Show me the Money!
Deriving the Pricing Power of Product Features by Mining Consumer Reviews.

Nikolay Archak, Anindya Ghose, Panagiotis Ipeirotis

Stern School of Business, NYU

April 1st, 2010

Presented by Mukund Jha

Overview

- Previous work has focused mainly on predicting polarity of the reviews (+ve or –ve).
- **Research Questions:**
  - How **important** is each product feature to customers?
    - ‘battery life’ vs ‘image quality’
  - What is the **pragmatic meaning** of customers’ evaluation?
    - ‘good battery life’ vs ‘nice battery life’

Approach

- Investigate how product feature evaluations in reviews affect the product demand.
  - Derive product feature weights and strength (polarity & intensity) of evaluations
- **Example:**

"Excellent Lenses" 6% **"Excellent battery life" 3%**

"Good Lenses" -2% **"Good battery life" -1%**

  - Weight of Lenses are twice as high as the weight of battery life
  - Excellent gets a score of +3% and Good a score of -1%

Hedonic Regression

- **Assumption**: Goods can be described as vectors of measured features and consumer's value of good can be decomposed into values of each feature.
  - Ex: A backpacking tent -> weight (w), capacity (c) & pole material (p) and utility = u(w,c,p,…)
  - Not all products can be treated this way. Ex *movies, books*

- Weakness of existing hedonic model is the need to manually identify product features and measurement scales.
  - Leads to biased judgments

Identifying Opinions

- Feature Identification
  - Use POS Tags to identify most frequent nouns
  - These nouns are assumed to be product features
- Consumer Opinion
  - Adjectives are assumed to carry opinion
  - Dependency parse on sentences with feature noun are used to determine the adjectives that modify a product feature.
- These noun-adjective pairs correspond to pairs of product features and their evaluations and are referred to as *opinion phrases.*
  - *Ex : Quality – High, Lens - Fantastic*

Reviews

- Express reviews as product of Feature & Evaluation

$$\mathscr{R} = \mathscr{F} \otimes \mathscr{E}$$

  - $\mathscr{F}$ = feature space (f1,f2,…,fn) – n product features
  - $\mathscr{E}$ = evaluation space (e1,e2,e3,…,em)
- Weight of opinion phrases

$$w(phrase, rev, prod) = \frac{N(phrase, rev, prod) + s}{\sum_{y \in \mathscr{V}} (N(y, rev, prod) + s)}$$

  - N(y,r,p) = Number of occurrences of opinion phrase y in review r for product p.
  - V = Vocabulary (set of all fi X ej)

Reviews

- Example
  - “The camera is of high quality and relatively easy to use. The lens are fantastic! I have been able to use the LCD viewfinder for some fantastic shots... To summarize, this is a very high quality product."
  - Review can be represented as

$$
\begin{aligned}
0.4 &\quad \cdot \quad (quality \otimes high) + \\
0.2 &\quad \cdot \quad (use \otimes easy) + \\
0.2 &\quad \cdot \quad (lens \otimes fantastic) + \\
0.2 &\quad \cdot \quad (shots \otimes fantastic)
\end{aligned}
$$

Econometric Model

- Models demand as function of product features and its price

$$\ln(D_{kt}) = a_k + \beta \ln(p_{kt}) + \varepsilon_{kt}$$

  - D_{kt} = Demand for product k at time t
  - p_{kt} = price of product k at time t
  - Product feature is captured by

$$a_k = \alpha + \Psi(\mathbf{W_{kt}})$$

    - W_{kt} captures the opinion of product k at time t, including all reviews before t.
    - Alpha is a constant – product and time invariant

Econometric Model

$$\Psi(\mathbf{W_{kt}}) = \sum_{phrase \in \mathcal{V}} \psi(x) \cdot w(phrase, reviews_t, product_k) =$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \psi(f_i \otimes e_j) \cdot w((f_i \otimes e_j), reviews_t, product_k).$$

- $\psi(x) = \psi(f_i \otimes e_j)$ is the value of the opinion phrase and w(phrase,reviews,product) is the weight
- By examining the demand and pricing of the products , parameters for $\psi(x)$ are learned.
- Since the number of parameters are large n*m and not enough data is available, Singular Value Decomposition (SVD) technique is used to reduce $\psi(x)$ to rank 1 matrix.
  - $\psi(x)$ = γ(feature) * δ(evaluation)

Econometric Model

- Using rank 1 approximation we get,

$$\ln(D_{kt}) = \alpha + \beta \cdot p_{kt} + \gamma^T \cdot \mathbf{W_{kt}} \cdot \delta + \varepsilon_{kt}$$

- Algorithm similar to EM is applied to learn parameters, feature weights (δ) and evaluation weights (γ) from the dataset.

Experiments

- Dataset used was collected from Amazon.com and covered two categories
  - "Audio & Vedio" (127 Products ; 1955 reviews)
  - "Camera & Photo" (115 Products; 2580 reviews)
- Since the actual demand was not known, sales rank were used instead
- From a set of frequent nouns ~ 30 nouns were picked manually as product features
- Hedonic regression technique was used to predict the sales rank of the products

Evaluation

- **Predicting Future Sales:** Coefficient obtained from the training data is used to predict sales rank of the test data.
  - 5% improvement in Root Mean Square Error (RMSE) and 3% improvement in Mean Absolute Error (MAE) compared to the model without any review text information.

Feature Weights and Evaluation Scores

Feature	Weight	Std.Err.
camera*	0.810	0.091
quality*	0.484	0.106
battery*	0.192	0.048
resolution*	0.129	0.018
size	0.096	0.063
color	0.086	0.052
photos	0.074	0.040
lens	0.046	0.033
screen	0.037	0.037

Table 1: Average Feature weight for Camera & Photo Category

Evaluation	Score	Std.Err.
great*	-2.460	0.353
good*	-1.693	0.211
best*	-0.914	0.154
excellent*	-0.442	0.180
perfect*	-0.433	0.146
nice	-0.006	0.051
decent	0.001	0.056
fantastic	0.085	0.050
bad*	0.206	0.038
amazing*	0.220	0.094
fine*	0.258	0.101
poor*	0.345	0.066

Table 2 : Evaluation Scores for Camera & Photo category

Feature Weights and Evaluation Scores

- Table 3 shows Partial Effects for Camera & Photo.
- Negative sign signifies decrease in sales rank.

Phrase	Effect	Phrase	Effect
great camera	-0.4235	excellent photos	0.0040
good camera	-0.1128	nice size	0.0045
great quality	-0.0931	decent photos	0.0062
good quality	-0.0385	fantastic photos	0.0066
great battery	-0.0138	amazing resolution	0.0069
great size	-0.0060	amazing photos	0.0073
great photos	-0.0060	fine photos	0.0075
great resolution	-0.0052	excellent battery	0.0089
good battery	-0.0051	decent battery	0.0139
great lens	-0.0037	amazing battery	0.0164
good size	-0.0027	fine battery	0.0168
great color	-0.0023	best quality	0.0170
good photos	-0.0022	excellent quality	0.0507
good resolution	-0.0017	nice quality	0.0817
good lens	-0.0016	decent quality	0.0822
great screen	-0.0012	fantastic quality	0.0882
good color	-0.0004	amazing quality	0.0979
good screen	-0.0004	poor quality	0.1067
nice screen	0.0014	best camera	0.2026
excellent lens	0.0020	excellent camera	0.3936
excellent color	0.0027	perfect camera	0.3973
perfect size	0.0027	nice camera	0.5703
nice lens	0.0032	decent camera	0.5731
decent lens	0.0032	fantastic camera	0.6071
fantastic lens	0.0035	bad camera	0.6547
amazing lens	0.0038	amazing camera	0.6619
fine lens	0.0039	fine camera	0.6770

Thank you