

Data Mining Techniques
Chapter 4:
Data Mining Applications in Marketing and Customer
Relationship Management

Prospecting	2
DM to choose the right place to advertise	3
Measuring fitness for groups of individuals	4
DM to improve direct marketing campaigns	5
Lift chart/cumulative gains or concentration chart	6
Response modeling	7
Differential response analysis	8
Using current customers to learn about prospects	9
DM for CRM I	10
DM for CRM II	11
Retention and churn/attrition	12

Prospecting

- Using DM to locate people who will be valuable customers.
- Identifying good prospects:
 - targeting: define what it means to be a good prospect and find rules that allow people with these characteristics to be targeted.
- Choosing a communication channel for reaching prospects:
 - public relations (not directed), advertising (target groups of people with common traits), direct marketing (allows customization of messages for individuals).
- Picking appropriate messages for different groups of prospects:
 - e.g., trade-off between price and convenience.

© Iain Pardoe, 2006

2 / 12

DM to choose the right place to advertise

- Look for customers who resemble current customers: *profiling*.
- Who fits the profile?
 - measure similarity (distance) between potential customer and profile, e.g., memory-based reasoning (a nearest-neighbor method) and automatic cluster detection.
- Who best fits the profile for this national newspaper?

	Alan	Beth	Readership	US pop	Yes index	No index
≥ BS	Yes	No	58%	20.3%	2.86 ^a	0.53 ^b
Prof/Exec	Yes	No	46%	19.2%	2.40	0.67
> \$75k	Yes	No	21%	9.5%	2.21	0.87
> \$100k	No	No	7%	2.4%	2.92	0.95

^a 58%/20.3%

^b 42%/79.7%

Alan's score = 2.86 + 2.40 + 2.21 + 0.95 = **8.42**

Beth's score = 0.53 + 0.67 + 0.87 + 0.95 = 3.02

© Iain Pardoe, 2006

3 / 12

Measuring fitness for groups of individuals

- Use census tract information (www.census.gov):
 - Census 2000 > American factfinder > Decennial census "get data" > SF3 quick tables;
 - Census tracts 45 (West Eugene), 48 (South-east Eugene), and 49 (East Eugene);
 - Tables DP-2 (social) and DP-3 (economic).

	Tract 45	Tract 48	Tract 49	Readership	45 score	48 score	49 score
≥ BS	43.9%	62.4%	80.6%	58%	0.76 ^a	1.00	1.00
Prof/Exec	34.4%	47.1%	61.5%	46%	0.75	1.00	1.00
> \$75k	6.7%	10.8%	34.5%	21%	0.32	0.51	1.00
> \$100k	3.4%	7.8%	21.5%	7%	0.49	1.00	1.00
Average					0.58	0.88	1.00

^a min(1, 43.9%/58%)

© Iain Pardoe, 2006

4 / 12

DM to improve direct marketing campaigns

- Improve targeting by selecting which people to contact (those most likely to respond).
- Response modeling allows prospect to be ranked by likelihood of response.
- Direct Marketing Association estimates it costs \$100k to mail 100k pieces (i.e., \$1 each).
- Optimizing response for a fixed budget can produce a different solution to maximizing overall profit.
- Receiver Operating Characteristic curves: *ignore*

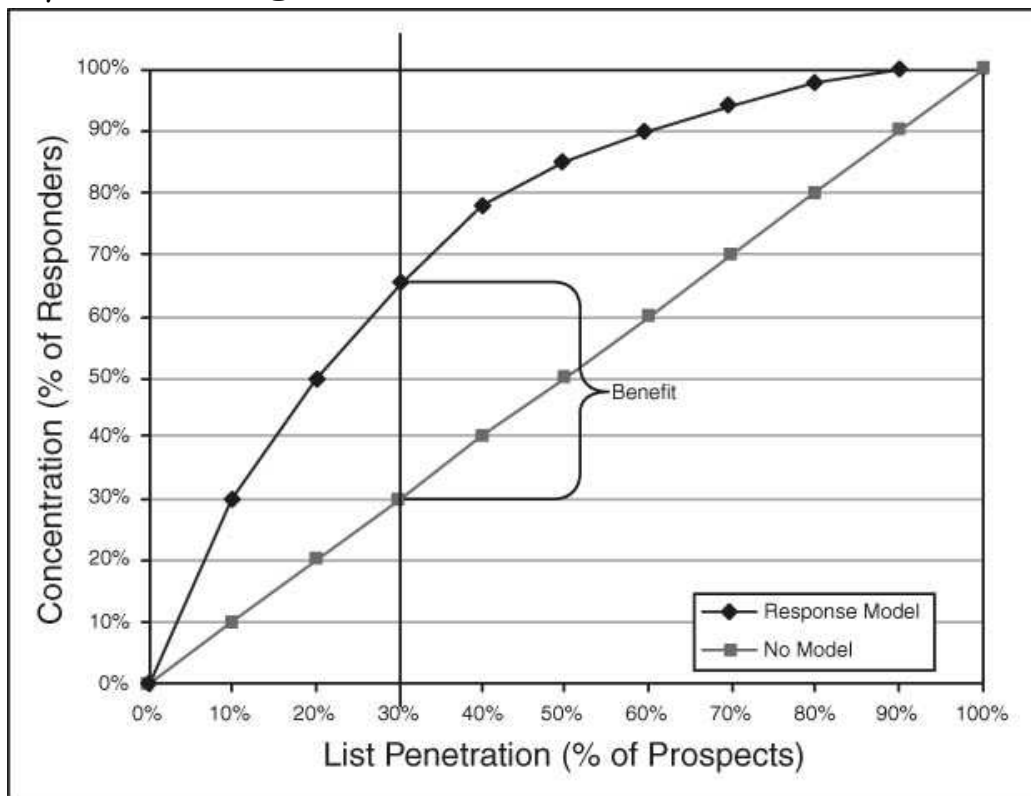
		Predict	
		Respond	Not respond
Reality	Respond	Correct	False negative
	Not respond	False positive	Correct

- High/low cutoff, few/many predicted responders, low/high false positives, high/low false negatives.

© Iain Pardoe, 2006

5 / 12

Lift chart/cumulative gains or concentration chart



© Iain Pardoe, 2006

6 / 12

Response modeling

- Lift is ratio of cumulative responses using model to cumulative responses with random mailing.
- Benefit is difference in response rates.
- Better to consider net profit if possible:
 - e.g., homework 1, question 5;
 - e.g., homework 3, question 4.
- Need to consider fixed campaign costs too.
- Reaching the people most influenced by the message:
 - control/test/holdout samples can identify people most likely to respond with or without being approached;
 - prospects with highest scores in response model may not provide greatest return on investment (they might have responded anyway).

© Iain Pardoe, 2006

7 / 12

Differential response analysis

- Identify prospects who are more likely to make purchases because of being contacted.
- Find segments where difference in response rate between a treated and untreated (control) group is greatest:

	control		treated	
	young	old	young	old
women	0.8	0.4	4.1	4.6
men	2.8	3.3	6.2	5.2

Which segment should be contacted to get the biggest return on investment? (also need to keep in mind size of segments in population)

© Iain Pardoe, 2006

8 / 12

Using current customers to learn about prospects

- Start tracking prospects before they become customers.
- Gather information from new customers (at the time they become customers).
- Model relationships between acquisition-time data and future outcomes of interest.
- Useful for evaluating existing acquisition channels, but not so good for identifying new ones.

© Iain Pardoe, 2006

9 / 12

DM for CRM I

- Response models used to tailor mix of marketing messages directed to existing customers:
 - beware focusing only on “responsive” customers (likely to be same ones each campaign), but rather send limited messages to each customer using scores, e.g.:

	score 1	score 2	score 3
Andrew	10	15	12
Bill	6	4	5

- Segmenting customer base:
 - mkt research/demographics (limited approach);
 - find behavioral segments, e.g., use decision trees to place every customer into an easily described segment (with respect to a specific marketing goal such as subscription renewal);
 - identify data patterns that correspond to pre-existing segments, e.g., deciding whether a residential phone likely to be used for business.

© Iain Pardoe, 2006

10 / 12

DM for CRM II

- Tie market research segments to behavioral data:
 - reducing exposure to credit risk—predicting who will default, terminating service for nonpayment (involuntary churn/forced attrition), improving collections;
 - determining customer value—careful financial definitions, retrospective/prospective customer value;
 - cross-selling, up-selling, and making recommendations (finding the right time to make an offer, association rules).

© Iain Pardoe, 2006

11 / 12

Retention and churn/attrition

- Increasing retention/reducing churn is a major application of DM.
- Recognizing churn easier in some businesses than others.
- Why churn matters—new customers expensive to acquire and initially less profitable, e.g.:
 - mailing cost \$1, new customer incentive \$20, if response rate is 5% then cost to reach 5 customers is $100 \times \$1 + 5 \times \$20 = \$200$

Response rate	5%	4%	3%	2%	1%
Cost per response	\$40	\$45	\$53	\$70	\$120

- Different kinds of churn: voluntary, involuntary, expected (customer no longer in target market).
- Build churn models to make retention offers only to high-value customers at risk for churning:
 - predict who will leave (within x days);
 - predict how long customers will stay (customer lifetime value models).

© Iain Pardoe, 2006

12 / 12