Analyzing Reviewer’s Credibility and Sentiment to Build A Profile Model for Product Recommendation of User

Dr. Vignesh Janarthanan¹, Esha Kumari Jain ², M. Sai Pavan Kumar ³, HanumanthaRedyy⁴, Madhukar Vemulawada ⁵
Professor¹, Students²,³,⁴,⁵ Malla Reddy Institute of technology and science ¹,²,³,⁴,⁵
Department of computer science and engineering, MRTS, Hyderabad-500100
vignesh.j@mritcs.ac.in¹
eshaandalisha@gmail.com²
pavan.mashapari@gmail.com³
hanuma3336@gmail.com⁴
madhukarvemulwada77@gmail.com⁵

Abstract
Understanding a particular customers product needs, likes, and dislikes and to make an automation based on it is a very convoluted job. This project augments heuristic-driven user interest profiling with reviewer credibility analysis and fine-grained feature sentiment analysis to conceive a vigorous recommendation methodology. The proposed credibility, interest and sentiment enhanced recommendation (CISER) model has five modules: candidate, feature extraction, reviewer credibility analysis, user interest mining, candidate feature sentiment assignment, and recommendation module. Review corpus is given as an input. The first module uses context and sentiment confidence to procure useful, crucial features. To detect the untrustworthy reviews and reviewers, reviewer credibility analysis proffers an approach to weigh reviews according to the parameters of credibility. The user interest mining module, uses fairness of review writing as heuristics for interest-pattern mining. The candidate feature sentiment assignment module compare existing features in review based on their fast Text sentiment polarity. The final module uses credible sentiment scoring for purchase recommendations. The proposed recommendation model utilizes not only numeric reviews but also uses sentiment expressions connected with components, customer preference profiles, and reviewer credibility for quantitative analysis of various alternative products. The mean average precision (MAP@1) for CISER is 93%, and MAP@3 is 49%, which is better than current state-of-the-art systems.

Keywords: Recommendation system, sentiment analysis, user credibility, user interest.

I. INTRODUCTION

The huge growth in digital information and the user traffic on the Internet has created compound disputes for people to unveil valuable information for long-term needs. The ever-expanding data flow is the result of this digital, networked economy. Big data is mostly changing how technology can help users and enterprises. By analysing a periscope-level view of the myriad interactions, patterns, and anomalies taking place within an industry and market, big data can be used to drive creative products and tools to market. But primary difficulties include kneading the data into a form that is useful for combining with Other sets of information. It is thus imperative to upgrade information filtering mechanisms for customized and personalized services enabled by big data analytics. Recommender systems are one of the most common and easily understandable big data applications. As a specialized information filtering system, a recommender system tries to make predictions based on user preferences and interests. Their use has been pervasive with interesting use-cases within various application domains that range from recommending products, movies, music, books, research articles, search queries, social tags, experts, Persons, jokes, restaurants, financial services and even Twitter followers. There are two significant paradigms of recommender systems, collaborative filtering and content-based methods. Collaborative filtering (CLF) is the process of filtering or evaluating items through the opinion of other people who share similar interests. CLF supports information filtering and retrieval based on shared preferences and opinions. On the other hand, content-based recommendation systems analyze item descriptions to identify items that are of particular interest to the user. These systems collect
descriptions of potential items, profiles of the user describing the types of items preferred by them and compare them to match user preferences to determine recommendation results.

**Related Work**
Product recommender systems find applications within e-commerce (Amazon, Flipkart, and Big-basket) and mediaservice (Amazon-Prime, Netflix) domains. Techniques for product recommendations like exploiting ratings for quantifying user-user and user-product adherence [content based and collaborative filtering], sentiment-analysis based recommendations, context-aware recommendations, user-preference and trust-oriented recommendations have been helpful. Primary studies on recommender systems (RS) have majorly focused on ratings, similarity patterns. The need to model human aesthetics, linguistics and emotions (sentiment analysis) and using it for product acceptability analysis is inevitable. Although sentiment analysis based RS is better than the numerical-ratings based RS, still they are not always precise. User-profile is extremely important in understanding preferences. Moreover, spam reviews or opinions add to the criticalities, impairing the merit of recommendations. The work proposed in this paper establishes a growing and credible data-driven model for product recommendations. The work combines user preference profiling, reviewer credibility assignment and fine-grained sentiment analysis for recommendations. The model considers product reviews as a dependable source of comprehending user preferences and uses text analytics, fine-grained sentiment analysis and feature extraction to model user interests. Further, to make the model robust to fake and unworthy reviews and reviewers, we proffer an approach of associating expertise, trust and influence scores with reviewers to weigh their opinion according to their credibility. This work aims to optimize the credibility weighted sentiment value of user-preferred features for recommended products. The proposed credibility, interest and sentiment enhanced recommendation (CISER) model is explained in the next section.

**Proposed System**
The proposed CISER model uses the strength of fine-grained sentiment analysis along with reviewer credibility analysis techniques and user interest patterns.

CISER mainly consists of six modules namely (1) candidate feature extraction (2) user product review mapping (3) reviewer credibility analysis (4) user interest mining (5) candidate feature sentiment scoring and (6) ranking and recommendation module.

**Dataset Acquisition**
We use the Amazon Camera review dataset which has reviews and ratings given by various users for different brands of camera on the e-commerce website amazon.com. In addition to review text and ratings, various other attributes like number of helpful votes, number of total votes, verified purchase information, vine user verification information are available in the dataset, which have been used as auxiliary features in CISER.
Candidate Feature Extraction
Not all words in a review contribute to user’s purchase preferences. It takes user reviews as input and outputs context-driven features (word lemmas) that further help in user preference modelling, review utility analysis and feature level sentiment mining. The methodology for candidate feature extraction can be broken down into two components:

- Sentence-level fine-grained sentiment analysis
- Unigram and bigram analysis

Sentence-Level Fine-Grained Sentiment Analysis
Sentiment Analysis is a quintessential text classification task which intends to categorize the opinion polarity.

Unigram and Bigram Analysis
Unigram Product Features: The product features comprising of a single word. Example: ‘lens’, ‘resolution’ etc.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Sentiment Confidence</th>
<th>Root Similarity Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lens</td>
<td>0.186</td>
<td>0.662</td>
</tr>
<tr>
<td>Viewfinder</td>
<td>0.059</td>
<td>0.5984</td>
</tr>
<tr>
<td>Photo</td>
<td>0.374</td>
<td>0.5354</td>
</tr>
<tr>
<td>Flash</td>
<td>0.6</td>
<td>0.5197</td>
</tr>
<tr>
<td>Mode</td>
<td>0.11</td>
<td>0.4129</td>
</tr>
</tbody>
</table>

Bigram Features: The product features comprising of two words. Example: ‘picture quality’, ‘red eye’. For the purpose of this study we limit our potential bigram features to 2-tuples possessing the following characteristics.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Sentiment Confidence</th>
<th>Root Similarity Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture quality</td>
<td>0.6</td>
<td>0.5668</td>
</tr>
<tr>
<td>Battery life</td>
<td>0.56</td>
<td>0.478</td>
</tr>
<tr>
<td>Flash photography</td>
<td>0.4</td>
<td>0.543</td>
</tr>
<tr>
<td>Auto-setting</td>
<td>0.3</td>
<td>0.42</td>
</tr>
<tr>
<td>Image processing</td>
<td>0.6</td>
<td>0.54</td>
</tr>
</tbody>
</table>

User Product Review Mapping Module
Product data is taken from the Amazon camera review dataset. The mapped data is input to the three modules namely user interest mining module, reviewer credibility analysis module and candidate feature sentiment scoring module.

User Interest Mining
Recommendation systems like Netflix, Amazon and Flipkart maintain user interest profiles to recommend products based on preferences. In this paper, we explore user reviews as source of mining user-interest patterns. The following heuristics help in modelling of the user-preference profile:

- Heuristic 1(H1): Users assess only a specified subset of product features based on requirement, taste and fondness. That is, some of the characteristics are more important than others. For example, cameras with ‘high resolution’ might be more important than its ‘battery-life’ for a user. So, it is likely that the user mentions resolution quality in most of his reviews, and says nothing much about the battery-life.

- Heuristic 2(H2): Numerical ratings seldom analyze the intricate details of products, but reviews present a rather detailed analysis. So, it can be assumed that different features of products can be assessed by analysing the reviews. On the whole, H2 suggests that reviews are a token to quantify product features quality, which can be used to enhance overall user experience. Following the heuristic (H1), we propose that, a user is interested in a specific candidate feature, if they make a frequent mention of it in his reviews. Frequency of allusion to specific features asserts their utility for users.

Reviewer Credibility Analysis
This module analyses the reliability of product reviewers for credibility score assignment. Reviewer credibility analysis (RCA) augments the recommendation capability of CISER and makes it robust to fake/unworthy reviews and reviewers. The following metrics are calculated for reviewer credibility scoring:
Trust Factor (calculated using review utility; review content; percentage coverage)
Expertise Factor (calculated using representative rating)

**Algorithm 1**
Recommendation Algorithm

**Input:** Product Features: $F$, Products: $P$, Set of Reviews: $Z$, Reviewer Credibility: $\gamma_{cred}$, User to recommend products to: $U$, Number of products to be recommended: $K$, User Interest Matrix: $UIM_u$

**Output:** Recommended Products: $\text{recommendedProducts}$

1. $g(\text{review}, \text{feature}) \leftarrow (1 : \text{feature} \in \text{review} 0 : \text{otherwise})$

2. $\forall p \in P : \text{rank}[p] \leftarrow P_{x \in F(r \in Z[p]} \times \text{fineGrainedSentiment}(x,r) \times \gamma_{cred}[user[r]] \times g(r,x) P_{r \in Z[p]} \gamma_{cred}[user[r]] \times g(r,x) \times UIM_u[x]$

3. $\text{recommendedProducts} \leftarrow \text{Top K products according to rank}$

**II. CONCLUSION**

In elementary words, we can say that in our algorithm, rather than recommending products based on the average rating of all users, the ratings provided by the creamy layer of the users (the 'best' users) are considered more relevant. But, this does not mean that we ignore the ratings provided by the other users. Every rating counts in the final recommendation. It's just that the extent of contribution from each rating depends on the reviewer's credibility. The scope of the future work would be to design an automated system that would collect the data used in the proposed algorithm from the user profile of every node in the network, which would help to test the algorithm with more extensive data.

**REFERENCES**


48. S.B. Jayabharathi, S. Manjula, E. Tamil Selvan, P. Vengatesh and V. Senthil Kumar” Semantic Risk Analysis Model Cancer Data Prediction” Journal on Science Engineering and Technology Vo1ume 5, No. 02, April 2018


58. Dr. V. Senthil kumar, Mr. P. Jeevanantham, Dr. A. Viswanathan, Dr. Vignesh Janarthanan, Dr. M. Umamaheswari, Dr. S. Sivaprakash Emperor Journal of Applied Scientific Research “Improve Design and Analysis of Friend-to-Friend Content Dissemination System ”Volume - 3 Issue - 3 2021