Framework for Data-Driven Marketing Analytics

서울대학교 경영대학
송인성 교수

Problem-Driven vs. Data-Driven Marketing

Problem-Driven Market Research

- 시장 조사 프로젝트는 대부분 특정 문제 해결에 초점 (예: 브랜드 XYZ의 시장점유율 감소. 그 이유는? 대책은?)
- 컨텐츠: 고객 니즈, 선호도, 태도 등 서베이 자료
- 주로 stated preference data에 의존
- Macro 마케팅에 초점

Data-Driven Analytics

- 데이터가 비즈니스를 드라이브: 비즈니스 기회 발굴; 신규 가치 창출; 신규 고객 확보 등
- 자료 소스: transaction data (은행, 신용카드, 이동전화, 포털, 항공 등), syndicated services (Nielsen, IRI, IMS 등),
- revealed preference data 활용
- Addressable consumers(개별 고객의 ID 파악, 고객 히스토리 확보) - Macro 및 Micro 마케팅 모두 가능

Data Driven Marketing Analytics

- Data–Driven Marketing : 데이터가 마케팅 활동의 근간
  - 데이터가 마케팅 value 창출의 근원

- 마케팅 데이터의 종류
  - 고객 태도 자료(stated preference data)
  - 고객 행동 자료(revealed preference data)
  - 빅데이터시대

- 분석
  - 데이터에 기반한 고객 구매 행동 분석
  - 마케팅 프로그램 평가(How profitable are my promotions?)
  - 최적 마케팅 프로그램 탐색(What price should I charge?)
  - Target marketing: 개별 고객별 특화된 마케팅 활동을 통해 최적의 가치 창출/획득

Data-Driven Changes in Standard Consumer Research

- Audit/share report: 자세한 시장 점유율 정보가 며칠 내로 획득
- 테스트 마케팅: 필드에서 이루어지는 standard test marketing
- 시장 세분화: 인구통계에서 행동으로 무게 중심 이동(ex. Kraft 치즈를 사는 사람은 어떤 사람이며 또한 그들은 어떤 제품을 주로 구매하는가?)
- 광고 테스트: “rented” 테스트 마켓, split cable
- Demand for accountability: “pay for performance”, 마케팅 ROI

Mass vs. Micro-Marketing

- mass marketing에서 micro-marketing으로
- revealed preference 기반으로 고객 파악 및 잠재 고객 발굴
- personalization, customization을 통한 제품 차별화(보험 등의 금융 상품), 가격 차별화, 제품 구색 차별화
- 광고 메시지 차별화 (split cable), targeted banner advertising (Doubleclick)

Customer Data as the Source of Business Opportunities

- 고객 데이터베이스가 신규 사업의 기회 창출
- Bloomingdale's 백화점은 고객 정보를 기반으로 해서 카탈로그 사업 진출
- AT&T - 신용카드 사업 진출(billing database)
- 데이터 기반 플랫폼 비즈니스: 넷플릭스, 우버, 에어비앤비, …

Data-driven Market Analytics 분야

- Pricing
- Promotion
- Micromarketing
- Targeting
- Direct Marketing
- …….

Marketing Analytics

데이터 집적 수준에 따른 구분

Market level analytics

- Aggregate data
- 시장 수준의 마케팅 활동 효과 측정과 최적화
- Continuous variables: regression based model
- Pricing, promotion, advertising, positioning, demand forecasting, etc.

Individual level analytics

- Customer level data
- 고객 이질성 (customer heterogeneity) 초점, 고객별 상이한 반응 행태 파악 및 타겟팅
- Discrete variables: classification based model
- Customer life time value, customer, choice, targeting, etc.

Tools used in Marketing Analytics

Analytical Tools

Aggregate Data

- Regression / Model Selection
- Demand forecasting
- Bayesian hierarchical models
- Lift Modeling
- Clustering for segmentation
- Perceptual mapping for positioning
- …

Individual level Data

- Conjoint analysis
- Choice modeling: Logit/Probit
- Classification: logistic, discriminant, tree, neural nets, SVM, etc
- Recommendation: collaborative filtering / market basket analysis
- Social Media and text: sentiment / topic modeling
- …

Objectives of Marketing Analytics

Model: $y = f(x, \theta)$

Descriptive analytics

- Interested in how X affect Y in terms of the functional forms and parameters
- Functional forms matter. We want to figure out the data generating process (relationship between X and Y) and sometimes want to test theories.
- Focus: $\theta, f(\cdot)$

Predictive analytics

- Interested in predicting Y for a given value of X
- Not much interest in the nature of the relationship between X and Y
- Focus: $\hat{y}(x_0)$ for a given value x_0
- Typical machine learning applications are interested in prediction.

Marketing analytics use both descriptive and predictive analytics

Marketing Analytics vs. Machine Learning

Typical marketing analytics projects suffer from data problems, unlike machine learning projects.

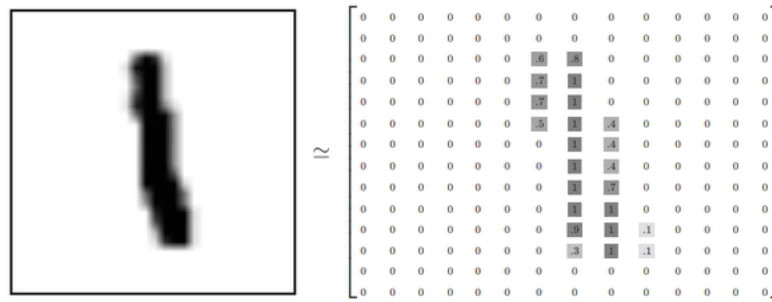
Example: MNIST project (recognizing handwritten digits)

Each image (handwritten digit) is represent by 784(=28*28) pixels. So we have 784 feature variables $x_1, x_2, \cdots, x_{784}$.

Each pixel is 8-bit gray scale (0,1,…,255) but normalized to a number between 0 and 1 (or 0/255, 1/255, … ,255/255).

Each image has label in the training data (i.e, $y = 5,0,4,1,\cdots$). So each y takes a value out of 10 labels (0,1,…,9).

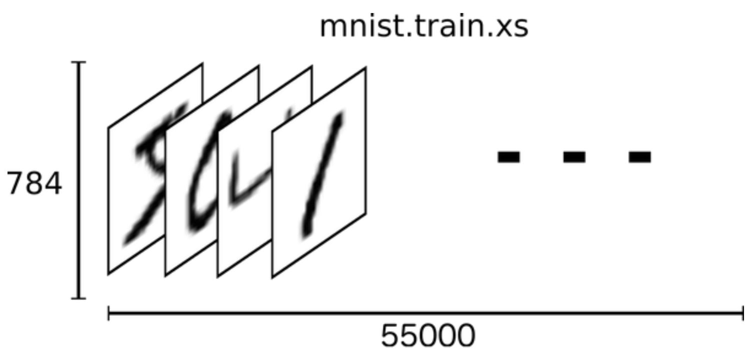
An image looks like

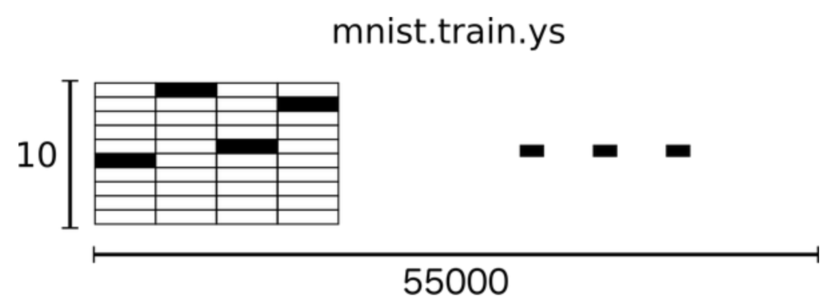


If we have 55000 images, the data set look like

$$\begin{array}{ll} y_1 & x_{1,1}, \cdots x_{784,1} \\ \vdots & \qquad \vdots \\ y_n & x_{n,1}, \cdots, x_{784,n} \end{array}$$

When n=55000, the feature (X) and label (Y) look like





Machine learning

- You fit you model using training data and apply the fitting result to predict the value of Y (label) for a new image. So in this case, the relationship between Y and a particular X variable is not a ,major interest.
- The point here that you have all pixels in your data. That is, the 784 pixels are sufficient to predict the label of each image.

Marketing analytics

- Consider a marketing analytics project where you want to predict a customer churn based on customers characteristics such as age, income, gender, prior purchase amount, etc.
- The fundamental problem here is that the features (age, income, etc) may not be complete. The worse thing here is that we do not know whether our feature data are complete. Marketing deals with customers (people) whose behaviors are probably affected by many factors. Much of them are not recorded in our database. Marketing analytics would utilize those (incomplete) data, unlike in machine learning.
- To overcome the incomplete data problem, marketers need to rely on some external information, “marketing theory”.