Received January 19, 2020, accepted January 27, 2020, date of publication February 3, 2020, date of current version February 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2971087

Reviewer Credibility and Sentiment Analysis Based User Profile Modelling for Online Product Recommendation

SHIGANG HU[1], AKSHI KUMAR[2], FADI AL-TURJMAN[3,4], SHIVAM GUPTA[2], SIMRAN SETH[2], AND SHUBHAM[2]

[1]School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China
[2]Department of Computer Science and Engineering, Delhi Technological University, New Delhi 110042, India
[3]Artificial Intelligence Department, Near East University, 99138 Nicosia, Turkey
[4]Research Center for AI and IoT, Near East University, 99138 Nicosia, Turkey

Corresponding author: Akshi Kumar (akshikumar@dce.ac.in)

This work was supported in part by the National Natural Science Foundation of China under Grant 61674056.

ABSTRACT Deciphering user purchase preferences, their likes and dislikes is a very tricky task even for humans, making its automation a very complex job. This research work augments heuristic-driven user interest profiling with reviewer credibility analysis and fine-grained feature sentiment analysis to devise a robust recommendation methodology. The proposed **c**redibility, **i**nterest and **s**entiment **e**nhanced **r**ecommendation (CISER) model has five modules namely candidate feature extraction, reviewer credibility analysis, user interest mining, candidate feature sentiment assignment and recommendation module. Review corpus is given as an input to the CISER model. Candidate feature extraction module uses context and sentiment confidence to extract features of importance. To make our model robust to fake and unworthy reviews and reviewers, reviewer credibility analysis proffers an approach of associating expertise, trust and influence scores with reviewers to weigh their opinion according to their credibility. The user interest mining module uses aesthetics of review writing as heuristics for interest-pattern mining. The candidate feature sentiment assignment module scores candidate features present in review based on their fastText sentiment polarity. Finally, the recommendation module uses credibility weighted sentiment scoring of user preferred features for purchase recommendations. The proposed recommendation methodology harnesses not only numeric ratings, but also sentiment expressions associated with features, customer preference profile 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.

INDEX TERMS Recommendation system, sentiment analysis, user credibility, user interest.

I. INTRODUCTION

The colossal growth in digital information and the number of visitors on the Internet has created intricate challenges for people to uncover potentially valuable information for long-time needs. The ever-widening flow of data is the by-product of this digital, networked economy. Big data is essentially changing how technology can serve consumers 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 new, creative products and tools to market. But primary difficulties include around 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 applications of big data. As a specialized information filtering system, a recommender system tries to make predictions on the basis of user preferences and interests [1]. Their use has been pervasive with interesting use-cases within variety of application domains that range from recommending products, movies, music, books, research articles, search queries, social tags, experts,

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Arafatur Rahman.

persons, jokes, restaurants, financial services and even twitter followers. There are two major 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 [2]. CLF supports information filtering and retrieval, based on shared preferences and opinion. Content-based recommendation systems, on the other hand analyse item descriptions to identify items that are of particular interest to the user [3]. These systems collect description of potential items, profile of the user describing the types of items preferred by them and compare them to match user preferences, to determine recommendation results.

Most of the present-day recommender systems tend to use rating-information to quantify user-user, user-item and item-item similarities. Several techniques like clustering, nearest-neighbour methods, matrix manipulations, point-of-interest modelling have been used to model user interest patterns so as to maximize purchase satisfaction. But user-ratings are biased by certain hidden factors like brand-adherence and product-prejudice. So, sole considerations of rating-oriented similarity and user-interest analysis is rendered useless in the complex modern-day setting. The limited capabilities of user-ratings have given way to heuristic-driven, context-driven, sentiment and emotion-driven user-interest profiling.

Simultaneously, it has been observed that not all reviews and corresponding product ratings contribute to helpful recommendations. Moreover, spammers exploit these review platforms illegally because of incentives involved in writing fake reviews. Currently, fake reviews and reviewers form a bulk of the review opus making review spamming an open research challenge. These spam reviews must be detected to nullify their contribution towards product ratings. Reviewer credibility analysis can be used to quickly address deceptive online reviews and reviewers. Pertinent research studies suggest various parameters that can affect reviewer credibility. These include: linguistic styles, review clarity and comprehensiveness, word count, sentence count in reviews, helpful vote count, evidence (speaker's degree of certainty using certain propositional attitudes for example, certainly, surely) [4], [5].

Motivated by the two-fold need to firstly ascertain latent factors that affect user's fondness for a product/product category and to secondly quantify reviewer's credibility and review's quality for recommendation, we propose a **c**redibility **i**nterest and **s**entiment **e**nhanced **r**ecommendation (CISER) model. The model augments heuristic-driven user interest profiling with reviewer credibility analysis and fine-grained feature sentiment analysis to devise a robust recommendation methodology. Generically, user preferences can be described in implicit or explicit ways (Fig. 1). In this study, we use aesthetics of review writing as heuristics for mining user preferences. Aesthetics would indicate some sort of definable set, a list of knowable criteria used to evaluate the stylistic, thematic, and structural choices of a piece of writing. These aesthetics also foster reviewer and review reliability evaluation. Simultaneously, it is also essential to retrieve relevant and context-sensitive features to determine whether the reviewer's attitude towards a particular product is positive, negative, or neutral. Sentiment analysis enables analysing people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [6].

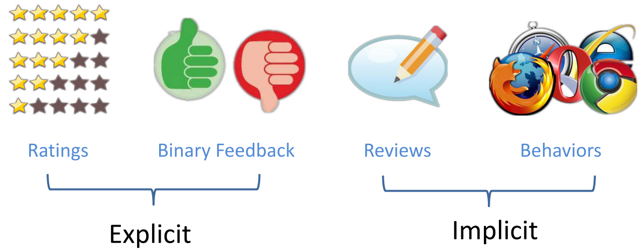

FIGURE 1. User preferences.

Hence, the proposed CISER model proffers novel techniques for candidate feature extraction, reviewer credibility analysis, feature sentiment scoring and user interest modelling. Sentiment confidence and context derive the candidate features. Reviewer credibility scoring is done by quantifying review utility, content and percentage coverage of candidate features (*trust factor*); manipulating rating trends (*expertise factor*) and recording the influence of reviewer in the network (*network importance factor*). Review writing aesthetics are used to devise heuristics for user-preference modelling. The product recommendation is based on feature sentiment value calculated using fastText and at the same time using credibility-weighted sentiment feedback (in terms of factors that interest the particular user). To the best of our knowledge, this is the first work that integrates credibility driven feature-based fine-grained sentiment analysis with user interest modelling for online product recommendation. The key contributions of this work are:

- A recommendation model that combines user-preference modelling, reviewer credibility analysis and feature sentiment scoring for robust product ranking and recommendations.
- User preferences and interests are modelled without social information (number of followers, marital status etc.). Review writing aesthetics have been used for deriving heuristics that help in user-preference profiling.
- A model for reviewer credibility analysis is proposed which assigns credibility values to weight user opinions. Three metrics namely trust factor, network importance factor and expertise factor have been proposed for reviewer scoring. Language semantics, sentiment, consistency, coherence, rating trends and reviewer influence are exploited for credibility analysis.

The CISER model has six modules namely, candidate feature extraction, user product review mapping, reviewer credibility analysis, user interest mining, candidate feature sentiment scoring and ranking and recommendation module. The Amazon's camera review dataset [7] is used to extract text and ratings along with some additional attributes such as number of helpful votes, number of total votes, verified purchase information, vine user verification information. Candidate feature extraction is done using context (spaCy[1]) and sentiment confidence (fastText[2]) to extract the features of importance. To make our model robust to fake and unworthy reviews and reviewers, reviewer credibility analysis proffers an approach of associating expertise, trust and influence scores with reviewers to weigh their opinion according to their credibility. The user interest mining module uses aesthetics of review writing as heuristics for interest-pattern mining. The candidate feature sentiment assignment module scores candidate features present in review based on their fastText sentiment polarity. Finally, the recommendation module uses credibility weighted sentiment scoring of user preferred features for purchase recommendations. This way, the proposed recommendation methodology harnesses not only numeric ratings, but also sentiment expressions associated with features, customer preference profile and reviewer credibility for quantitative analysis of various alternative products. The results have been evaluated using mean average precision (and compared with the current state-of-the-art systems).

The rest of the paper is organised as follows: in section II, we present related research within the domain of study. Section III describes the proposed CISER model in detail illustrating details of all its component modules. In section IV, a working example to support better understanding of the proposed model is given. Section V presents the experimentation and results followed by conclusion and future work in section VI. The frequently used abbreviations are tabulated in Table 1.

II. RELATED WORK

Product recommender systems find applications within e-commerce (Amazon, Flipkart, and Big-basket) and media-service (Amazon-Prime, Netflix) domains. Various 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 reported. Literature is well equipped with primary and secondary studies on state-of-the-art techniques for recommender systems [8]–[10].

Primary studies on recommender systems (RS) have majorly focussed on product-ratings for modelling user-user and product-product similarity patterns. These RS find applications in several state-of-the-art software platforms developed to recommend movies [11], [12], songs [13] and videos [14] etc. But numerical rating-based RS ignore valuable

[1]https://spacy.io/models/en

[2]https://github.com/facebookresearch/fastText

TABLE 1. List of abbreviations.

Abbreviation	Explanation
CBF	Content Based Filtering
CF	Coherence Factor
CFS	Candidate Feature Set
CISER	Credibility, Interest and Sentiment Enhanced Recommendation
CLF	Collaborative Filtering
EF	Expertise Factor
HV	Helpful Votes
NIF	Network Importance Factor
NIM	Network Importance Matrix
PC	Percentage Coverage
PSO	Particle Swarm Optimisation
RCA	Reviewer Credibility Analysis
RR	Representative Rating
RS	Recommender Systems
SC	Sentiment Confidence
TF	Trust Factor
TR	Total Ratings
TV	Total Votes
UIM	User Interest Matrix
VPI	Verified Purchase Information

review sentiment data and hence cannot model user interests completely. Therefore, research interests have started shifting to include human sentiments in rating products for recommendation. The need to model human aesthetics, linguistics and emotions (sentiment analysis) and using it for product acceptability analysis is inevitable.

Sentiment-analysis based RS have recently been reported in prominent natural language processing, information processing and business intelligence journals. Guerreiro and Rita [15] used a lexicon-based approach to search factors within text that uncover positive and negative triggers for direct recommendations. Chelliah and Sarkar [16] used several techniques for user-review exploitation. These included text mining methods for feature-specific sentiment analysis, topic models and distributed vocabulary representations. In "TRRuSST" [17], Gallege *et al.* used sentiment analysis for product ratings and extended CBF(Content-Based Filtering) and CLF(Collaborative Filtering) using user-reviews. External attributes were modelled for effective software product recommendations. Quian *et al.* [18] proposed "EARS"(Emotion-aware recommender system based on hybrid information fusion) that uses social information, sentiment analysis and gaussian distribution based user-behaviour analysis for recommendation. Da'u and Salim [19] proposed sentiment-aware deep recommender system with neural attention network (SDRA) that uses aspect based sentiment analysis for improving recommendation accuracies. The model reinforces semi-supervised topic modelling with LSTM (long short term memory) via neural attention mechanism for better modelling of user-item sentiments.

Although sentiment analysis based RS outperform the numerical-ratings based RS, still they do not always model user preferences accurately. User-profile plays a major role in understanding purchase preferences. Many research works focus on user-profile modelling and the role of user-product preference relationship in recommendation process. In "CUP" [20], Alan *et al.* presented the concept of inferred contextual profiles that describe a user in a given situation or context. Sanchez *et al.* [21] used ontological model to represent user and item profiles for advertisement recommendation. Wei *et al.* [22] proposed a method to determine user authority in a social tagging system for personalised recommendations. Xu and Liu [23] used semantic content and dynamic ontology modelling for user profiling to improve recommendation accuracies in event-based social networks. Covington *et al.* [14] used "DNN"(Deep Neural Networks) for video-recommendations using user-history and context. Another work by Liji *et al.* [24] used a network clustering model to cluster similar users and recommend items that the user may like. Xu *et al.* [25] presented user recommendation framework "UIS-MF" which captures interest and social factors to recommend users having similar interest and close social connection. Logesh *et al.* [26] proposed an activity and behaviour induced personalized RS to predict persuasive POI recommendations. Li *et al.* [27] used movie feature vector combined with the user rating matrix to generate the user interest vector. Tahmasbi *et al.* [28] gave an approach to model temporal dynamics of user preferences in movie recommendation systems based on a coupled tensor factorization framework. Middleton *et al.* [29] explored the acquisition of user profiles using browsing behaviour and presented an ontological representation to extract user preferences. Xie and Wang [30] proposed a web page recommendation based on two-fold clustering by considering both user behaviour and topic relation.

One of the key concerns with the use of customer reviews in RS is the quality of review and reviewer credibility. Several researchers have proposed criteria for opinion leader selection and review quality estimation. These include linguistic styles, review clarity and comprehensiveness; word count, sentence count in reviews; helpful vote count and evidence that quantifies the speaker's degree of certainty using certain propositional attitudes (certainly, surely). Al-Sharawneh and Williams *et al.* [31] made use of the "Follow the Leader" [32] model and proposed a novel technique for leader identification using social-profile information, trustability and expertise of reviewer. Li and Du [33] proposed an ontology-oriented model "BARR"(**B**log content, **A**uthors, **R**eaders and **R**elationship)that uses blog content, author, readers and centrality theory for hot-topic selection and credibility analysis. These hot-topics were further exploited for use in significant marketing policies. Chen *et al.* [34] explored the utility of trust relations to capture user attention. Poisson distribution is used to model trust relations; therefore the proposed model is robust to sparse data. Finally, Srivastava and Kalro [5] proposed certain latent factors that affect the credibility of reviews like readability, relevance, word-count etc. These works show that not all reviews and reviewers equally contribute to the assessment of product-acceptance in market. Moreover, spam reviews or opinions add to the criticalities, impairing the merit of recommendations.

The work proposed in this paper establishes a novel 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 optimise the credibility weighted sentiment value of user-preferred features for recommended products. The proposed **c**redibility, **i**nterest and **s**entiment **e**nhanced **r**ecommendation (CISER) model is explained in the next section.

III. THE PROPOSED CISER MODEL

The proposed CISER model reinforces the strength of fine-grained sentiment analysis in combination with reviewer credibility analysis techniques and user interest patterns to deal with the problem of product recommendations. 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. Fig.2 depicts the architecture of the proposed CISER model.

The following subsections elucidate the details of each module:

A. DATASET ACQUISITION

We have used the Amazon Camera review dataset [7] which contains reviews and ratings given by various users for different brands of camera on the e-commerce website amazon.com. This data was released to encourage further research in multiple disciplines related to understanding customer product experiences. Specifically, this dataset was constructed to represent a sample of customer evaluations and opinions, variation in the perception of a product across geographical regions, and promotional intent or bias in reviews. This dataset has been used to train our model and study its practical utility. An overview of the dataset is given in Table 2.

In addition to review text and ratings, some additional 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. The complete description of data attributes is given in Table 3.

TABLE 2. Dataset overview.

Number of Products	35
Total number of Users	22768
Total number of Reviews	23256

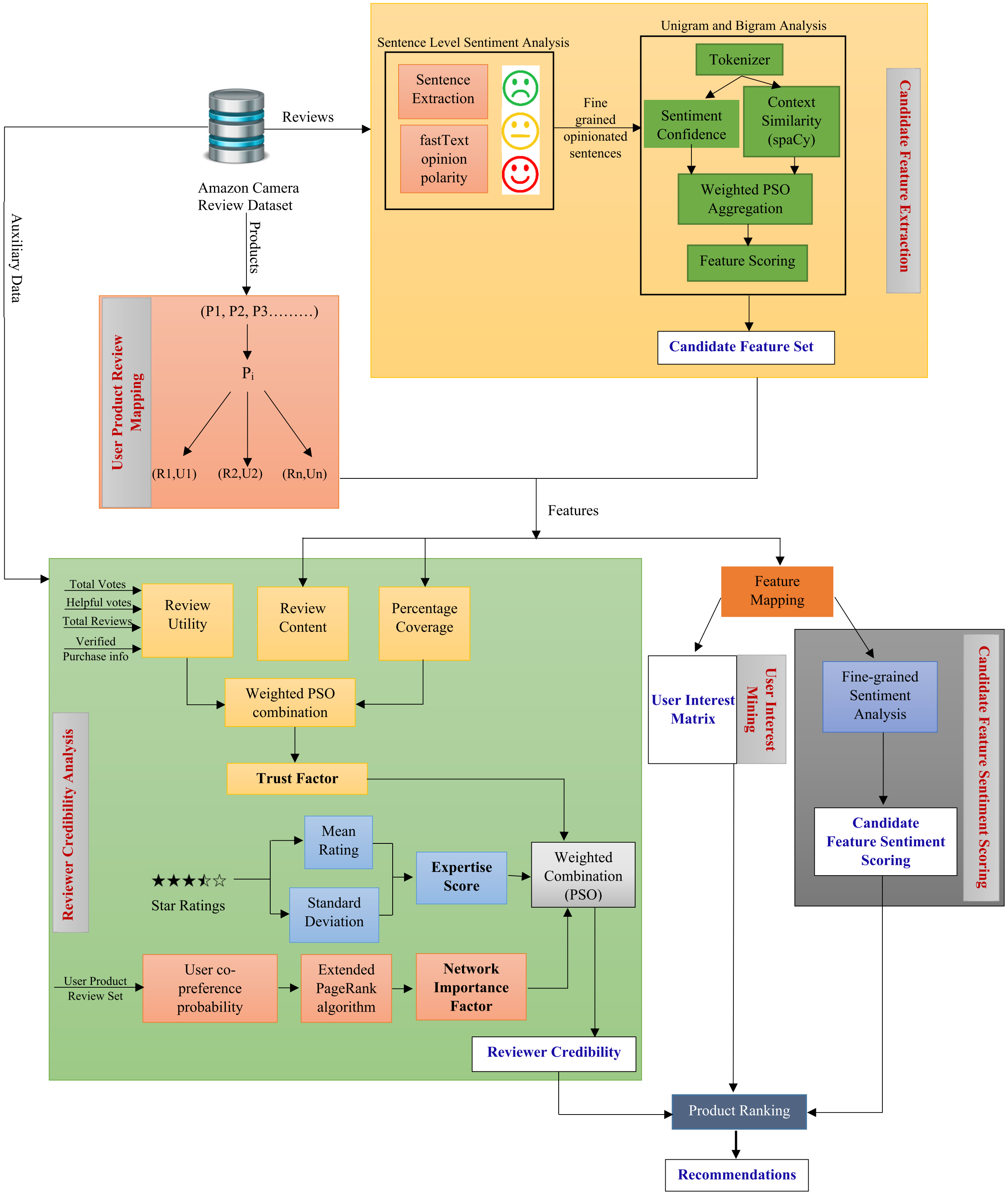

FIGURE 2. Architecture of CISER model.

B. CANDIDATE FEATURE EXTRACTION

User reviews comprise of multiple words, but not all words contribute to user's purchase preferences. A product feature is defined as 'any of the product's functionalities, capabilities, or its visual characteristics that may prove to be useful for its selection from alternatives'. For example, *'picture quality'*, *'battery-life'*, *'lens'*, *'resolution'* are some of the product features, in the context of a *'camera'*. This module uses

TABLE 3. Description of data attributes.

Column Name	Description
marketplace	2 letter country code of the marketplace where the review was written
customer_id	Random identifier that can be used to aggregate reviews written by a single author
review_id	The unique ID of the review
product_id	The unique Product ID the review pertains to. In the multilingual dataset the reviews for the same product in different countries can be grouped by the same product_id
product_parent	Random identifier that can be used to aggregate reviews for the same product
product_title	Title of the product
product_category	Broad product category that can be used to group reviews (also used to group the dataset into coherent parts)
star_rating	The 1-5 star rating of the review
helpful_votes	Number of helpful votes
total_votes	Number of total votes the review received
vine	Review was written as part of the Vine program
verified_purchase	The review is on a verified purchase
review_headline	The title of the review
review_body	The review text
review_date	The date the review was written

text-analytics for data reduction and extraction of features of importance, thereby enhancing the efficiency, speed and robustness of the system. 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

1) SENTENCE-LEVEL FINE-GRAINED SENTIMENT ANALYSIS

Sentiment Analysis is a quintessential text classification task which intends to categorize the opinion polarity of the incoming social media post to analyze the sentiment as positive, negative or neutral [35]–[38]. Text-driven sentiment analysis has been studied at several granularity levels. Recent literary works in this direction have focused on two key concepts which include aspect-based sentiment analysis [39] and fine-grained sentiment analysis [40]. Typically, aspect-based sentiment analysis goes one step further than sentiment analysis by automatically assigning sentiments to specific features or topics. It involves breaking down text data into smaller fragments, allowing more granular and accurate insights from your data. The fine-grained sentiment analysis involves sentiment scoring indicative of strong/weak sentiment intensities associated with the subtleties of human language. It breakdowns the sentiment into five discrete classes, namely, highly negative, negative, neutral, positive and highly positive.

In this study, fastText has been used for feature level fine grained sentiment analysis. fastText is a library for learning of word embeddings and text classification created by Facebook's AI Research lab. It has two major advantages. First, it takes into account the internal structure of words while learning word representations. Hence it is very useful for morphologically rich languages, and also for words that occur rarely. Second, fastText works well with n-grams, which makes it more suitable for our approach. Since we deal with real product reviews that capture complex human emotions, fastText provides the desired adaptability without compromising the space and time efficiency. The fine-grained sentiment rating is attached to all review sentences using fastText. Opinionated sentences are input to the next component, that is the unigram and bigram analysis.

2) UNIGRAM AND BIGRAM ANALYSIS

In this section, we explain how features are selected from the dataset. We focus on two types of product features [41]:

- **Unigram Product Features:** The product features comprising of a single word. Example: 'lens', 'resolution' etc.
- **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:
  - They should contain only nouns and adjectives. Specifically, 2-tuples of the form:
    - (Noun, Noun)
    - (Adjective, Noun) can only be considered as candidate features.
  - They should not contain any stop words or opinion lexicons.

Opinionated sentences (from section III-B1) are first tokenized to obtain word lemmas. These lemmas are scored by measuring their context similarity with the root ('camera' in our case), and probability of appearance in opinionated sentences (sentiment confidence). The context similarity is measured using spaCy, which is a library for advanced natural language processing in Python and Cython. Specifically, the "displaCy"[3] module of spaCy has been used for dependency parsing of sentences and target feature extraction. Fig. 3 depicts the dependency parsing of the sentence "The camera is very good" where opinion lexicon "good" maps to opinion target "camera".

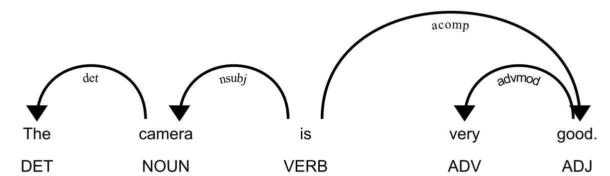

FIGURE 3. Dependency parsing and target feature extraction.

The sentiment confidence is derived using sentence sentiment ratings. For a given sentence, we first perform the opinion rating of the whole sentence (Section III-B1). From the sentence, target features are extracted (illustrated in fig. 3).

[3]https://spacy.io/universe/project/displacy

TABLE 4. Unigram features and feature scores.

Feature	Sentiment Confidence	Root Similarity Score
Lens	0.186	0.662
viewfinder	0.059	0.5984
photo	0.374	0.5354
Flash	0.6	0.5197
Mode	0.11	0.4129

TABLE 5. Bigram feature extraction and feature scores.

Feature	Sentiment Confidence	Root Similarity Score
Picture quality	0.6	0.5668
Battery life	0.56	0.478
Flash photography	0.4	0.543
Auto setting	0.3	0.42
Image processing	0.6	0.54

The target features are assigned confidence values based on the opinion rating. Equation (1) is used for measuring the sentiment confidence (SC):

$$SC_{f_i} = \frac{|OS_{f_i}|}{|OS|} \tag{1}$$

where,

SC_{f_i} = Sentiment Confidence for feature f_i
$|OS_{f_i}|$ = Number of Opinion sentences (Opinion Rating $\neq 3$) containing feature f_i
$|OS|$ = Total Number of Opinion Sentences

Context Similarity score and Sentiment Confidence score are linearly combined using PSO before feature ranking, for most optimising feature extraction results. Some examples of features, their corresponding sentiment confidence score and root-similarity score are tabulated in Table 4 (Unigrams) and Table 5 (Bigrams).

C. USER PRODUCT REVIEW MAPPING MODULE

This module processes the review-dataset. Product data is extracted from the Amazon camera review dataset. All the product reviews and corresponding reviewers are obtained and further mapped with the product. The mapped data is input to the three modules namely user interest mining module, reviewer credibility analysis module and candidate feature sentiment scoring module.

D. USER INTEREST MINING

Many state-of-the-art recommendation systems like Netflix, Amazon and Flipkart maintain user interest profiles so as to recommend products based on their preferences, likes and dislikes. 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 he makes a frequent mention of it in his reviews. Frequency of allusion to specific features asserts their utility for users. It is highly unlikely that a user acknowledges feature f_i, but never touches upon it while reviewing the product. Hence, each user is associated with a user-interest matrix (UIM) which is a matrix estimating user-feature adherence. We use probabilistic values to estimate user interest as against hard binary (yes/no) values, since probabilistic values are a better representative of user-feature affinity. Binary affinity estimates depict a simplistic view of user preference trends that may not always capture the subtleties of human aesthetics. Hard (yes/no) heuristics are useful in cases where frequency of product purchase is low (eg. Furniture) as the number of reviews per user are limited. In all other cases probabilistic estimates improve user-interest modelling. Therefore, UIM is populated with feature count probability (equation (3)).

The equations (2) & (3) gives the mathematical representation.

$$UIM_u = [a_1, a_2, a_3, \ldots a_n] \tag{2}$$

$$a_i = \frac{FC_u^i}{|\, R_u \,|} \tag{3}$$

FC_u^i = number of times user u mentions feature i in reviews
n = number of features in CFS(Candidate Feature Set)
$|R_u|$ = number of reviews written by user u

After user preference profiling, the next step is to quantify the quality of preferred (candidate) features for the products [H2]. This is done using credibility weighted feature sentiment scoring. The following sections detail reviewer credibility analysis and candidate feature sentiment scoring modules that ultimately lead to product ranking and recommendations.

E. 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)
- Network Importance Factor

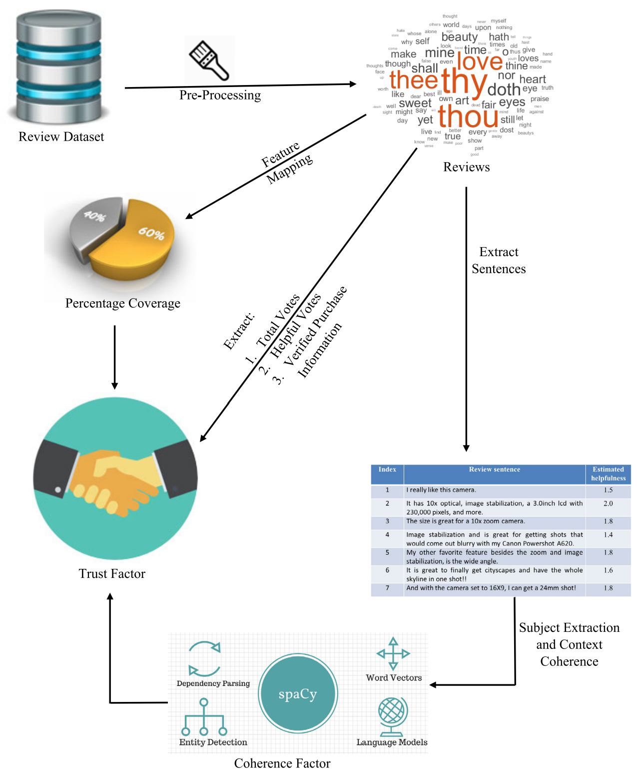

FIGURE 4. **Trust factor calculation.**

The following subsections expound the details of each of these factors.

1) TRUST FACTOR

Social psychology theory [42] proposes that role of a person in specific realm is indicative of the trust garnered by individual. This is the famous *"Follow The Leader"* [32] approach, whereby using some aspects of available information about individuals, we associate trust/credibility values with them. Following this strategy, we present a methodology for trust factor calculation for all reviewers. The following key-elements devise the trust factor of reviewer:

- Review Utility
- Review Content
- Percentage Coverage

Fig. 4 illustrates the trust factor calculation.

1) *Review Utility*
Review utility quantifies the user-perception about the review in consideration. It is measured using the following auxiliary factors extracted from dataset:
   - **Total Votes(γ_{TV}):** This attribute is indicative of total number of potential buyers that have read and reviewed the product review in consideration.
   - **Helpful Votes(γ_{HV}):** This value evocates the number of potential buyers who found the current product review useful in formulating purchase-decision. Metzger [43] suggests customers instinctively vote for reviewers having higher source credibility. Hence, higher the number of helpful votes possessed by a reviewer, higher is the trust that fellow customers associate with him.
   - **Verified Purchase Information(γ_{VPI}):** Generally, potential buyers attach greater importance to reviews where reviewer has actually used the product, as against reviewers commenting without prior-knowledge of product features. Verified Purchase Information is a Boolean attribute that assures a reader of purchase-status of reviewer.
   - **Total Ratings(γ_{TR}):** This value is a measure of total number of times the reviewer has rated products of the product-domain in consideration. It is a value derived from user product review mapping. This is a direct indicator of the association of the reviewer, with products of that type. Hence, it can easily be said that frequent reviewers are more likely to be domain experts, who can formulate general opinion about the product, when compared to occasional reviewers [5], [31].

2) *Review Content*
The quality of review content can be assessed using:
   - **Coherence Factor:** The coherence of text is contemplated by its syntactic and semantic structure and relations. Research and language experts suggest that language has a definitive assembly. Language expression says a lot about the intent of reviewer and quality of the review on the whole. As Sheng Tun Li [4] mentions in his work, several linguistic indicators like capitalisation, emoticons, spelling errors, semantics, regularity, consistency etc. play a major role in devising user credibility. We propose coherence factor as an indicator of relevance and hence quality of review. We focus on semantics, regularity and consistency of the review. Precisely, coherence factor measures the context-pertinence and consistency of sentences in reviews, which is a determinant factor for reviewer credibility assessment. It measures the context similarity of review sentences with candidate features, and thereby measures the consistency of context in reviews. Equations (4), (5) and (6) help calculate the coherence factor.

$$coherence_{review_{id}} = \sum_{i=1}^{|S|} \frac{cs\,(sub_i)}{|S|} \tag{4}$$

where,

$|S|$ = Total number of sentences in the review corresponding to review id
$cs(k)$ = context similarity of k with root context calculated using spaCy
sub_i = subject of i^{th} sentence

$$\mathrm{cs}\,(\mathrm{sub_i}) = \max\,(\mathrm{sim}\,(\mathrm{sub_i},\ \mathrm{feat})) \tag{5}$$

where,

$max(sim(sub_i, feat)) > sim(sub_i, feat)\ \forall$
feat ∈ feature set (section III-B)
sim(sub_i, $feat$) = context similarity of sub_i
with feat (feature in consideration)

$$CF_{user_{id}} = \sum_{\substack{x \in \{review_{id}: \\ reviewer(review_{id}) = user_{id}\}}} \left(\frac{coherence_x}{|r|}\right) \tag{6}$$

where,
$CF_{user_{id}}$ = Review coherence factor pertaining to current reviewer
|r| = Number of reviews of current reviewer

- *Percentage Coverage(PC)*
  This factor measures what percentage of candidate features are referred to by the reviewer in his reviews (running average measure). Expert reviewers critically analyse products over various baseline features (Hence their reviews possess high percentage coverage). In turn, their reviews influence readers and apprise them about the downfalls and salient perspectives of product. Mathematically, this is characterized using (7).

$$S_{noun} = \text{set of nouns proper nouns in review}$$
$$S_{features} = \text{set of candidate features (section III-B)}$$
$$PC_{review_{id}} = \frac{S_{noun} \bigcap S_{features}}{|S_{features}|} \times 100 \tag{7}$$

Using these 6 key factors, methodology for calculation of trust factor(TF^u) for each user(u) is mathematically illustrated in (8).

$$\begin{aligned} TF^u = w_{TR} \times \gamma_{TR_u} + &\sum_{x \in \{review_{id}: reviewer(review_{id}) = u\}} \\ &\times \Big\{ \Big(\gamma_{VPI_x} + |1 - \gamma_{VPI_x}| \times w_{VPI} \Big) \\ &\times \Big(w_{cc} \times coherence_x \\ &\times PC_x + w_v \times (2 \times \gamma_{HV_x} - \gamma_{TV_x}) \Big) \Big\} \end{aligned} \tag{8}$$

where,
w_{VPI}, w_{cc}, w_v, w_{TR} are weights assigned to unverified purchases (where the reviewer did not buy the product), Coherence-Coverage product, Useful (i.e. Helpful-Unhelpful) Votes and Total Ratings respectively.

2) EXPERTISE FACTOR

Expertise is a key aspect for user credibility analysis, it is defined as the degree of a user's competency to provide accurate ratings and exhibit high activity [44] There are various dimensions to this definition: (1) Expertise is directly related to *user's competency* to present the general opinion, by means of his reviews and ratings; (2) The second key aspect is *high activity exhibition*, which means domain experts are more active in reviewing newly launched products and their assessment is generally highly appraised by public.

In this study, we propose a novel measure for expertise measurement. The motivation for Expertise factor has been taken from [31], but it has been modified to meet some of its original shortcomings. The following terminology will be crucial for further calculations and derivations.

1) *Representative Rating*
   Representative rating (RR) is a quantitative indication of the opinion of masses. It is a quantity that represents product rating, such that it is inclusive of all the reviews for the product. In previous studies [31] mean rating is used as representative rating but there are several downfalls to this approach, the most important of them being mean rating ignores the standard deviation of distribution. Mean is a good representation of a set, if data is uniformly distributed. For skew data, mean is an approximation, but better estimates can be made. To deal with this problem, we add a quantity α to mean, to model each type of data in a more balanced manner, as given in (9):

$$\alpha_i = \frac{\left(\sum_{\gamma \in X_{1i}} \gamma\right) - \left(\sum_{\gamma \in X_{2i}} \gamma\right)}{2 \times sd_i \times |X_i|} \tag{9}$$

X_i = Set of Product Ratings for Product i
μ_i = Mean of Product Ratings for Product i
X_{1i} = $\{x \in X_i$ and $x > \mu_i\}$
X_{2i} = $\{x \in X_i$ and $x < \mu_i\}$
sd_i = standard deviation of product ratings for i
RR_i = $\mu_i + \alpha_i$

α shifts the mean representation towards the more heavily voted side i.e. mean is shifted towards majority opinion.

Thus, expertise factor for a user(u) for a specific product(i) (EF_i^u) is defined as proximity between user's perspective rating and the representative rating (RR_i). Mathematically, expertise factor is given by (10).

$$EF_i^u = 1 - \frac{abs\left(RR_i - R_i^u\right)}{5} \tag{10}$$

where,
5 is the scaling factor
$0 < RR_i \leq 5$ and $0 < R_i^u \leq 5$ hence,
$0 \leq abs(RR_i - R_i^u) \leq 5$
R_i^u = User(u) Rating for product i
(EF_i^u) is used to derive EF^u i.e. expertise factor for user u. EF^u is defined as average of (EF_i^u) over all products i, rated and reviewed by user u.

3) NETWORK IMPORTANCE FACTOR

Network Importance Factor (NIF) is a measure of the influence of a particular user in defining overall public assessment of a product. In this work, we modify PageRank algorithm [45] which is conventionally used for devising webpage ranks, to uncover the most influential reviewers. To understand the conception of NIF and it's relation to webpage ranking algorithm, readers need to have a basic understanding of PageRank algorithm. The underlying assumption in PageRank algorithm is "more important websites are likely to receive more links from other websites", on similar lines, the fundamental antecedent behind the conception of NIF is " The most influencing reviewer is likely to share product preferences, with those of people they influence" hence leading to the formula given in (11).

$$
\begin{aligned}
&NIM \text{(Network Importance Matrix)} \\
&\quad = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix} \\
&a_{ij} \\
&\quad = \begin{cases} \dfrac{\left|P[u_i] \cap P[u_j]\right|}{\left|\{x_k \mid P[x_i] \cap P[x_k] \neq \phi\}\right|} & : \left|\{x_k\}\right| \neq 0 \\ \dfrac{1}{N} & : \text{otherwise} \end{cases} \\
&NIF_0 \\
&\quad = \begin{bmatrix} 1.0 & 1.0 & \cdots & 1.0 \end{bmatrix}_{1\times N} \\
&NIF_k \\
&\quad = (1-d) \times d \times (NIM \times NIF_{k-1})
\end{aligned} \tag{11}
$$

where,

d is a damping factor between 0 and 1

a_{ij} is User co-preference probability of users i and j

$P[u_i]$ is the Set of Products reviewed by user(u_i)

$NIF_k[u_i]$ is the Network Importance Factor of user(u_i) at k^{th} iteration.

Thus, the three Factors i.e. Trust Factor, Expertise Factor and Network Importance Factor are linearly combined using PSO for final credibility scores assignment.

F. CANDIDATE FEATURE SENTIMENT SCORING

CISER attempts to combine user-preference and sentiment value for increasing the efficacy of recommendations. It aims to suggest those products that have a better sentiment value compared to the products previously purchased by the user. For this, product-level candidate feature sentiment scoring is done using fastText. The candidate features (present in CFS) are searched in product review, and then the sentence level fine-grained opinion (fastText) is attached to those features. These feature-sentiment scores are then combined with reviewer credibility for product ranking. The details of product ranking are discussed in the next section.

G. PRODUCT RANKING AND RECOMMENDATION METHODOLOGY

The last step of this study focusses on ranking each product among contextually similar products. This calls for combining reviewer credibility, user-interest and candidate feature sentiment for product scoring. Credibility weighted fine grained feature sentiment (for user-preferred features) is averaged over all product-reviews to calculate product rank. Products with highest ranks are recommended to a user. CISER is robust to user prejudice, since credibility normalises personal bias based on perspective knowledge (Expertise Factor). Also, fine grained feature sentiment analysis is in line with heuristic H2, as even the feature level acceptability-details are captured. User interest matrix stands consistent with the premise of heuristic H1. It ensures that only those features are considered for product ranking which are important to User U. Hence, CISER is an attempt to automate credibility driven feature quality assessment using electronic word of mouth. Algorithm 1 presents the recommendation algorithm.

Algorithm 1 Recommendation Algorithm

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

Output: Recommended Products: *recommendedProducts*

1 $g(review, feature) \leftarrow \begin{cases} 1 & : feature \in review \\ 0 & : \text{otherwise} \end{cases}$

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

3 $recommendedProducts \leftarrow$ Top K products according to *rank*

IV. WORKING EXAMPLE

In this section we will discuss an example to enable better understanding of the workflow of CISER model for product recommendation.

A. CANDIDATE FEATURE EXTRACTION

As discussed, this phase has two components, namely, sentence-level fine-grained sentiment analysis and Unigram & Bigram analysis.

1) SENTENCE LEVEL FINE GRAINED SENTIMENT ANALYSIS

This module takes a user review as an input, extracts sentences and then associates fine-grained sentiment values with each sentence as shown in table 6.

2) UNIGRAM AND BIGRAM ANALYSIS

Sentiment determining unigram (Noun / Proper - Noun) and bigram features [(Noun - Noun) / (Adjective - Noun)] are extracted in this phase. The extraction depends on sentiment confidence and similarity scores. Suppose, the unigram candidate feature is 'zoom'. Suppose 1000 review sentences contain 'zoom' out of total 20,000 opinion sentences. So, the calculated sentiment confidence of zoom is

TABLE 6. Fine-grained sentiment analysis.

Review	Sentence Extraction	Opinion Score
Amazing telescope! Main reasons why? Fully Automated with amazing zoom. Awesome Picture Quality. But the battery life was not up to the mark.	Amazing Telescope	5
	Main reasons why	3
	Fully Automated with amazing zoom	5
	Awesome Picture Quality	5
	But the battery life was not up to the mark.	2

TABLE 7. Coherence Factor (CF) calculation.

Review	Review Sentence	Subject	Closest Feature	CS(subj)
The picture quality is awesome. My wife really liked it.	The picture quality is awesome	picture-quality	picture-quality	1
	My wife really liked it	wife	photo	0.15

1000/20,000 = 0.05. The spaCy similarity of 'zoom' with root-context 'camera' is 0.56. These values are linearly combined using PSO for feature ranking and extraction.

B. USER INTEREST MINING

This module takes as input the candidate features (Section III-B) and the user product review mapped data (Section III-C), user interest matrix is produced as an output. Let the only review of user U101 be "The picture-quality is awesome. Also, the zoom, lens and battery-life are magnanimous". And the candidate feature set be {"picture-quality", "zoom", "lens", "battery-life", "photo"}. The pre-stated review adds {"picture-quality", "zoom", "lens", "battery-life"} to the set of user interest, hence leading to user interest vector for U101 as:

$$[1\ 1\ 1\ 1\ 0]$$

where the columns correspond to {"picture-quality", "zoom", "lens", "battery-life", "photo"} respectively. This calculation follows the heuristic H1. Following the same lines, user interest matrix is calculated for all the users and fed as an input to ranking and recommendation module.

C. REVIEWER CREDIBILITY ANALYSIS (RCA)

The credibility aspects related to review utility are extracted from data-set and described in section III-E1. The other associated aspects are:

1) COHERENCE FACTOR

For calculating review coherence, the review body and candidate feature set is given as input and the associated coherence factor is obtained as output as shown in table 7.

So, the CF for the review $= (1 + 0.15)/2 = 0.575$ CF_u is average CF over all the reviews written by user U that pertain to the particular product domain.

2) PERCENTAGE COVERAGE

This factor measures what percent of candidate features are mentioned in a review. The review and set of features are given as an input to this module, percentage coverage of the review is received as output. Table 8 summarizes the process.

TABLE 8. Percentage coverage calculation.

Review	Mapped Features	Total Features	Percentage Coverage
The picture-quality is awesome. Also, the zoom, lens and battery life is magnanimous.	picture-quality, zoom, lens, battery-life	37	4/37 = 0.108

TABLE 9. Product ratings.

User ID	Star Rating
U101	2
U102	5
U103	5
U104	4

3) EXPERTISE FACTOR

Expertise factor depends on the star-ratings given by users to products. As an illustration, let the star ratings for product P_i be as depicted in Table 9. The mean rating for product $P_i(\mu_i) = 4$ and standard deviation (sd_i) is 1.414. Number of ratings $< (\mu_i) = 1$ and number of ratings $\geq (\mu_i)$ are 3. The calculation of α_i is as follows using (9).

$$\alpha_i = \frac{\left(\sum_{\gamma \in X_{1_i}} \gamma\right) - \left(\sum_{\gamma \in X_{2_i}} \gamma\right)}{2 \times sd_i \times |X_i|} = \frac{3-1}{2 \times 1.414 \times 4} = 0.1767$$

The representative rating $(RR_i) = (\mu_i + \alpha_i) = 4 + 0.1767 = 4.1767$

$$EF_i^u = 1 - \frac{abs\left(RR_i - R_i^u\right)}{5}$$

So, for user U101, $EF_i^u = 1 - \frac{abs(4.1767-2)}{5} = 0.5646$
Similarly, expertise scores for all the users are evaluated.

4) NETWORK IMPORTANCE FACTOR

Network Importance Matrix (NIM) is needed for the calculation of NIF for all users. The calculation of each element of NIM is illustrated in table 10.

Taking the example as in table 10, it is clear that both U1 and U2 rated P2, Hence $P[U_1] \cap P[U_2] = 1$ and $|x_1| = 2$ as U2 and U3 share at least 1 review preference with U1.

$a_{12} = 1/2 = 0.5$

Similarly, the NIM matrix is calculated.

TABLE 10. Product review mapping.

User ID	Products Reviewed
U1	P1, P2, P3, P4
U2	P2, P5
U3	P3
U4	P5

$$\text{Let the final NIM Matrix be} \begin{bmatrix} 1 & 0.2 & 0.2 & 0.1 \\ 0.35 & 1 & 0.1 & 0.1 \\ 0.5 & 0.2 & 1 & 0.1 \\ 0.5 & 0.1 & 0.2 & 1 \end{bmatrix}$$

$$NIF_0 = [1.0 \;\; 1.0 \;\; \ldots \;\; 1.0]_{1\times N}$$
$$d = 0.8$$

The first iteration for calculation of NIF is as shown:

$$\begin{aligned} &NIF_1 \\ &= (1-d)/4 + d \times (NIM \times NIF_0^T) \\ &= \frac{0.2}{4} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 0.8 \cdot \begin{bmatrix} 1 & 0.2 & 0.2 & 0.1 \\ 0.35 & 1 & 0.1 & 0.1 \\ 0.5 & 0.2 & 1 & 0.1 \\ 0.5 & 0.1 & 0.2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1.25 \\ 1.29 \\ 1.49 \\ 1.49 \end{bmatrix} \end{aligned}$$

The other iterations can be done similarly for calculating Network Importance Factor for each User.

D. CANDIDATE FEATURE SENTIMENT SCORING

Fine-grained sentiments are associated with all the candidate features present in review. Table 11 illustrates the process:

TABLE 11. Candidate feature sentiment scoring.

Review	Feature in review	Feature Sentiment
The picture-quality is awesome. Also, the zoom, lens and battery-life are magnanimous.	picture-quality	4
	zoom	5
	lens	5
	battery-life	5

E. PRODUCT RANKING AND RECOMMENDATION

In this section we illustrate the contribution of one review towards product ranking as shown in table 12.

Table 12 represents a review for product P1 given by U101. It clearly illustrates that user in consideration (For whom product is being ranked) is interested in "picture-quality" and "zoom" and so only those feature sentiments contribute to product rank.

Review Contribution $= (0.7) \times (4+5)/2 = 3.15$

These review contributions are aggregated as presented in Algorithm 1 in section III-G for final product scoring and recommendation.

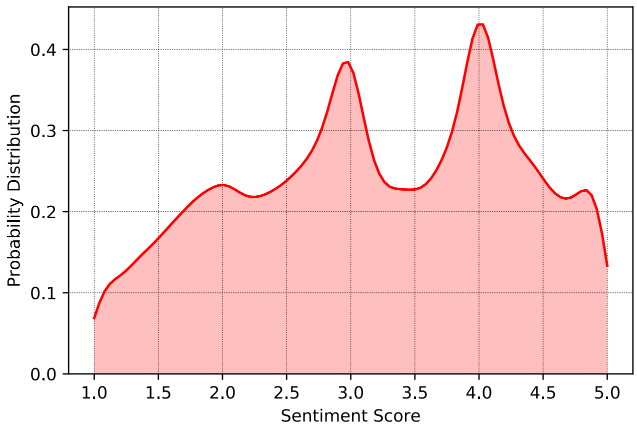

FIGURE 5. Sentiment score distribution.

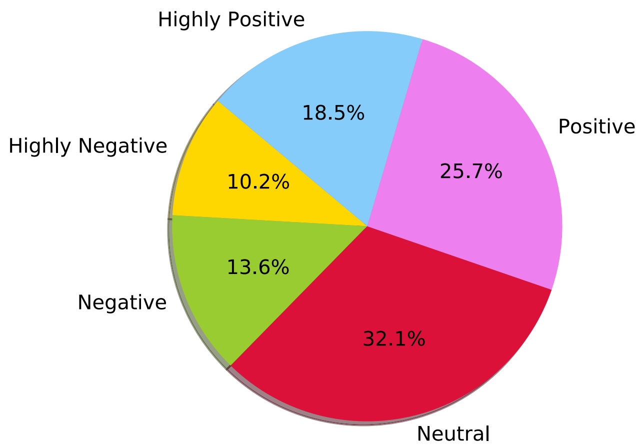

FIGURE 6. Percentage sentiment distribution.

V. EXPERIMENTATION AND RESULT

Amazon camera dataset has been used for evaluating the efficacy of our approach. 3500 reviews given by 1000 users over 10 products are used for the validation phase to predict recommendation accuracies. The sentiment distribution of validation data has been presented in fig. 5 and 6.

The following sub-sections discuss recommendation performance results, hyper-parameter tuning and comparison with baseline models.

A. RECOMMENDATION PERFORMANCE RESULTS

In order to assess the utility of our recommendation system, we have used Mean Average Precision (MAP) as the evaluation metrics. MAP is an extension of Average Precision (AP) where we take average of all AP's to calculate MAP. Average precision is a measure that combines recall and precision for ranked retrieval results. The mathematical equations for the metrics MAP and AP are listed in Table 13.

Specifically, we are using MAP@N as the metrics, where N is the number of products recommended to each user. Fig. 7 shows the variation of Mean Average Precision with Number of Recommendations. For single product recommendation,

TABLE 12. Review contribution to product rank.

Review	Reviewer Credibility	Feature in review	User Interest value for feature	Feature Sentiment
The picture-quality is awesome. Also, the zoom, lens and battery-life are magnanimous.	0.7	picture-quality	1	4
		zoom	0	5
		lens	1	5
		battery-life	0	5

TABLE 13. Metrics for recommendation efficiency.

Metrics	Formula
Average Precision@N	$\frac{1}{N} \times \sum_{k=1}^{N} (\text{Precision}(k) \times \Delta\text{Recall}(k))$
Mean Average Precision@N	$\frac{1}{\lvert U \rvert} \times \sum_{\text{user} \in U} (AP@N)_{\text{user}}$

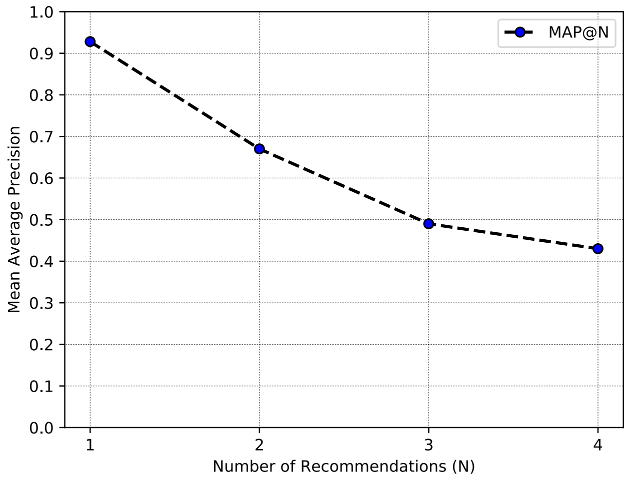

FIGURE 7. MAP vs number of recommendations.

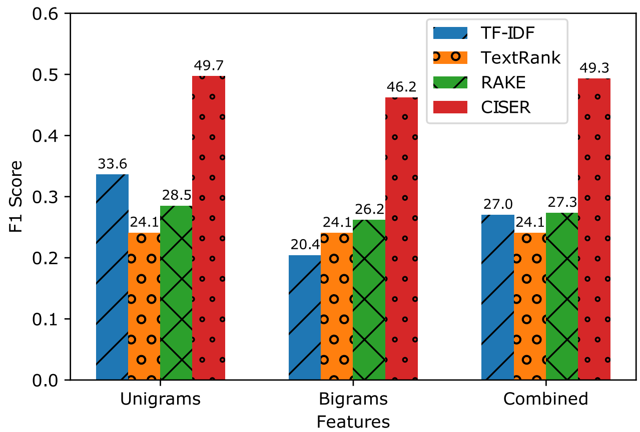

FIGURE 8. F1 score comparison with baselines.

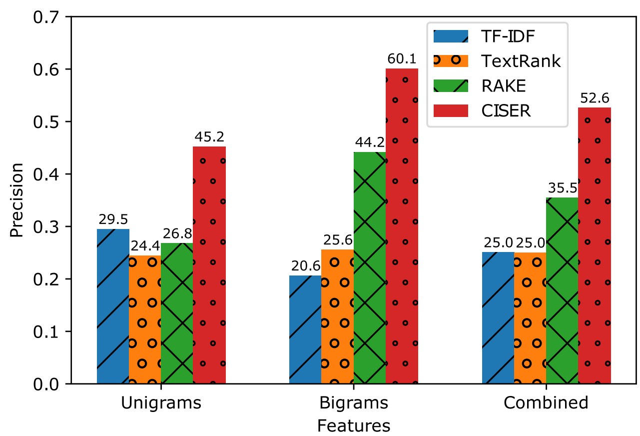

FIGURE 9. Precision score comparison with baselines.

MAP of our model is 0.93, which means that in 93% cases the user bought the recommended product. The decline in MAP@N values with increasing N can be explained by the fact that it is not very common for a buyer to buy many cameras. Hence, we limit our discussion to 4 products, so that the results remain meaningful. For different product domains, different number of recommendations can lead to remarkable results, depending on domain type and user perception.

Candidate Feature Extraction is a significant module of CISER. The feature selection results have been compared to basic NLP models like tf-idf, textRank [46] and RAKE [47]. These models were tested on Amazon's Camera review dataset. For Unigrams, Nouns were extracted from reviews and then fed to models so as to avoid stop-words and irrelevant words dominating top ranked frequent features. For Bigrams, Noun-Noun pairs and Adjective-Noun Pairs were extracted and only these were used for purpose of testing, to maintain morphological uniformity in results. The results obtained are shown in Figures 8, 9, 10. CISER has a significant precision recall and F1 score gain as compared to previous models.

One of the key contributions of this work is the reviewer credibility analysis given their opinions for a product. In order to augment the efficiency of RCA, we have used network importance factor, trust factor and expertise factor (section III-E), to integrate various latent factors for reviewer assessment. Fig. 11 shows a bar graph between reviewer rank and factor scores, demonstrating factor score distribution. It is also worth mentioning that reviewer rank is the combined total score of the reviewer, which is calculated as the weighted average of the three factors mentioned (section III-E). So, a reviewer R1 with a higher trust factor score as compared to another reviewer R2 can have a lower rank than R2, if expertise factor and network importance factor scores of R2 are sufficiently higher than that of R1.

Fig. 12 shows a graph between reviewer credibility (section III-E) and reviewer rank. Reviewer credibility is calculated as the weighted average of NIF, EF and TF.

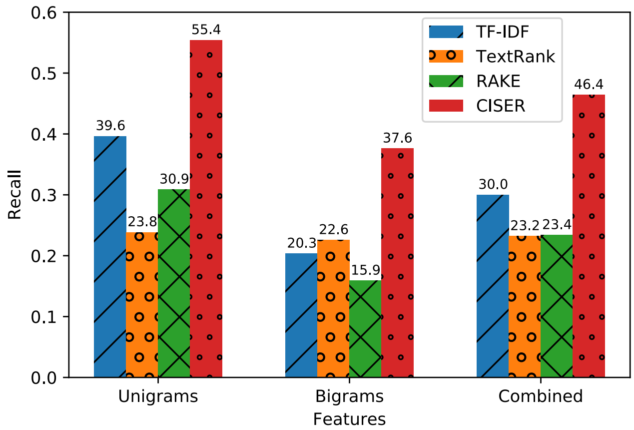

FIGURE 10. Recall score comparison with baselines.

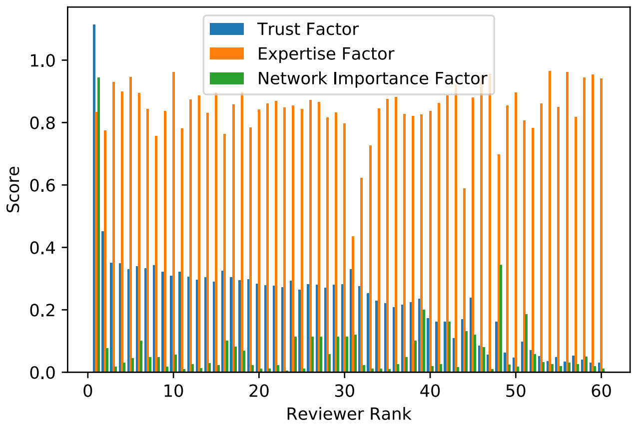

FIGURE 11. Reviewer rank vs factor Scores.

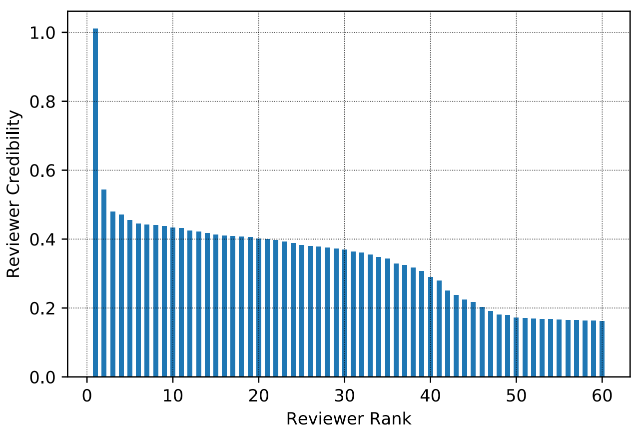

FIGURE 12. Reviewer rank vs reviewer credibility.

The weights are tuned so as to maximise recommendation efficiency.

The distributions demonstrated in figures 11 and 12 show the variance in factor scores and credibility scores for users, hence supporting our hypothesis that not all users are equally credible. To the best of our knowledge, this concept of reviewer credibility analysis without social network information (like number of followers, marital status, religion etc.) has not been used in any previous works in the field of product recommendation.

The comparative analysis of various latent factors contributing to reviewer credibility analysis is demonstrated in Figure 13. Using mean product ratings for evaluating expertise factor (ES_{Mean}) [31] does not give appreciable results. Addition of the proposed α factor (representative rating) to mean product rating (ES_{RR}) slightly improves the MAP performance of CISER. Hence, ES_{RR} better models reviewer expertise compared to ES_{Mean}. It is evident that trust factor and network importance factor greatly contribute towards recommendation performance. An efficient linear combination of these three factors leads to superior performance.

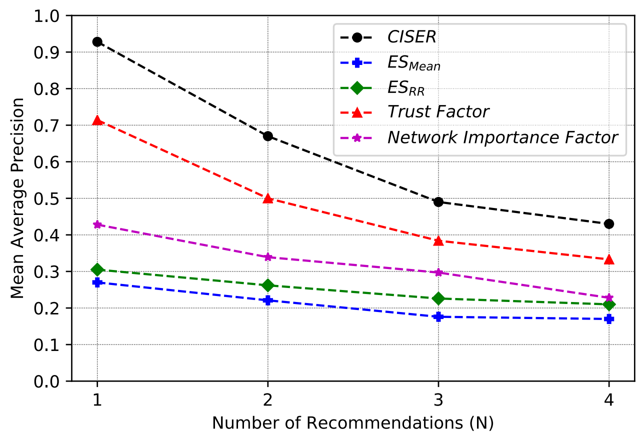

FIGURE 13. Comparative analysis of factors of RCA.

B. HYPER-PARAMETER TUNING

As discussed in Sections III-B and III-E, we use PSO for tuning coefficients for linear combination of feature scores and reviewer credibility factor scores. We empirically evaluate the tuning performances of PSO and other swarm optimization algorithms like Cuckoo Search, BAT and Artificial Bee Colony Optimisation. The comparative analysis is presented in fig. 14. The results support our choice of PSO over others, since PSO provides average MAP gain of 5%.

Various feature scores were combined using PSO. Optimal Parameter tuning is crucial for achieving remarkable performance. Since, PSO itself is a parameterised model, grid search has been performed to tune PSO parameters. The heatmap demonstrating the results of grid-search is presented in Figure 15 and final values for PSO parameters are enlisted in Table 14.

The hyper-parameters that were tuned using PSO are tabulated in Tables 15 and 16. Table 15 enlists weight

TABLE 14. PSO parameters.

Parameter	Value	Root Similarity Score
ω	0.9	0.5668
$c1$	1.6	0.478
$c2$	3.1	0.543

TABLE 15. Weights for trust factor.

Trust factor				
Unverified purchase		Verified purchase		total_ratings
coherence_coverage	helpful_votes	coherence_coverage	helpful_votes	
0.8028	0.1964	0.8963	0.2193	0.4269

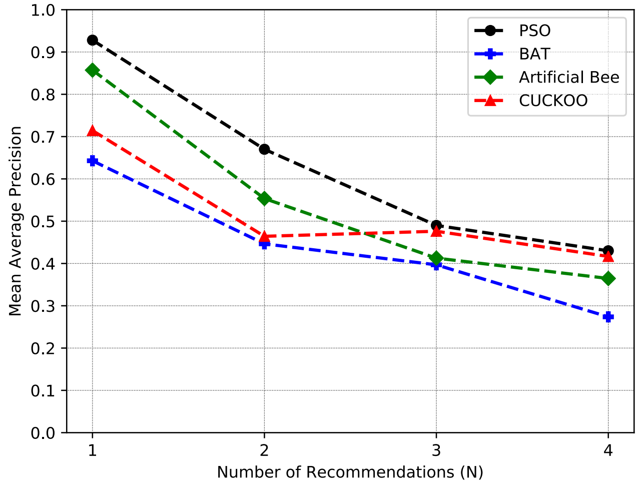

FIGURE 14. Comparison of various swarm optimisation algorithms.

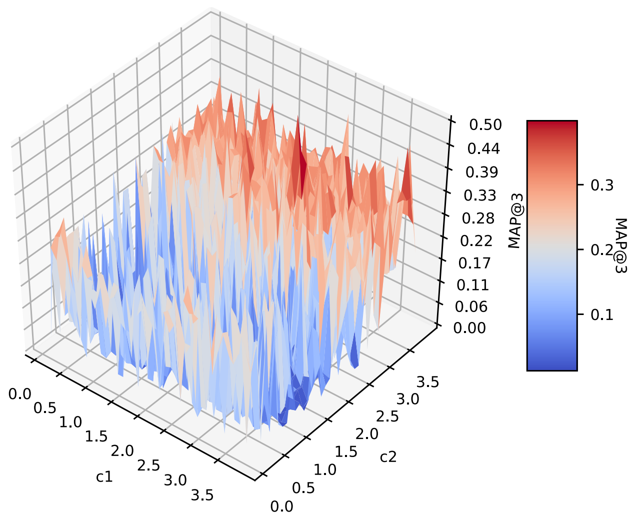

FIGURE 15. PSO parameter tuning.

TABLE 16. Factor weighting for credibility assignment.

Trust factor	Expertise factor	Network importance factor
0.6154	0.1431	0.2362

for calculating trust factor (Section III-E). Table 16 states weights for linear combination of trust factor, network importance factor and expertise factor for calculating reviewer credibility.

C. COMPARISON WITH BASELINE MODELS

To ascertain that our model performs substantially better than the current systems, we compared our results with some baseline recommendation systems. Content based

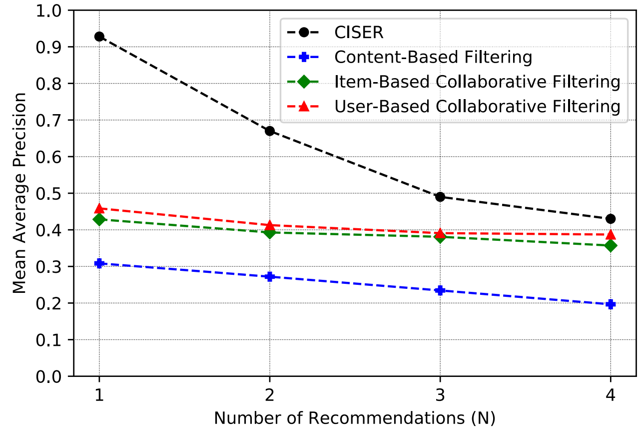

FIGURE 16. Comparing recommendation baselines with CISER.

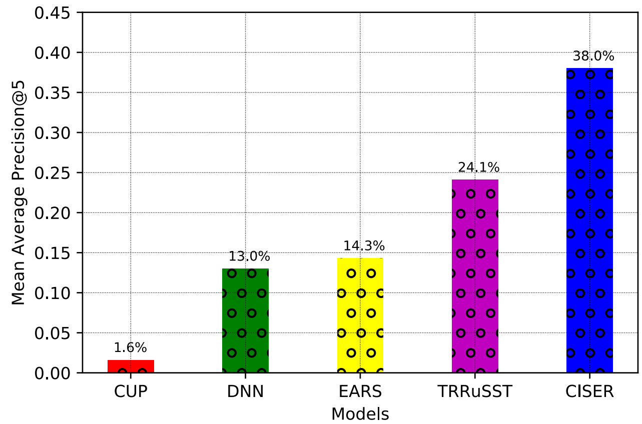

FIGURE 17. Comparison with state-of-the-art systems.

filtering [3], item-based collaborative filtering [2] and user-based collaborative filtering [2] manipulate numerical ratings to decipher user preference patterns. The proposed CISER model performs appreciably better than these, hence supporting our proposition that review sentiments greatly impact user purchase decisions. The results are presented in figure 16.

Figure 16 makes it evident that recommendation precision of CISER is commendably better than conventional content-based filtering (gain = 30%), item-based collaborative filtering (gain = 25%) and user-based collaborative filtering (gain = 24%).

Figure 17 compares the proposed CISER model with previous research models, namely, CUP [20], DNN [14], TRRuSST [17] and EARS [18].CUP and DNN model user-preference trends, EARS uses emotional and social-information for recommendations, whereas TRRuSST quantifies product quality by manipulating external attributes modelled on user-reviews. It is clear that CISER improves recommendation efficacy with MAP@5 = 38% i.e. a gain of approximately 14% (compared to TRRuSST [17]). Hence, the combination of user-interest modelling, sentiment analysis, product quality quantification aggregated with reviewer credibility analysis significantly improves upon the state-of-the art recommendation systems.

VI. CONCLUSION AND FUTURE WORKS

This research successfully counteracted the problem of reducing the product search space for customers. The proposed CISER model recommends products with commendable accuracy, without the need of collecting and processing alternative data from social networks, surveys etc. The model augments heuristic-driven user interest profiling with reviewer credibility analysis and fine-grained feature sentiment analysis to devise a robust recommendation methodology. To the best of our knowledge, this is the first work that integrates credibility driven feature-based fine-grained sentiment analysis with user modelling for online product recommendation. The remarkable performance of CISER suggests that recommender systems that combine user-profiling, reviewer credibility and feature sentiment scoring can prove to be a methodical successor to the present state-of-the-art systems. The work has also opened avenues of using sentiment analysis for product quality assessment and enhancing recommendation quality using reviewer trust, expertise and influence. Results were validated using the Amazon camera dataset and the proposed model outperformed the baselines. As a possible future direction, social network information and online activity logs can be used to enhance the user credibility model and build shopping cliques for product recommendation. Further, more sophisticated measures for representative rating and expertise can be devised.

REFERENCES

[1] J. A. Konstan and J. Riedl, "Recommender systems: From algorithms to user experience," *User Model. User-Adapted Interact.*, vol. 22, nos. 1–2, pp. 101–123, Apr. 2012, doi: 10.1007/s11257-011-9112-x.

[2] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," in *The Adaptive Web* (Lecture Notes in Computer Science), vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Cham, Switzerland: Springer, 2007.

[3] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," in *The Adaptive Web* (Lecture Notes in Computer Science), vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Germany: Springer, 2007.

[4] S.-T. Li, T.-T. Pham, and H.-C. Chuang, "Do reviewers' words affect predicting their helpfulness ratings? Locating helpful reviewers by linguistics styles," *Inf. Manage.*, vol. 56, no. 1, pp. 28–38, 2019.

[5] V. Srivastava and A. D. Kalro, "Enhancing the helpfulness of online consumer reviews: The role of latent (content) factors," *J. Interact. Marketing*, vol. 48, pp. 33–50, Nov. 2019.

[6] B. Liu, *Sentiment Analysis and Opinion Mining*. San Mateo, CA, USA: Morgan & Claypool Publishers, 2012.

[7] Amazon Web services, inc. *Amazon Customer Reviews Datase*. Accessed: Dec. 18, 2019. [Online]. Available: https://s3.amazonaws.com/amazon-reviews-pds/readme.html

[8] S. Raza and C. Ding, "Progress in context-aware recommender systems—An overview," *Comput. Sci. Rev.*, vol. 31, pp. 84–97, Feb. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1574013718302120

[9] L. Chen, G. Chen, and F. Wang, "Recommender systems based on user reviews: The state of the art," *User Model. User-Adapted Interact.*, vol. 25, no. 2, pp. 99–154, Jun. 2015, doi: 10.1007/s11257-015-9155-5.

[10] C. He, D. Parra, and K. Verbert, "Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities," *Expert Syst. Appl.*, vol. 56, pp. 9–27, Sep. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417416300367

[11] S. Reddy, S. Nalluri, S. Kunisetti, S. Ashok, and B. Venkatesh, "Content-based movie recommendation system using genre correlation," in *Smart Intelligent Computing and Applications*, S. C. Satapathy, V. Bhateja, and S. Das, Eds. Singapore: Springer, 2019, pp. 391–397.

[12] X. Li, W. Jiang, W. Chen, J. Wu, and G. Wang, "Haes: A new hybrid approach for movie recommendation with elastic serendipity," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, New York, NY, USA, 2019, pp. 1503–1512, doi: 10.1145/3357384.3357868.

[13] K. Benzi, V. Kalofolias, X. Bresson, and P. Vandergheynst, "Song recommendation with non-negative matrix factorization and graph total variation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 2439–2443.

[14] P. Covington, J. Adams, and E. Sargin, "Deep neural networks for youtube recommendations," in *Proc. 10th ACM Conf. Recommender Syst.*, New York, NY, USA, 2016, pp. 191–198, doi: 10.1145/2959100.2959190.

[15] J. Guerreiro and P. Rita, "How to predict explicit recommendations in online reviews using text mining and sentiment analysis," *J. Hospitality Tourism Manage.*, 2019, doi: 10.1016/j.jhtm.2019.07.001.

[16] M. Chelliah and S. Sarkar, "Product recommendations enhanced with reviews," in *Proc. Eleventh ACM Conf. Recommender Syst.*, New York, NY, USA, 2017, pp. 398–399, doi: 10.1145/3109859.3109936.

[17] L. S. Gallege and R. R. Raje, "Towards selecting and recommending online software services by evaluating external attributes," in *Proc. 11th Annu. Cyber Inf. Secur. Res.Conf. (CISRC)*, 2016, p. 23.

[18] Y. Qian, Y. Zhang, X. Ma, H. Yu, and L. Peng, "EARS: Emotion-aware recommender system based on hybrid information fusion," *Inf. Fusion*, vol. 46, pp. 141–146, Mar. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1566253518301234

[19] A. Da'u and N. Salim, "Sentiment-Aware Deep Recommender System With Neural Attention Networks," *IEEE Access*, vol. 7, pp. 45472–45484, 2019.

[20] A. Said, E. W. De Luca, and S. Albayrak, "Inferring contextual user profiles-improving recommender performance," in *Proc. 3rd RecSys Workshop Context-Aware Recommender Syst.*, Chicago, IL, USA, 2011.

[21] F. García-Sánchez, R. Colomo-Palacios, and R. Valencia-García, "A social-semantic recommender system for advertisements," *Inf. Process. Manage.*, vol. 57, no. 2, Mar. 2020, Art. no. 102153.

[22] J. Wei, F. Meng, and N. Arunkumar, "A personalized authoritative user-based recommendation for social tagging," *Future Gener. Comput. Syst.*, vol. 86, pp. 355–361, Sep. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X18303078

[23] M. Xu and S. Liu, "Semantic-enhanced and context-aware hybrid collaborative filtering for event recommendation in event-based social networks," *IEEE Access*, vol. 7, pp. 17493–17502, 2019.

[24] L. U, Y. Chai, and J. Chen, "Improved personalized recommendation based on user attributes clustering and score matrix filling," *Comput. Standards Inter.*, vol. 57, pp. 59–67, Mar. 2018.

[25] K. Xu, X. Zheng, Y. Cai, H. Min, Z. Gao, B. Zhu, H. Xie, and T.-L. Wong, "Improving user recommendation by extracting social topics and interest topics of users in uni-directional social networks," *Knowl.-Based Syst.*, vol. 140, pp. 120–133, Jan. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705117305002

[26] R. Logesh, V. Subramaniyaswamy, V. Vijayakumar, and X. Li, "Efficient User Profiling Based Intelligent Travel Recommender System for Individual and Group of Users," *Mobile Netw. Appl.*, vol. 24, no. 3, pp. 1018–1033, Jun. 2019, doi: 10.1007/s11036-018-1059-2.

[27] J. Li, W. Xu, W. Wan, and J. Sun, "Movie recommendation based on bridging movie feature and user interest," *J. Comput. Sci.*, vol. 26, pp. 128–134, May 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877750318300012

[28] H. Tahmasbi, M. Jalali, and H. Shakeri, "Modeling Temporal Dynamics of User Preferences in Movie Recommendation," in *Proc. 8th Int. Conf. Comput. Knowl. Eng. (ICCKE)*, Oct. 2018, pp. 194–199.

[29] S. E. Middleton, D. C. De Roure, and N. R. Shadbolt, "Capturing knowledge of user preferences: Ontologies in recommender systems," in *Proc. 1st Int. Conf. Knowl. Capture*, New York, NY, USA, 2001, pp. 100–107, doi: 10.1145/500737.500755.

[30] X. Xie and B. Wang, "Web page recommendation via twofold clustering: Considering user behavior and topic relation," *Neural Comput. Appl.*, vol. 29, no. 1, pp. 235–243, 2018, doi: 10.1007/s00521-016-2444-z.

[31] J. Al-Sharawneh and M.-A. Williams, "Credibility-based social network recommendation: Follow the leader," in *Proc. 21st Australas. Conf. Inf. Syst.*, 2010, pp. 1–11.

[32] D. Goldbaum, "Follow the leader: Simulations on a dynamic social network," UTS Finance Econ. Work. Paper No 155, 2008. [Online]. Available: http://www.business.uts.edu.au/finance/research/wpapers/wp155.pdf

[33] F. Li and T. C. Du, "Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs," *Decis. Support Syst.*, vol. 51, no. 1, pp. 190–197, Apr. 2011.

[34] J. Chen, C. Wang, Q. Shi, Y. Feng, and C. Chen, "Social recommendation based on users' attention and preference," *Neurocomputing*, vol. 341, pp. 1–9, May 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231219302693

[35] A. Kumar and G. Garg, "Sentiment analysis of multimodal Twitter data," *Multimedia Tools Appl.*, vol. 78, no. 17, pp. 24103–24119, Sep. 2019.

[36] A. Kumar and A. Jaiswal, "Particle swarm optimized ensemble learning for enhanced predictive sentiment accuracy of tweets," in *Proc. ICETIT*, P. K. Singh, B. K. Panigrahi, N. K. Suryadevara, S. K. Sharma, and A. P. Singh, Eds. Cham, Switzerland: Springer, 2020, pp. 633–646.

[37] A. Kumar and A. Joshi, "Ontology driven sentiment analysis on social Web for government intelligence," in *Proc. Special Collection EGovernment Innov. India*, New York, NY, USA, 2017, pp. 134–139, doi: 10.1145/3055219.3055229.

[38] A. Kumar and A. Jaiswal, "Empirical study of twitter and tumblr for sentiment analysis using soft computing techniques," in *Proc. World Congr. Eng. Comput. Sci.*, vol. 1, 2017, pp. 1–5.

[39] H. H. Do, P. Prasad, A. Maag, and A. Alsadoon, "Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review," *Expert Syst. Appl.*, vol. 118, pp. 272–299, Mar. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417418306456

[40] A. Dridi, M. Atzeni, and D. Reforgiato Recupero, "FineNews: Fine-grained semantic sentiment analysis on financial microblogs and news," *Int. J. Mach. Learn. Cyber.*, vol. 10, no. 8, pp. 2199–2207, Aug. 2019, doi: 10.1007/s13042-018-0805-x.

[41] R. Dong, M. P. O'Mahony, M. Schaal, K. Mccarthy, and B. Smyth, "Combining similarity and sentiment in opinion mining for product recommendation," *J. Intell. Inf. Syst.*, vol. 46, no. 2, pp. 285–312, Apr. 2016, doi: 10.1007/s10844-015-0379-y.

[42] G. Liu, Y. Wang, and M. Orgun, "Trust inference in complex trust-oriented social networks," in *Proc. Int. Conf. Comput. Sci. Eng.*, vol. 4, Aug. 2009, pp. 996–1001.

[43] M. J. Metzger, "Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research," *J. Am. Soc. Inf. Sci.*, vol. 58, no. 13, pp. 2078–2091, Nov. 2007.

[44] K. Kwon, J. Cho, and Y. Park, "Multidimensional credibility model for neighbor selection in collaborative recommendation," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 7114–7122, Apr. 2009, doi: 10.1016/j.eswa.2008.08.071.

[45] T. H. Haveliwala, "Topic-sensitive pagerank," in *Proc. 11th Int. Conf. World Wide Web*, New York, NY, USA, 2002, pp. 517–526, doi: 10.1145/511446.511513.

[46] R. Mihalcea and P. Tarau, "Textrank: Bringing order into text," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2004, pp. 404–411.

[47] S. Rose, D. Engel, N. Cramer, and W. Cowley, "Automatic keyword extraction from individual documents," *Text Mining: Appl. Theory*, vol. 1, pp. 1–20, Mar. 2010.

SHIGANG HU received the B.S. degree in electronic information engineering from the Hubei University of Technology, Wuhan, China, in 2003, and the Ph.D. degree in microelectronics from Xidian University, Xi'an, China, in 2009. Since 2018, he has been a Professor with the Electronic Information Engineering Department, Hunan University of Science and Technology, Xiangtan, China. He has authored over 30 research articles. His research interests include artificial intelligence technology and photoelectric information materials. He has received several grants, in recent years, from the National Natural Science Foundation of China.

AKSHI KUMAR received the B.Eng. degree (Hons.) in computer science and engineering from Maharishi Dayanand University, India, in 2003, the M.Tech. (Hons.) in computer science and engineering from Guru Gobind Singh Indraprastha University, New Delhi, in 2005, and the Ph.D. degree in computer engineering from the Faculty of Technology, University of Delhi, in 2011. She has been with the Delhi Technological University (formerly Delhi College of Engineering), for the past 11 years, where she is currently an Assistant Professor with the Department of Computer Science and Engineering. She has presented several papers in international conferences and published work in peer-reviewed and science cited journals. She has authored a monograph *Web Technology: Theory and Practice* (CRC Press, Taylor and Francis Group) and edited a book *A Roadmap to Industry 4.0: Smart Production, Sharp Business and Sustainable Development* (Springer). Her research interests are in the area of intelligent systems, Web technology, sentiment analysis, natural language processing, social media analytics, and soft computing. She is a recipient of the Commendable Research Award for Excellence In Research at Delhi Technological University, in 2018, and the VIWA Outstanding Women in Engineering (Computer Science & Engineering) Award, in 2020. She is an active reviewer of many top journals of the IEEE, IET, Springer, and Elsevier. In addition, she is currently serving as the Guest Editor for several journals, including the *Energy Systems* (Springer); *Future Internet*, the *IEEE Internet of Things Magazine*, the *Sustainable Cities & Society* (Elsevier), and the *Autosoft Journal (Intelligent Automation & Soft Computing)*.

FADI AL-TURJMAN received the Ph.D. degree in computer science from Queen's University, Kingston, ON, Canada, in 2011. He is currently a Full Professor and the Research Center Director with Near East University, Nicosia, Cyprus. He is a leading authority in the areas of smart/intelligent, wireless, and mobile networks' architectures, protocols, deployments, and performance evaluation. His publication history spans over 250 publications in journals, conferences, patents, books, and book chapters, in addition to numerous keynotes and plenary talks at flagship venues. He has authored and edited more than 25 books about cognition, security, and wireless sensor networks' deployments in smart environments, published by Taylor & Francis, Elsevier, and Springer. He has received several recognitions and best papers' awards at top international conferences. He also received the prestigious Best Research Paper Award from the *Elsevier Computer Communications Journal* (2015–2018), in addition to the Top Researcher Award for 2018 at Antalya Bilim University, Turkey. He has led a number of international symposia and workshops in flagship communication society conferences. He currently serves as an Associate Editor and the Lead Guest Editor for several well reputed journals, including the *Elsevier Computer Communications*, *Sustainable Cities and Society*, Springer MONET, and the *IET Wireless Sensor Systems journals*.

SHIVAM GUPTA is currently pursuing the B.Tech. degree in computer engineering with Delhi Technological University (formerly Delhi College of Engineering), New Delhi, India. He was a Software Engineer Intern at SAP. His research interests are in the areas of artificial intelligence, deep learning, natural language processing, sentiment analysis, recommendation systems, and cyber security.

SIMRAN SETH is currently pursuing the B.Tech. degree in computer engineering with Delhi Technological University, New Delhi, India. Her research interests include natural language processing, recommendation systems, cyber security, information retrieval, and text mining.

SHUBHAM is currently pursuing the B.Tech. degree in computer engineering with Delhi Technological University, New Delhi, India. His research interests include data mining, recommender systems, and artificial intelligence.