Exploring the Relationship between Customer Reviews and Prices

Lingjie Zhang, Lin Gong, Bo Man
Roadmap

- Introduction .............................................. Lingjie
- Methodology .............................................. Lin
- Experimental Results ................................. Bo
Customer Reviews Play an Important Role

90% of customers say buying decisions are influenced by online reviews.
Use of Customer Reviews

For customers
- Decision
- Recommendation

For retailers
- Feedback
- Marketing strategies

To what extend do they care about those reviews?
Motivation

Do customer reviews indirectly affect sale prices?
Related Work

✓ Classify reviews to help make decisions.
✓ Extract opinion features in customer reviews.
✓ Recommend products for customers.
✗ None of them combine customer reviews with prices.
Challenge

- Rating = Content?
- Relationship(Reviews, Prices)?
Methodology

Step 1: Collect Reviews

SNAP Amazon reviews:
- Products with over 100 reviews, in total 419 products.
Step 2: Assumption

User ratings == User reviews

Machine Learning Methods are adopted.
(Naive Bayes, Logistics Regression, Support Vector Machine)

Given contents -> predict ratings.
Compare final precisions and recalls.
Prediction Results:

Naive Bayes

<table>
<thead>
<tr>
<th>N</th>
<th>B</th>
<th>Sport Precision</th>
<th>Sport Recall</th>
<th>Tools Precision</th>
<th>Tools Recall</th>
<th>Home Precision</th>
<th>Home Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.924</td>
<td>0.476</td>
<td>0.91</td>
<td>0.447</td>
<td>0.949</td>
<td>0.795</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.947</td>
<td>0.907</td>
<td>0.929</td>
<td>0.833</td>
<td>0.962</td>
<td>0.881</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.947</td>
<td>0.981</td>
<td>0.946</td>
<td>0.889</td>
<td>0.967</td>
<td>0.905</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.927</td>
<td>0.994</td>
<td>0.906</td>
<td>0.961</td>
<td>0.893</td>
<td>0.567</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.981</td>
<td>0.977</td>
<td>0.963</td>
<td>0.993</td>
<td>0.865</td>
<td>0.979</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Beauty</th>
<th>Baby</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>0.923</td>
<td>0.773</td>
</tr>
<tr>
<td>0.898</td>
<td>0.808</td>
</tr>
<tr>
<td>0.933</td>
<td>0.957</td>
</tr>
<tr>
<td>0.968</td>
<td>0.952</td>
</tr>
<tr>
<td>0.976</td>
<td>0.992</td>
</tr>
</tbody>
</table>
Step 3: Crawl Prices

Price data:
- 221 items from previous 419 items
Step 4: Analysis

Scaling: 
\[ Y = (x - x_{\text{min}})/(x_{\text{max}} - x_{\text{min}}) \times 5 \]

Moving average: 
\[ RA_t = \frac{1}{k}(x_{t-k+1} + x_{t-k+2} + x_{t-k+3} + \ldots + x_t) \]

Shift Analysis:
- Compare against the prices ending \( L \) days later than the ratings.

Correlation Analysis:
- Pearson correlation coefficient is adopted.

\[ s(x, y) = \frac{\sum_{i=1}^{p} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{p} (x_i - \bar{x})^2 \times \sum_{i=1}^{p} (y_i - \bar{y})^2}} \]
Experimental Results

Sample Selection
Criteria:
Count (price changes) > 50, in 6 months

Sample size:
26 out of 221 items
Experimental Results

Scaling of prices

[Graphs showing price and score scaling]
Experimental Results

Moving Average & Tuning Parameter (window length)
Experimental Results

Shifting Analysis of prices and ratings (score)
Experimental Results

Correlation Analysis of prices and ratings

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely Strong</td>
<td>0.8-1.0</td>
</tr>
<tr>
<td>Strong</td>
<td>0.6-0.8</td>
</tr>
<tr>
<td>Medium</td>
<td>0.4-0.6</td>
</tr>
<tr>
<td>Weak</td>
<td>0.2-0.4</td>
</tr>
<tr>
<td>Extremely weak</td>
<td>0-0.2</td>
</tr>
<tr>
<td>No</td>
<td>&lt;0</td>
</tr>
</tbody>
</table>
Conclusion

- Relationship **exists** between prices and reviews.
- Reviews influence prices in most \(\frac{2}{3}\) of the items.
- Reviews often influence prices **after 7-30 days**.
- Categories with loose market forces fit this rule better.
  - like Home, Sport, Baby
Future Work

● Improvement on sample selection.

● Analyze relationship between prices and reviews.
  o For each separate category
  o With an expansion from single correlation calculation
  o Focus more on negative reviews

● Use our rules to predict prices.
References


Thank you!
Backup
Fun for adults too!

I really enjoy these scissors for my inspiration books that I am making (like collage, but in books) and using these different textures these give is just wonderful, makes a great statement with the pictures and sayings. Want more, perfect for any need you have even for gifts as well. Pretty cool!
### Logistics Regression Prediction Results

<table>
<thead>
<tr>
<th>L</th>
<th>R</th>
<th>Sport</th>
<th>Tools</th>
<th>Home</th>
<th>Beauty</th>
<th>Baby</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Precision</td>
<td>Recall</td>
<td>Precision</td>
<td>Recall</td>
<td>Precision</td>
</tr>
<tr>
<td>1</td>
<td>0.958</td>
<td>0.737</td>
<td>0.975</td>
<td>0.517</td>
<td>0.84</td>
<td>0.454</td>
</tr>
<tr>
<td>2</td>
<td>0.949</td>
<td>1.0</td>
<td>0.962</td>
<td>1.0</td>
<td>0.908</td>
<td>0.762</td>
</tr>
<tr>
<td>3</td>
<td>0.96</td>
<td>0.984</td>
<td>0.96</td>
<td>0.968</td>
<td>0.827</td>
<td>0.587</td>
</tr>
<tr>
<td>4</td>
<td>0.94</td>
<td>0.918</td>
<td>0.971</td>
<td>0.992</td>
<td>0.7</td>
<td>0.373</td>
</tr>
<tr>
<td>5</td>
<td>0.963</td>
<td>0.994</td>
<td>0.976</td>
<td>0.998</td>
<td>0.768</td>
<td>0.949</td>
</tr>
</tbody>
</table>
Support Vector Machine Prediction Results:

<table>
<thead>
<tr>
<th>SVM</th>
<th>Sport</th>
<th>Tools</th>
<th>Home</th>
<th>Beauty</th>
<th>Baby</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
<td>Precision</td>
<td>Recall</td>
<td>Precision</td>
</tr>
<tr>
<td>1</td>
<td>0.899</td>
<td>0.625</td>
<td>0.781</td>
<td>0.343</td>
<td>0.699</td>
</tr>
<tr>
<td>2</td>
<td>0.851</td>
<td>0.477</td>
<td>0.723</td>
<td>0.357</td>
<td>0.858</td>
</tr>
<tr>
<td>3</td>
<td>0.871</td>
<td>0.613</td>
<td>0.745</td>
<td>0.407</td>
<td>0.72</td>
</tr>
<tr>
<td>4</td>
<td>0.782</td>
<td>0.588</td>
<td>0.875</td>
<td>0.633</td>
<td>0.468</td>
</tr>
<tr>
<td>5</td>
<td>0.822</td>
<td>0.964</td>
<td>0.841</td>
<td>0.961</td>
<td>0.703</td>
</tr>
</tbody>
</table>