

Data Mining for CRM

한국데이터베이스학회
2001 춘계 Conference
과총회관

서울대학교 산업공학과
조성준

1

목차

- DM 및 CRM 소개
- DM of Biz Process
- DM 기법
- DM / CRM 효과 측정
- Case Studies
- 결론

서울대학교 산업공학과

조성준

2/72

Part I

DM 및 CRM 소개

3

Data Mining의 정의

- 대량의 데이터베이스로부터 탐색과 분석을 통하여 의미 있는 패턴이나 규칙을 찾아내는 과정
- 분류, 추정, 예측, 유사통합, 군집화, 기술
- 가설검정, 지식발견

서울대학교 산업공학과

조성준

4/72

Data Mining의 분야

- 생명과학, 공학
- 비즈니스
 - CRM, DB 마케팅
 - Finance

CRM의 배경

- 경쟁 기업 간 치열한 경쟁
- 제품 품질/가격의 차이 미미
 - 가전, 이동통신, 자동차, 예금, 신용카드
- 서비스와 같은 + alpha 중요: “prodices”
- 고객의 value, needs, concern 파악 필요

CRM의 정의

- 정의
 - 고객 “개개인에” 대하여 이해하여
 - 이를 토대로 고객의 loyalty 제고하여
 - 장기적인 수익을 올리려는 경영 활동
- Catch Phrase
 - Listen and learn (cf. Tell and sell)
- 전사적 활동 : 마케팅+영업+서비스
- Listen 한 내용을 Learn 으로 변환

CRM 의 예

- 누가 남고, 누가 이탈할 것인가 ?
- 누구에게 어떤 제품을 판매할 것인가 ?
- 누가 응답할 것인가 ?
- 신설 지점의 위치는 ?
- 고객이 다음에 원할 제품/서비스는 무엇인가 ?

Motivation

- 고객과의 관계가 중요한 이유?
 - 비용: 신규 고객 유치 >> 기존 고객 유지
 - 5% 유지 증가 → 50-60% 수익 증가
 - 신규 고객 유치에 6-7배 이상의 비용 지출

Part II

DM of Biz Process

DM과 비즈니스 프로세스와의 통합

- 아이디어: DM 은 solo 활동이 아니다.
- 4단계 Cycle
 - 1) 문제 파악 : 자체 도출, 인터뷰
 - 2) 데이터마이닝 적용 : 정보 추출
 - 3) 정보에 따른 조치: 모든 비즈니스 프로세스에 피드백 요청
 - 4) 결과측정 : Goto 문제파악

통합 예

- 상황: 고객을 잃고있는 은행 : attrition
- 대책 1: 예대마진 -- 하여 “경쟁력 있는 상품” 제공
 - 비용 + , 기회주의적인 고객 모집
- 대책 2: 손해 보는 서비스 중단
 - 고정 고객이 주로 사용
- 대책 3: 데이터마이닝 적용

통합 예

- 1) 문제파악: “우수 고객의 이탈”
- 2) 데이터마이닝 적용: 고객 군집화
 - 군집 C의 특징:
 - 이탈율 +, 저녁시간에 전화 & 대기, 타행 ATM 사용
- 3) 조치: 군집 C 고객에게 특별 전화번호 부여
- 4) 결과측정: 이탈율 감소, 영업이익 증가 확인

DM 사례 1 (이동 통신)

- 신제품을 구매할 고객은 ?
- 유사제품에 대한 구매 데이터를 마이닝
- 랜덤 샘플링에 비하여 5배 높은 판매 성공률
- Repeat !

DM 사례 2 (자동차)

- 새 모델의 우편 판촉
- 누구에게 보낼 것인가 ?
- 1M명 주소 리스트 + PRIZM DB + Sales Data (Who bought it)
- 신경망 + Decision Tree 로 구매자 예측
- Target Marketing 에 사용 가능

DM 사례 3 (금융)

- BOA: 집 담보 용자 고객 후보 찾기
- 기존 가설: 대학생 자녀의 학비를 위해서
- 불안정한 수입을 조정하기 위해서
- 고객 데이터 분석 : 군집화
- 군집 C 고객 분석 => 창업을 위해서

5 cases

- Response Modeling – 통신판매회사
- Cross Selling Prediction – Online 은행
- Churn Prediction – 이동통신 회사
- Call Detail Understanding – 전화 회사
- Purchase Behavior Understanding – 슈퍼마켓

DM이 활성화된 원인

- 대량의 데이터 생성:
 - 통신, 신용카드, 슈퍼, ATM, 홈쇼핑, 인터넷쇼핑, 보험, 호텔, 증권, 생산, 항공
- 대량 데이터의 저장: Warehouse, mart
 - 다양한 소스 (판매, 예약, 클레임, 주문입고)의 데이터를 단일 포맷으로 저장

DM이 활성화된 원인

- Processing 능력:
 - Powerful DM algorithms 수행 가능
- 소프트웨어 패키지 출현:
 - SAS Eminer, SPSS Clementine,
 - IBM Iminer, Oracle Darwin
 - 사용 편의성, 데이터분석 능력

Part III

DM 기법

Database

- 신용카드 고객 DB
- 대출 심사 DB
 - 연령, 직업, 자택, 연봉, 기존 대출금, 신청 금액
- 이동 통신 고객 DB
 - 성명, 연령, 직업, 최근 사용금액, 단말기

Regression

- 예: 대출 심사 DB의 fields
- 독립, 종속 변수 vs 입력, 출력 변수
- 선형 모델: 예, 의미
- 계수/모수, 모수 추정 (vs 학습), 에러
- 학습 에러, 테스트 에러
- 적합, 과대 적합, 과소 적합, 모델 복잡도
- Directed vs. Undirected / Supervised vs Unsupervised

Classification (분류)

- 종속변수가 범주변수인 회귀분석
- 학습 데이터, 테스트 데이터, 일반화
- Decision Tree, CBR, Neural Network
- 예
 - 남녀 = $f(\text{연령, source, 영화명})$
 - 대출 여부 = $f(\text{개인 신상정보})$
 - 주가 상승 여부 = $f(\text{최근 주가 추이, 경제환경})$

Estimation (추정)

- 종속변수가 연속변수인 회귀분석
- 입력변수는 연속 또는 범주
- CBR, Neural Network
- 예
 - 연령 = $f(\text{남녀, source, 영화명})$
 - 마케팅 캠페인에 반응할 확률
 - 옵션 가격 결정 모형

Prediction (예측)

- 미래에 대한 분류 또는 추정
- MBA + 기존 기법
- 예
 - 새로 나오는 영화는 어떠한 사람들이 볼 것인가?
 - 특정 고객이 향후 6개월 내에 떠날 가능성
 - 자동차를 구매한 사람이 구매할 상품/서비스?
 - 향후 주가 = $f(\text{최근 주가 추이, 경제 환경})$

Affinity Grouping, MBA

- 유사하게 구매되는 물품 파악
- A Priori algorithm
- 예
 - 영화 A를 본 사람은 영화 B를 본다?
 - 슈퍼에서 함께 팔리는 물품은?
 - 시간적인 관련성 파악

Clustering (군집화)

- 유사 데이터들의 군집화
- Unsupervised, Exploratory Knowledge Discovery
- C-means, Competitive Learning
- 예
 - 유사한 관객/영화들의 군집화
 - Cross selling을 위한 grouping
 - 이동통신 고객들의 세분화 (target marketing)
 - 유사한 주식 시장 상황

서울대학교 산업공학과

조성준

27/72

Part IV

DM / CRM 효과 측정

28

목표 설정

- DM의 “일반적” 정의
 - 예) “흥미로운 패턴을 찾는다”
- 구체적인 목표 필요 (최고 경영자 설득)
 - 예) “Churn Prediction”
 - 예) 고객의 Direct Mail 반응을 예측

모델 측정

- Regression / MLP / DT 간의 예측 성능 비교
- 정확도
- 기술력
- Confidence
- Comprehensibility

모델 측정

- 정확도: 모델 전체 vs. single prediction
 - 1) Descriptive Model: 예) rules - MDL
 - 2) Predictive Model
 - training error, test error (오분류율, MSE)
(가정: 동일한 분포)
 - cross validation (w/ training data only)
 - confidence / support (MBA)
 - distance

결과 측정 (“lift”) -1

- 전체 고객 Population 의 5% 가 churn 가능성
- DM을 통하여 “churn 가능성이 높은 사람들” 샘플 구성
- 샘플 내에서 실제로 churn할 사람의 구성비 50%
- 정의) $lift = 50/5 = 10.$

결과 측정 (“lift”) -2

- 샘플의 크기
- Small size -> High Lift : good ?
- Small size -> 실제 churn customer를 많이 miss

결과 측정 (“lift”) -3

- Type I, II error Tradeoff
- 한 가지 에러 감소는 다른 에러 증가
- Churn/No Churn (실제, 분류)
 - Miss = False Negative
 - False Alarm = False Positive
- Sample
 - high Lift + small Size = high FN
 - low Lift + large Size = high FP

영향력 측정 -1

- Bottom Line:
 - DM과 그에 따른 campaign의 ROI
- DM: Business Enterprise, not Scientific one
- Q) churn 후보들 중에 어느 선에서 끊나?
- Q) Mktg: Mass vs. Target (w/ DM)

영향력 측정 -2

- 비용과 수익 고려
- 고정비용: DM 비용, campaign 실시 비용
- 변동비용: special offer 비용 / customer
 - 예) “10 시간 무료 사용”
- 수익: 고객의 평생가치 (Lifetime value)
 - 특정고객과의 향후 모든 transaction에서 발생하는 수익(-비용)의 현재 할인액수

영향력 측정 -3

- 변동비용이 상대적으로 높으면
 - churn 가능성이 매우 높은 극소수의 고객에 대한 campaign
- 고정비용이 상대적으로 높으면
 - 대다수의 고객에 대한 campaign

ROI sales

- Top Management 설득에 필수적
- 고객 차별화
 - Transaction buyer vs Relation Buyer
 - RFM 분석 후 고수익 고객 중심
- Churn 방지 시 연 수익 계산
 - 카드 사례
- DM 비용 + 캠페인 비용 계산

Part VI

결론

39

Data Mining for CRM

- Data Mining
 - 기존의 AI, DB, Statistics 의 기법
 - “새로운 비즈니스” 문제에 적용
- CRM 성공의 필수 요소
 - Backend (cf. Frontend : call center, 영업, 서비스) 기능

Biz Tips

- DM-CRM 잘하는 방법
 - Business 문제 이해 (domain knowledge)
 - 입력 변수 선정 및 압축: 성패 90% 관건
 - 데이터 입수 및 청소: 70% 시간 소요
 - One ear to expert, one ear to data
- DM-CRM 프로젝트 시작하는 방법
 - ROI term 으로 대화 !!!

Q & A

조성준

zoon@snu.ac.kr