Personalized Recommendation with Confidence

Xiaqin Shelley Zhang, Sadhana Kuthuru, Ramya Mara and Brahmi Mamillapalli
Computer and Information Science Department
University of Massachusetts Dartmouth
North Dartmouth, MA 02747-2300
Email: x2zhang@umassd.edu

Abstract

This paper presents a personalized recommendation system mining online product reviews, fusing opinions together and providing a ranked order of a set similar products. We define three attributes of opinion summary: opinion coverage, opinion consistency and opinion consensus. Confidence factor is computed based on these attributes. A user specifies the relative importance of each product feature. The quantitative summary reflects the user’s preference, the opinion synopsis and the confidence measurement.

1. Introduction

Online product reviews generated by customers have been used by potential buyers to evaluate products before making a purchase decision. Opinion mining (OM) is to automatically extract opinions from such human-generated reviews. Many researchers have worked on OM and a comprehensive review is provided in [1]. Closely related to OM, information fusion (IF) is focusing on transforming information from different sources into useful knowledge to support decision making [2]. Making personalized purchase recommendation based on customer reviews is a task requires both OM and IF, especially when there is inconsistency among different customer opinions.

The goal of this work is to produce personalized purchase recommendation by mining product reviews, summarizing the results and providing a quantitative comparison of a set of similar products. The novelty of this work lay in two aspects. First, it adopts a personalized opinion mining approach, allowing a user to specify one’s unique requirement and preference. For example, some users consider print quality is more important than easiness to read, for a book, while some other users concern most about fast shipping. The recommendation needs to be made accordingly.

The second novel aspect of this work is to define a confidence measurement of the summarized result. Summarizing different review opinions is in fact an IF process since each reviewer is an independent information source. There are inconsistency and ambiguity in the data, also uncertainty and imprecise in the opinion extraction process. The user needs to be informed that how well-supported each summarized result is, in order to make a discerning decision.

Some research projects particularly related to these two aspects are discussed here. In an Opinion Relation Graph (ORG) [3], confidence is to measure how strong a dependency pattern of the opinion word and the target is supported, and this confidence measurement is used to select patterns to discover new opinion words and targets. In addition, Liu et al. [4] has proposed a graph-based co-ranking algorithm to estimate the confidence of each candidate on opinion relation graph. [5] provides a very good overview of different approaches to incorporate reliability into information fusion operators. It is assumed that a degree of belief is provide for each piece of data. How to obtain such information of degree of belief is a challenging task. Hattori and Takama [6] described a recommender system using personal-value-based user model with the focus on building such user model from pervious reviews; while our work focus on the usage of such user model in making recommendation. Verma and Dey [7] have presented a system to generate contextual recommendations based the user’s current interested content. This is personalized recommendation as our work is. The difference is that their work is to recommend content, while our work is to recommend products based on user’s different priorities of various product features.

In the rest of this paper, we first present the system work process in Section 2, and then in Section 3 we describe the details of measuring confidence, incorporating user’s preferences and confidence measurement in recommendation process. Testing results are demonstrated in Section 4, conclusion and future work are discussed in Section 5.

2. System Work Process

Figure 1 illustrates the system structure and below we describe three aspects of its work process.

2.1. Knowledge Bootstrap

The system is build with a knowledge base bootstrapped with q basic product categories \( \{ C_l | l = 1, \ldots, q \} \), such as book, liquid, and kitchen tools. Each category \( C_l \) is associated with a list of features (aspects) \( \{ F_j | j = 1, \ldots, m \} \). For example, category Book is associated with feature content, easy to read, fun, writing style, print/binding quality and shipping. Though some features, such as shipping, are common for different product categories, each product category usually has some unique features. For example, binding quality is unique for book, while smell is distinct for dish-wash liquid and easy to setup is special for electronic appliance. Therefore, maintaining category-dependent feature lists allows a user to specify preference more accurately and also helps mining
opinions more precisely. Table 1 shows six features for book category and five features for dish-wash liquid category. The number associated with each feature represents the relative importance of this feature to a user.

Additionally, a set of opinion words/phases \( \{o_k|k = 1, \ldots p \} \) are maintained for each feature; they are frequently used to describe a particular feature. Each opinion word/phase \( o_k \) is associated with a numeric sentiment rating \( r_k \), whose value is between 1 and 5; 1 is the most negative and 5 is the most positive assessment. For example, the following phases are used to describe writing style feature of a book: easy to understand (1), easy to grasp (4), great explanation (5), not easy to understand (1), vague explanation (2).

The initial bootstrap process is conducted manually so far. However, various mechanisms [1], may be used to expand this knowledge base automatically, such as supervised learning, semi-supervised learning, exploiting the relationship between features and opinion words using point-wise mutual information (PMI) score, information extraction techniques such as Conditional Random Fields (CRF) and Hidden Markov Models (HMM). We leave this as future work and currently only focus on making personalized recommendation.

### 2.2. User Input

When user provides product category to the system, a list of associated features are offered to the user, who may weight each feature \( F_j \) using a value \( w_j \) between 1 and 5 to represent the importance of each feature. Table 1 shows the examples of two users’ different feature weights for book and dish-wash liquid product. According to their preference profiles in book category, writing style of book is the most important feature for user A, while print/binding quality is also important. On the other hand, user B concerns most about print/binding quality and fun features of books. In dish-wash liquid category, user A concerns most about price, while user B cares most of shipping.

### 2.3. Opinion Mining

Product reviews are retrieved from online e-commerce website. In our experiment, these reviews are obtained from Amazon.com. The raw web page content goes through a pre-processing phase including tokenization, removal of stop words, and case normalization.

Each customer’s review opinion is considered as a document. We use a lexicon-based approach [8] for feature sentiment classification. The review text is processed to find matches to any opinion word/phase in the list for each feature \( F_j \) stored in the knowledge base. The same process is also used to identify matches to any feature and its synonyms. A match of an opinion word/phase in the review text is considered as an opinion for the closest feature appeared around, according to the nearest neighbor rule [9]. This approach is preliminary, and we recognize that more advanced text mining technologies are needed to improve the accuracy of this process.

After this process, a list of features and their matched opinion words/phases are identified from the review text. Based on this information, the system produces feature-based ratings for each product currently considered by this user and also analyze the confidence of this result. Finally a personalized recommendation is provided to user. We describe the feature-based rating procedure and the confidence analysis process in greater details in Section 3.

### 3. Feature-Based Rating With Confidence Measurement

For a given product \( P_j \), there is a set of features are associated with its category, and the above opinion mining process retrieves a set of opinion words/phases \( \{o_{jk}|k = 1, \ldots s \} \) for feature \( F_j \). Noted that there is a sentiment rating \( r_{jk} \) for \( o_{jk} \), with value 5 means the most positive assessment and value 1 means the most negative assessment. The average rating \( A_j \) for feature \( F_j \) is calculated as:

\[
A_j = \frac{\sum_{k=1}^{s} r_{jk}}{s}
\]

The standard deviation \( \sigma_j \) of the ratings for feature \( F_j \) is:

\[
\sigma_j = \sqrt{\frac{\sum_{k=1}^{s} (r_{jk} - A_j)^2}{s}}
\]

The average rating \( A_j \) and the standard deviation \( \sigma_j \) provide some basic information of the reviewers’ opinions about feature \( F_j \), more specifically, the mean value of the ratings and the difference among reviewers. However, other important information is not included here, such as the total number of reviewers who expressed opinions on this feature, the number

---

**TABLE 1: Product Feature and User Preference Examples**

<table>
<thead>
<tr>
<th>Features</th>
<th>Book (User A)</th>
<th>Book (User B)</th>
<th>Dish-wash Liquid (User A)</th>
<th>Dish-wash Liquid (User B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Easy to read</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Quality (print/binding)</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Content Presentation</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Fun</td>
<td>3</td>
<td>5</td>
<td>No-Leak 2</td>
<td>2</td>
</tr>
<tr>
<td>Writing Style</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
of reviewers with positive opinions \((r_{jk} \geq 3)\) and the number of reviewers with negative opinions \((r_{jk} \leq 2)\). Hence we introduce three confidence measures to represent these pieces of information.

1) **Opinion Coverage** \(C_1\). The ratio of the number of reviews who address this feature, \(s\), to the total number of reviews, \(n\).

\[
C_1 = \frac{s}{n}
\]

(3)

The greater this ratio is, the more convincing the summary result is. Consider two scenarios: out of 200 reviews, 180 of them address the quality of this book, versus only 10 of them comment on the same feature. Even in both scenario, the mean ratings are the same as 4.0, the first scenario brings more confidence with \(C_1 = \frac{180}{200} = 0.9\), because it is based on a much larger sample space than the second scenario with \(C_1 = \frac{10}{200} = 0.1\).

2) **Opinion Consistence** \(C_2\). It measures the consistence among reviewers’ opinions, normalized to the range of \([0,1]\) by dividing it with the range of the sentiment assessment value \(r\). The greater the deviation is, the less the consistence is.

\[
C_2 = 1 - \frac{\sigma_j}{\text{max}(r) - \text{min}(r)} = 1 - \frac{\sigma_j}{5 - 1} = 1 - \frac{\sigma_j}{4}
\]

(4)

3) **Opinion Consensus** \(C_3\). It measures how agreeable these reviewers’ opinions are. If one type of sentiment (whether positive or negative) is the majority, then it is more agreeable than another scenario, where a half of the opinions are positive and the other half are negative.

\[
C_3 = 1 - \frac{\text{min}(s_p, s_n)}{\text{max}(s_p, s_n)}
\]

(5)

\(s_p\) is the number of positive opinions, where \(s_n\) is the number of negative opinions on this feature: \(s_p + s_n = s\). For example, given \(s = 20\) opinions, 18 are positives and 2 are negative, the opinion consensus \(C_3\) is \(1 - \frac{2}{20} = 0.89\). If there are 10 positives and 10 negatives, then the opinion consensus \(C_3\) is \(1 - \frac{40}{40} = 0\).

For a specified feature, its confidence value \(C_v\) is the average value of these three confidence measures:

\[
C_v = \frac{C_1 + C_2 + C_3}{3}
\]

(6)

The algorithm below is used to calculate its confidence factor \(C_f\) based on confidence value \(C_v\).

When confidence value \(C_v\) is higher than a preset threshold \(ht\), confidence factor \(C_f\) is set as 1, meaning that there is sufficient confidence on the summarized opinion of this feature. On the other hand, if the confidence value is lower than a preset threshold \(lt\), there may not be enough confidence on the summarized opinion of this feature, hence we ignore this feature while rating this product by setting the confidence factor \(C_f\) as 0. If the confidence value \(C_v\) is in between of \(lt\) and \(ht\), then the confidence factor \(C_f\) is proportion to the confidence value \(C_v\).

Finally, the summarized rating \(SR_i\) for product \(P_i\) is calculated as:

\[
SR_i = \frac{\sum_{j=1}^{m} w_j * A_j * C_{f_j}}{10}
\]

(7)

where \(w_j\) is the weight value for feature \(F_j\) assigned by the user, \(A_j\) is the average rating for feature \(F_j\), and \(C_{f_j}\) is the confidence factor for feature \(F_j\). The overall rating \(SR_i\) is a weighted average rating of all features for product \(P_i\), considering confidence factor.

All products are then sorted according to their summarized ratings \(SR\) in decreasing order, the first one is the top recommendation for this user, as shown in Figure 3 and 4. In addition, the feature-based average rating and confidence values \(C_v\) of each product are also provided to the user to make a better-informed decision, as shown in Figure 2.

### 4. Example Results and Conclusion

We test our system in two categories: book and dishwashing liquid. These two categories are selected because each of them has quite distinguishing features from the other. Table 1 shows two users’ preference profiles for each category. Figure 3 and 4 present the final recommendations for each user. In each category, six products are ranked based on their summarized ratings \(SR\). Furthermore, Figure 2 shows the review summary of the six top ranked products in each category. For each product, the average rating \(A_j\) and confidence value \(C_{v_j}\) for each feature \(F_j\) are presented, which give the user detailed information to interpret the summarized rating of each product.

The top-ranked book for user A is *Head First Java*, which has high average rating with high confidence in *Writing Style* - the most important feature for user A. Another book *Head First Design Pattern* has a higher rating for *Writing Style*, however, it gets a lower summarized rating due to the very low confidence in *Content Presentation*. *Reusable OOPS* is recommended as the top choice for user B, partially due to its high average rating in *Fun*, which is one of the most concerned features for user B.

In dish-wash liquid category, the top recommended product *Downy Infusion* for user A is rated very high for *Price* - the
most important feature for user A. It also has high ratings for Smell and Cleans, which are relatively important for User A too. User B concerns most on Shipping followed by Smell and Cleans. The top ranked product for user B, Gain Liquid Detergent, has a very high rating for Shipping, and also quite high ratings for Smell and Cleans, all with high confidence, and are very important to user B too.

5. Conclusion and Future Work

We present a personalized recommendation system mining online product reviews, fusing opinions together and measuring the opinion coverage, opinion consistency and opinion consensus. User’s preference profile specifies the relative importance of each product feature. The quantitative summary is based on user’s profile, the opinion synopsis and the confidence measurement. In the future, we will explore different ways to measure confidence and examine various methods to incorporate confidence into decision-making process, possibly by allowing user choose how to use confidence. We would also like to utilize more advanced text mining methods to improve the accuracy of opinion mining. Automatically discovering features and opinion words are other future directions.

References