Journal of
Information Systems Engineering
and Business Intelligence

Vol.11, No.2, June 2025
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

Enhancing the Comprehensiveness of Criteria-Level Explanation in Multi-Criteria Recommender System

Rita Rismala [1)*], **Nur Ulfa Maulidevi** [2)], **Kridanto Surendro** [3)]

[1)] *School of Computing, Telkom University, Bandung, Indonesia*
[1)] ritaris@telkomuniversity.ac.id

[2)3)] *School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia*
[2)] ulfa@itb.ac.id, [3)] endro@itb.ac.id

Abstract

Background: The explainability of recommender systems (RSs) is currently attracting significant attention. Recent research mainly focus on item-level explanations, neglecting the need to provide comprehensive explanations for each criterion. In contrast, this research introduces a criteria-level explanation generated in a content-based pardigm by matching aspects between the user and item. However, generation may fall short when user aspects do not match perfectly with the item, despite possessing similar semantics.
Objective: This research aims to extend the aspect-matching method by leveraging semantic similarity. The extension provides more detail and comprehensive explanations for recommendations at the criteria level.
Methods: An extended version of the aspect matching (AM) method was used. This method identified identical aspects between users and items and obtained semantically similar aspects with closely related meanings.
Results: Experiment results from two real-world datasets showed that AM+ was superior to the AM method in coverage and relevance. However, the improvement varied depending on the dataset and criteria sparsity.
Conclusion: The proposed method improves the comprehensiveness and quality of the criteria-level explanation. Therefore, the adopted method has the potential to improve the explainability of multi-criteria RSs. The implication extends beyond the enhancement of explanation to facilitate better user engagement and satisfaction.

Keywords: Comprehensiveness, Content-Based Paradigm, Criteria-Level Explanation, Explainability, Multi-Criteria Recommender System

Article history: Received 21 November 2024, first decision 18 April 2025, accepted 09 June 2025, available online 22 July 2025

I. INTRODUCTION

Recommender systems (RSs) provide personalized recommendations by effectively filtering and sorting information based on the preferences of users. This reduces information overload and supports better decision-making [1], [2], [3], [4]. RSs have experienced a significant increase in the application across various domains, including e-commerce [5], [6] entertainment [7], [8] education [9], [10], and Internet of Things (IoT) [11], [12] to enhance user engagement and satisfaction. Research on RSs primarily focuses on improving the quality of recommended items. This comprises analyzing user preferences and behavior to provide more personal and relevant recommendations, including the incorporation of machine learning [1], [13], [14]. Currently, explainable RSs are gaining attention [15], [16] and the concept refers to personalized recommendation algorithms that tackle the "why" problem. In this context, users are provided with recommendation results and valuable information to explain the reasons for recommending items. Enhancing the transparency and persuasiveness of the RSs is important [17], [18]. This transparency builds trust and engagement among users, enhancing the overall experience. As demand for intuitive and user-friendly systems increases, developing explainable RSs will be essential to meet user expectations.

Explainable recommendation models can be categorized as model-intrinsic or model-agnostic [19], [20]. The model-intrinsic method focuses on developing interpretable models. This refers to RSs designed with built-in mechanisms for providing explanations for the recommendations. The method allows users to understand the reasoning behind the recommendations to enhance trust and usability. By integrating interpretability at the model level, these systems facilitate a more intuitive user experience [21]. The model-agnostic method, also known as the post-hoc explanation enables the decision-making process to remain a black box [22], [23]. The method does not rely

* Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print)
doi: http://dx.doi.org/10.20473/jisebi.11.2.160-172

on the specific details of the model but rather focuses on interpreting the output of various models to provide explanations for the recommendations. This flexibility allows for broader application across different systems and enhances the understanding of model behaviors without necessitating alterations [15].

In multi-criteria recommendation systems (MCRSs), the multi-criteria ratings can provide a criteria-based explanation [24]. This type of explanation helps users understand the reasoning behind specific recommendations by showing the importance of various criteria. By effectively presenting the insights, MCRSs enhance user satisfaction and trust in the system. Hou et al. [25] presented a quantitative explanation using a radar chart and the visual explanation was more attractive and challenging for users to understand. Similar to a criteria-based explanation is an aspect-based explanation. The difference is that when the criteria are clearly defined in the item profile, aspects must be extracted or learned from other sources, such as reviews [24]. Zhang et al. [21] introduced a template-based explanation that emphasized aspects. Meanwhile, N. Wang et al. [26] advanced templates that incorporated both aspects and opinions. A template-based explanation includes creating explanations using a predefined sentence structure, which is customized with various words for each user [24]. This method is straightforward and capable of generating systematic, structured, and easily understandable explanations [21], [26]. A template-based explanation facilitates a more efficient communication of information since users can quickly understand key insights without becoming overwhelmed by complexity. However, the explanation focused solely on the item level, overlooking the necessity of delivering a comprehensive explanation for each criterion. Since aspects can be considered as sub-criteria, this research proposed an aspect-based explanation at the criteria level. The objective was to deliver a more detailed explanation in line with the established criteria.

An aspect-based explanation is closely related to content-based recommendation methodes. This is explained by matching key aspects between the user profile and the content features of candidate items, referred to as the aspect-matching method [21]. The method is straightforward and interpretable but may fall short when the key aspects of the user and the item are not perfectly balanced. This shows a research gap since existing aspect-based explanation methods often rely on exact aspect matches, which limits the effectiveness when user and item aspects differ in wording. The shortcoming affects the generation of truly comprehensive and meaningful explanations at the criteria level in cases with sparse or diverse linguistic expressions. Identical aspects between users and items were identified with the nearest aspects sharing similar semantic meanings by extending the aspect matching. The objective is to enhance the richness of information and ensure the comprehensiveness of the explanation at the criteria level. Users have access to more comprehensive information to make more precise decisions. The research question considered is "How does the extended aspect matching method enhance the comprehensiveness of criteria-level explanations?" The main contribution of this research was the semantic extension of aspect matching, which enabled the identification of semantically similar aspects between users and items, supporting the generation of more detailed and comprehensive explanations at the criteria level.

II. LITERATURE REVIEW

Explanations are crucial for users to assess a recommendation system. In an initial research on explanation in RSs, 21 different types of explanation interfaces were assessed for a collaborative filtering-based system. The research also examined the impact of each interface on users' acceptance of the recommendations [27]. Research on explanation have significantly increased in recent years, reporting the importance of transparency in intelligent systems [15]. The increasing focus on transparency shows the necessity for algorithms to deliver valuable recommendations and explain the rationale behind the suggestions. Therefore, users are expected to trust and engage with systems prioritizing clear and comprehensible explanations.

Friedrich & Zanker [28] introduced a taxonomy of explanation methodes in RSs. The three dimensions of the taxonomy included the reasoning model, the recommendation paradigm, and the exploited information categories. According to the reasoning model, explanations can be classified into black-box and white-box categories. Black-box explanations justify the reason for making a recommendation without disclosing the underlying process. For example, Musto et al. [15] generated post hoc natural language justifications derived from the review. These justifications provided users insight into the reasoning behind the recommendations while keeping the confidentiality of details. In contrast, white-box explanations disclosed the decision-making process, enhancing transparency. Bilgic & Mooney [29] justified system recommendations using neighborhood information, while Coyle & Smyth [30] generated text explanations based on the search histories of online users. These methodes balance transparency and privacy, enhancing user trust in the system. Additionally, continuous improvement is conducted by providing feedback on the reasoning processes. This research is also classified as a white box explanation since a multi-criteria rating is exploited to generate an explanation. The provision of insights into the decision-making process allows users to understand the factors influencing the outcomes more comprehensively.

The generation of explanations leverages the relationships among users, items, and properties. However, the derivation of these relational instances varies based on the paradigm used. The three fundamental paradigms of recommendations are collaborative filtering, content-based filtering, and knowledge-based recommendations. Collaborative filtering relies on known preferences between users and items. Furthermore, neighborhood-based collaborative filtering enhances these models by analyzing the similarity relationships between users or items [28]. Different methodologies of collaborative explanation have been proposed, including research by Bilgic & Mooney [29]. Several research by N. Wang et al. [21], [26] and Zhang et al. [21], [26] have proposed methodologies for this paradigm. The third paradigm, knowledge-based, is defined by the incorporation of additional domain characteristics. This includes abstract user requirements or preferences with the various relationships. For instance, Bai et al. [31] integrated external knowledge from Wikipedia to provide a more detailed description of specific aspects of an item. This content-based explanation relies on descriptive aspects of items extracted from user reviews. Previous results focused on providing content-based explanations for single-criterion recommendation systems at the item level. Meanwhile, this research proposed a content-based explanation for MCRSs at the criteria level.

The third dimension comprises the information categories used to generate explanations, namely the user model, recommended items, and alternatives. User models are explanations generated by exploiting the available information, such as ratings, preferences, reviews, or demographics. Recommended items are explanations generated by exploiting the specific characteristics of the recommended item. Alternatives are explanations that argue in favor of or against the recommended item [28]. Even though the paradigm is a content-based explanation, this research did not exploit descriptive aspects of items directly but rather extracted the concept from reviews. Other relevant research that also implemented a content-based paradigm and exploited reviews were N. Wang et al. and Zhang et al. [21], [26]. Zhang et al. [21] generated an explanation using the following template to inform users about recommended and non-recommended aspects, *You might be interested in **[aspect]**, on which this product performs well. You might be interested in **[aspect]**, on which this product performs poorly.* The template used a personalization algorithm to select specific aspects, resulting in a personalized explanation. N. Wang et al. [26] provided a template-based explanation using aspects and opinion words, such as, *Its **decor** is **[neat] [good] [nice]**. Its **sandwich** is **[grilled] [cajun] [vegan]**. Its **sauce** is **[good] [green] [sweet]**.* The words enclosed in brackets are selected by the model to describe the corresponding aspect of the item. In this context, aspects and content-based paradigms are closely related. By matching key aspects between the user profile and the content features of candidate items, the content-based paradigm generates an explanation [24]. However, an explanation may not be generated when the key aspects of the user and the item are not perfectly balanced, despite having similar semantic meanings. Extending the aspect matching method by finding the nearest aspects that share the same semantic meaning is proposed to address this issue, as well as to improve the comprehensiveness of an explanation.

III. METHODS

This research introduced an aspect-based explanation at the criteria level with an enhanced version of the AM method, referred to as AM+. The methods were proposed to improve the comprehensiveness of the explanation and process multi-criteria ratings as input, generating sentence-based explanations as output. The main stages included explaining the criteria and providing details about using an aspect-based explanation.

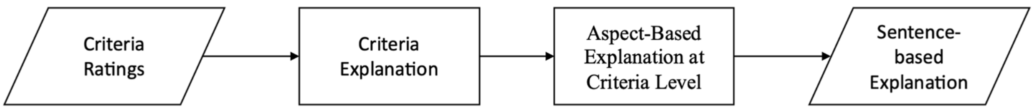


Fig. 1 The framework of AM+.

The purpose of a criteria explanation was to translate multi-criteria ratings into coherent explanations presented in sentence form. For each recommended criterion, a more detailed explanation would be provided regarding the aspects and opinions supporting the criteria. Generally, the explanation was presented in a template-based sentence, as described in Fig. 2.

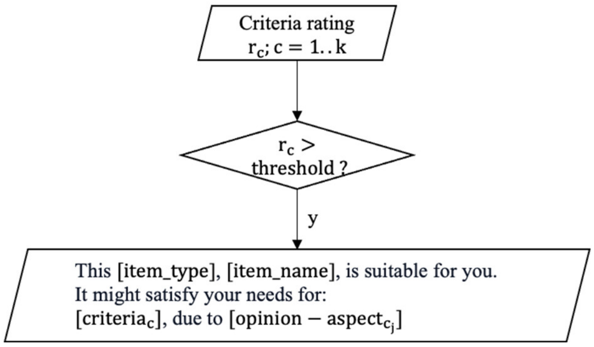


Fig. 2 Explanation template of AM+.

A. Datasets

Experiments were conducted on two domains of real-world datasets, namely TripAdvisor and BeerAdvocate. The datasets were selected based on the relevance to the problem and the availability of open access. These datasets contained item ID, user ID, review, overall rating, and criteria rating. TripAdvisor was reported as a hotel domain dataset, consisting of six criteria, namely service (c_1), cleanliness (c_2), value (c_3), sleep quality (c_4), rooms (c_5), and location (c_6) [32], [33]. BeerAdvocate was a beer domain dataset, including appearance (c_1), aroma (c_2), palate (c_3), and taste (c_4) [34], [35]. These datasets had at least five user and item interactions. The ratings were presented on a scale of 1 to 5, with 5 being the highest, as reported in Table 1.

TABLE 1
DATASETS SUMMARY

	TripAdvisor	BeerAdvocate
#Users	7,175	2,186
#Items	3,272	2,622
#Criteria	6	4
#Reviews	88,617	85,054
#Overall ratings	88,617	85,054
#c_1 ratings	81,507	85,054
#c_2 ratings	81,914	85,054
#c_3 ratings	81,795	85,054
#c_4 ratings	46,692	85,054
#c_5 ratings	75,073	-
#c_6 ratings	71,392	-
Sparsity level	99.62%	98.52%
Rating scales	[1, 5]	[1, 5]

B. Criteria Explanation

Multi-criteria ratings were translated into explanations using structured template-based sentences. This explanation was generated for each recommended item. To provide a personalized explanation, the $[criteria_c]$ $[criteria_c]$were filled with the recommended criteria for the user. The criteria were only recommended when the rating surpassed the threshold.

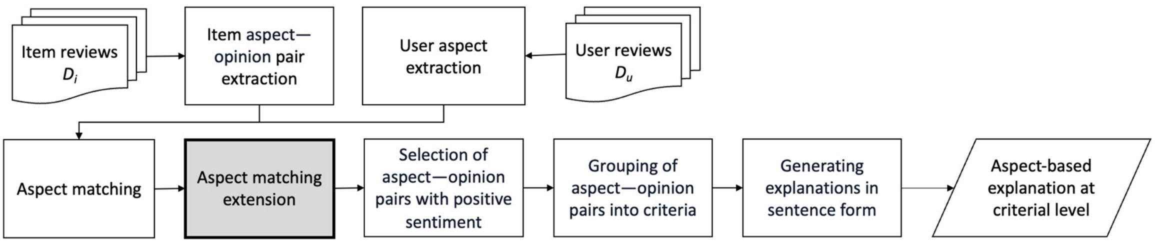


Fig. 3 Aspect-based explanation at the criteria level using AM+.

C. Aspect-Based Explanation at Criteria Level

A more detailed explanation was provided regarding the aspects and opinions supporting the criteria. Aspects were paired with opinions to clearly express sentiments. Therefore, this research proposed the aspect-based explanation at the criteria level using AM+, as reported in Fig. 3. The grey-reported part showed the proposed aspect-matching extension based on semantic similarity, which was the key process distinguishing AM+ from the original method.

1) Item aspect-opinion pair extraction

Item aspect-opinion pair extraction was conducted to obtain a comprehensive list of phrases used for generating detailed explanations. These pairs were derived from user reviews and the process started with text preprocessing to refine tokenization, expansion of contractions, conversion to lowercase, and removal of special characters such as non-ASCII symbols, punctuation, and numerical digits. Subsequently, aspect-opinion pair extraction was performed using a rule-based method leveraging pattern-based linguistic knowledge. The method used predefined syntactic patterns namely adjective-noun combinations (e.g., 'delicious food') to identify meaningful aspect-opinion pairs from the reviews. These patterns were defined based on part-of-speech (POS) tag sequences, particularly an adjective (JJ) followed by a noun (NN or NNS), which typically indicated an opinion-aspect relationship in user-generated text [36].

As reported in Fig. 3, item reviews D_i represented the collection of all reviews for item i. To identify item-level aspect-opinion pairs, all noun phrases were first extracted from D_i. A noun phrase is a group of words that functions as a noun in a sentence and typically consists of a noun and modifiers, such as adjectives or determiners. For example, in the sentence "*The hotel offers nice rooms at reasonable prices*," the phrases "nice rooms" and "reasonable prices" are noun phrases. An aspect-opinion pair is formed by locating a noun (as the aspect) and an adjective (as the opinion) that appear in proximity within the same phrase. From the noun phrase "nice rooms", the adjectives "nice" and "rooms" are the opinion and aspect, forming the aspect-opinion pair (rooms, nice).

2) User aspect extraction

User aspect extraction was conducted to identify and compile a list of aspects that reflected the primary concerns of users. Similar to item aspect-opinion pairs, user aspects were also extracted from review texts. Accordingly, text preprocessing was performed beforehand. Subsequently, aspects were identified using a rule-based method that applied syntactic patterns to the processed text. As shown in Fig. 3, user reviews D_u denoted the collection of all reviews written by user u. To extract aspects associated with user u, all noun phrases were identified from D_u, and considered as user aspects. From the sentence "*The hotel offers nice rooms and friendly service*," the noun phrases "nice rooms" and "friendly service" could be extracted. Meanwhile, nouns "rooms" and "service" were identified as user aspects.

3) Aspect matching

User aspects were reported to be balanced with the item. The corresponding aspect-opinion pair would be the candidate of recommendation explanation when the user and item share the same aspect. For example, a match was identified when an item was associated with the aspect-opinion pair (rooms, nice) and the user expressed a preference for the aspect 'rooms'. Therefore, the aspect-opinion pair (rooms, nice) would be considered as a candidate for an explanation.

4) Aspect matching extension

Users and items did not mention identical aspects but shared similar meanings. Therefore, this research introduced an method to extend the user aspect by identifying the nearest aspects with similar semantics. The similarity described the resemblance of two pieces of text in meaning [37] and could be measured by semantic distance [38]. This research used Euclidean distance, which provided a straightforward way to measure the "closeness" of the texts by calculating the geometric distance between the vectors in an embedding space. Euclidean distance also embodied both simplicity and interpretability [39]. For example, when the user mentioned 'vanilla' and the item included 'caramel' represented in vector form, Euclidean distance could quantify the difference in the semantic space. Even though cosine similarity was also commonly used in text-related tasks, Euclidean distance was selected to reflect directional similarity and incorporate magnitude relevant for capturing fine-grained distinctions in semantic proximity.

Each aspect was represented as a word vector using GloVe including a global context applied across various domains. GloVe provided widely available pre-trained models on large corpora [40]. The use was essential in embedding words into a high-dimensional vector space. In this context, semantically similar words were positioned close to each other. Subsequently, Euclidean distance operated over these GloVe vectors to measure the difference between the two aspects in meaning. This combination allowed the model to identify exact matches and semantically related aspects enabling more informative explanations. Therefore, aspect matching was reexamined between the extended user and item aspects. The matched aspect-opinion pairs were used to generate candidate explanations for recommendations.

For instance, when a user expressed a preference for the aspect 'mattress', and the item explicitly included 'pillow', the terms would be semantically related. The system could identify that 'pillow' was a semantically close aspect to 'mattress' using Euclidean distance. This showed that 'pillow' became a candidate aspect for an explanation since a similar meaning was shared with 'mattress'.

5) Selection of positive sentiment aspect—opinion pairs

Explanations were generated for the recommended items. Therefore, the selected aspect-opinion pairs for explanations needed to be indicative of positive sentiment [26]. This selection was conducted using a lexicon-based method through the TextBlob library. TextBlob also assigned sentiment polarity scores ranging from -1 (very negative) to +1 (very positive) [41]. Only aspect-opinion pairs with positive sentiment polarity (i.e., > 0) were retained for an explanation. For example, in considering the two aspect-opinion pairs (bed, comfortable) and (service, terrible), TextBlob would assign a positive and negative polarity to 'comfortable' and 'terrible', respectively. Therefore, (bed, comfortable) was selected as a candidate explanation, while (service, terrible) was filtered out. A refinement process was carried out to eliminate any redundant aspect-opinion pairs with exact matches.

6) Grouping of aspect—opinion pairs into criteria

The selected aspect-opinion pairs were grouped according to the respective criteria to obtain the criteria-level explanation. The grouping was performed based on the word distance between the aspect and the centroid of criteria. The calculation of distance was based on the Euclidean distance, using a GloVe vector to represent each word. In considering two criteria, namely rooms and service, the aspect 'bed' and 'staff' were semantically closer to the criterion rooms and service. The Euclidean distance between 'bed' and the centroid of rooms would be smaller than the distance to service using GloVe vectors. The aspect 'staff' would be assigned to service due to the proximity in the embedding space. Therefore, the aspect-opinion pairs such as (bed, comfortable) and (staff, friendly) would be grouped under rooms and service, respectively. This process obtained a set of aspect-opinion pairs for each criterion, enabling the generation of criterion-level explanations.

7) Generating explanations in sentence form

Explanations were intended to be presented to the user of the RSs. Continuing from the grouping process, the resulting aspect-opinion pairs for each criterion were organized into sentence structures for recommendation explanations using the established template shown in Fig. 2. The $[opinion - aspect_{c_j}]$ $[opinion - aspect_{c_j}]$ was filled with the aspects and opinions supporting the criteria. For each criterion, a maximum of 10 pairs of aspects and opinions reflecting the most positive sentiment were presented. Fig. 4 shows the process of generating explanations at the criteria level based on the corresponding criteria ratings. Explanations were provided only for recommended criteria. In practice, RSs typically classified ratings above the median as a positive rating since the item was suitable for recommendation [42]. Therefore, this research used a rating threshold of 3 to determine criteria considered recommended and eligible for explanation.

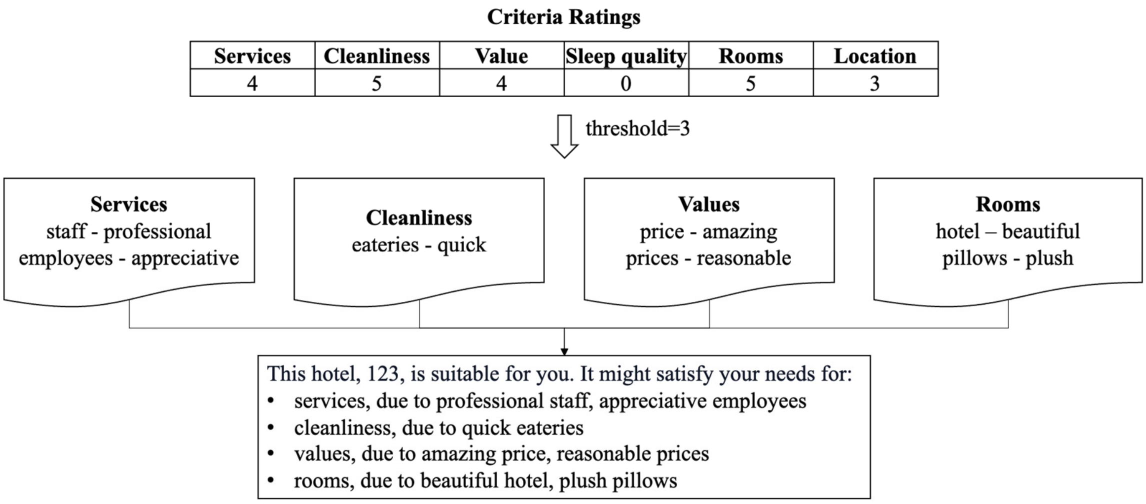


Fig. 4 Illustration of criteria-level explanation.

D. Evaluation Metrics

This research improved the comprehensiveness of the explanation for each criterion. Therefore, coverage served as the primary metric for evaluating system performance. Relevance-based measurements were used to evaluate the quality of the explanation. The metrics for coverage and relevance comprised Feature Coverage Ratio (FCR), Criteria

Explanation Coverage Ratio (CECR), and Criteria-level Recall (CR). FCR was used to measure feature coverage at the corpus level. The calculation was based on the different features displayed in the generated explanations, as reported in (1). F represented all aspects of the dataset, while N_g was the number of unique aspects in the explanation. CECR was used to measure the coverage of explanation at the criteria level. N_{ce} represented the total number of criteria to be explained, while N_{cg} showed the criteria successfully explained. CR measured the quality of explanations by comparing the generated criteria to actual reviews (ground truth). For each pair of user u and item i, $S_{u,i}$ represented the set of criteria in the explanation, and $T_{u,i}$ was the set of criteria in the ground truth.

$$FCR = \frac{N_g}{|F|} \quad (1)$$

$$CECR = \frac{N_{ce}}{N_{cg}} \quad (2)$$

$$CR_{u,i} = \frac{|S_{u,i} \cap T_{u,i}|}{|T_{u,i}|} \quad (3)$$

IV. RESULTS

A. Experimental Setups

1) Parameters

Each aspect was classified into a criterion based on the level of closeness with the centroid. Table 2 presents the words used as the criterion centroid for each dataset.

2) Baseline for comparison

The traditional content-based explanation paradigm using AM was used as the baseline to assess the performance of the proposed method. This choice was rooted in the paradigm's historical significance and wide adoption in recommendation systems. Content-based explanation methodes relied on matching user preferences with item aspects derived from metadata or explicit user feedback. These provided a straightforward and interpretable method for generating recommendations and explanations [21].

TABLE 2
WORD CENTROID FOR EACH CRITERION

Dataset	Criterion	Word Centroid
Tripadvisor	Service	service
	Cleanliness	cleanliness
	Value	value
	Sleep Quality	sleep
	Rooms	room
	Location	location
BeerAdvocate	Appearance	appearance
	Aroma	aroma
	Palate	palate
	Taste	taste

B. Experimental Results

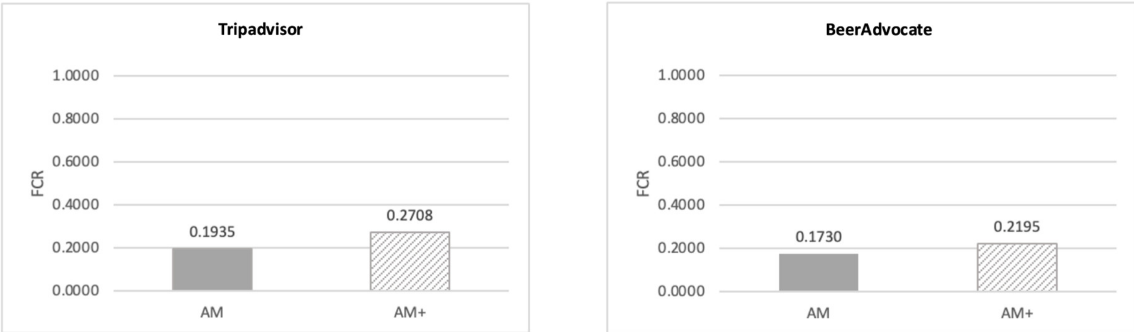
The proposed method was compared to the baseline on two public datasets to show performance. The datasets were split into 80% and 20% training and test data to evaluate the model.

1) Experiment 1: Coverage

Coverage quantified the comprehensiveness with which aspects and criteria were addressed within the explanation. This variable was evaluated using FCR and CECR metrics. Higher values showed that the explanation included a greater number of aspects and criteria.

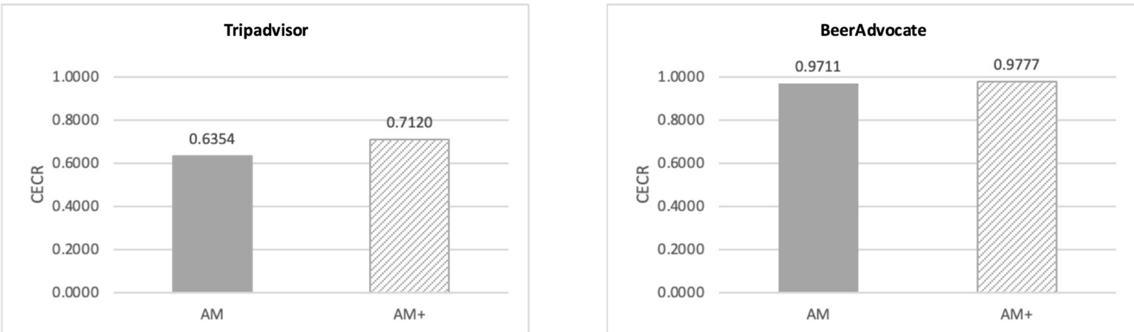
According to Fig. 5, the explanation generated by the AM+ covered more aspects than the baseline method. This was shown by the remarkable increases in FCR, with a 40% and 27% rise in the TA and BA datasets, respectively. The AM+ explained more criteria than the baseline method. This was clear from the 12% and 1% increase in CECR for the TA and BA datasets (Fig. 6). The results showed that the proposed method surpassed the baseline method concerning the coverage.

The proposed method improved the comprehensiveness of criteria-level explanations compared to the baseline method (Fig. 7). The improvement reported the efficacy of AM+ in providing richer and more informative feedback for users. This was particularly valuable for recommender systems in domains such as e-commerce, education, travel, and entertainment**,** where users frequently evaluated multiple factors before making decisions. For instance, in e-commerce, AM+ could explain recommendations based on diverse aspects such as price, quality, or shipping and delivery, improving trust and reducing uncertainty. In education, recommendations were balanced with user goals by considering criteria such as course difficulty and relevance. Similarly, in travel and entertainment, users considered different options by addressing factors including price, location, or amenities. AM+ could lead to enhanced user confidence, satisfaction, and trust across various applications by offering richer and more comprehensive explanations.



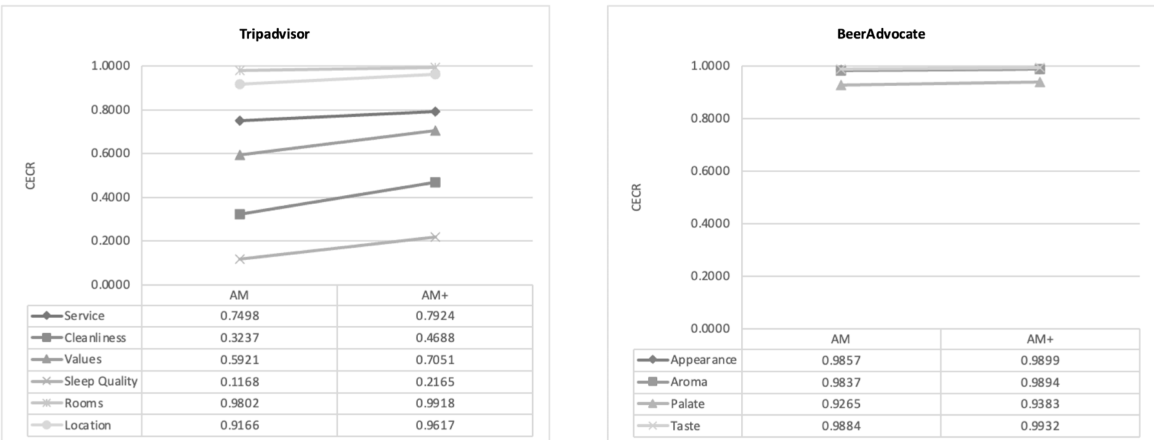
(a) (b)

Fig. 5 FCR comparison of AM and AM+ methods on (a) Tripadvisor dataset; and (b) BeerAdvocate dataset.



(a) (b)

Fig. 6 CECR comparison of AM and AM+ methods on (a) Tripadvisor dataset; and (b) BeerAdvocate dataset.

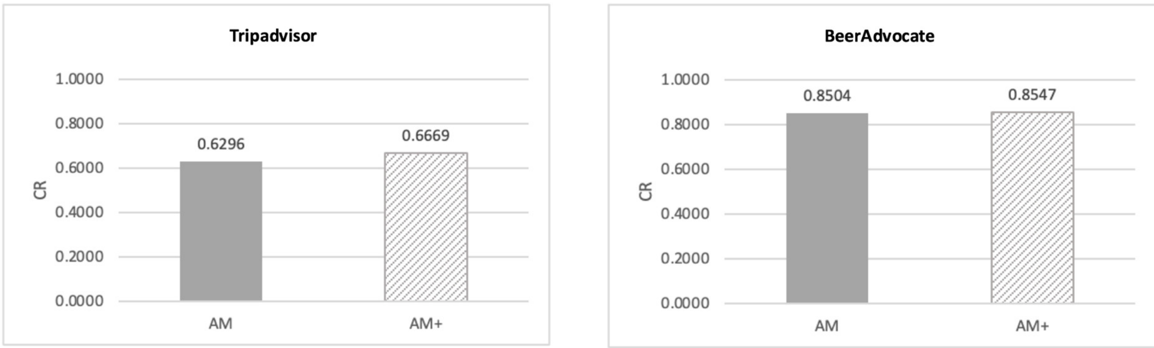


(a) (b)

Fig. 7 CECR comparison of AM and AM+ methods for each criterion on (a) the Tripadvisor dataset; and (b) the BeerAdvocate dataset.

2) Experiment 2: Quality

The quality of the explanations was evaluated through relevance-based metrics, namely CR. Criteria for ground truth used as a benchmark were derived from the review. Fig. 8 shows the results of the explanation quality experiment.



(a) (b)

Fig. 8 CR comparison of AM and AM+ methods on (a) Tripadvisor dataset; and (b) BeerAdvocate dataset.

According to Fig. 8, AM+ provided more relevant explanations compared to the baseline method. This was evidenced by a 6% and 1% increase in CR for TA and BA datasets, respectively. The improvement in explanation relevance made the AM+ method particularly beneficial for recommender systems where decision-making heavily relied on clear and accurate justifications. For example, in e-commerce platforms, more relevant explanations could help users better understand the recommendation of specific products to increase trust and purchase likelihood. In educational systems, relevant explanations balanced with user goals improved engagement. In travel and hospitality, personalized, relevant explanations about destinations or accommodations could also reduce decision fatigue and enhance user confidence. The increased relevance provided by AM+ improved user satisfaction and strengthened the ability of the recommender system to support informed and confident decision-making across diverse applications.

V. DISCUSSION

The proposed method, AM+, outperforms AM in terms of coverage and quality. Therefore, AM+ improves the comprehensiveness of the criteria-level explanation compared to the baseline method. The extent of this performance improvement differs across each dataset. The performance improvement on the Tripadvisor dataset surpasses BeerAdvocate. This discrepancy arises from the higher sparsity level of the Tripadvisor dataset and varies across different criteria. As the level of sparsity increases, the number of reviews and discussions concerning the criteria decreases significantly. Fewer discussions increase the challenges of identifying the precise aspects between users and items. The AM+ method addresses the challenge of sparsity by extending the aspect matching. The method allows semantically similar aspects to be uncovered by leveraging the relationships between word vectors. This is particularly effective for datasets and criteria characterized by high sparsity levels. For example, "caramel" is not explicitly mentioned by the user, but "vanilla". In the baseline method, “caramel” and “vanilla” are considered unmatched [21]. However, with AM+, “caramel” is identified as semantically similar to “vanilla” based on the proximity of the word vectors [40]. In this context, “caramel” is incorporated as part of the explanation for the aroma criterion, significantly enriching the explanation of recommendations with relevant and meaningful aspects. The intuition behind the improvement of AM+ is in the ability to expand the exploration space of aspects, effectively uncovering hidden relationships in sparse datasets. This leads to more personalized and relevant explanations for users, enhancing the overall experience. The improved clarity in the explanations enhances greater trust and satisfaction in the recommendation system.

The coverage and relevance offered by AM+ can significantly impact user decision-making processes in recommender systems. In e-commerce platforms, the ability of the method to uncover semantically similar aspects such as "caramel" for "vanilla" allows for richer and more diverse explanations, helping users make informed purchase decisions even when preferences are sparsely expressed. Furthermore, the practical value of AM+ lies in its ability to handle data sparsity, a common challenge in real-world systems where user reviews are limited or incomplete. By capturing semantic similarities, AM+ ensures that meaningful aspects are included in explanations, providing a better user experience even in sparse datasets. This improvement is particularly relevant for platforms where user feedback is minimal. Therefore, the enhanced comprehensiveness and relevance of explanations achieved by AM+ can be directly translated into more confident decision-making, increased user engagement, and stronger trust in recommender systems.

Table 3 shows examples of the recommendation explanation presented at the criteria level. The aspects derived from the AM+ are consistent with the established criteria but some aspects remain misaligned. For example, "eateries" is currently categorized as cleanliness criteria when the word can be more accurately classified under service criteria. In terms of distance, the word vector for "eateries" is closer to "cleanliness" than "service". Practically, these misalignments affect the clarity and accuracy of explanations provided to users. In a travel recommendation system, categorizing *"eateries"* under cleanliness may mislead users with the expectation of service quality, such as food delivery efficiency. This discrepancy shows that the distance metrics do not fully capture the nuanced meanings of the words or the contextual relationships relevant to the domain [40], [43]. To address the issue, enhancements can be made to the distance calculation method or the word vector. For instance, cosine similarity is used to measure semantic closeness more effectively [44], [45]. Context-aware word representations such as BERT embeddings are used to better capture meaning [46].

This research has several limitations despite the promising results. First, the performance of the proposed method varies across datasets, particularly due to differences in sparsity levels. Second, the semantic similarity used to expand aspect matching relies on word vector distances, which may not fully capture domain-specific contextual meanings, leading to potential misclassifications. Third, the evaluation of explanation quality is limited to aspect balance and does not include user-centric validation [47]. These factors affect the internal and external validity of the results and are addressed in future work.

TABLE 3
EXAMPLES OF RECOMMENDATION EXPLANATION AT CRITERIA LEVEL

Criteria Ratings	Criteria Ground Truth from Reviews	Method	
		AM (Baseline)	AM+ (Proposed)
4 5 5 5 0 0 (Tripadvisor)	*service, cleanliness, value, sleep quality, rooms*	*This hotel, 970, is suitable for you. It might satisfy your needs for:* *service, due to comfortable service, available times, real staff* *cleanliness* *value, due to good deal, good value, amazing price* *sleep quality*	*This hotel, 970, is suitable for you. It might satisfy your needs for:* *service, due to comfortable service, available times, real staff, best employees, great services* *cleanliness, due to quick eateries* *value, due to a good deal, good value, amazing price, good fee, right amount, reasonable prices, reasonable cost* *sleep quality*
4 4 4 4 (BeerAdvocate)	*appearance, aroma, taste*	*This beer, 25978, is suitable for you. It might satisfy your needs for:* ● *appearance, due to a sweet finish, real head, clear body* ● *aroma, due to modest carbonation* ● *palate* ● *taste*	*This beer, 25978, is suitable for you. It might satisfy your needs for:* ● *appearance, due to a sweet finish, real head, clear body* ● *aroma, due to modest carbonation, sweet caramel* ● *palate* ● *taste, due to the strong flavor*

VI. CONCLUSIONS

In conclusion, previous literature primarily focused on item-level explanations. In contrast, this research introduces a novel method for generating aspect-based recommendation explanations at the criteria level. Aspects are extracted from reviews, followed by a matching process in line with the user and item. However, user aspects may not be balanced perfectly with item aspects when semantically similar. Therefore, this study introduces an extended version of the aspect matching method, referred to as AM+. The method identifies match and semantically similar aspects to enrich information and improve the comprehensiveness of recommendation explanations at the criteria level. Experiment results from two real-world datasets show that AM+ outperforms the baseline in coverage and relevance. The extent of performance improvement varies depending on the dataset and criteria sparsity. These results show that the proposed method improves the comprehensiveness as well as the quality of the criteria-level explanation. Furthermore, the method has a positive impact on improving the explainability of MCRSs. This improvement may lead to greater user trust and satisfaction, making the recommendation process more transparent and user-friendly.

Further improvements are necessary to mitigate the current limitations of this method. For example, exploring alternative methods for distance calculation or using word vectors that incorporate both local and global context can more effectively capture domain-relevant contextual relationships. Combining template-based explanation with natural language generation (NLG) is also interesting to improve the naturalness of the explanation. Moreover, the incorporation of user research into the evaluation process provides valuable insights into the effectiveness and perceived usefulness of the explanations from the end-user perspective.

Author Contributions: *Rita Rismala*: Conceptualization, Methodology, Software, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing. *Nur Ulfa Maulidevi*: Supervision, Validation. *Kridanto Surendro*: Supervision, Validation.

All authors have read and agreed to the published version of the manuscript.

Funding: This research received no specific grant from any funding agency.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: Datasets used in this study can be accessed at https://www.cs.virginia.edu/~hw5x/dataset.html and https://cseweb.ucsd.edu/~jmcauley/datasets.html#multi_aspect.

Informed Consent: There were no human subjects.

Institutional Review Board Statement: Not applicable.

Animal Subjects: There were no animal subjects.

ORCID:
Rita Rismala: https://orcid.org/0000-0002-4674-2400
Nur Ulfa Maulidevi: https://orcid.org/0000-0001-6624-4153
Kridanto Surendro: https://orcid.org/0000-0003-1705-1202

REFERENCES

[1] R. Rismala, N. U. Maulidevi, and K. Surendro, "Personalized neural network-based aggregation function in multi-criteria collaborative filtering," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 1, p. 101922, Jan. 2024, doi: 10.1016/j.jksuci.2024.101922.

[2] S. M. Al-Ghuribi and S. A. Mohd Noah, "Multi-Criteria Review-Based Recommender System–The State of the Art," *IEEE Access*, vol. 7, pp. 169446–169468, 2019, doi: 10.1109/ACCESS.2019.2954861.

[3] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," *Egyptian Informatics Journal*, vol. 16, no. 3, pp. 261–273, Nov. 2015, doi: 10.1016/j.eij.2015.06.005.

[4] S. Sahu, R. Kumar, P. MohdShafi, J. Shafi, S. Kim, and M. F. Ijaz, "A Hybrid Recommendation System of Upcoming Movies Using Sentiment Analysis of YouTube Trailer Reviews," *Mathematics*, vol. 10, no. 9, Art. no. 9, Jan. 2022, doi: 10.3390/math10091568.

[5] A. Shankar *et al.*, "An intelligent recommendation system in e-commerce using ensemble learning," *Multimed Tools Appl*, vol. 83, no. 16, pp. 48521–48537, May 2024, doi: 10.1007/s11042-023-17415-1.

[6] P. Keikhosrokiani and G. M. Fye, "A hybrid recommender system for health supplement e-commerce based on customer data implicit ratings," *Multimed Tools Appl*, vol. 83, no. 15, pp. 45315–45344, May 2024, doi: 10.1007/s11042-023-17321-6.

[7] S. V. Reddy, V. R. Krishna, R. J. Sapkal, J. Dhanke, S. P. Waghmare, and K. Kumar, "A cutting-edge artificial intelligence paradigm for entertainment-infused music recommendations," *Entertainment Computing*, vol. 51, p. 100717, Sep. 2024, doi: 10.1016/j.entcom.2024.100717.

[8] G. K. Shaw *et al.*, "Entertainment Recommendation and Rating System Based on Emotions," in *Building Embodied AI Systems: The Agents, the Architecture Principles, Challenges, and Application Domains*, P. Raj, A. Rocha, S. P. Singh, P. K. Dutta, and B. Sundaravadivazhagan, Eds., Cham: Springer Nature Switzerland, 2024, pp. 443–462. doi: 10.1007/978-3-031-68256-8_20.

[9] N. Dehbozorgi, M. T. Kunuku, and S. Pouriyeh, "Personalized Pedagogy Through a LLM-Based Recommender System," in *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky*, A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, and I. I. Bittencourt, Eds., Cham: Springer Nature Switzerland, 2024, pp. 63–70. doi: 10.1007/978-3-031-64312-5_8.

[10] J. Du, K. F. Hew, and L. Zhang, "Designing a recommender system to promote self-regulated learning in online contexts: A design-based study," *Educ Inf Technol*, vol. 30, no. 2, pp. 2389–2428, Feb. 2025, doi: 10.1007/s10639-024-12867-w.

[11] A. Yashudas, D. Gupta, G. C. Prashant, A. Dua, D. AlQahtani, and A. S. K. Reddy, "DEEP-CARDIO: Recommendation System for Cardiovascular Disease Prediction Using IoT Network," *IEEE Sensors Journal*, vol. 24, no. 9, pp. 14539–14547, May 2024, doi: 10.1109/JSEN.2024.3373429.

[12] C.-L. Ho, W.-L. Chen, and C.-H. Ou, "Constructing a personalized travel itinerary recommender system with the Internet of Things," *Wireless Netw*, vol. 30, no. 7, pp. 6555–6567, Oct. 2024, doi: 10.1007/s11276-023-03453-y.
[13] Y. Xian *et al.*, "EX3: Explainable Attribute-aware Item-set Recommendations," in *Fifteenth ACM Conference on Recommender Systems*, Amsterdam Netherlands: ACM, Sep. 2021, pp. 484–494. doi: 10.1145/3460231.3474240.
[14] R. Rismala, N. U. Maulidevi, and K. Surendro, "Multi-Attribute BERT for Preferences Completion in Multi-Criteria Recommender System," in *Proceedings of the 2023 12th International Conference on Software and Computer Applications*, in ICSCA '23. New York, NY, USA: Association for Computing Machinery, Jun. 2023, pp. 315–320. doi: 10.1145/3587828.3587875.
[15] C. Musto, M. De Gemmis, P. Lops, and G. Semeraro, "Generating post hoc review-based natural language justifications for recommender systems," *User Model User-Adap Inter*, vol. 31, no. 3, pp. 629–673, Jul. 2021, doi: 10.1007/s11257-020-09270-8.
[16] L. Li, Y. Zhang, and L. Chen, "Generate Neural Template Explanations for Recommendation," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, Virtual Event Ireland: ACM, Oct. 2020, pp. 755–764. doi: 10.1145/3340531.3411992.
[17] N. Tintarev and J. Masthoff, "Explaining Recommendations: Design and Evaluation," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, and B. Shapira, Eds., Boston, MA: Springer US, 2015, pp. 353–382. doi: 10.1007/978-1-4899-7637-6_10.
[18] C. Musto, G. Rossiello, M. de Gemmis, P. Lops, and G. Semeraro, "Combining text summarization and aspect-based sentiment analysis of users' reviews to justify recommendations," in *Proceedings of the 13th ACM Conference on Recommender Systems*, in RecSys '19. New York, NY, USA: Association for Computing Machinery, Sep. 2019, pp. 383–387. doi: 10.1145/3298689.3347024.
[19] Z. C. Lipton, "The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery.," *Queue*, vol. 16, no. 3, pp. 31–57, Jun. 2018, doi: 10.1145/3236386.3241340.
[20] C. Molnar, *Interpretable Machine Learning*. Leanpub, 2024. Accessed: Oct. 02, 2024. [Online]. Available: https://christophm.github.io/interpretable-ml-book/
[21] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma, "Explicit factor models for explainable recommendation based on phrase-level sentiment analysis," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, Gold Coast Queensland Australia: ACM, Jul. 2014, pp. 83–92. doi: 10.1145/2600428.2609579.
[22] G. Peake and J. Wang, "Explanation Mining: Post Hoc Interpretability of Latent Factor Models for Recommendation Systems," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, in KDD '18. New York, NY, USA: Association for Computing Machinery, Jul. 2018, pp. 2060–2069. doi: 10.1145/3219819.3220072.
[23] X. Wang, Y. Chen, J. Yang, L. Wu, Z. Wu, and X. Xie, "A Reinforcement Learning Framework for Explainable Recommendation," in *2018 IEEE International Conference on Data Mining (ICDM)*, Nov. 2018, pp. 587–596. doi: 10.1109/ICDM.2018.00074.
[24] Y. Zhang and X. Chen, "Explainable Recommendation: A Survey and New Perspectives," *FNT in Information Retrieval*, vol. 14, no. 1, pp. 1–101, 2020, doi: 10.1561/1500000066.
[25] Y. Hou, N. Yang, Y. Wu, and P. S. Yu, "Explainable recommendation with fusion of aspect information," *World Wide Web*, vol. 22, no. 1, pp. 221–240, Jan. 2019, doi: 10.1007/s11280-018-0558-1.
[26] N. Wang, H. Wang, Y. Jia, and Y. Yin, "Explainable Recommendation via Multi-Task Learning in Opinionated Text Data," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, Jun. 2018, pp. 165–174. doi: 10.1145/3209978.3210010.
[27] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, Philadelphia Pennsylvania USA: ACM, Dec. 2000, pp. 241–250. doi: 10.1145/358916.358995.
[28] G. Friedrich and M. Zanker, "A Taxonomy for Generating Explanations in Recommender Systems," *AI Magazine*, vol. 32, no. 3, Art. no. 3, Jun. 2011, doi: 10.1609/aimag.v32i3.2365.
[29] M. Bilgic and R. J. Mooney, "Explaining Recommendations: Satisfaction vs. Promotion," *eyond personalization workshop, IUI*, vol. 5, p. 153, 2005.
[30] M. Coyle and B. Smyth, "Explaining Search Results," *International Joint Conference on Artificial Intelligence*, vol. 19, p. 1553, Jul. 2015.
[31] P. Bai, Y. Xia, and Y. Xia, "Fusing Knowledge and Aspect Sentiment for Explainable Recommendation," *IEEE Access*, vol. 8, pp. 137150–137160, 2020, doi: 10.1109/ACCESS.2020.3012347.
[32] H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis on review text data: a rating regression approach," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, in KDD '10. New York, NY, USA: Association for Computing Machinery, Jul. 2010, pp. 783–792. doi: 10.1145/1835804.1835903.
[33] H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis without aspect keyword supervision," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, in KDD '11. New York, NY, USA: Association for Computing Machinery, Aug. 2011, pp. 618–626. doi: 10.1145/2020408.2020505.
[34] J. McAuley, J. Leskovec, and D. Jurafsky, "Learning Attitudes and Attributes from Multi-aspect Reviews," in *2012 IEEE 12th International Conference on Data Mining*, Dec. 2012, pp. 1020–1025. doi: 10.1109/ICDM.2012.110.
[35] J. J. McAuley and J. Leskovec, "From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews," in *Proceedings of the 22nd international conference on World Wide Web*, Rio de Janeiro Brazil: ACM, May 2013, pp. 897–908. doi: 10.1145/2488388.2488466.
[36] N. R. Ahamed Kabeer, K. H. Gan, and E. Haris, "Domain-Specific Aspect-Sentiment Pair Extraction Using Rules and Compound Noun Lexicon for Customer Reviews," *Informatics*, vol. 5, no. 4, Art. no. 4, Dec. 2018, doi: 10.3390/informatics5040045.
[37] D. Chandrasekaran and V. Mago, "Evolution of Semantic Similarity—A Survey," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–37, Mar. 2022, doi: 10.1145/3440755.
[38] M. A. Hadj Taieb, T. Zesch, and M. Ben Aouicha, "A survey of semantic relatedness evaluation datasets and procedures," *Artif Intell Rev*, vol. 53, no. 6, pp. 4407–4448, Aug. 2020, doi: 10.1007/s10462-019-09796-3.
[39] C. De Boom, S. Van Canneyt, S. Bohez, T. Demeester, and B. Dhoedt, "Learning Semantic Similarity for Very Short Texts," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, Nov. 2015, pp. 1229–1234. doi: 10.1109/ICDMW.2015.86.
[40] J. Pennington, R. Socher, and C. Manning, "Glove: Global Vectors for Word Representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar: Association for Computational Linguistics, 2014, pp. 1532–1543. doi: 10.3115/v1/D14-1162.

[41] O. Abiola, A. Abayomi-Alli, O. A. Tale, S. Misra, and O. Abayomi-Alli, "Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser," *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, p. 5, Jan. 2023, doi: 10.1186/s43067-023-00070-9.

[42] S. Bag, S. K. Kumar, and M. K. Tiwari, "An efficient recommendation generation using relevant Jaccard similarity," *Information Sciences*, vol. 483, pp. 53–64, May 2019, doi: 10.1016/j.ins.2019.01.023.

[43] M. Eminagaoglu, "A new similarity measure for vector space models in text classification and information retrieval," *Journal of Information Science*, vol. 48, no. 4, pp. 463–476, Aug. 2022, doi: 10.1177/0165551520968055.

[44] R. Singh and S. Singh, "Text Similarity Measures in News Articles by Vector Space Model Using NLP," *Journal of The Institution of Engineers (India): Series B*, vol. 102, pp. 329–388, 2021, doi: https://doi.org/10.1007/s40031-020-00501-5.

[45] MD. A. Iqbal, O. Sharif, M. M. Hoque, and I. H. Sarker, "Word Embedding based Textual Semantic Similarity Measure in Bengali," *Procedia Computer Science*, vol. 193, pp. 92–101, Jan. 2021, doi: 10.1016/j.procs.2021.10.010.

[46] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[47] X. Chen, Y. Zhang, and J.-R. Wen, "Measuring 'Why' in Recommender Systems: a Comprehensive Survey on the Evaluation of Explainable Recommendation," Feb. 13, 2022, *arXiv*: arXiv:2202.06466. Accessed: Jan. 25, 2024. [Online]. Available: http://arxiv.org/abs/2202.06466