

1. Introduction to Data Mining

금융 데이터마이닝
2022 Summer
김아현

데이터 마이닝 개요

◎데이터 마이닝 개요

• 데이터 마이닝이란?

▫ 데이터 마이닝 정의

- ♦ 대량의 데이터 집합으로부터 유용한 정보를 추출하는 것 (Hand et al., 2001)
- ♦ 의미있는 패턴과 규칙을 발견하기 위해서 자동화되거나 반자동화된 도구를 이용하여, 대량의 데이터를 탐색하고 분석하는 과정(Berry and Linoff, 2000)
- ♦ 통계 및 수학적 기술 뿐만 아니라 패턴인식 기술들을 이용하여 데이터 저장소에 저장된 대용량의 데이터를 조사함으로써 의미있는 새로운 상관관계, 패턴, 추세 등을 발견하는 과정 (Gartner Group, 2004)

▫ 데이터 마이닝 특징

- ♦ 대용량의 관측 가능한 자료를 다룸.
- ♦ 컴퓨터 중심의 기법으로서 경험적 방법을 중시함.
- ♦ 데이터 탐색, 분석 및 예측을 위한 다양한 통계기법과 더불어, 복잡성이 높은 데이터를 예측하기 위한 머신러닝 이론이 주로 적용됨.

데이터 마이닝 개요

◎데이터 마이닝 개요

• 데이터 마이닝 기법

- 통계 및 머신러닝 분야의 다양한 기법이 활용되고 있음.
- ♦ 데이터를 탐색하고 모델을 구축하는 다양한 기법들은 전통적인 통계학에서 오랜 시간 연구되어 왔던 주제이기도 함.
- ♦ 그러나 동일한 모델이라도, 통계학에서는 모델에 적용되는 가정으로 응용이 제한적이고, 가설에 의거하여 검증하는 방식임에 반해, 데이터마이닝에 적용되는 기법은 가정과 제약이 적고, 예측 성능의 우수성을 더 중요하게 본다는 차이가 있음.
- ♦ 최근, 데이터의 용량이 커지고, 계산 능력이 확보됨에 따라 데이터마이닝을 위해 다양한 머신러닝 기법이 주로 활용되고 있음.
- ♦ 데이터마이닝은 미래에 대한 정확한 예측 뿐 아니라 데이터에 대한 이해를 궁극적인 목표로 하고 있음. 따라서, 보다 해석 가능한 주제에 더욱 초점을 맞추고 있다는 점에서 일반적인 머신러닝 모델링과 차이가 있음.

$$y \approx f(x)$$

$$y = \underbrace{\alpha + \beta x}_{\text{단순 } f(x)} + \epsilon \rightarrow \text{가정 : } \begin{cases} \beta = 0 \\ \beta \neq 0 \end{cases}$$
$$y = f(x) + \epsilon \rightarrow \text{가정 X.}$$

데이터 마이닝 개요

◎데이터 마이닝 개요

• 데이터마이닝 수행 단계



- **목적 결정** : 프로젝트의 목적을 계획하고 설정
- **데이터 수집** : 무작위로 추출하거나 전수조사
- **데이터 탐색 및 정제** : 분석에 적절한 형태로 데이터를 가공하고 점검하는 단계
- **데이터 마이닝 방법 결정** : 데이터마이닝 문제 및 기법을 선택
- **모형 선택** : 데이터마이닝 프로세스의 여러 단계를 반복적으로 수행하여 가장 좋은 모형을 찾는 단계
- **성능평가** : 구축된 모형의 일반화 성능을 평가.
- **적용** : 구축된 모형을 운영 시스템에 탑재하여 실제 의사결정에 적용하는 단계.

데이터 마이닝 개요

◎데이터 마이닝 개요

• 데이터 마이닝 활용 분야

- 고객관계관리 (CRM) : 목표마케팅, 고객세분화, 고객성향 변동분석, 장바구니분석 등
- 신용평가 : 불량채권과 대손 추정하여 이를 최소화
- 품질개선 : 불량품을 찾고, 그 원인을 밝혀서 이를 궁극적으로 예방
- 부정거래 적발 : 고도의 사기행위를 발견할 수 있는 패턴을 알아내는 것
- 이미지분석 : 디지털화된 사진으로부터 패턴을 추출하는 기법
- 이커머스: 추천 시스템, 재고 관리, 수요 예측.

금융과 데이터 마이닝

데이터 마이닝 + 마케팅

◎금융업계에서의 데이터 마이닝 활용사례

• 신용평가

▣ 미국의 벤처 기업 Sofi (Social Finance)

- ♦ 2011년 스탠포드 대학원생 4명이 시작한 스타트업.
- ♦ 졸업생 중 성공한 동문으로부터 자금을 조달하여, 장래 고소득일 가능성이 높은 재학생 대상으로 온라인 P2P 방식으로 낮은 금리의 학자금 대출을 제공하고, 채무의 조기상환이 가능하도록 하는 서비스를 제공.
- ♦ Sofi는 학자금 대출에서 간과되었던 패턴을 찾아내는 알고리즘을 개발. 대출 이자는 낮지만, 신용보강비용 및 채무 불이행에 따른 비용이 절감되어 양호한 수익을 달성할 수 있었음.
- ♦ 고객 충성도를 높이고 재이용률을 높이는 전략으로 주목받음.



금융과 데이터 마이닝

◎금융업계에서의 데이터 마이닝 활용사례

• 신용평가

- 중국 앤트파이낸셜 그룹의 '즈마신용'
 - ♦ 알리바바 그룹이 보유한 대량의 개인 데이터를 바탕으로 다양한 지표의 조합으로 신용평점을 산출하여 평가함.
 - 연령, 학력, 직업 뿐 아니라, 알리페이의 결제이력, 공공요금 지불이력, 인터넷 쇼핑몰 구입이력, 공유경제 서비스에서의 평가, 제테크 상품 가입현황, SNS 교류관계, 취미생활 등
 - ♦ 데이터 제공 시 신용평점이 높아지기 때문에 이용자는 적극적으로 개인 데이터를 제공하고 있으며, 신용도가 하락하지 않도록 하여 개인 건전성을 높이는 효과가 있음.



구글 이미지



芝麻信用是合法独立的信用评估及信用管理机构，除法律法规另有规定或经你授权的情况下，我们不会向他人透露你的信用相关的隐私信息。



Company Data, BNK투자증권

금융과 데이터 마이닝

◎금융업계에서의 데이터 마이닝 활용사례

• 보험 상품 및 서비스

▣ 자동차 보험 관련

- ◆ 계약자의 자동차에 센서를 탑재하여 운전 상황을 모니터링하고, 그 데이터로부터 개인 별로 상이한 보험료를 책정함.
- ◆ 자동차 사고가 발생 시, 스마트폰으로 사고현장을 촬영해서 보내면 신속하게 손해정도를 판단해주는 시스템을 구축.

손상심도 인식결과				예상수리비 산출		
				20200318007970 48하1076 <div>사진견적아력</div>		
				부품 소계	공임 소계	사진견적 금액 합계
				87,900	215,829	303,729
				2020-04-17 결과값수정 사진보기		
				작업 내용	작업 시간	부품금액 공임금액
☑ 프론트범퍼 교환				도장 1.43		40,600원 47,480원
☑ 프론트범퍼커버				교환 2.93		0원 97,280원
☑ 커버 어셈블리 - 프론트 범퍼				1		87,900원 0원
☑ 라디에이터그릴				탈착 0.27		0원 8,960원
☑ 라디에이터그릴상부커버				탈착 0.17		0원 5,640원
☑ 가열건조비				도장 0		0원 15,860원
손상부품명	수리 클래스	손상_결과_위치	손상_결과	보험개발원 제공		
프론트범퍼	교환					

금융과 데이터 마이닝

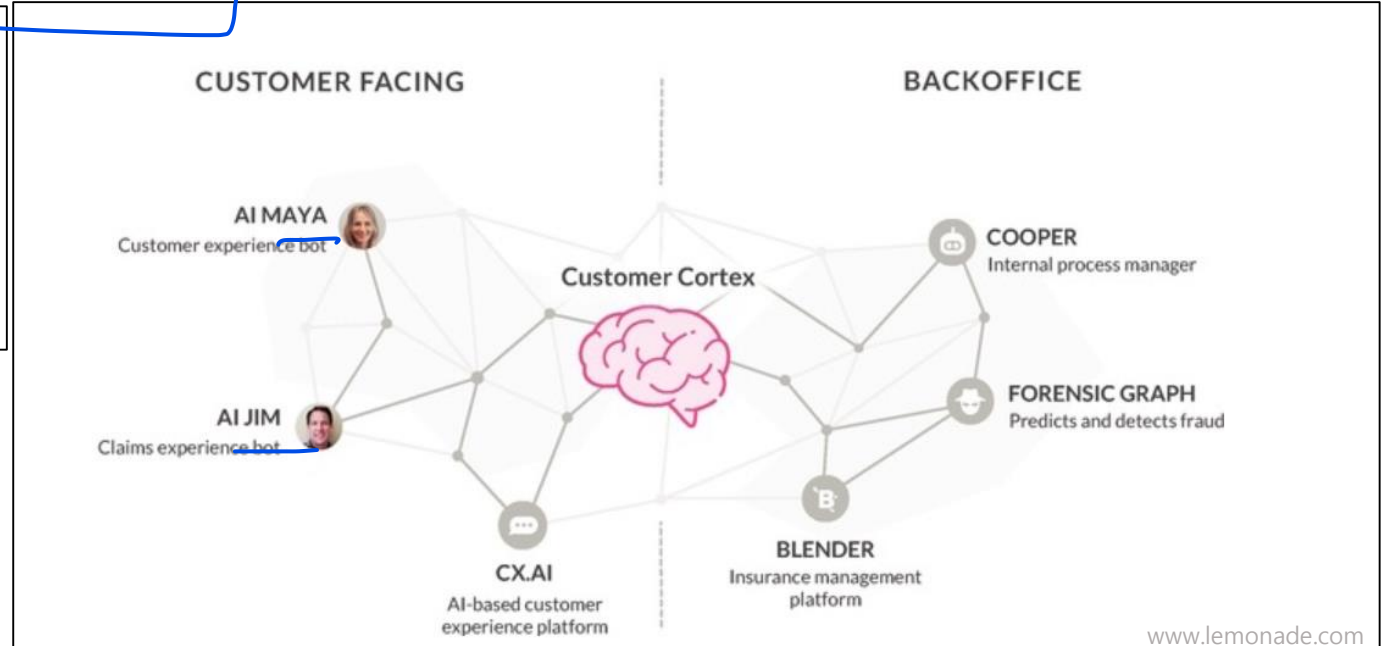
◎금융업계에서의 데이터 마이닝 활용사례

• 보험 상품 및 서비스

▫ 미국의 온라인 P2P 보험사 레모네이드(Lemonade) 사

◆ 주택소유자/임차인 보험

- 모바일 및 인공지능 기술을 적극 활용하여 가입 절차부터 보험료 지급까지의 프로세스를 획기적으로 단축시키고 비용을 절감함.
- 인공지능 챗봇이 보험 전반 업무와 고객 상담을 담당.
- 보험사기를 예측, 탐지, 차단하는 '포렌식 그래프(Forensic Graph)'를 구축.



금융과 데이터 마이닝

◎금융업계에서의 데이터 마이닝 활용사례

• 투자 및 자산관리

▫ American Century Investment

- ◆ 자연어처리와 감성분석 등이 가능한 인공지능 기반 로보 애널리스트를 활용해 기업들의 실적 컨퍼런스 콜 내용을 분석.
- ◆ 경영진 목소리, 감정, 뉘앙스 등의 데이터를 분석하여, 실적 과장이나 의도적인 복잡한 설명, 외부변수 핑계 여부 등을 판별함.

▫ JPMorgan Asset Management

- ◆ 인공지능 기반 분석모델 'Themebot'을 통해 비정형 데이터에서 키워드를 분석하여 투자테마와 종목을 발굴하는데 적용.



금융과 데이터 마이닝

◎금융업계에서의 데이터 마이닝 활용사례

• 이상금융거래 탐지

▣ 미국의 PayPal

- ◆ 글로벌 온라인 결제 시스템 업체
- ◆ 2001년 세계 최초로 자체적으로 FDS를 개발
- ◆ 고객의 이상거래 판별을 위해 딥러닝을 적용해 자사 FDS 시스템을 고도화해서 활용
- ◆ 이상거래 패턴을 탐지하여 사고를 사전에 차단할 수 있는 '이상거래 관리 필터' 서비스 제공

▣ 피처스페이스(Featurespace)

- ◆ 카드 및 결제 산업을 위한 자동심층 행동 네트워크를 출시하여, 부정행위, 계정 탈취, 카드 및 결제 사기 등으로부터 소비자를 보호할 방어 수단을 제공
- ◆ 특징 탐색을 자동화하고, 순환신경망 알고리즘을 이용하여 거래 흐름의 시간을 맥락적으로 이해하여 행동을 예측. 기존 적응형 행동 분석의 성능을 한층 향상시킴.



FEATURE
SPACE

OUTSMART RISK

각 사 홈페이지