Customer Wallet and Opportunity Estimation: Analytical Approaches and Applications

Saharon Rosset
Tel Aviv University
(work done at IBM T. J. Watson Research Center)

Collaborators: Claudia Perlich, Rick Lawrence, Srujana Merugu, et al.
Project evolution and my roles

- **Business problem definition**: Targeting, Sales force mgmt.
- **Modeling problem definition**: Wallet / opportunity estimation
- **Statistical problem definition**: Quantile est., Latent variable est.
- **Modeling methodology design**: Quantile est., Graphical model

**Model generation & validation**
- Programming, Simulation, IBM Wallets

**Implementation & application development**
- OnTarget, MAP

**Role Indicators**
- Red: Minor role
- Yellow: Major role
- Green: Leading contributor
Outline

- **Introduction**
  - Business motivation and different wallet definitions

- **Modeling approaches for conditional quantile estimation**
  - Local and global models
  - Empirical evaluation

- **MAP (Market Alignment Program)**
  - Description of application and goals
  - The interview process and the feedback loop
  - Evaluation of Wallet models performance in MAP
What is Wallet (AKA Opportunity)?

- Total amount of money a company can spend on a certain category of products.

IBM sales ≤ IT wallet ≤ Company revenue
Why Are We Interested in Wallet?

- **Customer targeting**
  - Focus on acquiring customers with high wallet
  - Evaluate customers’ growth potential by combining wallet estimates and sales history
  - For existing customers, focus on high wallet, low share-of-wallet customers

- **Sales force management**
  - Make resource assignment decisions
    - Concentrate resources on untapped
  - Evaluate success of sales personnel and sales channel by share-of-wallet they attain
Wallet Modeling Problem

- **Given:**
  - customer firmographics $\mathbf{x}$ (from D&B): industry, employee number, company type etc.
  - customer revenue $r$
  - IBM relationship variables $\mathbf{z}$: historical sales by product
  - IBM sales $s$

- **Goal:** model customer wallet $w$, then use it to “predict” present/future wallets

*No direct training data on $w$ or information about its distribution!*
Historical Approaches within IBM

- **Top down**: this is the approach used by IBM Market Intelligence in North America (called ITEM)
  - Use econometric models to assign total “opportunity” to segment (e.g., industry × geography)
  - Assign to companies in segment proportional to their size (e.g., D&B employee counts)

- **Bottom up**: learn a model for individual companies
  - Get “true” wallet values through surveys or appropriate data repositories (exist e.g. for credit cards)

- **Many issues with both approaches (won’t go into detail)**
  - We would like a predictive approach from raw data
Traditional Approaches to Model Evaluation

- Evaluate models based on surveys
  - Cost and reliability issues

- Evaluate models based on high-level performance indicators:
  - Do the wallet numbers sum up to numbers that “make sense” at segment level (e.g., compared to macro-economic models)?
  - Does the distribution of differences between predicted Wallet and actual IBM Sales and/or Company Revenue make sense? In particular, are the % we expect bigger/smaller?
  - Problem: no observation-level evaluation
Proposed Hierarchical IT Wallet Definitions

- **TOTAL**: Total customer available IT budget
  - Probably not quantity we want (IBM cannot sell it all)

- **SERVED**: Total customer spending on IT products covered by IBM
  - Share of wallet is portion of this number spent with IBM?

- **REALISTIC**: IBM sales to “best similar customers”
  - This can be concretely defined as a high percentile of: \( P(\text{IBM revenue} | \text{customer attributes}) \)
  - Fits typical definition of opportunity?

\[
\text{REALISTIC} \leq \text{SERVED} \leq \text{TOTAL}
\]
An Approach to Estimating SERVED Wallets

- **Wallet** is unobserved, all other variables are
- Two families of variables --- firmographics and IBM relationship are conditionally independent given given wallet
- We develop inference procedures and demonstrate them
- Theoretically attractive, practically questionable

(Will not discus further)
REALISTIC Wallet: Percentile of Conditional

- Distribution of IBM sales to the customer given customer attributes:  \( s|r,x,z \sim f_{\theta,r,x,z} \)
  
  E.g., the standard linear regression assumption:

  \[ s \mid x, r, z \sim N(\alpha x + \beta r + \gamma z, \sigma^2) \]

What we are looking for is the \( p^{th} \) percentile of this distribution
Estimating Conditional Distributions and Quantiles

- Assume for now we know which percentile p we are looking for

- First observe that modeling well the complete conditional distribution \( P(s|r,x,z) \) is sufficient

  \[ \Rightarrow \text{If have good parametric model and distribution assumptions can also use it to estimate quantiles} \]

  – E.g.: linear regression under linear model and homoskedastic iid gaussian errors assumptions

- Practically, however, may not be good idea to count on such assumptions

  – Especially not a gaussian model, because of statistical robustness considerations
Modeling REALISTIC Wallet Directly

- REALISTIC defines wallet as $p^{th}$ percentile of conditional of spending given customer attributes
  - Implies some $(1-p)\%$ of the customers are spending full wallet with IBM

- Two obvious ways to get at the $p^{th}$ percentile:
  - Estimate the conditional by integrating over a neighborhood of similar customers
    $\Rightarrow$ Take $p^{th}$ percentile of spending in neighborhood
  - Create a global model for $p^{th}$ percentile
    $\Rightarrow$ Build global regression models
Local Models: K-Nearest Neighbors

- **Design distance metric, e.g.:**
  - Same industry
  - Similar employees/revenue
  - Similar IBM relationship

- **Neighborhood sizes ($k$):**
  - Neighborhood size has significant effect on prediction quality

- **Prediction:**
  - Quantile of firms in the neighborhood

![Diagram showing the concept of K-Nearest Neighbors with IBM Sales and Employees axes, a target company's quantile, and the universe of IBM customers with D&B information.](image-url)
Global Estimation: the Quantile Loss Function

- Our REALISTIC wallet definition calls for estimating the $p^{th}$ quantile of $P(s|x,z)$.

- Can we devise a loss function which correctly estimates the quantile on average? Answer: yes, the quantile loss function for quantile $p$.

\[ L_p(y, \hat{y}) = \begin{cases} 
    p \times (y - \hat{y}) & \text{if } y > \hat{y} \\
    (1 - p) \times (\hat{y} - y) & \text{if } \hat{y} > y 
\end{cases} \]

- This loss function is optimized in expectation when we correctly predict REALISTIC:

\[
\arg\min_{\hat{y}} E(L_p(y, \hat{y}) \mid x) = p^{th} \text{ quantile of } P(y \mid x)
\]
Quantile Regression

- Squared loss regression:
  - Estimation of conditional **expected value** by minimizing sum of squares
  \[
  \min_\beta \sum_{i=1}^{n} (s_i - f(z_i, x_i, \beta))^2
  \]

- Quantile regression:
  - Minimize Quantile loss: \[
  \min_\beta \sum_{i=1}^{n} L_p (s_i, f(z_i, x_i, \beta))
  \]
  \[
  L_p (y, \hat{y}) = \begin{cases} 
  p \times (y - \hat{y}) & \text{if } y > \hat{y} \\
  (1 - p) \times (\hat{y} - y) & \text{if } \hat{y} > y
  \end{cases}
  \]

- Implementation:
  - assume linear function in some representation \( y = \beta^t f(x, z) \), solution using **linear programming**
  - Linear quantile regression package in R (Koenker, 2001)
Quantile Regression Tree – Local or Global?

**Motivation:**
- Identify a locally optimal definition of neighborhood
- Inherently nonlinear

**Adjustments of M5/CART for Quantile prediction:**
- Predict the quantile rather than the mean of the leaf
- Empirically, splitting/pruning criteria do not require adjustment
Aside: Log-Scale Modeling of Monetary Quantities

- Due to exponential, very long tailed typical distribution of monetary quantities (like Sales and Wallet), it is typically impossible to model them on original scale, because e.g.:
  - Biggest companies dominate modeling and evaluation
  - Any implicit homoskedasticity assumption in using fixed loss function is invalid

- Log scale is often statistically appropriate, for example if % change is likely to be “homoskedastic”

- Major issue: models ultimately judged in dollars, not log-dollars...
Empirical Evaluation: Quantile Loss

- **Setup**
  - Four domains with relevant quantile modeling problems: direct mailing, housing prices, income data, IBM sales
  - Performance on test set in terms of $0.9^{th}$ quantile loss
  - Approaches: Linear quantile regression, Q-kNN, Quantile trees, Bagged quantile trees, Quanting (Langrofd et al. 2006 -- reduces quantile estimation to averaged classification using trees)

- **Baselines**
  - Best constant model
  - Traditional regression models for expected values, adjusted under Gaussian assumption ($+1.28\sigma$)
Performance on Quantile Loss

<table>
<thead>
<tr>
<th>Approach</th>
<th>KDD98</th>
<th>California</th>
<th>Adult</th>
<th>IBM</th>
<th>Log-IBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.51 (0.144)</td>
<td>25217 (259)</td>
<td>5976 (34.9)</td>
<td>412810 (86793)</td>
<td>0.543 (0.0047)</td>
</tr>
<tr>
<td>Lin</td>
<td>1.95 (0.064)</td>
<td>14543 (250)</td>
<td>3564 (32.6)</td>
<td>555360 (21701)</td>
<td>0.363 (0.0035)</td>
</tr>
<tr>
<td>CART</td>
<td>1.86 (0.0831)</td>
<td>12774 (235)</td>
<td>3348 (36.1)</td>
<td>199280 (83711)</td>
<td>0.372 (0.0035)</td>
</tr>
<tr>
<td>kNN</td>
<td>2.22 (0.118)</td>
<td>14036 (258)</td>
<td>37144 (36.8)</td>
<td>175120 (73683)</td>
<td>0.361 (0.0032)</td>
</tr>
<tr>
<td>LinQuantReg</td>
<td>1.223 (0.061)</td>
<td>13901 (245)</td>
<td>3332 (30.7)</td>
<td>66049 (13720)</td>
<td>0.279 (0.0036)</td>
</tr>
<tr>
<td>Quanting</td>
<td>1.383 (0.094)</td>
<td>9797 (204)</td>
<td>2837 (32.4)</td>
<td>108370 (46663)</td>
<td>0.251 (0.0036)</td>
</tr>
<tr>
<td>Q-kNN</td>
<td>1.630 (0.115)</td>
<td>12532 (235)</td>
<td>3232 (35.5)</td>
<td>244910 (67661)</td>
<td>0.279 (0.0039)</td>
</tr>
<tr>
<td>LSE Tree</td>
<td>1.321 (0.086)</td>
<td>11695 (254)</td>
<td>2915 (34.7)</td>
<td>92831 (29411)</td>
<td>0.259 (0.0039)</td>
</tr>
<tr>
<td>QL Tree</td>
<td>1.325 (0.087)</td>
<td>11681 (231)</td>
<td>2979 (33.8)</td>
<td>94920 (29636)</td>
<td>0.257 (0.0036)</td>
</tr>
<tr>
<td>Bagged LSE Tree</td>
<td>1.319 (0.092)</td>
<td>9989 (217)</td>
<td>2830 (30.1)</td>
<td>91032 (28353)</td>
<td>0.254 (0.0036)</td>
</tr>
</tbody>
</table>

- **Conclusions**
  - Standard regression models are not competitive
  - If there is a time-lagged variable, LinQuantReg is best
  - Otherwise, bagged quantile trees (and quanting) perform best
  - Q-kNN is not competitive
Residuals for Quantile Regression

Total positive holdout residuals: 90.05% (18009/20000)
Market Alignment Project (MAP): Background

- **MAP - Objective:**
  - Optimize the allocation of sales force
  - Focus on customers with growth potential
  - Set evaluation baselines for sales personal

- **MAP – Components:**
  - Web-interface with customer information
  - Analytical component: wallet estimates
  - Workshops with Sales personal to review and correct the wallet predictions
  - Shift of resources towards customers with lower wallet share
MAP Tool Captures Expert Feedback from the Client Facing Teams

The objective here is to use expert feedback (i.e. validated revenue opportunity) from last year’s workshops to evaluate our latest opportunity models.
MAP Workshops Overview

- Calculated 2005 opportunity using naive Q-kNN approach
- 2005 MAP workshops
  - Displayed opportunity by brand
  - Expert can accept or alter the opportunity
- Select 3 brands for evaluation: DB2, Rational, Tivoli
- Build ~100 models for each brand using different approaches
- Compare expert opportunity to model predictions
  - Error measures: absolute, squared
  - Scale: original, log, root
Initial Q-kNN Model Used

- **Distance metric**
  - Identical Industry
  - Euclidean distance on size (Revenue or employees)

- **Neighborhood sizes 20**

- **Prediction**
  - Median of the non-zero neighbors
  - (Alternatives Max, Percentile)

- **Post-Processing**
  - Floor prediction by max of last 3 years revenue
Expert Feedback (Log Scale) to Original Model (DB2)

Experts accept opportunity (45%)

Increase (17%)

Experts change opportunity (40%)

Decrease (23%)

Experts reduce opportunity to 0 (15%)
Observations

- Many accounts are set for external reasons to zero
  - Exclude from evaluation since no model can predict this

- Exponential distribution of opportunities
  - Evaluation on the original (non-log) scale suffers from huge outliers

- Experts seem to make percentage adjustments
  - Consider log scale evaluation in addition to original scale and root as intermediate
  - Suspect strong “anchoring” bias, 45% of opportunities were not touched
Evaluation Measures

- **Different scales to avoid outlier artifacts**
  - Original: $e = \text{model} - \text{expert}$
  - Root: $e = \sqrt{\text{model}} - \sqrt{\text{expert}}$
  - Log: $e = \log(\text{model}) - \log(\text{expert})$

- **Statistics on the distribution of the errors**
  - Mean of $e^2$
  - Mean of $|e|$

- **Total of 6 criteria**
Model Comparison Results

We count how often a model scores within the top 10 and 20 for each of the 6 measures:

<table>
<thead>
<tr>
<th>Model</th>
<th>Rational</th>
<th>DB2</th>
<th>Tivoli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Displayed Model (kNN)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max 03-05 Revenue</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Linear Quantile 0.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Tree</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Q-kNN 50 + flooring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decomposition Center</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Quantile Tree 0.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Anchoring) (Best)
MAP Experiments Conclusions

- Q-kNN performs very well after flooring but is typically inferior prior to flooring.
- 80th percentile Linear quantile regression performs consistently well (flooring has a minor effect).
- Experts are strongly influenced by displayed opportunity (and displayed revenue of previous years).
- Models without last year’s revenue don’t perform well.

*Use Linear Quantile Regression with q=0.8 in MAP 06*
MAP Business Impact

- MAP launched in 2005
  - In 2006 420 workshops held worldwide, with teams responsible for most of IBM’s revenue

- Most important use is segmentation of customer base
  - Shift resources into “invest” segments with low wallet share

- Extensive anecdotal evidence to success of process
  - E.g., higher growth in “invest” accounts after resource shifts

- MAP recognized as 2006 IBM Research Accomplishment
  - Awarded based on “proven” business impact
Summary

- Wallet estimation problem is practically important and under-researched

- **Our contributions:**
  - Propose Wallet definitions: SERVED and REALISTIC
  - Offer corresponding modeling approaches:
    - Quantile estimation methods
    - Graphical latent variable model
  - Evaluation on simulated, public and internal data
  - Implementation within MAP project

- **We are interested in extending both theory and practice to other domains than IBM**
Thank you!
saharon@post.tau.ac.il