과거에 유사한 구매 패턴을 보였다고 하자. 이 때, 오른쪽 사용자가 콜라를 구매하였으므로, 왼쪽 사용자에게 콜라를 추천해주는 시스템이 협업 필터링의 기초적인 메커니즘임



추천 시스템의 분류

· 아마존(Amazon.com)의 아이템 간 협업 필터링(Item-to-item collaborative filtering)[Linden et al 2003]

- 협업 필터링 중 가장 유명한 방식 중 하나로 아마존에서 고객들에게 제품을 추천해 줄 때 기반이 되는 메커니즘임

- "*People who buy X also buy Y*"

- 아이템 간 유사도 계산과 아이템 추천 크게 두 단계로 이루어짐

● 1단계 - (아이템 간) 유사도 계산: 모든 아이템 쌍 간의 유사도를 계산함. 유사도를 계산하는 방식은 상관계수(correlation), 벡터 간의 코사인(cosine) 등 다양한 방식을 택할 수 있음

● 2단계 - (아이템) 추천: 사용자가 과거에 평가한 기록을 바탕으로, 유사한 고객층을 찾아내고 그들이 높게 평가한 아이템 중 사용자가 사용하지 않은 아이템을 추천해줌

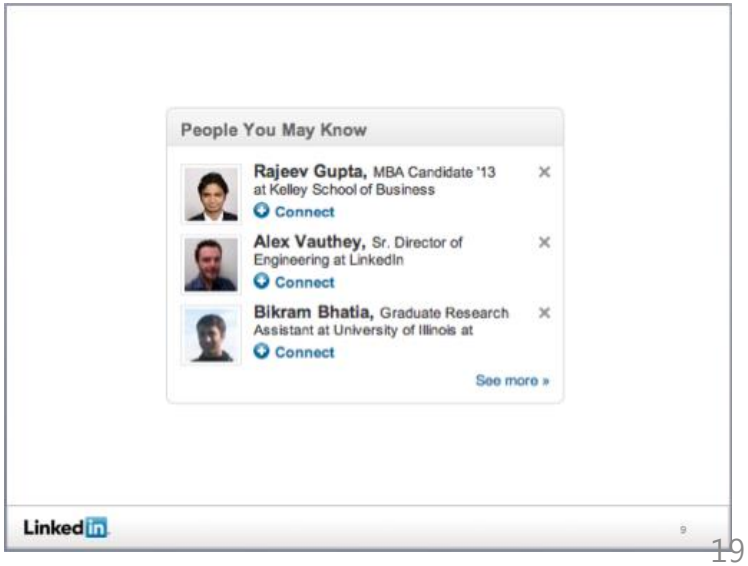
추천 시스템의 분류

· 협업 필터링 메커니즘의 활용 예시[source: https://en.wikipedia.org/wiki/Recommender_system]

- 페이스북, 마이스페이스, 링크드인 등 많은 소셜 네트워크 서비스(Social Network Service)는 협업 필터링을 활용해 사용자들에게 추천 서비스를 제공함

- 사용자들에게 새로운 친구, 그룹 등을 추천할 때 기존에 사용자가 가지고 있는 소셜 네트워크(Social network)를 분석해 협업 필터링을 거쳐 추천함

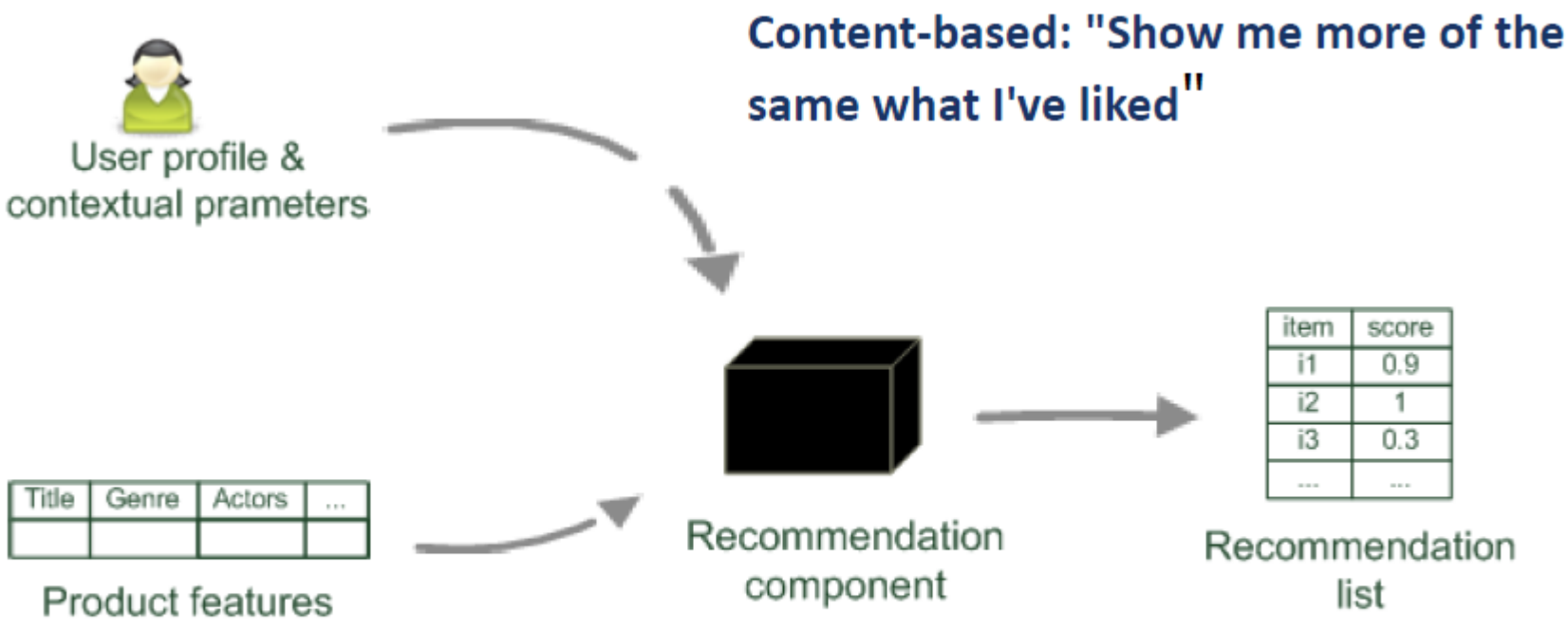
● 페이스북의 “You Might Also Like **THIS** because someone likes **THIS**”



추천 시스템의 분류

· 컨텐츠 기반 필터링(Content-based filtering): 사용자가 과거에 후한 평가를 내린 아이템과 유사한 속성을 가진 아이템들을 찾아 이들을 추천
[source: https://en.wikipedia.org/wiki/Recommender_system, Jannach and Friedrich 2013]

- “Show me more of the same what I’ve liked”



추천 시스템의 분류

· 컨텐츠 기반 필터링(Content-based filtering): 사용자가 과거에 구매한 아이템과 유사한 속성을 가진 아이템들을 찾아 이들을 추천
[source: https://en.wikipedia.org/wiki/Recommender_system, Jannach and Friedrich 2013]

- **“Show me more of the same what I’ve liked”**

- 컨텐츠 기반 필터링 방식의 핵심은 각 아이템의 구별되는 특성(feature)를 뽑아내는 데에 있음

- 예를 들어, 아래 사용자가 코카콜라를 과거에 구매하였으면 이와 유사한 Duff 콜라를 추천해주는 시스템이 컨텐츠 기반 필터링의 기초적인 메커니즘임

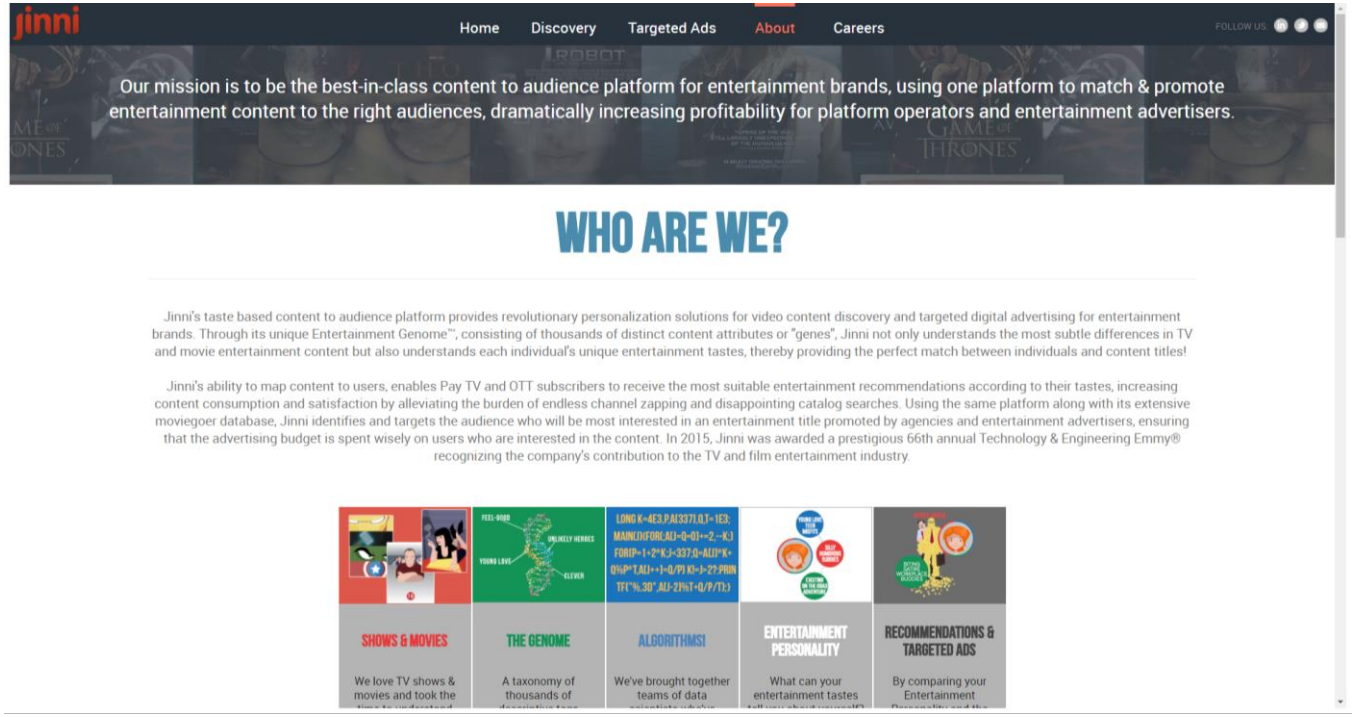
Example of content based recommendation

추천 시스템의 분류

· 컨텐츠 기반 필터링 메커니즘의 활용 예시: 지니닷컴(jinni.com)

- jinni는 컨텐츠를 기반으로 추천과 타겟 마케팅 서비스를 제공함

- 시청자들에 관한 방대한 데이터베이스를 가지고 영화와 TV 프로그램에 대한 추천 서비스를 제공해 시청자들 뿐만 아니라 배급자들에게도 높은 효용을 가져다 주고 있음

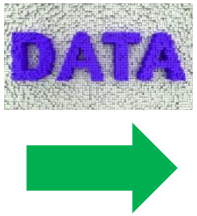


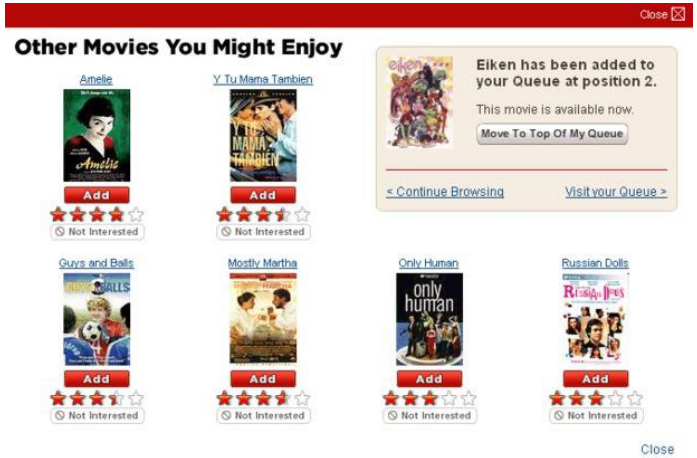
추천 시스템의 의의

· 이제는 매스 마케팅(Mass marketing)에서 1-1 마케팅(Personalized marketing)으로

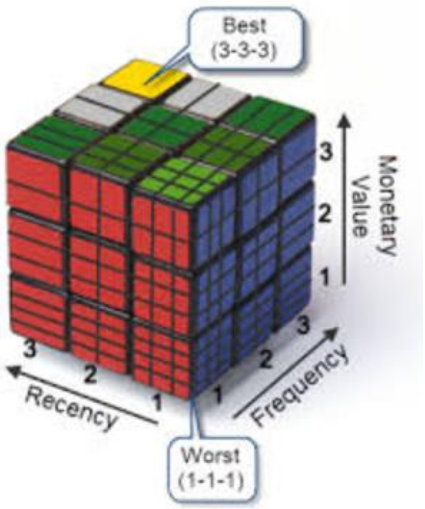
- 이전에는 일반 대중을 대상으로 한 무차별적인 마케팅이었다면, 이제는 각 개인에게 맞춤화 된 1-1 마케팅을 제공함으로써 비용을 절감하고 높은 효과를 기대할 수 있음

- 이 모든 것을 가능하게 한 것은 바로 즉시 현업에서 사용할 수 있게 마련된 “빅 데이터”와 이를 성공적으로 활용할 수 있도록 한 제반 기술임





RFM 모형(RFM Model)

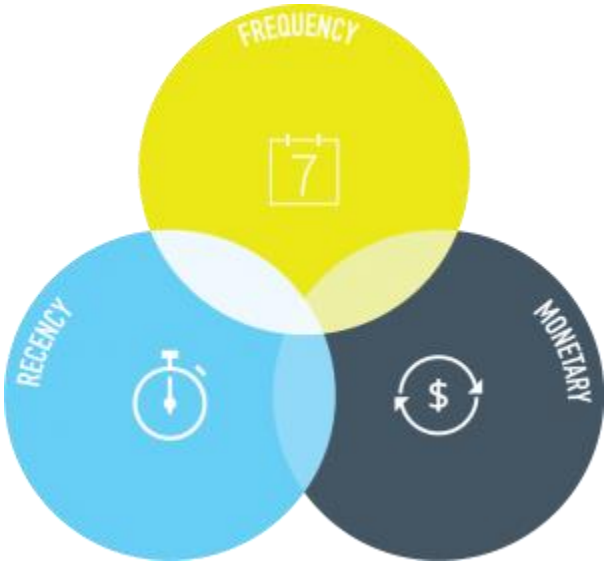


RFM 모형이란?

· RFM 모형[source: https://en.wikipedia.org/wiki/RFM_(customer_value)]

- 소비자 가치(customer value)를 진단하기 위한 간단한 소비자 데이터베이스(customer database) 모형으로 소매업과 전문 서비스 산업에서 널리 활용되고 있음

- 구매 최근성(Recency), 구매 빈도(Frequency), 구매 규모(Monetary Value) 세 가지 지표를 고려해 고객의 수익 기여도를 계산함



RFM 모형이란?

· RFM 모형[Bartocha 2015]

- 측정 가능하고 반복되는 특정 활동(activity)과 관련된 모든 데이터에 활용 가능함

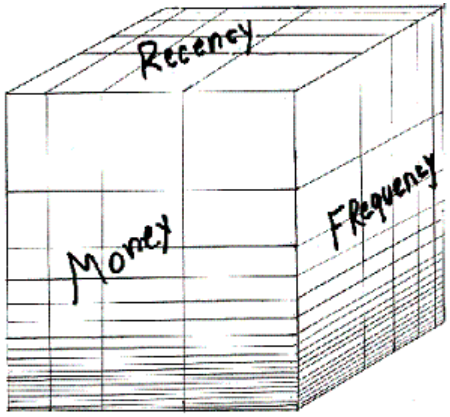
- RFM 모형 분석 적용 분야 예시

● 전자상거래(e-commerce): 상품 주문, 사이트 방문 등

● 소셜 미디어(Social media): 공유, 좋아요, 댓글 등

● 게임(games): 앱 내 구매(in-app purchases), 게임 접속 등

● 토론(discussion board): 포스팅, 투표(up-votes) 등



RFM 모형이란?

· 구매 최근성(Recency)

- 소비자가 최근에 언제 구매를 했는가(How recently did the customer purchase)?
- 예를 들어, 어떤 소비자의 구매 최근성은 아래와 같이 계산할 수 있음

● **"10 – 가장 최근에 구매한 날짜로부터 지난 개월 수"**와 **1** 중에 큰 값

● 만약 특정 소비자가 3달 전에 구입했다면, 그 소비자의 구매 최근성 지수는 다음과 같다

$$Recency = max(10 - 3, 1) = max(7, 1) = 7$$

RFM 모형이란?

· 구매 빈도(Frequency)

- 소비자가 얼마나 자주 구매를 하는가(How often do they purchase)?

- 예를 들어, 어떤 소비자의 구매 빈도는 아래와 같이 계산할 수 있음

● **"지난 12개월 동안 소비자가 구매한 횟수"**와 **1** 중에 큰 값

● 만약 특정 소비자가 지난 12개월 동안 어떤 상품을 9회 구입했다면, 그 소비자의 구매 빈도 지수는 다음과 같다

$$Frequency = max(9, 1) = 9$$

RFM 모형이란?

· 구매 규모(Monetary)

- 소비자가 얼마 만큼의 규모로 구매를 하는가(How much do they spend)?

- 예를 들어, 어떤 소비자의 구매 규모는 아래와 같이 계산할 수 있음

● **"특정 기간 동안 구매 중 가장 큰 지출액 규모"**

● 만약 특정 소비자가 어떤 상품을 3회 구입했는데 지출액이 300, 100, 20이었다면 그 소비자의 구매 규모 지수는 다음과 같다

$$Monetary = max(300, 100, 20) = 300$$

RFM 분석

· RFM 분석은 소비자를 세 가지 기준으로 세분화(segmentation)하고 세분화 기준에 따른 상이한 마케팅 전략을 수립하기 위해 수행됨

- 각 지수의 선형결합인 RFM 지수를 계산해 소비자 가치를 계산

$$RFM = \gamma_1 * Recency + \gamma_2 * Frequency + \gamma_3 * Monetary$$

● 각 γ_i는 각각의 지수에 할당된 가중치

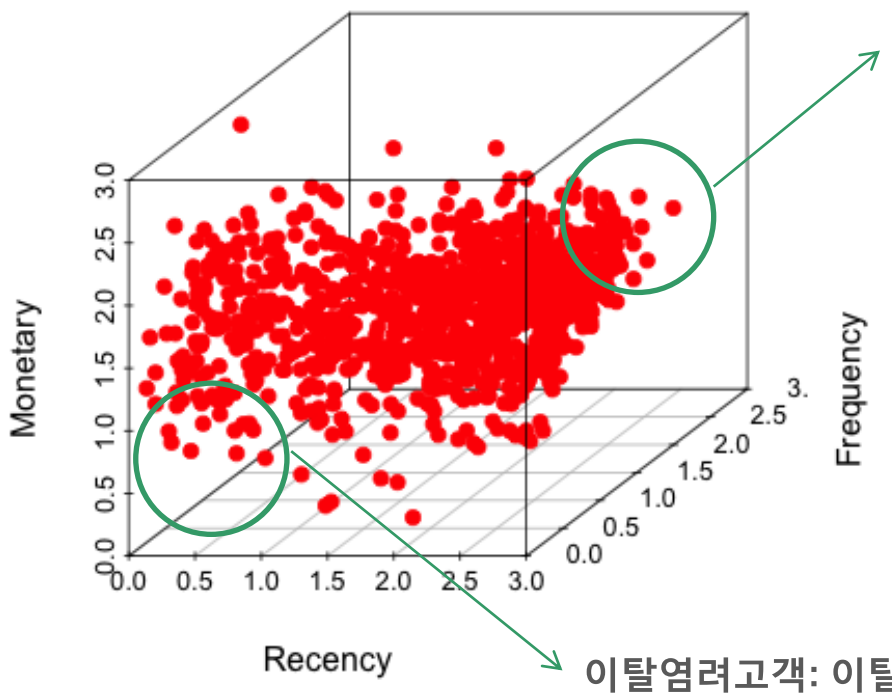
● 가중치의 값은 마케터가 정하기에 따라 달라짐. 똑같은 상품에 대해서도 시즌이나 경기 등 상황에 따라 달라질 수 있는 수치임

- 각각의 가중치가 상황에 따라 달라질 수 있듯이, R, F, M 지표는 앞서 예시를 든 것과 다른 방법(평균, 최대값, 최빈값 등)으로 계산될 수 있음

RFM 분석

· RFM 분석은 소비자를 세 가지 기준으로 세분화(segmentation)하고 세분화 기준에 따른 상이한 마케팅 전략을 수립하기 위해 수행됨

- 혹은, 세 가지 기준에 따라 구매성향을 분석함으로써 각기 다른 마케팅 전략을 펼침



우수고객: 업셀(upsell) 시도

이탈염려고객: 이탈방지를 위한 인센티브 및 프로모션 제공

RFM 분석

· RFM 분석은 소비자를 세 가지 기준으로 세분화(segmentation)하고 세분화 기준에 따른 상이한 마케팅 전략을 수립하기 위해 수행됨

- 단순히 RFM 모형만 수립하는 것이 아니라 RFM 모형을 통해 만들어진 새로운 정보를 가지고 군집 분석(clustering analysis), 회귀 분석(regression analysis) 등 데이터 마이닝 기법을 활용해 추가적인 분석을 수행해 인사이트를 도출할 수 있음

● 고객의 과거 구매 기록을 나타내는 R, F, M 지표를 독립 변수로 하고, 고객 응답률을 종속변수로 하여 미래의 고객 응답률을 예측하는 다중회귀모형을 수립할 수 있음

$$y_i = \beta + \alpha_1 R_i + \alpha_2 F_i + \alpha_3 M_i$$

RFM 분석결과를 활용한 고객 응답률 예측($\beta = 3, \boldsymbol{\alpha_1 = 10}, \boldsymbol{\alpha_2 = 8}, \boldsymbol{\alpha_3 = 5}$일때)

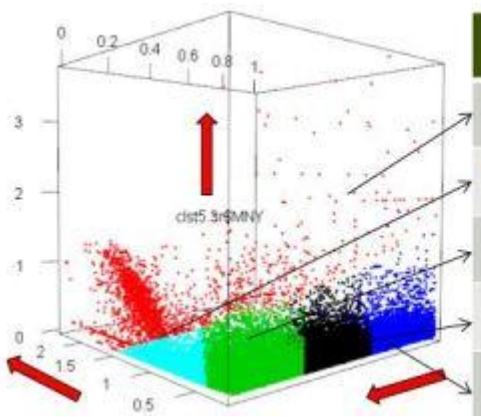
고객번호	R 점수	α_1 가중치	F 점수	α_2 가중치	M 점수	α_3 가중치	총점	예측
1	5	10	4	8	3	5	100	우량
2	5	10	4	8	4	5	105	우량
…								
100	1	10	2	8	1	5	31	불량

RFM 분석

· RFM 분석은 소비자를 세 가지 기준으로 세분화(segmentation)하고 세분화 기준에 따른 상이한 마케팅 전략을 수립하기 위해 수행됨

- 단순히 RFM 모형만 수립하는 것이 아니라 RFM 모형을 통해 만들어진 새로운 정보를 가지고 군집 분석(clustering analysis), 회귀 분석(regression analysis) 등 데이터 마이닝 기법을 활용해 추가적인 분석을 수행해 인사이트를 도출할 수 있음

● RFM 지표를 입력변수로 군집분석을 수행해 시장 세분화를 정교하게 할 수 있음



REC	FREQ	MNY	size	DESC
high	high	high	0.1%	VIP
high, mid	mid	mid, low	9.0%	Big client
high	low	mid, low	35.5%	Potential, active client
mid	low	mid, low	29.6%	Potential client
low	low	mid, low	25.8%	Normal client

RFM 분석

· RFM 모형을 활용한 소비자 세분화(Customer segmentation)[Bartocha 2015]

- 1단계: 각 고객의 RFM 지표 계산에 활용될 수 있는 데이터를 불러온다

고객 ID	Recency	Frequency	Monetary
1	53 days	3 transactions	$230
2	120 days	10 transactions	$900
3	10 days	1 transactions	$20

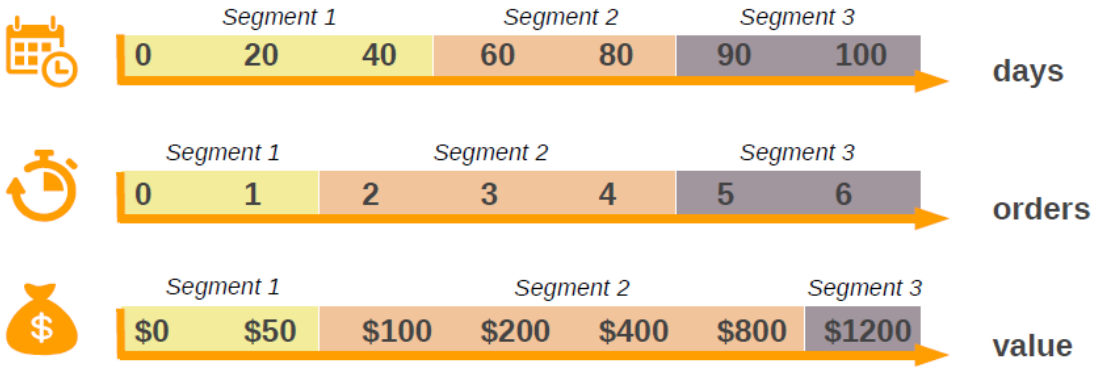
● 위와 같은 표를 RFM 테이블(RFM Table)이라 부름

● 계산에 포함시키는 데이터를 선택하는 기준은 상황에 따라 달라질 수 있음(예를 들어, 12개월 내/6개월 내/3개월 내의 거래 내역 포함)

RFM 분석

· RFM 모형을 활용한 소비자 세분화(Customer segmentation)[Bartocha 2015]

- 2단계: 데이터의 분포를 확인하고 세분화 기준을 정의한다



● 위의 예시는 R, F, M 각각의 차원을 세 개의 segment로 분류함

● segment의 개수는 정의하기에 따라 달라질 수 있음. 때로는 R, F, M 각각의 segment 개수를 달리 줄 수도 있음

● 가장 기초적인 방법은 4분위수(quantiles)를 이용하는 방법임 – 직관적이고 이해하기 쉬움

RFM 분석

· RFM 모형을 활용한 소비자 세분화(Customer segmentation)[Bartocha 2015]

- 3단계: 각 고객들을 세분화 기준에 따라 분류한다

고객 ID	Recency	Frequency	Monetary
1	2	2	2
2	3	3	2
3	1	1	1

● 이를 세분화된 RFM 테이블(Segmented RFM Table)이라 함

RFM 분석

· RFM 모형을 활용한 소비자 세분화(Customer segmentation)[Bartocha 2015]

- RFM 모형을 활용하기 위해서 일반적으로 3차원 정육면체를 만들어 시각화한다

● 단점이 있다면, 3차원 공간상의 도형을 사람이 인지하기 어렵다는 것!

● 인간의 뇌는 3차원 이상의 공간을 직관적으로 받아들이기 힘들어함

RFM 분석

· RFM 모형을 활용한 소비자 세분화(Customer segmentation)[Bartocha 2015]

- 이러한 단점을 보완하기 위하여, 소비자 세분화를 할 때 다중분할표(Stacked contingency table)을 활용할 수 있음

		M			
R	F	1	2	3	4
1	1				
	2				
	3				
	4				
2	1				
	2				
	3				
	4				
3	1				
	2				
	3				
	4				

● R, F, M 지표로 다중분할표를 만들어 각 분할에 속하는 소비자 명수와 통계량 등을 쉽게 계산할 수 있음

RFM 분석

· RFM 모형을 활용한 소비자 세분화(Customer segmentation)[Bartocha 2015]

- Recency 지표를 활용해 소비자의 구매 활성도를 점검해볼 수 있음

		M			
R	F	1	2	3	4
1	1				
	2				
	3				
	4				
2	1				
	2				
	3				
	4				
3	1				
	2				
	3				
	4				

ACTIVE(활발)

AT RISK(위험군)

CHURNED(이탈)

● 이탈 위험 고객을 판별해낼 수 있음

RFM 분석

· RFM 모형을 활용한 소비자 세분화(Customer segmentation)[Bartocha 2015]

- Frequency와 Monetary 지표를 활용해 고객의 장기적 가치(Long-term value)를 계산할 수 있음

		M			
R	F	1	2	3	4
1	1			SILVER	SILVER
	2		SILVER	SILVER	GOLD
	3	SILVER	SILVER	GOLD	GOLD
	4	SILVER	GOLD	GOLD	PREMIUM

● F, M 지표에 따라 프리미엄(Premium), 골드(Gold), 실버(Silver) 등으로 높은 가치의 고객들을 찾아낼 수 있음

RFM 분석

· RFM 분석 활용 예시[Birant 2011]

- 터키에 있는 온라인 스포츠 상점 데이터베이스를 활용함

- 고객(customers)과 상품(products), 그리고 고객의 주문(orders) 데이터를 수집해 이를 RFM 모형에 맞게 코딩함

Customers

CID	Age	Sex	State	Education	Marital S.	Child	Year	...
5	Teenager	M	Aegean	Middle	NeverM	0	4	...
8	Adult	M	Marmara	HighSchool	Married	0	3	...
19	Adult	F	BlackSea	HighSchool	Married	3+	4	...
...	...	...	...	...	...	...	...	...

Products

PID	PName	Price	Brand	Group	Type	Color	Sex	...
100	NK DRI FIT PO	42	Nike	TShirt	Running	NK10	Male	...
106	PM AIKI JR	81	Puma	Sneaker	Soccer	PM03	Child	...
110	AD MALV OH	125	Adidas	Jersey	Soccer	AD05	Male	...
...	...	...	...	...	...	...	...	...

Orders

TID	PID	CID	Date	Quantity	Discount	Total	Type	...
T1	106	19	2008.12.2	1	0	81	SS	...
T2	100	8	2008.12.2	1	0	42	YS	...
T3	110	5	2008.12.3	1	0	125	SS	...
...	...	...	...	...	...	...	...	...

Table 3. An example data from customers, products and orders tables

CID	Recency (Day)	Frequency (Number)	Monetary (TL)	R	F	M	RFM
5	95	4	237	4	3	4	434
8	269	10	790	5	5	5	555
19	321	1	81	1	1	2	112
...	...	...	...	...	...	...	...

1-5점 스케일로 변환한 각각의 R, F, M 지수

RFM 분석

· RFM 분석 활용 예시[Birant 2011]

- Customers 테이블: 각 고객에 대한 인구통계학적인 정보 포함

● CID: 고객 ID (각 고객에게 하나씩 부여된 unique한 primary key)

● Age: 나이를 4가지 클래스로 분류(child 0-12; teenager 13-19; adult 20-59; senior 60+)

● Sex: 남성(M) 혹은 여성(F)

● State: 거주지

● Education: 최종 학력

● Marital S.: 결혼 여부

● Child: 슬하 자녀(0, 1, 2, 3+)

● Year: 가입 연수

Customers

CID	Age	Sex	State	Education	Marital S.	Child	Year	...
5	Teenager	M	Aegean	Middle	NeverM	0	4	...
8	Adult	M	Marmara	HighSchool	Married	0	3	...
19	Adult	F	BlackSea	HighSchool	Married	3+	4	...
...	...	...	...	...	...	...	...	...

RFM 분석

· RFM 분석 활용 예시[Birant 2011]

- Products 테이블: 각 제품에 대한 설명 포함

● PID: 제품 ID (각 제품에 하나씩 부여된 unique한 primary key)

● PName: 제품 이름

● Price: 제품의 가격

● Brand: 제품 브랜드(Nike, Adidas 등)

● Group: 제품 종류(티셔츠, 저지, 스니커즈 등)

● Type: 용도(런닝, 축구 등)

● Color: 색상

● Sex: 성별(남성용, 여성용, 아동용 등)

Products

PID	PName	Price	Brand	Group	Type	Color	Sex	...
100	NK DRI FIT PO	42	Nike	TShirt	Running	NK10	Male	...
106	PM AIKI JR	81	Puma	Sneaker	Soccer	PM03	Child	...
110	AD MALV OH	125	Adidas	Jersey	Soccer	AD05	Male	...
...	...	...	...	...	...	...	...	...

RFM 분석

· RFM 분석 활용 예시[Birant 2011]

- Orders 테이블: 각 주문(order)에 대한 설명 포함

● TID: 주문 ID (각 주문에 하나씩 부여된 unique한 primary key)

● PID: 주문된 제품 ID

● CID: 주문한 고객 ID

● Date: 주문한 날짜

● Quantity: 주문 수량

● Discount: 할인 여부

● Total: 총 가격

Orders

TID	PID	CID	Date	Quantity	Discount	Total	Type	...
T1	106	19	2008.12.2	1	0	81	SS	...
T2	100	8	2008.12.2	1	0	42	YS	...
T3	110	5	2008.12.3	1	0	125	SS	...
...	...	...	...	...	...	...	...	...

RFM 분석

· RFM 분석 활용 예시[Birant 2011]

- RFM 지표 계산

● 구매 최근성(R): 현재 시간과 가장 최근에 구매한 날짜(last transaction date) 사이의 시간 간격으로 계산

● 구매 빈도(F): 특정 기간 동안 소비자가 구매한 횟수(number of transactions)

● 구매 규모(M): 특정 기간 동안 소비자의 총 구매액(total money spent)

- 예를 들어, 아래 표에서 고객 5(CID = 5)는 가장 최근에 구입한 날짜가 95일 전(Recency)이며, 특정 기간 동안 4회 구입했으며(Frequency), 총 구입액(Monetary)은 237임

● 이를 내부적인 기준에 맞추어 R, F, M 지수로 변환한 결과가 **4, 3, 4** 임

CID	Recency (Day)	Frequency (Number)	Monetary (TL)	R	F	M	RFM
5	95	4	237	4	3	4	434
8	269	10	790	5	5	5	555
19	321	1	81	1	1	2	112
…	…	…	…	…	…	…	…

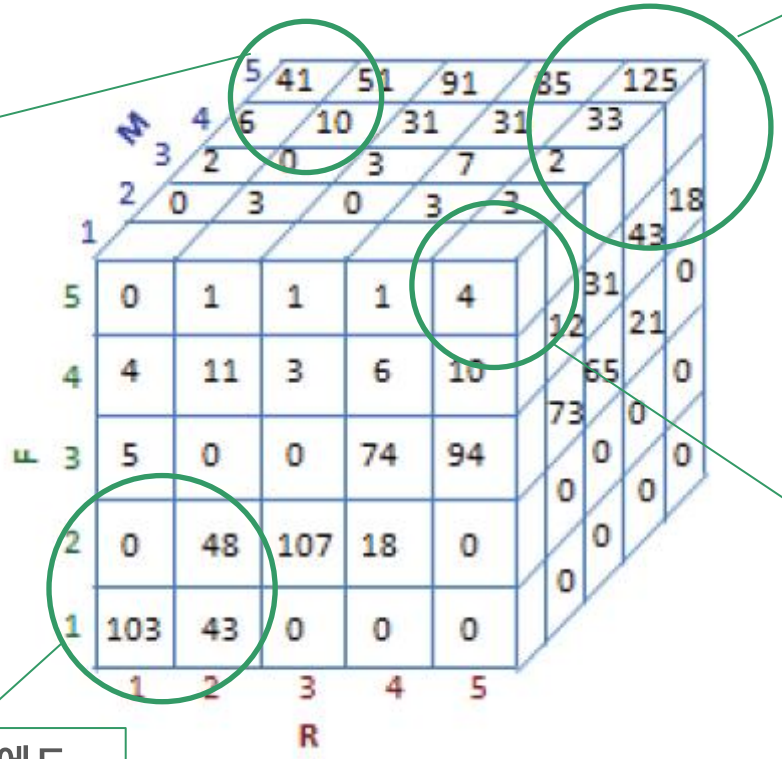
RFM 분석

· RFM 분석 활용 예시[Birant 2011]

- RFM 분포도: 125가지의 경우의 수가 존재함

우수고객: 최근에, 자주, 그리고 큰 규모로 구매함

(과거의) 우수고객: 예전에는 자주, 그리고 큰 규모로 구매했으나 최근에는 구입한 기록이 없음



최근에 자주 구매하나 큰 규모로 구매하지는 않음

이탈염려고객: 과거에도 소규모로 가끔 구매했으며, 최근에 구매한 적도 없음

RFM 분석

· RFM 분석 활용 예시[Birant 2011]

- *k*-군집화 알고리즘(*k*-means clustering)을 통해 8개의 군집으로 분류(*k*=8)

	Cluster	Size	Recency (Avg.)		Frequency (Avg.)		Monetary (Avg.)		RFM Pattern	Customer Type
			Day	R	#	F	TL	M		
군집1	C1	309	65.4	4.57	6.28	4.89	485.1	4.79	R↑F↑M↑	Best
군집2	C2	392	83.5	4.32	1.52	3.44	146.8	3.42	R↑F↑M↑	Valuable
군집3	C3	415	75.1	4.44	1.18	3.05	70.1	1.49	R↑F↑M↓	Shopper
군집4	C4	300	202.4	2.86	1.01	2.02	69.5	1.47	R↑F↓M↓	FirstTime
군집5	C5	425	247.8	2.22	4.27	4.51	387.4	4.67	R↓F↑M↑	Churn
군집6	C6	135	325.8	1.38	2.26	3.76	137.5	2.94	R↓F↑M↓	Frequent
군집7	C7	381	290.1	1.86	1.00	1.41	138.1	3.33	R↓F↓M↑	Spenders
군집8	C8	309	339.1	1.35	1.00	1.00	69.5	1.53	R↓F↓M↓	Uncertain
	Overall	2666		2.85		3.01		2.95		

RFM 분석

· RFM 분석 활용 예시[Birant 2011]

- 소비자의 미래 행동을 예측하기 위한 분류 규칙(classification rule)

Rule 1: **if** *State*=Aegean **and** *EducationLevel*=Bachelors **and** *MaritalStatus*=Married **and** *Gender*=M **then** R↑F↑M↑

Rule 2: **if** *State*=Aegean **and** *MaritalStatus*=Married **and** *Gender*=M **and** *Age*=Adult **then** R↑F↑M↑

Rule 3: **if** *State*=Marmara **and** *Membership*=3 **and** *EducationLevel*=HighSchool **and** *Children*=0 **then** R↓F↑M↑

Rule 4: **if** *State*=CentralAnatolia **and** *Age*=Teenager **and** *MaritalStatus*=NeverMarried **and** *Membership*=3 **then** R↑F↑M↓

Rule 5: **if** *State*=SouthEasternAnatolia **and** *Children*=3+ **and** *Gender*=M **then** R↓F↑M↓

Rule 6: **if** *State*=Mediterranean **and** *EducationLevel*=Middle **then** R↓F↓M↑

Rule 7: **if** *State*=EasternAnatolia **and** *Gender*=F **then** R↑F↓M↓

Rule 8: **if** *State*=BlackSea **and** *Children*=3+ **and** *Gender*=F **then** R↓F↓M↓

RFM 분석

· RFM 분석 활용 예시[Birant 2011]

- 소비자의 미래 행동을 예측하기 위한 분류 규칙(classification rule)

● **규칙1:** Aegean주에 살고, 최종학력이 학사(Bachelor)이고, 결혼하였으며, 성별이 남성이면 우수고객일 가능성이 높음(R, F, M 지표가 모두 높음)

● **규칙2:** Aegean주에 살고, 결혼하였으며, 성별이 남성이고, 성인이면 우수고객일 가능성이 높음(R, F, M 지표가 모두 높음)

● **규칙3:** Marmara주에 살고, 멤버십 등급이 3등급이며, 최종학력이 고교(High school)이며, 자식이 없으면 R 지표는 낮지만 F, M 지표는 높을 가능성이 높음

● **규칙4:** Central Anatolia주에 살고, 10대(teenager)이며, 결혼하지 않았고 멤버십 등급이 3등급이면 R,F 지표는 높지만 M 지표는 낮을 가능성이 높음

● **규칙5**: South Eastern Anatolia주에 살고, 자식이 3명 이상이고 남성이면 R,M 지표는 낮지만 F 지표는 높을 가능성이 높음

● **규칙6:** Mediterranean 주에 살고, 최종 학력이 중학교이면 R, F 지수는 낮지만 M 지표는 높을 가능성이 높음

● **규칙7:** Eastern Anatolia 주에 살고, 여성이면 R 지표는 높지만 F, M 지표는 낮을 가능성이 높음

● **규칙8:** Black Sea 주에 살고, 자식이 3명 이상이며 여성이면 R, F, M 지표가 모두 낮을 가능성이 높음

RFM 분석

· RFM 분석 활용 예시[Birant 2011]

- 상품 추천(product recommendation) 규칙

Rule 1: {C1, Adidas soccer jersey (man), Adidas soccer jersey (woman)} → {Adidas soccer jersey (child)}
Rule 2: {M>3, Adidas Sneaker (child)} → {Adidas Socks, Adidas Equipment Bag}
Rule 3: {C3, Adidas Soccer ball} → {Adidas Backpack (unisex), Adidas Water Bottles}
Rule 4: {C4, Adidas Soccer ball} → {Nike of-kick}
Rule 5: {C5, Converse Sneaker (woman), Puma Sneaker (man)} → {Nike Cap (unisex)}
Rule 6: {C6, Adidas T-Shirt (male)} → {Adidas Short (male), Adidas Training Bag}
Rule 7: {R<=3, F<=3, M>3} → {Reebok Sneakers}
Rule 8: {R<=3, F<=3, M<=3} → {Converse Shoes}

RFM 분석

· RFM 분석 활용 예시[Birant 2011]

- 상품 추천(product recommendation) 규칙

● **규칙1:** 군집1(C1)에 속하고, 남/여용 아디다스 축구 저지를 구입한 고객 → **아동용 아디다스 축구 저지** 추천

● **규칙2:** M 지표가 3보다 크고, 아동용 아디다스 스니커즈를 구입한 고객 → **아디다스 양말과 아디다스 이퀍먼트 가방**을 추천

● **규칙3:** 군집3(C3)에 속하고, 아디다스 축구공을 구입한 고객 → **아디다스 백팩과 아디다스 물병**을 추천

● **규칙4:** 군집4(C4)에 속하고, 아디다스 축구공을 구입한 고객 → **나이키 킥 신발**을 추천

● **규칙5**: 군집5(C5)에 속하고, 여성용 컨버스 스니커즈와 남성용 푸마 스니커즈를 구입한 고객 → **나이키 모자**를 추천

● **규칙6:** 군집6(C6)에 속하고, 남성용 아디다스 티셔츠를 구입한 고객 → **남성용 아디다스 반바지와 아디다스 트레이닝 가방**을 추천

● **규칙7:** R지표가 3 이하, F 지표가 3 이하, M 지표가 4 이상인 고객 → **리복 스니커즈**를 추천

● **규칙8:** R 지표가 3 이하, F 지표가 3 이하, M 지표가 3 이하인 고객 → **컨버스 신발**을 추천

RFM 분석의 의의

· 근거 기반의(evidence-based) 의사결정 가능

- 이전에는 단순히 마케터, 혹은 영업 사원의 직감과 '촉'에 기반하여 마케팅 및 영업 전략을 펼쳤다면, 이제는 데이터를 바탕으로 충분한 근거 있는 의사결정이 가능해짐

- 객관적이고 일관된 마케팅 및 영업 전략 수립 가능

● 지속적으로 수행된다면 이를 통해 다른 경쟁자들에 비해 경쟁우위(competitive advantage) 형성 가능

References

· Bartocha, K. (2015). RFM Customer Segmentation. https://www.slideshare.net/WhiteRavenPL/rfm-segmentation

· Birant, D. (2011). Data Mining Using RFM Analysis. In Knowledge-oriented applications in data mining. InTech..

· Jafarkarimi, H., Sim, A. T. H., & Saadatdoost, R. (2012). A naive recommendation model for large databases. International Journal of Information and Education Technology, 2(3), 216.

· Jannach, D., Friedrich, G. (2013). Tutorial: Recommender systems. International Joint Conference on Artificial Intelligence, Beijing, August 2013.

· Linden, G., Smith, B., & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. IEEE Internet computing, 7(1), 76-80.

· Schwartz, B. (2004). The paradox of choice.