User Modeling and User-Adapted Interaction
https://doi.org/10.1007/s11257-024-09398-x

Understanding user intent modeling for conversational recommender systems: a systematic literature review

Siamak Farshidi[1,2] · Kiyan Rezaee[3] · Sara Mazaheri[3] · Amir Hossein Rahimi[3] · Ali Dadashzadeh[3] · Morteza Ziabakhsh[3] · Sadegh Eskandari[3] · Slinger Jansen[1,4]

Received: 6 August 2023 / Accepted in revised form: 17 May 2024

Abstract

User intent modeling in natural language processing deciphers user requests to allow for personalized responses. The substantial volume of research (exceeding 13,000 publications in the last decade) underscores the significance of understanding prevalent models in AI systems, with a focus on conversational recommender systems. We conducted a systematic literature review to identify models frequently employed for intent modeling in conversational recommender systems. From the collected data, we developed a decision model to assist researchers in selecting the most suitable models for their systems. Furthermore, we conducted two case studies to assess the utility of our proposed decision model in guiding research modelers in selecting user intent modeling models for developing their conversational recommender systems. Our study analyzed 59 distinct models and identified 74 commonly used features. We provided insights into potential model combinations, trends in model selection, quality concerns, evaluation measures, and frequently used datasets for training and evaluating these models. The study offers practical insights into the domain of user intent modeling, specifically enhancing the development of conversational recommender systems. The introduced decision model provides a structured framework, enabling researchers to

Kiyan Rezaee, Sara Mazaherim, Amir Hossein Rahimi and Sadegh Eskandari have contributed equally to this work.

✉ Siamak Farshidi
s.farshidi@uva.nl

✉ Sadegh Eskandari
eskandari@guilan.ac.ir

✉ Slinger Jansen
slinger.jansen@uu.nl

1 Department of Information and Computer Science, Utrecht University, Utrecht, The Netherlands

2 Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands

3 Department of Computer Science, University of Guilan, Rasht, Iran

4 Lappeenranta University of Technology, Lappeenranta, Finland

Published online: 06 June 2024

navigate the selection of the most apt intent modeling methods for conversational recommender systems.

Keywords User intent modeling · User behavior · Query intent · Conversational recommender systems · Personalized recommendation · Machine learning models

1 Introduction

User intent modeling is a fundamental process within natural language processing models, with the primary aim of discerning a user's underlying purpose or objective (Carmel et al. 2020). The comprehension and prediction of user goals and motivations through user intent modeling hold great significance in optimizing search engines and recommender systems (Zhang et al. 2019). This alignment of the user experience with preferences and needs contributes to enhanced user satisfaction and engagement (Oulasvirta and Blom 2008). Notably, state-of-the-art models such as ChatGPT have generated substantial interest for their potential in search engines and recommender systems (Cao et al. 2023), as they exhibit the capability to understand user intentions and engage in meaningful interactions.

The realm of user intent modeling finds extensive applications across diverse domains, spanning e-commerce, healthcare, education, and social media. In e-commerce, it plays a pivotal role in providing personalized product recommendations and detecting fraudulent product reviews (Tanjim et al. 2020; Wang et al. 2020; Guo et al. 2020; Paul and Nikolaev 2021). The healthcare sector leverages user intent modeling for delivering personalized health recommendations and interventions (Zhang et al. 2016; Wang et al. 2022). Similarly, within the educational sphere, it contributes to the tailoring of learning experiences to individual student goals and preferences (Liu et al. 2021; Bhaskaran and Santhi 2019). Furthermore, user intent modeling proves invaluable in comprehending user interests, preferences, and behaviors on social media, driving personalized content and targeted advertising delivery (Ding et al. 2015; Wang et al. 2019). Additionally, it plays a pivotal role in virtual assistants by aiding in the understanding of user queries and the provision of relevant responses (Penha and Hauff 2020; Hashemi et al. 2018).

User intent modeling approaches generally encompass a blend of models, including machine learning algorithms, to analyze various aspects of user input, such as words, phrases, and context. This approach enables the delivery of personalized responses within conversational recommender systems (Khilji et al. 2023). As a result, the domain of user intent modeling encompasses a diverse array of machine learning models, comprising Support Vector Machine (SVM) (Xia et al. 2018; Hu et al. 2017), Latent Dirichlet Allocation (LDA) (Chen et al. 2013; Weismayer and Pezenka 2017), Naive Bayes (Hu et al. 2018; Gu et al. 2016), as well as deep learning models like Bidirectional Encoder Representations from Transformers (BERT) (Yao et al. 2022), Word2vec(Da'u and Salim 2019; Ye et al. 2016), and Multilayer Perceptron (MLP) (Xu et al. 2022; Qu et al. 2016). A comprehensive examination of these models and their characteristics provides a holistic understanding of their advantages and limitations, thereby offering valuable insights for future research and development.

However, selecting the most suitable models within this domain presents challenges due to the multitude of available models and approaches (Zhang et al. 2019; Ricci et al. 2015). This challenge is further compounded by the absence of a clear classification scheme (Portugal et al. 2018), making it challenging for researchers and developers to navigate this diverse landscape and leading to uncertainties in model selection for specific requirements (Allamanis et al. 2018; Hill et al. 2016). Although user intent modeling in conversational recommender systems has been extensively explored, the research is dispersed across various sources, making it difficult to gain a cohesive understanding. The extensive array of machine learning models, concepts, datasets, and evaluation measures within this domain adds to the complexity.

To streamline and synthesize this wealth of information, we conducted a systematic literature review, adhering to established protocols set forth by Kitchenham et al. (2009), Xiao and Watson (2019), and Okoli and Schabram (2015). Furthermore, we developed a decision model derived from the literature review data, designed to assist in selecting user intent modeling methods. The utility of this decision model was evaluated through two academic case studies, structured in accordance with Yin's guidelines (Yin 2009).

The study is structured as follows: we begin by defining the problem statement and research questions, followed by an outline of the employed research methods in Sect. 2. Subsequently, we delve into the methodology of the Systematic Literature Review (SLR) in Sect. 3, with findings and analysis presented in Sect. 4. In Sect. 5, we shift our focus to the practical application of the collected data through the introduced decision model. Furthermore, we present academic case studies validating our research in Sect. 6. The outcomes, lessons learned, and implications of our findings are discussed in Sect. 7, with a comparative analysis of related studies provided in Sect. 8. Finally, Sect. 9 offers a summary of our contributions and outlines future research directions.

2 Research approach

This study uses a systematic research approach, combining an SLR and Case Study Research, to investigate approaches in user intent modeling for conversational recommended systems. The SLR was instrumental in collating relevant information from existing literature, and the case studies provided a means to evaluate the application of our findings.

2.1 Problem statement

Conversational recommender systems can be considered a multidisciplinary field, residing at the intersection of various domains including Human–Computer Interaction (HCI) (de Barcelos Silva et al. 2020; Rapp et al. 2021), Conversational AI (Zaib et al. 2022; Saka et al. 2023), Conversational Search Systems (Keyvan and Huang 2022; Yuan et al. 2020), User Preference Extraction & Prioritization (Pu et al. 2012; Liu et al. 2022), and Contextual Information Retrieval Systems (Tamine-Lechani et al.

2010; Chen et al. 2015). To develop an effective conversational recommender system empowered by user intent modeling, a comprehensive understanding of models and approaches recognized in other domains such as "topic modeling" (Vayansky and Kumar 2020), "user intent prediction" (Qu et al. 2019), "conversational search" (Zhang et al. 2018), "intent classification" (Larson et al. 2019), "intent mining" (Huang et al. 2018), "user response prediction" (Qu et al. 2016), "user behavior modeling" (Pi et al. 2019), and "concept discovery" (Sun et al. 2015) is essential. These concepts are discussed prominently within the context of search engines and recommender systems. This study acknowledges the evolving nature of modern search engines, increasingly incorporating conversational features, blurring the lines between traditional search engines and conversational recommender systems (Beel et al. 2016; Jain et al. 2015). Consequently, the research scope extends to encompass the analysis of search engines as conversational recommender systems, enabling exploration of how they engage with users, with a particular emphasis on accurately modeling and predicting user intent. It is essential to emphasize that developing conversational recommender systems requires a holistic understanding of key domains contributing to their efficacy and relevance. This study centers on integrating distinct yet interconnected fields pivotal for advancing conversational recommender systems, with each domain playing a unique role in shaping their design, functionality, and user engagement aspects.

An analysis of current approaches in user intent modeling reveals significant challenges, necessitating a systematic literature review and the development of a decision model:

Scattered knowledge: Concepts, models, and characteristics of intent modeling are widely dispersed in academic literature (Portugal et al. 2018), requiring systematic consolidation and categorization to advance conversational recommender systems.

Model integration: Combining different user intent modeling models is complex, demanding an analysis of their compatibility and synergistic potential (von Rueden et al. 2020).

Trends and emerging patterns: Understanding the evolving field of user intent modeling requires a comprehensive review of current trends and emerging patterns (Chen et al. 2015; Jordan and Mitchell 2015).

Assessment criteria: Selecting suitable evaluation measures for intent modeling is complex, necessitating tailored metrics for effective assessment (Telikani et al. 2021; Singh et al. 2016).

Dataset selection: Identifying appropriate datasets reflecting diverse intents and user behaviors for training and evaluating intent modeling approaches is a significant challenge (Yuan et al. 2020).

Decision-making framework: The absence of a decision model in current literature covering various intent modeling concepts and offering guidelines for model selection and evaluation highlights the critical need for such a model (Farshidi et al. 2020; Farshidi 2020).

These challenges form the basis of our research, leading to our systematic literature review and the development of a decision model to assist researchers in addressing these complexities in conversational recommender systems and user intent modeling.

2.2 Research questions

The research questions, formulated in response to the identified challenges, are as follows:
RQ_1: What models are commonly used in intent modeling approaches for conversational recommender systems?
RQ_2: What are the key characteristics and features supported by these models used in intent modeling?
RQ_3: What trends are observed in employing models for intent modeling in conversational recommender systems?
RQ_4: What evaluation measures and quality attributes are commonly used to assess the performance and efficacy of intent modeling approaches?
RQ_5: Which datasets are typically considered in the literature for training and evaluating machine learning models in intent modeling approaches?
RQ_6: How can a decision model be developed to assist researchers in selecting appropriate models for intent modeling approaches?

2.3 Research methods

A mixed research method was utilized to address these research questions, combining SLR and Case Study Research (Jansen 2009; Johnson and Onwuegbuzie 2004). The SLR provided an in-depth understanding of user intent modeling approaches, and the case studies evaluated the practical application of the proposed decision model.

The SLR followed guidelines by Kitchenham et al. (2009), Xiao and Watson (2019), and Okoli and Schabram (2015) to identify models, their definitions, combinations, supported features, potential evaluation measures, and relevant concepts. A decision model was developed from the SLR findings, influenced by previous studies on multi-criteria decision-making in software engineering (Farshidi 2020).

Two case studies were conducted to evaluate the decision model's practicality, following Yin's guidelines (Yin 2017). These studies tested whether the decision model effectively aided researchers in selecting models for their projects.

This mixed research method, encompassing SLR and case studies, provided insights and practical solutions for advancing intent modeling in conversational recommender systems.

2.4 Research methods

A mixed research method was utilized to address these research questions, combining SLR and Case Study Research (Jansen 2009; Johnson and Onwuegbuzie 2004). The SLR provided an in-depth understanding of user intent modeling approaches, and the case studies evaluated the practical application of the proposed decision model.

The SLR followed guidelines by Kitchenham et al. (2009), Xiao and Watson (2019), and Okoli and Schabram (2015) to identify models, their definitions, combinations, supported features, potential evaluation measures, and relevant concepts. A decision

model was developed from the SLR findings, influenced by previous studies on multi-criteria decision-making in software engineering (Farshidi 2020).

Two case studies were conducted to evaluate the decision model's practicality, following Yin's guidelines (Yin 2017). These studies tested whether the decision model effectively aided researchers in selecting models for their projects.

This mixed research method, encompassing both SLR and case studies, provided insights and practical solutions for advancing intent modeling in conversational recommender systems.

3 Systematic literature review methodology

This section outlines the review protocol employed in this study to systematically collect data from the literature on user intent modeling approaches for conversational recommender systems. The SLR review protocol, as depicted in Fig. 1, systematically collects and analyzes data relevant to this area.

(1) Problem formulation: The review protocol began by defining the problem and formulating research questions, followed by identifying research methods suitable for these questions. The procedures of Xiao and Watson (2019) were followed for defining the problem statement and formulating research questions, as detailed in Sects. 2.1 and 2.2. Analysis showed the first five research questions were suitable for exploration via an SLR. The outcomes of this SLR informed the development of a decision model. The final research question, focusing on the decision model's development and application, was addressed through case study research.

(2) Initial hypotheses: A set of keywords was initially selected to locate primary studies relevant to the research questions. These keywords helped identify potential

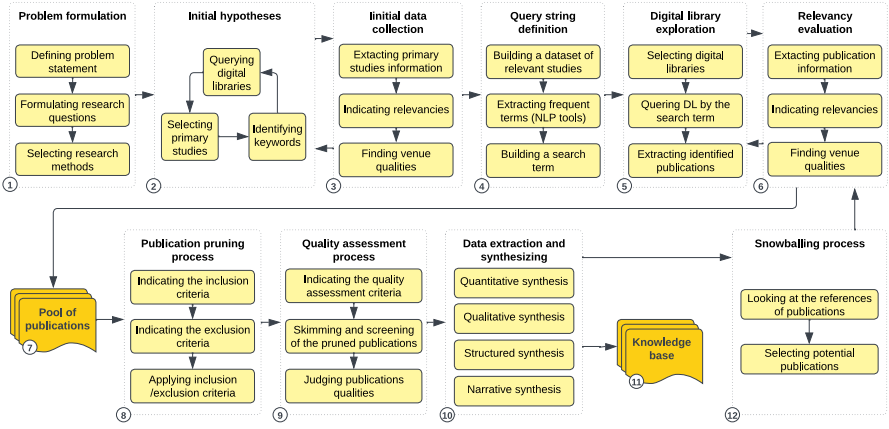


Fig. 1 The review protocol used in this study, following the guidelines by Kitchenham et al. (2009), Xiao and Watson (2019), and Okoli and Schabram (2015). The protocol consists of 12 elements for systematically collecting and extracting data from relevant studies, ensuring rigorous investigation and adherence to scientific standards. For details on how the protocol aligns with the guidelines, see Appendix A

seed papers, marking the start of our literature review and facilitating a systematic exploration of relevant publications.
(3) Initial data collection: Primary studies' characteristics, including source, URL, title, keywords, abstract, venue, venue quality, publication type, number of citations, publication year, and relevancy level, were manually collected. This process aided in focusing the review and establishing inclusion/exclusion criteria.
(4) Query string definition: The search query was developed by analyzing keywords, abstracts, and titles from primary studies, focusing on terms prevalent in relevant and high-quality papers. This method refined our search to include pertinent publications.
(5) Digital library exploration: Digital libraries such as ACM, ScienceDirect, and Elsevier were searched using the formulated query. This exploration ensured a thorough coverage of relevant publications.
(6) Relevancy Evaluation: Publications' characteristics were evaluated for relevance to our research questions and challenges, confirming the inclusion of pertinent publications in our review.
(7) The pool of publications: The selected papers formed the basis of our review. This collection was expanded through the snowballing process, providing a thorough examination of the literature.
(8) Publication pruning process: Inclusion/exclusion criteria were strictly applied to the pool of publications, filtering out irrelevant content and focusing on relevant and high-quality studies.
(9) Quality assessment process: The quality of remaining publications was evaluated based on criteria like clarity of research questions and findings, ensuring the inclusion of only high-quality studies.
(10) Data extraction and synthesizing: Systematic data extraction from selected publications facilitated the identification and summarization of key information.
(11) Knowledge base: The final selection of publications formed a knowledge base, with extracted data linking findings and sources. This base serves as a resource for future research and further analysis.
(12) Snowballing process: Additional relevant papers were identified by reviewing references in selected publications, enhancing the review's comprehensiveness.

This systematic review protocol ensured rigorous standards in collecting and analyzing literature on user intent modeling approaches, ensuring the validity and reliability of our study.

3.1 Review protocol

This section details the implementation of the review protocol, as depicted in Fig. 1, for our SLR.

3.1.1 Pool of publications

In our systematic literature review, the *manual search* phase, comprising the Initial Hypothesis and Initial Data Collection stages, preceded the *automatic search* phase. During the manual phase, we initially collected publications and extracted keywords

indicative of common terms in noteworthy high-quality papers. These keywords were subsequently used as the foundation for the *automatic search* phase, beginning with the Query String Definition step.

In the manual search phase, we initially gathered a set of primary studies using search terms to identify relevant publications addressing our research questions. These terms were refined based on our domain understanding, considering publication abstracts, keywords, and titles. This process led to the identification of 314 highly relevant and high-quality publications. A publication was deemed 'relevant' if it addressed at least one of our research questions (Sect. 2.2). We evaluated quality based on criteria like publication venue reputation (CORE Rankings Portal[1] and Scimago Journal & Country Rank (SJR)[2]), citation count, and recency.

Subsequently, we employed Sketch Engine (Kilgarriff et al. 2014), a topic modeling tool, to extract frequently mentioned keywords from these 314 primary studies. We considered keywords that appeared at least three times and used them to formulate our search query for the *automatic search* phase, focusing on topics related to user intent modeling in search engines and recommender systems, including intent detection, prediction, interactive modeling, conversational search, classification, and user behavior modeling. Our search query combined keywords using logical operators "AND" and "OR," resulting in the following query:

("user intent" OR "user intent modeling" OR "topic model" OR "user intent detection" OR "user intent prediction" OR "interactive intent modeling" OR "conversational search" OR "intent classification" OR "intent mining" OR "conversational recommender system" OR "user response prediction" OR "user behavior modeling" OR "interactive user intent" OR "intent detection" OR "concept discovery") AND ("search engine" OR "recommender system")

In the *automatic search* phase, we assessed the relevance of these papers by examining their titles, abstracts, keywords, and conclusions, classifying them as 'highly relevant' (addressing at least three research questions), 'medium relevant' (addressing two questions), 'low relevant' (addressing one question), or 'irrelevant' (not addressing any questions). After this evaluation, we excluded irrelevant publications from the pool, leaving 3,828 relevant publications out of the initial 13,168 search results.

The publications underwent rigorous screening, adhering strictly to our predefined inclusion and exclusion criteria, ensuring the selection of only the most pertinent and high-quality publications for data extraction and analysis. We assessed the effectiveness of our search query by comparing results with those from a manual search, confirming the consistency and accuracy of our approach. This evaluation verified that our query included publications identified as high-quality and highly relevant during the manual phase, affirming the successful retrieval of publications relevant to user intent modeling in search engines and recommender systems.

[1] https://www.core.edu.au.

[2] https://www.scimagojr.com/.

3.1.2 Publication pruning process

In systematic literature reviews or meta-analyses, inclusion/exclusion criteria play a pivotal role as definitive guidelines for determining study relevance and eligibility. These criteria guarantee the selection of high-quality studies that directly address the research question.

For our study, we implemented stringent inclusion and exclusion criteria to eliminate irrelevant and low-quality publications. These criteria considered several factors, including the publication venue's quality, publication year, citation counts, and relevance to our research topic. We precisely defined and consistently applied these criteria to include only high-quality and relevant publications.

We categorized publications based on their quality using assessments from the CORE Rankings Portal and SJR:

- Publications with "*A**" or "*Q1*" indicators were classified as "*Excellent*."
- Those with "*A*" or "*Q2*" were deemed "*Good*."
- "*B*" or "*Q3*" were categorized as "*Average*."
- "*C*" or "*Q4*" were labeled as "*Poor*."
- Publications without quality indicators on these platforms were marked as "*N/A*."

Publications classified as "*Poor*" or "*N/A*" were excluded from further consideration. Additional exclusion criteria encompassed publications with low citation counts, older publication dates, or classification as Gray literature (e.g., books, theses, reports, and short papers).

After applying our predefined inclusion/exclusion criteria, we identified and selected 1067 publications from the initial pool of 3828 publications.

3.1.3 Quality assessment process

During the SLR, we assessed the quality of the selected publications after applying the inclusion/exclusion criteria. Several factors were taken into consideration to evaluate the quality and suitability of the publications for our research:

Research method: We evaluated whether the chosen research method was appropriate for addressing the research question. The clarity and transparency of the research methodology were also assessed.

Research type: We considered whether the publication presented original research, a review article, a case study, or a meta-analysis. The relevance and scope of the research in the field of machine learning were also taken into account.

Data collection method: We evaluated the appropriateness of the data collection method in relation to the research question. The adequacy and clarity of the reported data collection process were also assessed.

Evaluation method: We assessed whether the chosen evaluation method was suitable for addressing the research question. The transparency and statistical significance of the reported results were considered.

Problem statement: We evaluated whether the publication identified the research problem and provided sufficient background information. The clarity and definition of the research question were also taken into account.

Research questions: We assessed the relevance, clarity, and definition of the research questions in relation to the research problem.
Research challenges: We considered whether the publication identified and acknowledged the challenges and limitations associated with the research.
Statement of findings: We evaluated whether the publication reported the research results and whether the findings were relevant to the research problem and questions.
Real-world use cases: We assessed whether the publication provided real-world use cases or applications for the proposed method or model.

Based on these assessment factors, a team of five researchers involved in the SLR evaluated the publications' quality. Each researcher independently assessed the publications based on the established criteria. In cases where there were discrepancies or differences in evaluating a publication's quality, the researchers engaged in discussions to reach a consensus and ensure a consistent assessment.

Through this collaborative evaluation process, a final selection of 791 publications was made from the initial pool of 1,067 publications. These selected publications demonstrated high quality and relevance to our research question, meeting the predefined inclusion/exclusion criteria. The consensus reached by the research team ensured a rigorous and reliable selection of publications for further analysis and data extraction in the SLR.

3.1.4 Data extraction and synthesizing

During the data extraction and synthesis phase of the SLR, our primary objective was to address the identified research questions and gain insights into the foundational models commonly employed by researchers in their intent modeling approaches. We aimed to understand the features of these models, the associated quality attributes, and the evaluation measures utilized by research modelers to assess their approaches. Furthermore, we explored the potential combinations of models that researchers incorporated into their research papers.

We extracted relevant data from the papers included in our review to achieve these objectives. In our perspective, evaluation measures encompassed a range of measurements and key performance indicators (KPIs) used to evaluate the performance of the models. Quality attributes represent the characteristics of models that are not easily quantifiable and are typically assigned values using Likert scales or similar approaches. For example, authors may assess the performance of a model as high or low compared to other models. On the other hand, features encompassed any characteristics of models that authors highlighted to demonstrate specific functionalities. These features played a role in the selection of models by research modelers. Examples of features include ranking and prediction capabilities.

In this study, 'models' are conceptualized as structured, mathematical, or computational frameworks employed for simulating, predicting, or classifying phenomena within user intent modeling in conversational recommender systems. These models are organized into a variety of categories, reflecting diverse methodologies. This includes Supervised Learning, where models are trained on labeled data for accurate predictions; Unsupervised Learning, which uncovers patterns in unlabeled data; and Collaborative Filtering, among others, each offering unique insights into user interac-

Table 1 An overview of the systematic search process for identifying relevant publications on user intent modeling for conversational recommender systems

	#hits	Phase 1	Phase 2	Phase 3	Phase 4
Google scholar	3940	314	96	96	68
ACM DL	2152	586	311	311	243
IEEE Xplore	89	82	9	9	7
ScienceDirect	1528	921	246	246	190
Springer	5459	1896	379	379	263
Snowballing	N/A	29	26	26	20
	13,168	3828	1067	1067	**791**

tions. Furthermore, the study emphasizes the critical role of development metrics such as Cosine similarity (Gunawan et al. 2018) and KL Divergence (Bigi 2003), which are not just evaluation tools but are fundamental in refining and optimizing the functionality of these models. Algorithmic and computational techniques like ALS (Takács and Tikk 2012) and BM25 (Robertson et al. 2004) also play an integral part in the implementation and efficacy of these categorized models (refer to Sect. 4.1).

By extracting and analyzing this data, we aimed to comprehensively understand the existing literature, including popular open-access datasets used for training and evaluating the models. This knowledge empowered us to contribute insights and recommendations to the academic community, supporting them in selecting appropriate models and approaches for their intent modeling research endeavors.

3.2 Search process

In this study, we followed the review protocol presented in this section (see Fig. 1) to gather relevant studies.

The search process involved an automated search phase, which utilized renowned digital libraries such as ACM DL, IEEE Xplore, ScienceDirect, and Springer. However, Google Scholar was excluded from the automated search due to its tendency to generate numerous irrelevant studies. Furthermore, Google Scholar significantly overlaps the other digital libraries considered in this SLR. Table 1 provides an overview of the sequential phases of the search process, outlining the number of studies encompassed within each stage. It provides insights into the search process conducted in four phases:
Phase 1 (Pool of Publications): We initially performed a manual search, resulting in 314 relevant publications from Google Scholar. Additionally, automated searches from ACM DL, IEEE Xplore, ScienceDirect, and Springer contributed to the pool of publications with 586, 82, 921, and 1,896 relevant papers, respectively.
Phase 2 (Publication pruning process): In this phase, the inclusion/exclusion criteria were applied to the collected publications, ensuring the selection of high-quality and relevant studies. The numbers were reduced to 311 in ACM DL, 9 in IEEE Xplore, 246 in ScienceDirect, and 379 in Springer.

Phase 3 (Quality assessment process): Quality assessment was conducted for the publications based on several criteria, resulting in a final selection of 1067 studies from all sources.
Phase 4 (Data extraction and synthesizing + Snowballing process): During this phase, data extraction and synthesis were performed to gain insights into foundational intent modeling models, quality attributes, evaluation measures, and potential combinations of models used by researchers. Additionally, snowballing, involving reviewing references of selected publications, led to an additional 20 relevant papers. Applying the review protocol and snowballing, we retrieved 791 high-quality studies for our comprehensive analysis and synthesis in this systematic literature review.

4 Findings and analysis

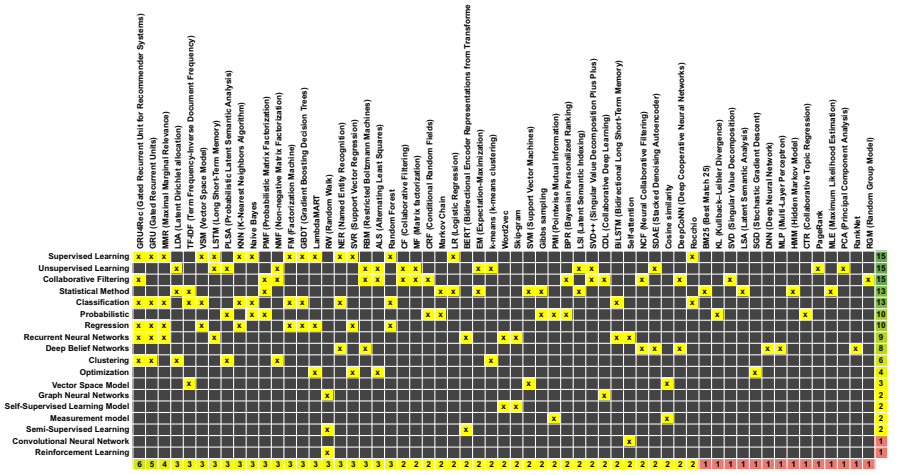
In this section, we present the SLR results and provide an overview of the collected data[3], which were analyzed to address the research questions identified in our study.

4.1 Models

This study defines a 'model' as a structured, mathematical, or computational framework specifically designed for simulating, predicting, or classifying phenomena within user intent modeling in conversational recommender systems. These models have been organized into distinct categories, each representing a unique approach to comprehending and interpreting user interactions.
Model categories: Our categorization includes various methodologies such as Supervised Learning, Unsupervised Learning, Collaborative Filtering, and others. For instance, models under Supervised Learning rely on labeled data for training, enabling them to make informed predictions or classifications. Unsupervised Learning models, in contrast, derive insights autonomously from unlabeled data, revealing underlying patterns without explicit guidance.
Development metrics: To measure and refine model performance, development metrics like Cosine similarity (Gunawan et al. 2018) and Kullback–Leibler (KL) Divergence (Bigi 2003) are employed. These metrics are not just evaluative tools; they are pivotal in enhancing system functionality and optimization throughout the development process. In the development and assessment of conversational recommender systems, it is essential to differentiate between metrics used for system development and those applied for model evaluation. Metrics such as Cosine similarity and KL Divergence are integral during the development phase, where they contribute significantly to system functionality and optimization. These metrics help fine-tune the system by assessing similarity measures and information loss. Conversely, the evaluation of model performance relies on a distinct set of measures, which are crucial for understanding the efficacy and accuracy of models in real-world applications. These

[3] For access to the complete lists of datasets, quality attributes and evaluation measures, models, and features related to this study, please refer to the supplementary materials available on *Mendeley Data* (Farshidi 2024).

Table 2 The mapping of 59 models to their respective categories in user intent modeling, showcasing models featured in at least six publications



The complete list is available on Mendeley Data (Farshidi 2024)

evaluation measures are detailed in Sect. 4.5, providing insights into how well the models perform regarding user intent prediction and recommendation accuracy.

Algorithmic and computational techniques: The study also underscores the importance of various algorithmic and computational techniques, such as ALS (Takács and Tikk 2012) and BM25 (Robertson et al. 2004). These techniques are integral to the practical implementation of the categorized models, aiding in critical tasks like data processing and system optimization.

The SLR conducted reveals a multifaceted landscape of models used in user intent modeling, each marked by its distinct methodology and application. Detailed information about these models and their categorizations can be found in the appendix (Appendix C).

Key categories such as Classification (Qu et al. 2019; Zhang et al. 2016) and Clustering (Zhang et al. 2021; Agarwal et al. 2020) models, Convolutional Neural Network (CNN)(Wang et al. 2020; Zhang et al. 2016), Deep Belief Networks (DBN)(Zhang et al. 2018; Hu et al. 2017), and Graph Neural Networks (GNN) (Yu et al. 2022; Lin et al. 2021) are highlighted. These categories, detailed in our SLR, represent a spectrum of techniques and approaches within user intent modeling.

Table 2 presents an overview of the 59 most frequently mentioned models in the SLR on user intent modeling. The table showcases the models appearing in at least six publications (columns) and their corresponding 18 categories (rows). Each model in user intent modeling can often be categorized into multiple categories, highlighting their versatility and diverse functionalities. For example, GRU4Rec (Hidasi and Karatzoglou 2018), a widely recognized model in the field (cited in 10 publications included in our review), exhibits characteristics that align with various categories. GRU4Rec falls under Supervised Learning, as it uses labeled examples during training to pre-

dict user intent. Additionally, it incorporates Collaborative Filtering techniques by analyzing user behavior and preferences to generate personalized recommendations, associating it with the Collaborative Filtering category (Latifi et al. 2021). Moreover, GRU4Rec can be classified as a Classification model as it categorizes input data into specific classes or categories to predict user intent (Park et al. 2020). It also demonstrates traits of Regression models by estimating and predicting user preferences or ratings based on the available data. Considering its reliance on recurrent connections, GRU4Rec can be associated with the Recurrent Neural Networks (RNN) category, enabling it to process sequential data and capture temporal dependencies (Ludewig and Jannach 2018). Lastly, GRU4Rec's ability to cluster similar users or items based on their behavior and preferences places it within the Clustering category. This clustering capability provides valuable insights and recommendations to users based on their respective clusters.

4.2 Features

In our research, we analyzed user intent modeling within conversational recommender systems. This involved the identification of 74 distinct features, each frequently mentioned in a minimum of six publications. These features provide an alternative means of categorizing models based on the specific functions they are designed to serve, as described by the authors in their studies. Subsequently, we categorized the models utilized in these systems based on the features they support, presenting the results systematically in Table 3.

We grouped these features into 20 categories, each reflecting specific contexts and applications. Features such as historical data references (Zhou et al. 2020; White et al. 2013; Zou et al. 2022) enable models to leverage past interactions for future predictions, while algorithm-agnostic models (Zhou et al. 2019; Musto et al. 2019; Mandayam Comar and Sengamedu 2017) offer flexibility in selecting the most suitable algorithms for specific tasks. Model-based features (Ding et al. 2022; Pradhan et al. 2021; Yu et al. 2018), which rely on statistical methods (Schlaefer et al. 2011; Kim et al. 2017) and semantic analysis (Zhang and Zhong 2016; Xu et al. 2015), are used to provide predictions based on predefined models.

The categorization includes various focus areas: 'Rule-Based Approaches' use pattern and template methods to interpret user intent, while 'Query Processing' models specialize in refining user queries to improve interaction quality. In 'Predictive Modeling', the focus is on forecasting user preferences using techniques such as Prediction and Ratings Prediction. 'Text Analytics' involves models that perform Topic Modeling, Text Similarity, and Semantic Analysis, which are crucial for analyzing user dialogues. Personalization features, ranging from 'User-Based Personalization' and 'Temporal Personalization' to 'Content-Based Personalization' and 'Interaction-Based Personalization', adapt recommendations according to user activity, time factors, content characteristics, and user interactions. Finally, 'Recommendation Techniques' cover a broad spectrum of models optimized for tasks like Item Recommendation, Hybrid Recommendation, and Ranking.

Table 3 Maps 74 features to corresponding models in user intent modeling, dividing them into 20 categories. For comprehensive definitions and explanations, see Appendix D

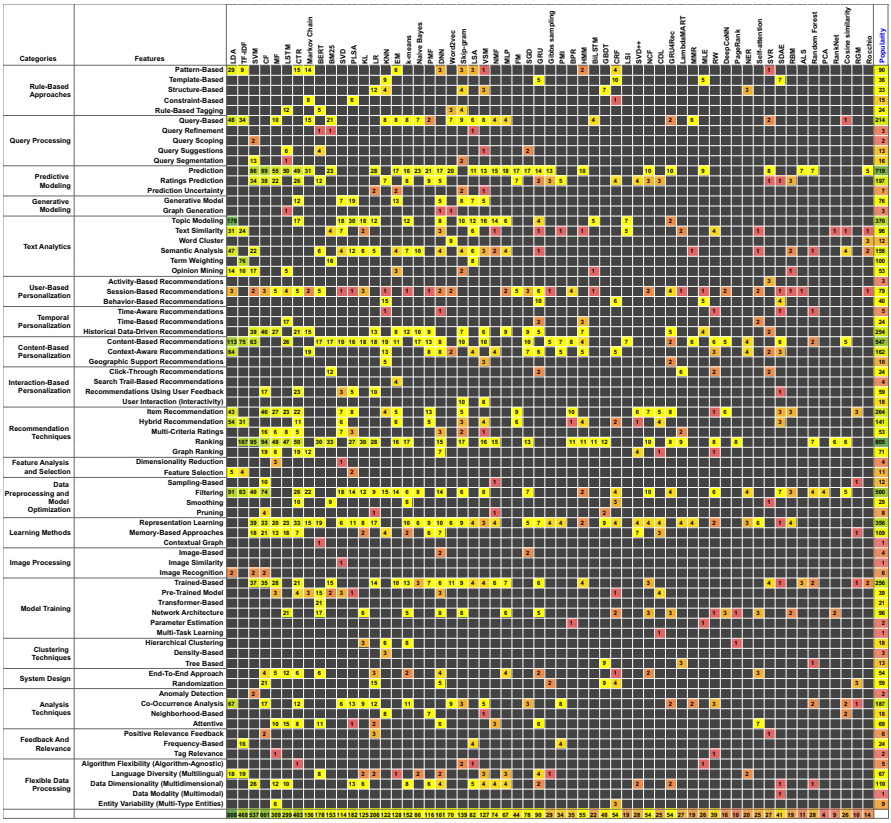


Table 3 not only illustrates the mapping of features to models in user intent modeling but also highlights the frequency of explicit mentions of these features in relevant publications. The color coding in each cell indicates the level of support for each feature by the models, with gray cells denoting an absence of evidence supporting the feature's compatibility with a particular model, based on our comprehensive review of 791 papers. For example, LDA is frequently mentioned in the context of pattern-based approaches within rule-based methods (Tang et al. 2010; Li et al. 2014), underscoring its applicability in scenarios where patterns are analyzed to extract meaningful insights.

The process of mapping features to models in user intent modeling requires an in-depth understanding of the particular features and the capabilities of the available models. For instance, in text analysis and natural language processing, models like LDA, TF-IDF, and BERT are often chosen for their effectiveness in semantic analysis and topic modeling. Similarly, for predictive modeling tasks, SVM, Random Forest, and Gradient Boosted Decision Trees (GBDT) are preferred due to their accuracy in classification and regression tasks. In cases where temporal dynamics are significant,

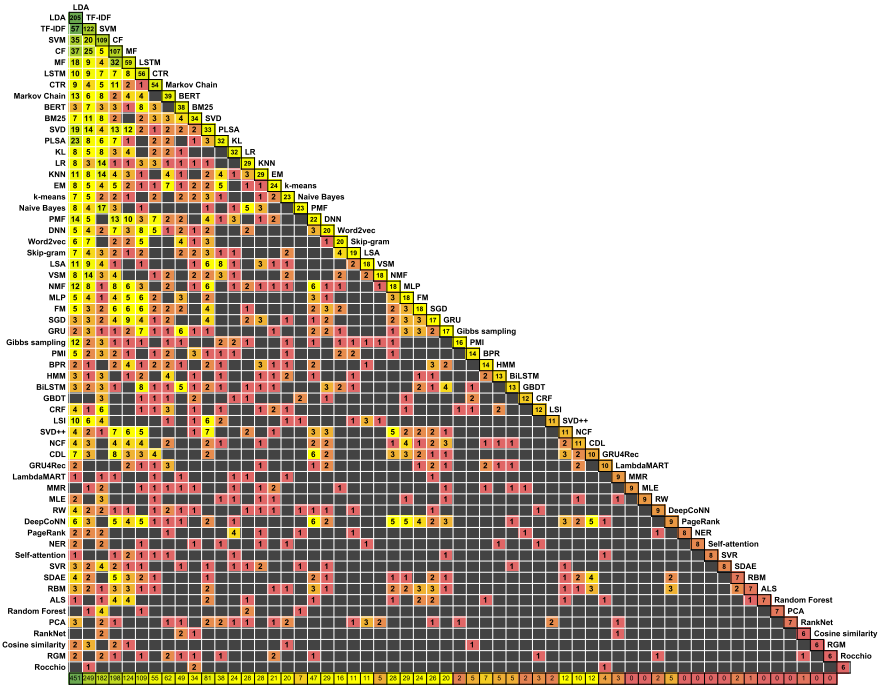


Fig. 2 Shows the matrix representation of model combinations in user intent modeling research. The matrix shows the combinations of 59 models, with each cell indicating the number of publications discussing the model combination. Diagonal cells show the count of publications discussing each model individually. Green cells represent a higher number of research articles, yellow and red cells indicate a lower number, and gray cells show areas with no evidence of valid combinations. The last row indicates the frequency of publications where models were combined with others. For example, 451 publications mentioned LDA in combination with other models. This combination matrix offers insights into the frequency and popularity of model combinations, helping researchers identify existing combinations and potential research areas

models like LSTM, GRU, and Markov Chains are utilized for their ability to handle sequential data effectively. Furthermore, for tasks involving recommendation systems, models like Matrix Factorization (MF), Collaborative Filtering (CF), and Neural Collaborative Filtering (NCF) are often employed for their efficiency in capturing user preferences and generating personalized recommendations.

4.3 Model combinations

The data extraction and synthesis phase of the SLR identified 59 models, each referenced in a minimum of six publications. These models were often integrated to address various research considerations, such as feature requirements, quality attributes, and evaluation measures, as illustrated in Fig. 3. The selected publications discussed combinations of models based on the authors' research and evaluated the outcomes of these combinations.

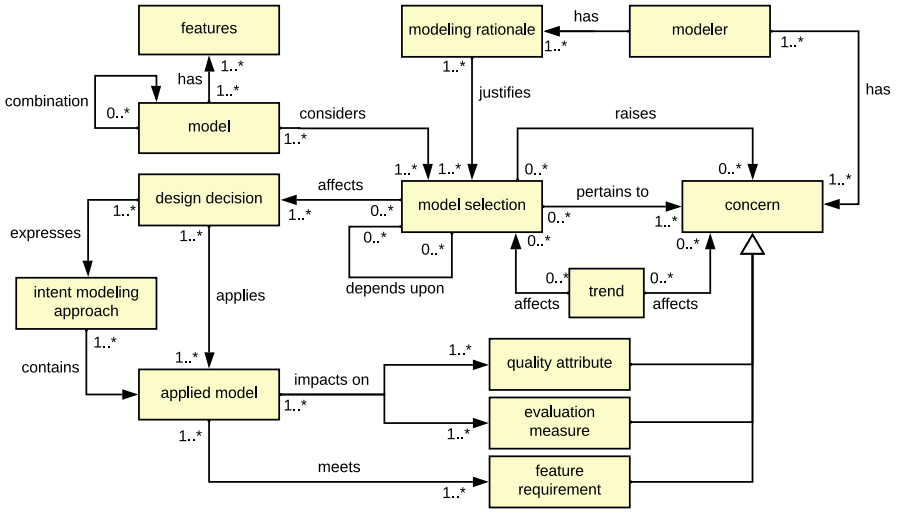


Fig. 3 The decision-making process researchers employ in selecting intent modeling approaches within the academic literature

To analyze model combinations, a matrix similar to a symmetric adjacency matrix was created, with models as nodes and combinations as edges in a graph. This matrix, shown in Fig. 2, includes 59 models. Diagonal cells indicate the count of publications discussing each model independently, such as 205 papers on LDA (Chen et al. 2013; Weismayer and Pezenka 2017) and 122 on TF-IDF (Binkley et al. 2018; Izadi et al. 2022).

Matrix cells show the number of papers discussing model combinations. For instance, 57 papers explored the LDA and TF-IDF combination (Venkateswara Rao and Kumar 2022), and 35 examined SVM and LDA (Yu and Zhu 2015).

The matrix uses color coding to indicate the research volume associated with each combination. Green cells represent higher research volumes, yellow and red lower volumes, and gray cells indicate areas lacking evidence of valid combinations. These gray areas present opportunities for future research.

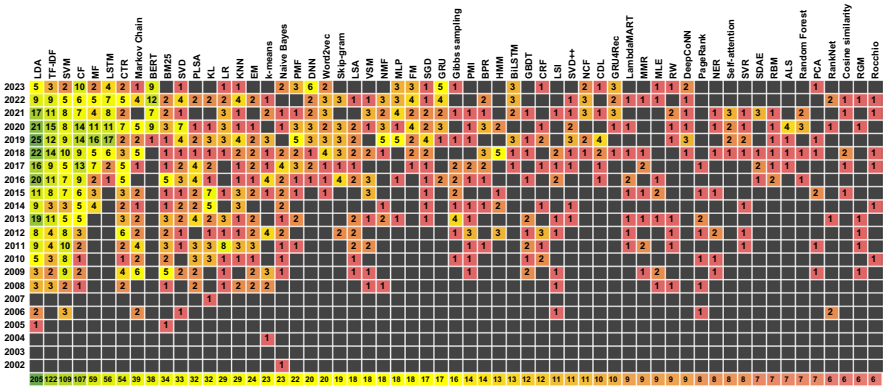
The combination matrix provides an overview of model combinations in user intent modeling research, highlighting the frequency of their use in literature and serving as a resource for identifying existing combinations and potential research areas.

Combining various models, often termed as 'ensemble' or 'hybrid' modeling (Sagi and Rokach 2018), can enhance the predictive power (Beemer et al. 2018) and accuracy of conversational recommender systems. However, this approach is subject to certain constraints and requires careful consideration.

Firstly, it's crucial to acknowledge that while combining models is possible, it's not always straightforward or advantageous. The feasibility of integrating multiple models depends on several factors:

Compatibility: The models to be combined must be compatible in terms of input and output data formats, scale, and the nature of predictions they make (Srivastava et al. 2020). For instance, combining a probabilistic model with a neural network requires

Table 4 The trend of models mentioned in user intent modeling research over publication years, highlighting the popularity and emergence of various models



a harmonious interface where the output of one can effectively serve as the input for another.

Complexity and overfitting: Increasing model complexity can lead to overfitting, where the model performs well on training data but poorly on unseen data (Sagi and Rokach 2018). It is essential to balance the complexity with the generalizability of the model.

Computational resources: More complex ensembles demand greater computational power and resources. This can be a limiting factor, especially in real-time applications (Bifet et al. 2009).

Interpretability: Combining models can sometimes lead to a loss of interpretability, making it challenging to understand how predictions are made, which is crucial for certain applications (Wang and Lin 2021).

Regarding a more in-depth analysis, combining models indeed necessitates a thorough evaluation of their individual and collective performance. This includes assessing how they complement each other, their synergistic potential, and the trade-offs involved.

4.4 Model trends

In recent studies, machine learning models have witnessed significant advancements across various fields, leading to notable trends in their development and application. However, it is worth noting that our study goes beyond recent years. By using the term "models," we refer to a wide range of models that research modelers can employ in user intent modeling.

To gain insights into the usage patterns of these models, we organized the 59 selected models (mentioned in at least six publications) based on the publication years of the studies that referenced them. The span of these publications ranges from 2002 to 2023. Table 4 provides an overview of these trends.

Among the selected models, LDA, TF-IDF, SVM, CF, and MF emerged as the top five most frequently mentioned models, appearing in over 500 papers. It is important to note that while some recently gained substantial attention, such as BERT (Yao et al. 2022), CF (Yadav et al. 2022), LSTM (Xu et al. 2022; Gozuacik et al. 2023), DNN (Yengikand et al. 2023), and GRU (Chen and Wong 2020; Elfaik 2023), our study encompasses models from various time periods.

These trends shed light on the popularity and usage patterns of different models in user intent modeling. By identifying frequently mentioned models and observing shifts in their prevalence over time, researchers and practitioners can stay informed about the evolving landscape of user intent modeling and make informed decisions when selecting models for their specific applications (Zaib et al. 2022; Ittoo and van den Bosch 2016).

4.5 Quality models and evaluation measures

In AI-based projects, selecting high-quality models and using evaluation measures is crucial. Quality attributes, defined in studies (de Barcelos Silva et al. 2020; Hernández-Rubio et al. 2019), reflect a model's performance, effectiveness, and user-centric features in conversational recommender systems. These attributes are essential for a comprehensive evaluation but are not straightforward to measure empirically. They often require subjective assessment or indirect methods. "Novelty," for example, relates to the uniqueness of recommendations (Cremonesi et al. 2011). Although challenging to quantify, methods like user studies or item distribution analysis can offer insights into a model's novelty. Conversely, evaluation measures, as discussed in literature (Zaib et al. 2022), provide a quantitative assessment of model outputs. These attributes and measures are pivotal in delivering accurate and reliable results, as various studies demonstrate (Pan et al. 2022; Pu et al. 2012; Hernández-Rubio et al. 2019).

While accuracy is a commonly employed evaluation measure, it may not adequately represent the model's performance, especially in imbalanced classes. Alternative measures such as precision (Salle et al. 2022; Baykan et al. 2011), recall (Wang et al. 2022; Phan et al. 2010), and F1-score (Yu et al. 2019; Ashkan et al. 2009) are used to evaluate model performance, particularly when dealing with imbalanced data. Additionally, evaluation measures like the area under the curve (AUC) (Xu et al. 2016; Liu et al. 2022) and receiver operating characteristic (ROC) (Wu et al. 2019; Wang et al. 2020) curve are frequently used to assess binary classifiers. These measures provide insights into the model's ability to differentiate between positive and negative instances, particularly when the costs of false positives and false negatives differ.

For ranking problems, evaluation measures such as mean average precision (MAP) (Mao et al. 2019; Ni et al. 2012) and normalized discounted cumulative gain (NDCG) (Liu et al. 2020; Kaptein and Kamps 2013) are commonly employed. These measures evaluate the quality of the ranked lists generated by the model and estimate its effectiveness in predicting relevant instances.

When evaluating regression models, measures such as root mean squared error (RMSE) (Cai et al. 2014; Colace et al. 2015) and mean absolute error (MAE) (Yao

Table 5 An overview of quality models and evaluation measures used in machine learning, including performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, as well as other evaluation techniques such as cross-validation, holdout validation, and confusion matrices

#F	Quality attributes	#F	Quality attributes	#F	Quality attributes	#F	Evaluation measures
675	Performance	89	Validity	22	Predictability	233	Precision
363	Effectiveness	72	Novelty	19	Clarification	200	Recall
252	Diversity	48	Resource Utilization	18	Classification	150	F1-Score
206	Usefulness	43	Flexibility	15	Query Accuracy	120	Accuracy
203	Stability	42	Computational Cost	14	Appropriateness	120	Normalized Discounted Cumulative Gain
158	Scalability	42	Interpretability	10	Comparability	81	Mean Average Precision (MAP)
139	Recommendation	41	Retrieval Performance	10	Retrieval Accuracy	76	Mean Reciprocal Rank (MRR)
128	Satisfaction	40	Convergence	9	Readability	54	Area Under the ROC Curve (AUC)
125	Coverage	40	Recommendation	8	Persuasiveness	27	Mean Absolute Error (MAE)
122	Robustness	39	Classification Accuracy	8	Scrutability	24	Hit Rate (HR)
114	Resource Efficiency	35	Transparency	6	Unexpectedness	22	Discounted Cumulative Gain (DCG)
111	Simplicity	31	Informativeness	6	Memory Efficiency	19	Root Mean Squared Error (RMSE)
99	Reliability	24	Recommendation Efficiency			9	Normalized Mutual Information (NMI)

et al. 2017; Yadav et al. 2022) are used to quantify the discrepancy between predicted values and actual values of the target variable.

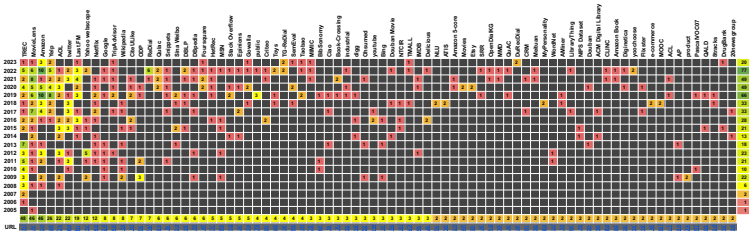
The selection of appropriate evaluation measures is crucial to ensure the accuracy and reliability of machine learning models. The suitable measure(s) choice depends on the specific problem domain, data type, and project objectives. These factors are pivotal in selecting the most appropriate quality attributes and evaluation measures. Table 5 presents the quality attributes and evaluation measures identified in at least six publications. Performance, Effectiveness, Diversity, Usefulness, and Stability are among the top five quality attributes. Precision, recall, F1-score, accuracy, and NDCG are among the top five evaluation measures identified in the SLR. For detailed explanations of the identified quality attributes and evaluation measures, please refer to Appendix E.

4.6 Datasets

Datasets are fundamental to machine learning and data science research, as they provide the raw material for training and testing models and enable the development of solutions to complex problems. They come in various forms and sizes, ranging from small, well-curated collections to large, messy datasets with millions of records. The quality of datasets is crucial (Pan et al. 2022), as high-quality data ensure the accuracy and reliability of models, while poor-quality data can introduce biases and inaccuracies. Data quality encompasses completeness, accuracy, consistency, and relevance, and ensuring data quality involves cleaning, normalization, transformation, and validation.

The size and complexity of datasets pose challenges in terms of storage, processing, and analysis. Big datasets require specialized tools and infrastructure to handle the volume and velocity of data. On the other hand, complex datasets, such as graphs, images, and text, may require specialized techniques and models for extracting meaningful information and patterns.

Table 6 Datasets commonly used for user intent modeling approaches



Furthermore, the availability of datasets is a vital consideration in advancing machine learning research and applications. Open datasets that are freely accessible and well-documented foster collaboration and innovation, while proprietary datasets may restrict access and impede progress (Zhang et al. 2016; Teevan et al. 2008; Ittoo and van den Bosch 2016). Data sharing and ethical considerations in data use are increasingly recognized, leading to efforts to promote open-access and responsible data practices.

In this study, we identified 80 datasets that researchers have utilized in the context of intent modeling approaches, and these datasets were mentioned in at least two publications. Table 6 provides an overview of these datasets and their frequency of usage from 2005 to 2023. Notably, TREC, MovieLens, Amazon, Yelp, and AOL emerged as the top five datasets commonly used in evaluating intent modeling approaches for recommender systems (Wang et al. 2021; Papadimitriou et al. 2012; Wang et al. 2020) and search engines (Fan et al. 2022; Liu et al. 2022; Konishi et al. 2016). These datasets have been utilized in over 200 publications, highlighting their significance and wide adoption in the field.

The datasets selected for this study cover a broad range of scenarios in user intent modeling for conversational recommender systems. This diversity aligns with the comprehensive nature of the research. Each dataset contributes unique insights into user behaviors, preferences, and interactions, which are crucial for understanding and effectively modeling user intent within conversational interfaces.

The variety of datasets reflects the complexity of conversational recommender systems, which need to address varied user needs, contexts, and interaction modes. Including datasets that differ in size, structure, and origin ensures the study captures a broad spectrum of user interactions and system responses, providing a solid foundation for developing and evaluating intent modeling approaches.

5 Decision-making process

This section describes how researchers make decisions when selecting intent modeling approaches. It illustrates a systematic approach to choosing intent modeling methods based on academic literature.

5.1 Decision meta-model

Research modelers face the challenge of selecting the most suitable combination of models to develop an intent modeling approach for a conversational recommender system. In this section, we present a meta-model for the decision-making process in the context of intent modeling. Adopting this meta-model is based on the principles outlined in the ISO/IEC/IEEE standard 42010 (ISO 2011), which provides a framework for conceptual modeling of Architecture Description. This process requires a systematic approach to ensure that the chosen models effectively capture and understand users' intentions. Let's consider a scenario where research modelers encounter this challenge and go through the decision-making process:

Goal and concerns: The research modelers aim to build an intent modeling approach for a conversational recommender system. Their goal is to accurately determine the underlying purposes or goals behind users' requests, enabling personalized and precise responses. The modelers have concerns regarding quality attributes and functional requirements, and they aim to achieve an acceptable level of quality based on their evaluation measures.

Identification of models and features: To address this problem, the modelers consider various models that can capture users' intentions in the conversational context. They identify essential features, such as *user intent prediction* or *context analysis* based on their concerns. They explore the available models and techniques, such as *Supervised Learning*, *Unsupervised Learning*, *Recurrent Neural Networks*, *Deep Belief Networks*, *Clustering*, and *Self-Supervised Learning Models*. The modelers also consider the recent trends in employing models for intent modeling.

Evaluation of models: The modelers review the descriptions and capabilities of several models that align with capturing users' intentions in conversational interactions. They analyze each model's strengths, limitations, and applicability to the intent modeling problem. They consider factors such as the models' ability to handle natural language input, understand context, and predict user intents accurately. This evaluation allows them to shortlist a set of candidate models that have the potential to address the intent modeling challenge effectively.

In-depth analysis: The research modelers conduct a more detailed analysis of the shortlisted models. They examine the associated techniques for each model to ensure their suitability in the conversational recommender system. They assess factors such as training data requirements, model complexity, interpretability, and scalability. Additionally, they explore the possibility of combining models to identify compatible combinations or evaluate the existing literature on such combinations. If necessary, further study may be conducted to assess the feasibility of model combinations. This step helps them identify the optimal combination of models that best capture users' intentions in the conversational setting and address their concerns.

5.2 A decision model for intent modeling selection

Decision theories have wide-ranging applications in various fields, including e-learning (Garg et al. 2018) and software production (Xu and Brinkkemper 2007;

Fitzgerald and Stol 2014; Rus et al. 2003). In the literature, *decision-making* is commonly defined as a process involving problem identification, data collection, defining alternatives and selecting feasible solutions with ranked preferences (Fitzgerald et al. 2017; Kaufmann et al. 2012; Garg 2020; Garg et al. 2017; Sandhya et al. 2018; Garg 2019). However, decision-makers approach decision problems differently, as they have their priorities, tacit knowledge, and decision-making policies (Doumpos and Grigoroudis 2013). These differences in judgment necessitate addressing them in decision models, which is a primary focus in the field of multiple-criteria decision-making (MCDM).

MCDM problems involve evaluating a set of alternatives and considering decision criteria (Farshidi et al. 2023). The challenge lies in selecting the most suitable alternatives based on decision-makers' preferences and requirements (Majumder 2015). It is important to note that MCDM problems do not have a single optimal solution, and decision-makers' preferences play a vital role in differentiating between solutions (Majumder 2015). In this study, we approach the problem of model selection as an MCDM problem within the context of intent modeling approaches for conversational recommender systems.

Let $Models = m_1, m_2, \ldots, m_{\|Models\|}$ be a set of models found in the literature (decision space), such as *LDA*, *SVM*, and *BERT*. Let $Features = f_1, f_2, \ldots, f_{\|Features\|}$ be a set of features associated with the models, such as ranking, prediction, and recommendation. Each model $m \in Models$ supports a subset of the set $Features$ and satisfies a set of evaluation measures ($Measures = e_1, e_2, \ldots, e_{\|Measures\|}$) and quality attributes ($Qualities = q_1, q_2, \ldots, q_{\|Qualities\|}$). The objective is to identify the most suitable models, or a combination of models, represented by the set $Solutions \subset Models$, that address the concerns of researchers denoted as $Concerns$, where $Concerns \subseteq \{Features \cup Measures \cup Qualities\}$. Accordingly, research modelers can adopt a systematic strategy to select combinations of models by employing an MCDM approach. This approach involves taking $Models$ and their associated $Features$ as input and applying a weighted combination to prioritize the $Features$ based on the preferences of decision-makers. Subsequently, the defined $Concerns$ are considered, and an aggregation method is utilized to rank the $Models$ and propose fitting $Solutions$. Consequently, the MCDM approach can be formally expressed as follows:

$$MCDM : Models \times Features \times Concerns \rightarrow Solutions$$

The decision model developed for intent modeling, using MCDM theory and depicted in Fig. 3, is a valuable tool for researchers working on conversational recommender systems. This approach helps researchers explore options systematically, consider important factors for conversational interactions, and choose the best combination of models to create an effective intent modeling approach. The decision model suggests five steps for selecting a combination of models for conversational recommender systems:

(1) **Models:** In this phase, researchers should gain insights into best practices and well-known models employed by other researchers in designing conversational

recommender systems. Appendix B can be used to understand the definitions of models, while Appendix C can help become familiar with the categories used to classify these models. Table 2 illustrates the categorization of models in this study, and Table 4 presents the trends observed among research modelers in utilizing models to build their conversational recommender systems.

(2) **Feature requirements elicitation:** In this step, researchers need to fully understand the core aspects of the intent modeling problem they are studying. They should carefully analyze their specific scenario to identify the key characteristics required in the models they seek, which may involve using a combination of models. For instance, researchers might consider prediction, ranking, and recommendation as essential feature requirements for their conversational recommender systems. Researchers can refer to Appendix D to gain a better understanding of feature definitions and model characteristics, which will help them select the most suitable features for their intent modeling project.

(3) **Finding feasible solutions:** In this step, researchers should identify models that can feasibly fulfill all of their feature requirements. Table 3 can be used to determine which models support specific features. For example, the table shows that 99 publications explicitly mentioned Collaborative Filtering as a suitable model for applications requiring predictions, and 94 publications indicated CF's applicability for ranking. Moreover, 46 studies employed CF for item recommendation. Based on these findings, if a conversational recommender system requires these three feature requirements, CF could be selected as one of the potential solutions. If the number of feature requirements increases, the selection problem can be converted into a set covering problem (Caprara et al. 2000) to identify the smallest subcollection of models that collectively satisfy all feature requirements.

(4) **Selecting feasible combinations:** In this phase, researchers need to assess whether the identified models can be integrated or combined. Figure 2 provides information on the feasibility of combining models based on the reviewed articles in this study. If the table does not indicate a potential combination, it does not necessarily imply that the combination is impossible. It means no evidence supports its feasibility, and researchers should investigate the combination independently.

(5) **Performance analysis:** After identifying a set of feasible combinations, researchers should address their remaining concerns regarding quality attributes and evaluation measures. Table 5 and Appendix E can be used to understand the typical concerns other researchers in the field employ. Additionally, Table 6 provides insights into frequently used datasets across domains and applications. Researchers can then utilize off-the-shelf models from various libraries, such as TensorFlow[4] and scikit-learn,[5] to build their solutions (pipelines). These solutions can be evaluated using desired datasets to assess whether they meet all the specified concerns. This phase of the decision model differs from the previous four phases, as it requires significant ad-hoc efforts in developing, training, and evaluating the models.

By employing this decision-making process, research modelers can develop an intent modeling approach that accurately captures and understands users' inten-

[4] https://www.tensorflow.org/.

[5] https://scikit-learn.org/.

Table 7 An overview of the feature requirements considered by the case study participants (Manzoor and Jannach 2022; Tanjim et al. 2020) during their decision-making process for developing their conversational recommender systems

	Case study 1	Case study 2
Domain	Conversational recommender systems	Recommender system
Approach	CRB-CRS	ASLI
Dataset	MovieLens, ReDial	Etsy, Alibaba
Year	2022	2020
Research Institute	University of California San Diego	University of Klagenfurt
Country	Austria	US
CORE ranking	A*	A*
Venue	Information systems	WWW
Design decisions by case study participants	TF-IDF, BERT	Self-attention
Decision model results	TF-IDF, BERT	GRU, LDA
#Feature requirements	6	8
Pattern-based		✓
Prediction		✓
Semantic analysis	✓	
Term weighting	✓	
Historical data-driven recommendations		✓
Content-based recommendations	✓	
Click-through recommendations		✓
Item recommendation		✓
Ranking	✓	
Transformer-based	✓	✓
Network architecture		✓
End-to-end approach	✓	
Attentive		✓

The selected feature requirements were instrumental in guiding the participants' model selection. We employed the decision model based on the defined feature requirements to identify feasible combinations of models. The results of the decision model are also presented in this table, showcasing how it aligns with the participants' choices, validating the effectiveness of the decision-making process for developing innovative and effective conversational recommender systems

tions in the conversational recommender system. This enables personalized and precise responses, enhancing the overall user experience and satisfaction.

6 Evaluation of findings: case studies

In this section, we detail the evaluation of our proposed decision model (see Sect. 5) through two scientific case studies. These studies were conducted by a team of eight researchers from the University of California San Diego, USA, and the University of Klagenfurt, Austria. The primary goal was to test the decision model's applicability in the participants' projects and to understand their decision-making processes better.

During the case studies, participants specified their unique feature requirements, which we recorded in Table 3. Essentially, after reviewing the features listed in the Table, the participants defined their requirements. Using these data, we pinpointed suitable models from the extensive information in Table 2 and Table 3. We then examined potential combinations of these models, as depicted in Fig. 2.

To evaluate the significance and recognition of the chosen models in academic circles, we undertook a detailed analysis, referencing Table 4. This examination yielded insights into the models' popularity and relevance over time within the research community. The most notable and trending combinations were then presented to the case study participants. Figure 3 provides a schematic of the typical decision-making process researchers follow when selecting models for intent modeling.

Table 7 offers a thorough summary of the conducted case studies. This table outlines the specific contexts of each study, the feature requirements identified by the participants, the model selections made by the researchers based on these requirements, and the outcomes from applying our decision model in each scenario. The following sections will delve deeper into these case studies, discussing the addressed concerns, the results achieved using the decision model, and the conclusions drawn from our comprehensive analysis.

6.1 Case study method

Case study research is an empirical research method (Jansen 2009) that investigates a phenomenon within a particular context in the domain of interest (Yin 2017). Case studies can be employed to describe, explain, and evaluate a hypothesis. They involve collecting data regarding a specific phenomenon and applying a tool to evaluate its efficiency and effectiveness, often through interviews. In our study, we followed the guidelines outlined by Yin (2009) to conduct and plan the case studies.

Objective: The main aim of this research was to conduct case studies to evaluate the effectiveness of the decision model and its applicability in the academic setting for supporting research modelers in selecting appropriate models for their intent modeling approaches.

The cases: We conducted two case studies within the academic domain to assess the practicality and usefulness of the proposed decision model. The case studies aimed to evaluate the decision model's effectiveness in assisting research modelers and researchers in selecting models for their intent modeling tasks.

Methods: For the case studies, we engaged with research modelers and researchers actively involved in intent modeling approaches. We collected data through expert interviews and discussions to gain a comprehensive understanding of their specific

requirements, preferences, and challenges when selecting models. The case study participants provided valuable insights into the decision-making process and offered feedback on the suitability of the decision model for their intent modeling needs.

Selection strategy: In line with our research objective, we employed a multiple case study approach (Yin 2009) to capture a diverse range of perspectives and scenarios within the academic domain. This selection strategy aimed to ensure the credibility and reliability of our findings. We deliberately selected two publications from highly regarded communities with an A* CORE rank. We verified the expertise of the authors, who actively engage in selecting and implementing intent modeling models. Their knowledge and experience allowed us to consider various factors in different application contexts, including quality attributes, evaluation measures, and feature requirements.

By conducting these case studies, our research aimed to validate the practicality of the decision model and demonstrate its value in supporting research modelers and researchers in their intent modeling endeavors. The insights gained from the case studies provided valuable feedback for refining the decision model and contributed to advancing the intent modeling field within the academic community.

6.2 Case study 1:

The first case study presented in our paper revolves around a research project conducted at the University of Klagenfurt in Austria. The study focused on investigating a retrieval-based approach for conversational recommender systems (CRS) (Manzoor and Jannach 2022). The primary objective of the researchers was to assess the effectiveness of this approach as an alternative or complement to language generation methods in CRS. They conducted user studies and carefully analyzed the results to understand the potential benefits of retrieval-based approaches in enhancing user intent modeling for conversational recommender systems.

Throughout the project, the case study participants made two important design decisions (models), TF-IDF and BERT, to develop the CRS. They evaluated their approach using MovieLens and ReDial datasets to measure its performance.

By applying the decision model presented in our paper (in Sect. 5.2), the case study participants identified six essential features that were crucial in guiding their decision-making process for selecting the most suitable models and datasets. These features provided valuable insights into designing and implementing an effective retrieval-based approach for conversational recommender systems, contributing to improving user intent modeling in this context.

6.2.1 Feature requirements

In this section, we outline the feature requirements that the case study participants considered during their decision-making process for the research project. Each feature requirement was carefully chosen based on its relevance and potential to enhance the retrieval-based approach for CRS. Below are the feature requirements and their rationale for selection:

Semantic analysis: The case study participants recognized the importance of analyzing the meaning and context of words and phrases in natural language data. Semantic analysis helps the model understand user intents more accurately, leading to more relevant and contextually appropriate recommendations.
Term weighting: Assigning numerical weights to terms or words in a document or dataset helps the machine learning model comprehend the significance of different terms in the data. The participants adopted term weighting to improve the model's ability to identify relevant features and make better recommendations.
Content-based recommendations: This feature involves utilizing item characteristics or features to recommend similar items to users. The participants valued this approach, allowing the system to tailor recommendations based on users' past interactions and preferences.
Ranking: The case study participants sought a model capable of ranking items or entities based on their relevance to specific queries or users. By incorporating ranking, the system ensures that the most relevant recommendations appear at the top, enhancing user satisfaction.
Transformer-based: Transformer-based models, such as neural networks, excel at learning contextual relationships in sequential data like natural language. The participants chose this approach to effectively leverage the model's ability to understand and process conversational context.
End-to-end approach: The case study participants preferred an end-to-end modeling strategy, where a single model directly learns complex tasks from raw data inputs to desired outputs. By avoiding intermediate stages and hand-crafted features, the participants aimed to simplify the model and improve its performance in CRS tasks.

6.2.2 Results and analysis

During the expert interview session with the case study participants, we systematically followed the decision model presented in Sect. 5.2 to identify appropriate combinations of models that align with the defined feature requirements for their conversational recommender systems. In the initial steps (Steps 1 and 2), we collaboratively established the essential feature requirements for their CRS, carefully considering the critical aspects that would enhance their system's performance. Subsequently, we referred to Table 3 (Steps 3 and 4) to evaluate which models could fulfill these specific feature requirements.

Upon analyzing the table, both the case study participants and we discovered that BERT offered support for Semantic Analysis, Content-Based Recommendations, Ranking, Transformer-Based, and End-To-End Approaches. Additionally, TF-IDF was found to be supportive of Term Weighting, Content-Based Recommendations, and Ranking. This insightful information made us realize that combining these two models would adequately address all six feature requirements for their CRS. Consequently, the case study participants confirmed that combining BERT and TF-IDF would be a suitable choice to fulfill their CRS needs. This combination was validated as a compatible and valid option, consistent with the guidance provided by the decision model.

The data presented in Table 4 further reinforce the popularity and relevance of BERT and TF-IDF as widely used models for conversational recommender systems. The case study participants were well aware of these trends and acknowledged that their model choices aligned with prevailing practices. This alignment provides additional validation to their model selections, demonstrating their dedication to adopting the latest technologies in their research project to create an effective CRS.

Furthermore, Table 6 provides valuable insights into the popularity and significance of various datasets, including MovieLens and ReDial. These datasets have been cited and utilized in over 50 publications, underscoring their recognition within the research community. The case study participants acknowledged the widespread use of these datasets by other researchers, reflecting an interesting trend in dataset selection. This awareness further highlights their commitment to utilizing well-established and reputable datasets in their research, contributing to the credibility and reliability of their study findings.

6.3 Case study 2

The second case study presented in our paper focuses on a research project conducted at the University of California San Diego in the United States (Tanjim et al. 2020). The study introduces the Attentive Sequential model of Latent Intent (ASLI) to enhance recommender systems by capturing users' hidden intents from their interactions.

Understanding user intent is essential for delivering relevant recommendations in conventional recommender systems. However, user intents are often latent, meaning they are not directly observable from their interactions. ASLI addresses this challenge by uncovering and leveraging these latent user intents.

Using a self-attention layer, the researchers (case study participants) designed a model that initially learns item similarities based on users' interaction histories. They incorporated a Temporal Convolutional Network (TCN) layer to derive latent representations of user intent from their actions within specific categories. ASLI employs an attentive model guided by the latent intent representation to predict the next item for users. This enables ASLI to capture the dynamic behavior and preferences of users, resulting in state-of-the-art performance on two major e-commerce datasets from Etsy and Alibaba.

By utilizing the decision model presented in our paper (in Sect. 5.2), the case study participants identified eight essential features crucial in guiding their decision-making process for selecting the most suitable models and datasets.

6.3.1 Feature requirements

In this section, we present the feature requirements that were crucial considerations for the case study participants during their decision-making process for the research project. The following are the feature requirements and the reasons behind their selection:

Pattern-based: In the case study, the researchers aimed to improve conversational recommender systems by capturing users' hidden intents from their interactions. By

identifying user interactions and behavior patterns, the ASLI model can make informed guesses about users' intents and preferences, leading to more accurate and relevant recommendations.

Prediction: The ASLI model predicts the next item for users based on their latent intents derived from their historical interactions within specific categories. The model can deliver personalized and effective recommendations by predicting users' preferences and future actions.

Historical data-driven recommendations: The researchers used previously collected data from users' interactions to train the ASLI model. By analyzing historical data, the model can identify patterns, relationships, and trends in users' behaviors, which inform its predictions and recommendations for future interactions.

Click-through recommendations: In the case study, the ASLI model considers users' clicks on items to understand their preferences and improve the relevance and ranking of future recommendations. The model can adapt and refine its recommendations by utilizing click-through data to meet users' needs better.

Item recommendation: The ASLI model suggests items to users based on their previous interactions, enabling it to offer personalized recommendations tailored to individual users' preferences and behaviors.

Transformer-based: ASLI is a neural network model based on the Transformer architecture. Transformers are well-suited for learning context and meaning from sequential data, making them suitable for capturing the dynamic behavior and preferences of users in conversational recommender systems.

Network architecture: The ASLI model's network architecture is crucial in guiding information flow through the model's layers. By designing an effective network architecture, the researchers ensure that the model can capture and leverage users' latent intents to make accurate recommendations.

Attentive: ASLI utilizes attention mechanisms to focus on the most relevant parts of users' interactions and behaviors. The model can better understand users' intents and preferences by paying attention to critical information, leading to more attentive and accurate recommendations.

6.3.2 Results and analysis

During the expert interview session with the case study participants, we used the decision model (outlined in Sect. 5.2) to identify suitable combinations of models that align with the defined feature requirements for their conversational recommender systems. In Steps 1 and 2, we collaboratively established the essential feature requirements for the ASLI, carefully considering critical aspects to enhance system performance. Then, in Steps 3 and 4, we referred to Table 3 to evaluate models that could fulfill these specific feature requirements.

According to the table, both the case study participants and ourselves found that the GRU model supports Prediction, Historical Data-Driven Recommendations, Click-Through Recommendations, Network Architecture, and Attentive features. Additionally, the LDA model supports Pattern-Based and Item Recommendation features. We also discovered that BERT is the only model in our list supporting Transformer-Based features, and the case study participants agreed with this com-

bination, considering these models as the baseline of their approach. However, after performance analysis, they found that GRU's performance was unsatisfactory in their setting. Consequently, they chose to develop their own model from scratch, modifying the self-attentive model. It's worth noting that the Self-attentive model only supports Network Architecture and Attentive features, making it a suitable baseline in combination with other models for their solutions. The case study participants mentioned considering LDA and BERT as potential models for their upcoming research project due to their similar requirements, although they were not previously aware of this combination. As per Step 5 of the decision model, researchers should address any remaining concerns about quality attributes and evaluation measures after identifying feasible combinations. Thus, the decision model provided valid models in this case study, but in real-world scenarios, model combinations may be modified based on other researchers' concerns, such as quality attributes and evaluation measures.

The case study participants emphasized the value of the data presented in Table 4 and their intention to incorporate it into their future design decisions. Understanding trends in model usage is crucial to identify models that may perform well in conversational recommender systems, considering similar concerns and requirements from other researchers.

Furthermore, Table 6 indicates that Etsy and Alibaba datasets are not widely known in the context of user intent modeling, although the case study participants clarified that these datasets are well-known in e-commerce services, aligning with their project's specific domain of focus. Nonetheless, they expressed their intention to utilize the data presented in this table to explore potential datasets for evaluating their approach and comparing their work against other approaches in the literature.

7 Discussion

7.1 SLR outcomes

Code sharing: Our review of 791 publications revealed that only 68 (8.59%) explicitly shared their code repositories, such as GitHub. This observation underscores a significant gap in code sharing among researchers, posing challenges to replicating experiments and advancing scientific knowledge. Open access to code is imperative for ensuring transparency and reproducibility in machine learning research (Haefliger et al. 2008).

Singleton models: The systematic literature review yielded 600 models, with 352 (58.66%) being singletons. This trend indicates a preference for developing unique models tailored to specific research questions. However, an overreliance on singletons might hinder the generalizability of findings and the ability to compare different methods. Promoting the use of common models or establishing standard evaluation benchmarks is essential to enhance reproducibility and comparability in machine learning research (Amershi et al. 2019).

Model combination: The methodology for combining models in some publications was not clearly articulated, making it difficult to understand the techniques employed and their efficacy. Clear documentation of model combination techniques and their

underlying rationale is crucial for ensuring transparency and facilitating the replication and extension of research findings (Kuwajima et al. 2020). The challenge lies in determining the effectiveness of integrating different models without extensive contextual information. The current approach, based on literature and general requirements, provides a foundational framework but may not capture the specific nuances needed for particular applications. Future research should involve detailed analyses of model combinations in specific scenarios, using case studies or empirical evaluations to provide insights into the interactions and complementarity of different models, thereby enhancing the practical applicability of intent modeling methods in conversational recommender systems.

Model variations: Our analysis identified a diverse range of model variations, such as BERT4Rec (Chen et al. 2022), SBERT (Garcia and Berton 2021), BERT-NeuQS (Hashemi et al. 2020), BioBERT (Carvallo et al. 2020), ELBERT (Gao and Lam 2022), and RoBERTa (Wu et al. 2021), primarily derived from BERT (Devlin et al. 2018). Despite the utility of these variations in addressing different tasks, their extensive use complicates model comparison and experiment replication. Establishing standardized categories for model variations would aid researchers in discerning model differences and similarities, thereby promoting model sharing, reuse, and collaborative progress in machine learning research (Sarker 2021).

Trends: As depicted in Fig. 2, LDA is a predominant model in user intent modeling approaches (Table 4). Although traditional models like LDA have significantly contributed to the field and inspired the development of advanced models such as BERT (Devlin et al. 2018), the adoption of traditional models has possibly declined due to the emergence of sophisticated models like BERT. The bidirectional contextual embeddings and transformer architecture of BERT have demonstrated remarkable performance across various NLP tasks, attracting considerable attention from the research community. The preference for modern models is also influenced by the trade-off between the interpretability of traditional models and the complexity of advanced models like BERT, as well as the diversity of NLP applications (Ribeiro et al. 2016).

Datasets: Only 394 out of 791 publications (49.81%) utilized public, open-access datasets, indicating a reliance on proprietary datasets by more than half of the publications. This limitation hinders data reuse and poses challenges to research reproducibility and credibility. While 253 public open-access datasets were identified, 173 (68.37%) were mentioned in only one publication and not reused, highlighting deficiencies in dataset-sharing practices. The limited availability of datasets impedes the reproduction and validation of results, comparison, and benchmarking of models, and identification of state-of-the-art techniques. Moreover, the lack of diverse and openly accessible datasets may result in biased model development and evaluation, limiting the applicability of models to real-world scenarios and diverse user populations. Addressing these issues necessitates fostering a culture of openness and collaboration within the research community.

7.2 Case study participants

The case study participants showed a careful and thorough approach to decision-making by conducting extensive research and literature reviews. This method allowed them to select models for their research project carefully, showcasing the effectiveness of the decision model in helping researchers make well-informed and compatible model choices for developing conversational recommender systems.

Both case study participants emphasized the value of using the decision model and the knowledge gained during this study. They expressed their intention to use this information to make informed decisions when selecting the appropriate combinations of models for user intent modeling approaches.

Furthermore, the case study participants recognized that the decision model serves as a valuable tool for generating an initial list of models to develop their approaches. However, they acknowledged that Step 5 of the decision model highlights the importance of further analysis, such as performance testing, to identify the right combinations of models that work well for specific use cases. This recognition underscores the need for practical testing and validation to ensure the chosen model combinations are effective and suitable for their particular research goals.

The use of well-known datasets, such as MovieLens and ReDial in the first case study and Etsy and Alibaba datasets in the second case study, underlines the researchers' commitment to using credible data sources for evaluation. The decision model allowed researchers to consider dataset popularity and relevance, enhancing the credibility and reliability of their study findings.

The decision model provided valuable insights into the trends in model usage, as presented in Table 4. Both case study participants expressed interest in incorporating these trends into their future research decisions, ensuring they stay up-to-date with the latest advancements in intent modeling approaches.

Throughout the case studies, the discussion highlighted the dynamic nature of the decision-making process. While the decision model offered feasible model combinations based on feature requirements, the final choices were influenced by additional factors such as model performance, quality attributes, and evaluation measures. This adaptability showcased the decision model's flexibility in accommodating researchers' unique priorities and preferences.

Both case studies effectively demonstrated that the decision model offers a systematic approach to model selection and helps researchers explore various options and combinations of models. This exploratory nature allowed researchers to consider novel solutions and build upon existing models, creating innovative intent modeling approaches.

The success of the decision model in assisting researchers in their model selection process holds promising implications for the broader academic community. By providing a structured and comprehensive methodology, the decision model can streamline the development of conversational recommender systems with accurate intent modeling capabilities, ultimately enhancing user experience and satisfaction.

7.3 Threat to validity

Validity evaluation is essential in empirical studies, encompassing SLRs and case study research (Zhou et al. 2016). This paper's validity assessment covers various dimensions, including Construct Validity, Internal Validity, External Validity, and Conclusion Validity. Although other types of validity, such as Theoretical Validity and Interpretive Validity, are relevant to intent modeling, they are not explicitly addressed in this context due to their relatively limited exploration.

Construct validity pertains to the accuracy of operational measures or tests used to investigate concepts. In this research, we developed a meta-model (refer to Fig. 3) based on the ISO/IEC/IEEE standard 42010 (ISO 2011) to represent the decision-making process in intent modeling for conversational recommender systems. We formulated comprehensive research questions by utilizing the meta-model's essential elements, ensuring an exhaustive coverage of pertinent publications on intent modeling approaches.

Internal validity concerns verifying cause-effect relationships within the study's scope and ensures the study's robustness. We employed a rigorous quasi-gold standard (QGS) (Zhang et al. 2011) to minimize selection bias in paper inclusion. Combining manual and automated search strategies, the QGS provided an accurate evaluation of sensitivity and precision. Our search spanned four major online digital libraries, widely regarded to encompass a substantial portion of high-quality publications relevant to intent modeling for conversational recommender systems. Additionally, we used snowballing to complement our search and mitigate the risk of missing essential publications. The review process involved a team of researchers, including three principal investigators and five research assistants. Furthermore, the findings were validated by real-world researchers in intent modeling to ensure their practicality and effectiveness.

External validity pertains to the generalizability of research findings to real-world applications. This study considered publications discussing intent modeling approaches across multiple years. Although some exclusions and inaccessibility of studies may impact the generalizability of SLR and case study results, the proportion of inaccessible studies (less than 2%) is not expected to affect the overall findings significantly. The knowledge extracted from this research can be applied to support the development of new theories and methods for future intent modeling challenges, benefiting both academia and practitioners in this field.

Conclusion validity ensures that the study's methods, including data collection and analysis, can be replicated to yield consistent results. We extracted knowledge from selected publications, encompassing various aspects such as *Models*, *Datasets*, *Evaluation Metrics*, *Quality Attributes*, *Combinations*, and *Trends* in intent modeling approaches. The accuracy of the extracted knowledge was safeguarded through a well-defined protocol governing the knowledge extraction strategy and format. The authors proposed and reviewed the review protocol, establishing a clear and consistent approach to knowledge extraction. A data extraction form was employed to ensure uniform extraction of relevant knowledge, and the acquired knowledge was validated against the research questions. All authors independently determined quality assess-

ment criteria, and crosschecking was conducted among reviewers, with at least three researchers independently extracting data, thus enhancing the reliability of the results.

8 Related work

The development of conversational recommender systems is significantly influenced by the findings from SLRs in various related research domains, each contributing to the collective understanding of user intent modeling. These SLRs are pivotal in gathering and analyzing data to interpret user needs within conversational interfaces.

In the field of Human–Computer Interaction, key SLRs conducted by de Barcelos Silva et al. (2020), Rapp et al. (2021), Iovine et al. (2023), Jiang et al. (2013), and Jindal et al. (2014) have systematically collected and analyzed data to understand how user-friendly interfaces can enhance user engagement and satisfaction in conversational systems, a core aspect of user intent modeling.

Similarly, in Conversational AI, the SLRs by Zaib et al. (2022) and Saka et al. (2023) have aggregated research findings focusing on simulating natural, human-like interactions, a key component in understanding and modeling user intent in conversational recommender systems.

The research in Conversational Search Systems, notably synthesized by Keyvan and Huang (2022), and Yuan et al. (2020), represents comprehensive reviews of the dynamics of user-system interaction for information retrieval. These studies align with user intent modeling by providing insights into how conversational systems can better parse and understand user queries.

For User Preference Extraction & Prioritization, SLRs by Pu et al. (2012), Liu et al. (2022), Zhang et al. (2019), and Hernández-Rubio et al. (2019) have methodically reviewed the literature to inform how conversational recommender systems can more accurately and contextually tailor their recommendations.

In the realm of Contextual Information Retrieval Systems, the systematic reviews by Tamine-Lechani et al. (2010), Chen et al. (2015), and Latifi et al. (2021) have contributed to understanding the impact of explicit and implicit user queries and contextual factors, crucial for refining user intent modeling in conversational systems.

Our SLR encapsulates these efforts, covering a total of 791 publications. We highlight the collective contribution of these SLRs to the field of user intent modeling in conversational recommender systems. Table 8 summarizes these efforts, offering a comparative analysis and showcasing the contributions of our study. Notably, our review reveals that while there is a substantial amount of literature on individual aspects of user intent modeling, a comprehensive, integrated approach in the form of an SLR is less common. The synthesis of findings from HCI, Conversational AI, Conversational Search Systems, User Preference Analytics, and Contextual Information Retrieval forms the foundation for advancing user intent modeling in conversational recommender systems (Dodeja et al. 2024; Zhang et al. 2024).

In Table 8: Column 1 shows the authors of the studies, Column 2 indicates the year of publication, and Column 3 indicates the type of publications, which could be either academic or gray literature. Column 4 highlights the research methods that the publications employed. Column 5 signifies the main focus of the topic of the

Table 8 Positions our study within the existing body of literature on conversational recommender systems, highlighting user intention understanding from various perspectives

Study	Year	Review type	Research method	Main focus	Application / Domain	# Reviewed publications	Decision model	Trend	Datasets	Model categories	Model combinations	Feature/Model Mapping	# Quality attributes	#Features	# Evaluation measures	#Models	# Common quality attributes	# Common features	# Common evaluation measures	# Common models	Coverage (%)
This study	2023	Academic literature	SLR Case study	User intent modeling	Conversational recommender	791	Yes	Yes	Yes	Yes	Yes	Yes	38	74	13	59	N/A	N/A	N/A	N/A	100
Silva et al.	2020	Academic literature	SLR	Human-computer interaction	Virtual Personal Assistants	58	No	No	No	Yes	No	No	7	4	8	6	6	3	5	2	64
Rapp et al.	2021	Academic literature	SLR	Human-computer interaction	Virtual Personal Assistants	83	No	No	No	Yes	No	No	8	4	5	9	6	5	3	2	61.54
Keyvan et al.	2022	Academic literature	Survey	Conversational search systems	Virtual Personal Assistants	N/A	No	No	Yes	No	No	No	8	12	14	15	5	9	10	7	63.27
Zaib et al.	2022	Academic literature	Survey	Conversational AI	Conversational question answering	88	No	Yes	Yes	No	No	No	5	8	9	18	4	7	5	5	52.50
Iovine et al.	2023	Academic literature	Survey	Human-computer interaction	Virtual Personal Assistants	116	No	No	No	No	No	No	6	7	5	11	6	7	2	8	79.31
Saka et al.	2023	Academic literature	Survey	Conversational AI	Focus Group Discussion	21	No	No	No	No	No	No	4	7	9	11	4	4	9	5	70.97
Pu et al.	2012	Academic literature	Survey	User preference extraction &	Item recommendation	N/A	Yes	Yes	No	No	No	No	8	7	5	11	8	7	5	11	100
Liu et al.	2022	Academic literature	Survey	User preference extraction &	e-learning environments	N/A	No	No	No	Yes	No	No	2	7	3	8	2	7	1	4	70
Tamine-Lechani et al.	2010	Academic literature	Survey	Contextual information retrieval	Context modeling approaches	N/A	No	No	No	No	No	No	5	4	10	8	3	4	4	4	55.56
Chen at al.	2015	Academic literature	Survey	Contextual information retrieval	User reviews	N/A	No	Yes	No	No	No	No	7	8	10	19	5	5	9	9	63.64
Latifi et al.	2021	Academic literature	Review	Contextual information retrieval	Session-aware recommendation	N/A	No	No	Yes	No	No	No	3	5	10	14	2	3	7	8	62.50
Zhang et al.	2019	Academic literature	Survey	User preference extraction &	Item recommendation	N/A	No	Yes	No	Yes	No	No	6	8	6	15	6	7	3	8	68.57
Jiang et al.	2013	Gray literature	Survey	Human-computer interaction	Query Understanding	N/A	No	No	No	No	No	No	6	7	7	13	5	5	5	10	75.76
Hernández-Rubio et al.	2019	Gray literature	Review	User preference extraction &	User reviews	N/A	No	No	Yes	Yes	Yes	No	4	11	6	12	4	11	3	9	81.82
Jindal et al.	2014	Academic literature	Review	Human-computer interaction	Semantic search	15	No	No	Yes	Yes	No	No	5	4	2	3	5	4	1	2	85.71
Yuan et al.	2020	Gray literature	Review	Conversational search systems	Conversational question answering	N/A	No	No	Yes	Yes	Yes	No	3	8	1	15	2	5	1	8	59.26

publications, and Column 6 indicates the Application or Domain that the publication conducted research on. Column 7 (# Reviewed publications) indicates the number of publications that each study reviewed in its research. Column 8 (Decision model) indicates whether a selected publication offered a decision model based on its findings from the data captured in the literature. Column 9 (Trend) shows if the selected study reported on the trends in employing models that it found. Column 10 (Datasets) shows if the researchers reported on the training or evaluation datasets. Column 11 (Model categories) indicates whether the publications reported on categories of the models (if they categorized the models). Column 12 (Model combinations) indicates if they reported on model combinations and integration. Column 13 (Feature/Model Mapping) shows if they offered the features that the models support. The subsequent four columns (Columns 14, 15, 16, and 17) show the number of quality attributes, features, evaluation measures, and models that each study reported. Columns 18, 19, 20, and 21 indicate how many quality attributes, features, evaluation measures, and models that the selected publications reported are in common with the ones in our study. Finally, Column 22 (Coverage (%)) shows the percentage of common concepts between our study and each selected publication.

Academic literature reviews dominate the selected studies, representing over 80 percent of the reviewed literature, aligning with our primary focus on academic sources. The research methods in these studies include SLR, Case Study, Survey, and Review.

However, none of the reviewed SLRs employed case studies to evaluate their findings, relying solely on the SLR process. Our study adopts a more comprehensive approach by incorporating case studies into our research methods, offering a holistic perspective on decision-making in user intent modeling.

Our study places a significant emphasis on decision-making processes and decision models. Among the reviewed SLRs, only one paper (Pu et al. 2012) focused on this aspect, while our study introduces a decision model based on existing literature. This model serves as a valuable tool for research modelers to make informed decisions and identify suitable models or combinations for specific scenarios.

In terms of trends within models, four studies (Zaib et al. 2022; Pu et al. 2012; Chen et al. 2015; Zhang et al. 2019) (23.52%) reported on this aspect. Additionally, seven studies (Latifi et al. 2021; Yuan et al. 2020; Jindal et al. 2014; Hernández-Rubio et al. 2019; Zaib et al. 2022; Keyvan and Huang 2022; Pan et al. 2022) (41.17%) provided insights into open-access datasets, valuable for training or evaluating models.

Furthermore, our study categorizes models similar to eight other SLRs (de Barcelos Silva et al. 2020; Rapp et al. 2021; Pan et al. 2022; Liu et al. 2022; Zhang et al. 2019; Hernández-Rubio et al. 2019; Jindal et al. 2014; Yuan et al. 2020) (47.05%). However, only two publications (Hernández-Rubio et al. 2019; Yuan et al. 2020) (11.76%) reported on model combinations, suggesting a research gap in effective model integration.

9 Conclusion and future work

In this paper, the investigation focused on the decision-making process involved in selecting intent modeling approaches for conversational recommender systems. The primary aim was to tackle the challenge encountered by research modelers in determining the most effective model combination for developing intent modeling approaches.

To ensure the credibility and reliability of our findings, we conducted a systematic literature review and carried out two academic case studies, meticulously examining various dimensions of validity, including Construct Validity, Internal Validity, External Validity, and Conclusion Validity.

Drawing inspiration from the ISO/IEC/IEEE standard 42010 (ISO 2011), we devised a meta-model as the foundational framework for representing the decision-making process in intent modeling. By formulating comprehensive research questions, we ensured the inclusion of relevant studies and achieved an exhaustive coverage of pertinent publications.

Our study offers a holistic understanding of user intent modeling within the context of conversational recommender systems. The SLR analyzed over 13,000 papers from the last decade, identifying 59 distinct models and 74 commonly used features. These analyses provide valuable insights into the design and implementation of user intent modeling approaches, contributing significantly to the advancement of the field.

Building on the findings from the SLR, we proposed a decision model to guide researchers and practitioners in selecting the most suitable models for developing conversational recommender systems. The decision model considers essential factors such as model characteristics, evaluation measures, and dataset requirements, facil-

itating informed decision-making and enhancing the development of more effective and efficient intent modeling approaches.

We demonstrated the practical applicability of the decision model through two case studies, showcasing its usefulness in real-world scenarios. The decision model aids researchers in identifying initial model sets and considering essential quality attributes and functional requirements, streamlining the process and enhancing its reliability.

The significance of contributions in user intent modeling cannot be overstated in the current landscape of scientific research. Whether actively advancing the fundamentals or exploring its applications within their respective domains, scientists are undeniably conscious of this field. Amidst this crucial juncture, our study is essential as it consolidates the field's foundations. We envision our research to become an integral component of essential literature for newcomers, fostering the promotion of this vital field and streamlining researchers' efforts in selecting suitable models and techniques. By solidifying the understanding and relevance of User Intent Modeling, we aim to facilitate future advancements and innovation in this study area.

To ensure the longevity and up-to-dateness of the knowledge base constructed from our SLR, we are enthusiastic about taking the necessary steps to maintain its relevance and value for future researchers embarking on similar projects. We plan to establish a collaborative platform or repository, inviting researchers to contribute their latest findings and studies pertaining to the addressed research challenges. By fostering a community-driven approach, we aim to create an engaging environment that encourages regular and meaningful contributions. To streamline the process, we intend to develop user-friendly interfaces and implement effective content moderation to ensure the knowledge base's scientific integrity.

Additionally, we aim to extend the current methodology by introducing more detailed criteria and context-specific frameworks for the selection and integration of intent modeling methods in conversational recommender systems. This involves developing nuanced frameworks that assess model compatibility and integration potential, tailored to address the unique challenges and requirements of specific domains and conversational scenarios. By deepening the analysis of how different models interact and complement each other in varying contexts, future research will not only refine the decision-making process for method selection but also enhance the overall effectiveness and user-centricity of conversational recommender systems.

Moreover, we are excited to explore implementing an automated data crawling mechanism, periodically and systematically searching reputable literature sources and academic databases. This technology will enable seamless integration of the latest research into the knowledge base. Additionally, we are committed to maintaining a record of changes and updates to the knowledge base, including precise timestamps and new information sources. This transparent documentation will empower future researchers to follow the knowledge base's evolution and confidently leverage it for their specific research needs. By embracing these proactive measures, we envision establishing a continuously updated and robust knowledge base that serves as a valuable resource for researchers in the dynamic domain of user intent modeling and recommender systems.

A Review protocol

See Table 9.

Table 9 Mapping of review protocols: alignment with Okoli et al., Xiao et al., and Kitchenham's Review Steps

	Okoli and Schabram (2015)	Xiao and Watson (2019)	Kitchenham et al. (2009)
Problem formulation	Purpose of the literature review	Formulate the Problem	The need for a systematic review
Initial hypotheses		Keywords used for the search	Manual Search and Selection
Initial data collection		Sampling strategy	Generating a search strategy
Query string definition		Refining results with additional restrictions	
Digital library exploration	Searching for the literature	Search the Literature	
Relevancy Evaluation		Criteria for inclusion/exclusion	Study selection criteria
The pool of publications			Documenting the Search
Publication pruning process	Practical screen	Screening procedure	
Quality assessment process	Quality appraisal	Quality assessment procedure	Study Quality Assessment
Data extraction and synthesizing	Data extraction / Synthesis of studies	Extracting Data / Analyzing and Synthesizing Data	Data Extraction / Data Synthesis
Knowledge base	Writing the review	Report Findings	Reporting the review
Snowballing process			References and Appendices

B Models

See Table 10.

Table 10 Model definitions

Name	Definition
ALS *(Alternating least squares)*	A type of matrix factorization algorithm that alternates between solving for user and item factors to minimize the squared error between observed and predicted ratings
BERT *(Bidirectional Encoder Representations from Transformers)*	A deep learning model used for natural language processing tasks, such as text classification and question answering
BiLSTM *(Bidirectional Long Short-Term Memory)*	A type of recurrent neural network used for modeling sequential data, where information flows in both forward and backward directions
BM25 *(Best Match 25)*	A ranking function used in information retrieval to rank documents based on their relevance to a query
BPR *(Bayesian Personalized Ranking)*	A ranking algorithm used in recommender systems that is based on Bayesian inference and models the preferences of individual users
CDL *(Collaborative Deep Learning)*	A technique that combines deep learning models with collaborative filtering algorithms for recommendation systems
CF *(Collaborative Filtering)*	A technique used in recommender systems to make predictions by leveraging the similarity between users or items
Cosine similarity	A measure of similarity between two vectors that calculates the cosine of the angle between them
CRF *(Conditional Random Fields)*	A probabilistic model used for structured prediction tasks, such as named entity recognition and part-of-speech tagging
CTR *(Collaborative Topic Regression)*	A probabilistic model that combines topic modeling and collaborative filtering
DeepCoNN *(Deep Cooperative Neural Networks)*	A type of neural network architecture that jointly learns the user and item embeddings to model user behavior for personalized recommendations
DNN *(Deep Neural Network)*	A type of artificial neural network with multiple hidden layers used for learning representations of complex data such as images, audio, and natural language
EM *(Expectation-Maximization)*	An iterative algorithm used to find maximum likelihood or maximum a posteriori estimates of parameters in statistical models
FM (Factorization Machine)	A type of machine learning model used for predicting interactions between pairs of variables, often used in recommendation systems

Table 10 continued

Name	Definition
GBDT *(Gradient Boosting Decision Trees)*	A machine learning algorithm that combines multiple decision trees in an ensemble, used for supervised learning tasks such as classification and regression
Gibbs sampling	A technique used for generating samples from complex probability distributions, often used in Bayesian inference
GRU *(Gated Recurrent Units)*	A type of recurrent neural network used for modeling sequential data, particularly for applications such as natural language processing and speech recognition
GRU4Rec *(Gated Recurrent Unit for Recommender Systems)*	A type of neural network architecture that uses gated recurrent units to model user behavior for personalized recommendations
HMM *(Hidden Markov Model)*	A probabilistic model used to model sequential data, where the underlying states of the system are hidden but can be inferred from observed outputs
k-means *(k-means clustering)*	An unsupervised machine learning algorithm used to group similar data points into k clusters
KL (Kullback–Leibler Divergence)	A measure of how different two probability distributions are from each other
KNN *(K-Nearest Neighbors Algorithm)*	A supervised machine learning algorithm used for classification and regression by finding the k closest data points to the new data point and making predictions based on their labels
LambdaMART	A ranking algorithm used in information retrieval based on a gradient-boosting framework that combines multiple decision trees
LDA *(Latent Dirichlet allocation)*	A generative statistical model used for topic modeling
LR (Logistic Regression)	a supervised machine learning algorithm used for binary classification
LSA *(Latent Semantic Analysis)*	A technique used to uncover the underlying topics in a corpus by decomposing a term-document matrix using singular value decomposition
LSI *(Latent Semantic Indexing)*	A technique used for dimensionality reduction in text data based on matrix factorization methods such as SVD
LSTM *(Long Short-Term Memory)*	A type of recurrent neural network that can handle long-term dependencies in sequence data
Markov Chain	A mathematical model used to describe a sequence of events where the probability of each event depends only on the state attained in the previous event
MF *(Matrix factorization)*	A technique used in recommender systems to decompose a user-item matrix into two lower-rank matrices, which can then be used to make recommendations

Table 10 continued

Name	Definition
MLE *(Maximum Likelihood Estimation)*	A statistical method that estimates the parameters of a probability distribution by maximizing the likelihood of the observed data
MLP *(Multi-Layer Perceptron)*	A type of neural network consisting of multiple layers of perceptrons used for supervised learning tasks such as classification and regression
MMR *(Maximal Marginal Relevance)*	A ranking algorithm used in information retrieval that maximizes the relevance of a set of documents while minimizing redundancy
Naive Bayes	A probabilistic algorithm used for classification based on Bayes' theorem and the assumption of independence between features
NCF *(Neural Collaborative Filtering)*	A type of collaborative filtering algorithm that uses neural networks to learn the user and item representations to predict the user-item interaction
NER *(Named Entity Recognition)*	A task in natural language processing that involves identifying and classifying named entities in text, such as people, organizations, and locations
NMF *(Non-negative Matrix Factorization)*	A matrix factorization technique used for dimensionality reduction and feature extraction
PageRank	A ranking algorithm used in web search engines that assigns scores to web pages based on the structure of the web graph
PCA (Principal Component Analysis)	A dimensionality reduction technique that finds the principal components of a dataset by decomposing the covariance matrix of the data
PLSA *(Probabilistic Latent Semantic Analysis)*	A statistical technique used to uncover the latent topics within a set of documents
PMF *(Probabilistic Matrix Factorization)*	A probabilistic model used in recommender systems to learn low-dimensional representations of users and items
PMI *(Pointwise Mutual Information)*	A measure of the association between two terms in a corpus, often used in natural language processing tasks such as text classification and information retrieval
Random Forest	A type of ensemble learning method that combines multiple decision trees to make predictions
RankNet	A type of neural network architecture that learns to rank items by minimizing the pairwise ranking loss between items
RBM *(Restricted Boltzmann Machines)*	A type of unsupervised learning algorithm that models probability distributions over inputs by minimizing the free energy of the system
RGM *(Random Group Model)*	A type of matrix factorization algorithm that models the interactions between users and items using random group effects

Table 10 continued

Name	Definition
Rocchio	A content-based filtering algorithm calculates the similarity between a target item and the user's previously rated items using a vector space model. It updates the user's profile based on the similarity scores
RW *(Random Walk)*	A graph traversal algorithm is used in network analysis and ranking algorithms such as PageRank
SDAE *(Stacked Denoising Autoencoder)*	A type of autoencoder neural network that learns to remove noise from input data by encoding and decoding it through multiple hidden layers
Self-attention (Intra Attention)	A mechanism used in neural networks, particularly in natural language processing tasks, that allows the model to focus on different parts of the input sequence selectively
SGD *(Stochastic Gradient Descent)*	A popular optimization algorithm used for training machine learning models, particularly neural networks
SVD *(Singular Value Decomposition)*	A technique used for matrix factorization and dimensionality reduction. SVD can be used for tasks such as collaborative filtering, which helps identify latent features or factors in a dataset
Skip-gram	A variant of word2vec that focuses on predicting the context words given a target word
SVD++ *(Singular Value Decomposition Plus Plus)*	An extension of SVD used in collaborative filtering algorithms for recommendation systems that incorporate user-item interactions and user/item biases
SVM *(Support Vector Machines)*	A supervised machine learning algorithm that classifies data by finding the best hyperplane to separate two classes
SVR *(Support Vector Regression)*	A type of regression algorithm that uses support vector machines to minimize the margin error between predicted and actual values
TF-IDF *(Term Frequency-Inverse Document Frequency)*	A numerical statistic that reflects how important a word is in a document or corpus
VSM *(Vector Space Model)*	A mathematical model used in information retrieval to represent documents as vectors in a high-dimensional space
Word2vec	A deep learning model used to learn word embeddings that capture the meaning of words based on their context in a corpus

C Categories

See Table 11.

Table 11 Categories

Name	Definition
Classification	Classification is a supervised machine learning technique that predicts a categorical target variable based on input features. the algorithm learns from labeled examples during training and creates a model to classify new, unseen data
Clustering	Clustering is a machine learning technique that groups similar objects into clusters based on their similarities and differences without the need for predefined labels or output
Collaborative Filtering	Collaborative filtering is a technique used in recommender systems that predicts the preferences and interests of a user by analyzing the behavior and choices of a large number of similar users
Convolutional Neural Network (CNN)	A convolutional neural network (CNN) is a type of deep neural network widely used for image and video recognition tasks. it uses a series of convolutional layers to automatically learn and extract features from input data, followed by fully connected layers for classification or regression. cnns are trained using backpropagation to minimize the error between predicted and actual output
Deep Belief Networks (DBN)	Deep belief networks (DBN) are deep learning neural networks that consist of multiple layers of hidden units and are trained unsupervised using a generative model. They have been used for tasks involving complex, hierarchical data structures such as speech recognition, image recognition, and natural language processing
Graph Neural Networks (GNN)	Graph neural networks (GNNs) are deep learning models designed to operate on data with a graph structure, which iteratively pass information between connected nodes in a graph to learn representations that can be used for various tasks on the graph
Measurement model	A measurement model is a statistical model used to assess the relationship between observed variables and underlying constructs or latent variables. It is often used in psychology, sociology, and marketing research to quantify the relationship between observable data and the underlying constructs they represent
Optimization	Optimization models are mathematical or computational models that are used to find the best solution to a problem while taking into account constraints and objectives. they involve identifying the optimal value of one or more variables within a given set of constraints

Table 11 continued

Name	Definition
Probabilistic	A probabilistic model is a mathematical framework for representing uncertainty and quantifying the probability of different outcomes or events based on available information and assumptions about the underlying mechanisms
Recurrent Neural Networks (RNN)	Recurrent neural networks (RNNs) are a type of artificial neural network designed to process sequential data by maintaining an internal state or "memory" of previous inputs, enabling the network to make predictions about future inputs. RNNs have variants such as LSTM, GRU, and bi-RNNs and are used in applications such as language translation, speech recognition, and image captioning
Reinforcement Learning (RL)	Reinforcement learning is a type of machine learning where an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties, with the goal of maximizing the total reward over time
Self-Supervised Learning Model	Self-supervised learning is a type of machine learning in which a model learns to represent data by predicting certain attributes or properties of the data without explicit supervision, using unlabeled data
Semi-Supervised Learning	Semi-supervised learning is a type of machine learning that uses both labeled and unlabeled data to train algorithms for making predictions or classifications. it is useful when labeled data is limited or costly to obtain and is commonly used in applications such as speech recognition, image classification, and natural language processing
Statistical Method	Statistical methods are tools, techniques, and procedures used to collect, analyze, interpret, and present data to make conclusions or inferences about a population or sample
Supervised Learning	Supervised learning is a machine learning technique in which an algorithm learns to predict output values based on input data and their corresponding labeled examples
Unsupervised Learning	Unsupervised learning is a type of machine learning where the algorithm learns patterns and relationships in unlabeled data without specific guidance or supervision
Vector space model	A vector space model is a mathematical model used to represent text documents as vectors of numerical values, where each dimension corresponds to a particular term or word in the document collection

D Features

See Table 12.

Table 12 Features

Name	Definition
Activity-Based Recommendations	Refers to the features that capture the activities or behavior patterns of users. for example, in recommender systems, activity-based features may include the number of times a user has viewed, rated, purchased, or liked a particular item or category. these features can provide valuable insights into user preferences and be used to make personalized recommendations
Algorithm Flexibility (Algorithm-Agnostic)	Refers to the ability of a model or system to work with different types of algorithms or methods without being specific to any one of them
Anomaly Detection	Anomaly detection is examining specific data points and detecting rare occurrences that seem suspicious because they're different from the established pattern of behaviors
Attentive	Refers to the ability of the model to focus on certain parts of the input that are most relevant to the task at hand. this is typically achieved through mechanisms such as attention networks or attention heads in neural networks
Behavior-Based Recommendations	Refers to capturing user behavior and patterns to make predictions or recommendations
Click-Through Recommendations	Refers to using user clicks on search results to improve the relevance and ranking of search results for future queries
Co-Occurrence Analysis	Refers to the identification and analysis of the frequency of occurrence of two or more items or concepts together in a given context
Constraint-Based	Refers to incorporating domain-specific constraints into the learning algorithm to improve its performance
Content-Based Recommendations	Refers to a recommendation system that uses the characteristics or features of an item to recommend similar items to users
Context-Aware Recommendations	Consider the context or environment in which a particular task is performed to improve the accuracy of the model's predictions or recommendations
Contextual Graph	It is a directed acyclic graph with one input and one output that provides a uniform representation of elements of reasoning and of contexts in problem-solving
Data Dimensionality (Multidimensional)	The algorithm's ability to handle data with multiple input features or dimensions
Data Modality (Multimodal)	Refers to a type of data that includes multiple types of information or input modalities. this means that the data being used for the model contains features that come from different sources, such as text, audio, images, video, or other types of sensory data. multimodal machine learning models are designed to handle and learn from this diverse set of features, and to combine them in a meaningful way to achieve better performance than models that use only a single modality of data
Density-Based	Refers to machine learning techniques that identify dense regions of data points in a dataset, which are then used to cluster similar data points together

Table 12 continued

Name	Definition
Dimensionality Reduction	Dimensionality reduction, or dimension reduction, is the transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data, ideally close to its intrinsic dimension
End-To-End Approach	In machine learning, the end-to-end approach refers to a modeling strategy that involves training a single model to perform a complex task directly from raw data inputs to desired outputs without relying on intermediate stages or hand-crafted features. this approach is in contrast to the traditional pipeline approach, where different modules or stages are designed and trained independently to perform specific sub-tasks and then combined to form the final system
Entity Variability (Multi-Type Entities)	Refer to situations where there are multiple types of entities in the dataset, such as people, organizations, and locations. this feature is relevant for tasks such as named entity recognition and entity disambiguation, where the model must identify and differentiate between various entities
Feature Selection	Feature selection refers to selecting a subset of relevant features (or variables) from a larger set of features in a dataset to use as input for a machine learning model. The goal of feature selection is to improve model performance, reduce computational complexity, and increase interpretability by eliminating irrelevant, redundant, or noisy features
Filtering	Refers to selecting or excluding certain data points based on criteria or conditions
Frequency-Based	Refers to a type of approach in machine learning where the frequency of occurrence of certain items or events is used to derive patterns, associations, or predictions
Generative Model	"Generative models" are a type of machine learning model that can be used to generate new data that is similar to the data it was trained on
Geographic Support Recommendations	Refers to using geographic or location information as a feature in the model. this information can be used to improve the model's accuracy in predicting outcomes or providing recommendations based on the user's location or the location of relevant data
Graph Generation	Graph generation, whose purpose is to generate new graphs from a distribution similar to the observed graphs
Graph Ranking	"Graph ranking" is a feature in machine learning models that involves ranking or scoring nodes in a graph based on their importance or relevance

Table 12 continued

Name	Definition
Hierarchical Clustering	Involves grouping data points into nested clusters based on their similarity. it involves a series of iterative steps that build a hierarchy of clusters, where smaller clusters are combined into larger ones until all data points are in a single cluster
Historical Data-Driven Recommendations	Refers to the use of previously collected data to train machine learning models. this data is typically used to identify patterns, relationships, and trends that can inform future predictions or decisions
Hybrid Recommendation	Refers to a feature that combines different recommendation approaches, such as content-based and collaborative filtering, to generate personalized recommendations for users
Image Recognition	The process of identifying an object or a feature in an image or video
Image Similarity	"Image similarity" is a feature used in machine learning models that involves comparing two or more images to determine their similarity. this is typically done by computing a distance metric between the images based on their pixel values or features extracted from the images
Image-Based	Refers to using visual content, such as images or videos, as a basis for machine learning models
Item Recommendation	Refers to the ability of a model to suggest items to a user based on their previous interactions with a system or similar users' behaviors
Language Diversity (Multilingual)	Refers to the ability of the algorithm to work with and understand multiple languages
Memory-Based Approaches	Memory-based approaches in machine learning refer to algorithms that store and retrieve training data directly, without explicit model training. these methods often involve computing similarities between new input data and the stored training data to make predictions or classifications
Multi-Criteria Ratings	Refers to the ability of a model to consider multiple criteria or factors when assigning a rating or score to an item or entity
Multi-Task Learning	Refers to a single shared machine learning model that can perform multiple different (albeit related) tasks
Neighborhood-Based	Refers to using information about the items or users in the local neighborhood of a given item or user to make recommendations. it is based on the idea that users with similar preferences rate items similarly, and items rated similarly tend to have similar properties
Network Architecture	Refers to the structure of a neural network. it comprises layers of artificial neurons, each connected to the layers above and below
Opinion Mining	Involves analyzing and categorizing people's opinions, sentiments, emotions, and attitudes expressed in text data such as reviews, social media posts, and online forums
Parameter Estimation	Parameter estimation computes a model's parameter values from measured data

Table 12 continued

Name	Definition
Pattern-Based	Involves identifying patterns in the data that can be used to make predictions or classifications
Positive Relevance Feedback	"Positive relevance feedback" is a feature of some information retrieval systems. The idea behind relevance feedback is to take the results that are initially returned from a given query, to gather user feedback, and to use information about whether or not those results are relevant to performing a new query. the system takes user feedback into account to refine and improve future search results
Pre-Trained Model	Pre-trained models are machine learning models trained on a large data dataset and can be used as a starting point for training a new model on a different dataset. they can save a lot of time and effort and can be very effective. however, they may not be as accurate as models that are trained from scratch on a specific dataset
Prediction	Refers to the ability of a machine learning model to make informed guesses or forecasts about future outcomes based on patterns and trends it has learned from past data
Prediction Uncertainty	Refers to the ability of a model to estimate the uncertainty associated with its predictions
Pruning	In machine learning, pruning is a technique used to reduce the size of a decision tree by removing sections of the tree that provide little power to classify instances. pruning reduces the complexity of the final classifier and hence improves predictive accuracy by the reduction of overfitting
Query Refinement	Improving a search query by modifying, expanding, or narrowing its terms or parameters to retrieve more relevant results
Query Scoping	Involves identifying and segmenting user queries into specific categories or topics
Query Segmentation	Involves breaking down a user's query or input into smaller, more manageable parts to extract relevant information and provide more accurate results or recommendations
Query Suggestions	Involves generating and presenting a list of recommended queries to a user based on their initial search query or input
Query-Based	Refers to models responding to user queries, such as in search engines or question-answering systems. these systems use machine learning models to interpret the user's query and retrieve relevant information from a large database or corpus
Randomization	In machine learning, randomization refers to introducing randomness into the learning algorithm to improve its performance. randomness is often used to break any patterns in the data that might cause the model to overfit or underfit

Table 12 continued

Name	Definition
Ranking	Refers to the ability of a model to rank items or entities based on their relevance to a specific query or user
Ratings Prediction	"Rating prediction" is a feature in machine learning models that involves predicting the rating or preference of a user for a particular item or product. this is a common application in recommender systems, where the goal is to predict the rating that a user would give to a particular item based on their past behavior or preferences. this feature can be implemented using various approaches, such as collaborative filtering, matrix factorization, or content-based filtering
Recommendations Using User Feedback	Involves incorporating user feedback into the recommendation process to improve the relevance and personalization of the recommendations provided
Representation Learning	Refers to automatically learning a representation or a set of features that capture the underlying patterns and structure in user intent data
Rule-Based Tagging	Involves the use of predefined rules to assign tags or labels to input data automatically
Sampling-Based	Sampling-based is a feature in machine learning models that involves randomly selecting a subset of data from the entire dataset. this is done to make computations and analysis more efficient and faster, especially when dealing with a large dataset
Search Trail-Based Recommendations	Refers to the ability of a model to analyze a user's search history or search behavior to provide more personalized and relevant search results
Semantic Analysis	Refers to analyzing the meaning and context of words and phrases in natural language data
Session-Based Recommendations	Refers to using a user's current session behavior, such as search queries and clicks, to generate personalized recommendations
Smoothing	Refers to a technique to handle unknown or rare events in probabilistic models. it involves adjusting the probability estimates of events based on their frequency of occurrence in the training data. the basic idea is to redistribute probability mass from more frequent to less frequent events. this helps prevent overfitting and improves the accuracy of the model's predictions
Structure-Based	Refers to using structural information or prior knowledge about a problem domain to guide the learning process
Tag Relevance	Refers to the use of tags or keywords that are deemed relevant to a particular item or content
Template-Based	Refers to using pre-defined templates or rules to generate responses in natural language processing tasks such as chatbots or virtual assistants

Table 12 continued

Name	Definition
Term Weighting	Term weighting is a feature in machine learning models that assigns a numerical weight to each term or word in a document or dataset. the purpose of term weighting is to help the machine learning model better understand the importance of different terms or words in the data
Text Similarity	Measures how similar two pieces of text are to each other. it involves analyzing the text to extract its key features and comparing those features between the two pieces of text to determine their similarity. this feature is often used in tasks such as document classification, clustering, and information retrieval
Time-Aware Recommendations	Refers to considering the temporal dimension of data when building the model. this means that the model considers the sequence in which events occur over time and can make predictions based on this information
Time-Based Recommendations	Refers to the inclusion of time or temporal information in the data used for a machine learning task to model time-dependent patterns or to make predictions based on changes over time
Topic Modeling	Refers to the ability of the algorithm to automatically identify topics or themes in a collection of text documents. this feature is particularly useful in natural language processing applications, where the goal is to extract insights or understand the content of large text datasets
Trained-Based	Refers to an approach in which a mathematical model is used to learn from training data and make predictions or decisions based on new data. this approach involves selecting a model appropriate for the specific task, training it on the available data, and then using it to make predictions on new, unseen data
Transformer-Based	A neural network that learns context and thus meaning by tracking relationships in sequential data like the words in this sentence
Tree Based	The "tree-based" feature in machine learning refers to algorithms that use decision trees to make predictions. decision trees are tree-like structures where each node represents a feature or attribute, and each branch represents a possible value for that feature
User Interaction (Interactivity)	Refers to the ability of a model to interact with a user or other systems in real time
Word Cluster	"Word cluster" is a feature used in natural language processing and machine learning models that groups similar words into clusters based on their semantic meaning. this feature is used to represent the meaning of a word in a more abstract and generalized way, which can improve the performance of machine learning models that deal with natural language data

E Quality attributes and evaluation measures

See Table 13.

Table 13 Quality attributes and evaluation measures

Name	Definition
Accuracy	It measures the proportion of correctly classified instances in a binary classification problem
Area Under the ROC Curve (AUC)	It stands for the area under the roc curve and is used to evaluate the performance of a binary classification model. it measures the ability of the model to distinguish between positive and negative instances
Discounted Cumulative Gain (DCG)	It stands for discounted cumulative gain and is used to evaluate the quality of a ranking of items. it measures the usefulness of each item in the ranking, considering its position in the list
F1-Score	It is the harmonic mean of precision and recall and is used as a single metric to evaluate the performance of a recommendation system
Mean Absolute Error (MAE)	It stands for mean absolute error and is used to evaluate the performance of a regression model. it measures the average difference between the predicted and actual values
Mean Average Precision (MAP)	It stands for mean average precision and is used to evaluate the quality of a ranking of items. it measures the average precision across all relevant items
Mean Reciprocal Rank (MRR)	It stands for mean reciprocal rank and is used to evaluate the quality of a ranking of items. it measures the average of the reciprocal rank of the first relevant item
Normalized Discounted Cumulative Gain (NDCG)	It is an evaluation metric commonly used in information retrieval and recommendation systems to assess the quality and ranking of the recommended items or search results
Normalized Mutual Information (NMI)	It stands for normalized mutual information and is used to evaluate the similarity between two data clusterings
Precision	It measures the accuracy of the positive predictions made by the model, indicating how well it identifies true positive instances

Table 13 continued

Name	Definition
Recall	It quantifies the proportion of actual positive instances correctly classified as positive by the model
Root Mean Squared Error (RMSE)	It stands for root mean squared error and is used to evaluate the performance of a regression model. it measures the square root of the average squared differences between the predicted and actual values
Computational Cost	It measures the amount of computational resources, such as processing time or energy consumption, required by a software system to perform a particular operation or function, which can impact its performance, scalability, and cost-effectiveness
Convergence	It is a quality measure that evaluates how quickly an algorithm can find a solution or reach an optimal state
Coverage	Is a quality measure that evaluates the extent to which an algorithm or system can offer comprehensive and diverse information about a specific topic or data set by assessing the amount of captured and included information in the results or output
Diversity	It decreases the redundancy in the training data, the learned model, and the inference and provides more information for the machine learning process
Effectiveness	It measures the degree to which a system meets its functional requirements and achieves its intended goals
Flexibility	It refers to the ability of a software system to be easily modified or adapted to changing requirements or environments, which can reduce costs and risks associated with major software rewrites or redesigns
Informativeness	It is a quality attribute that refers to the degree to which a software system provides useful and relevant information to its users, which can aid in understanding and using the system effectively
Interpretability	It refers to the ability of a software system, particularly in the field of machine learning and artificial intelligence, to explain its decisions and actions clearly and understandably to both technical and non-technical users

Table 13 continued

Name	Definition
Novelty	It is a quality attribute in recommendation systems that measures its ability to suggest new and diverse content to users
Performance Efficiency	It is a software quality attribute that measures the ability of a system to use computing resources effectively to meet performance requirements
Predictability	A system or software can produce expected and consistent results given specific conditions or inputs
Recommendation Effectiveness	It measures how well a software system can provide accurate and useful recommendations to users, which is important for increasing user engagement and satisfaction
Recommendation Efficiency	It refers to the ability of a recommendation system to generate relevant and accurate recommendations for a user efficiently
Recommendation Performance	It measures how well a recommendation system suggests relevant items to users
Reliability	It measures the ability of a system to perform its intended functions consistently and predictably without unexpected or erroneous behavior
Resource Efficiency	It measures the ability of a system to use computing resources, such as memory, processing power, and storage, efficiently and effectively
Resource Utilization	It is a software quality attribute that measures the efficient and effective use of computing resources, such as memory, processing power, and storage, by a software system, which can impact its performance, scalability, and cost-effectiveness
Retrieval Performance	It measures how effectively and efficiently a software system can retrieve relevant information from a large data collection
Robustness	It measures the ability of a system to remain stable and reliable under various abnormal or unexpected conditions, such as invalid inputs, system failures, or attacks from malicious users
Satisfaction	It measures the extent to which a system meets or exceeds the expectations and needs of its users, resulting in a positive user experience
Scalability	It measures the ability of a system to handle increasing amounts of work or users without experiencing a degradation in performance

Table 13 continued

Name	Definition
Simplicity	It refers to the quality or state of being simple, straightforward, or easy to understand. In various domains and contexts, simplicity is often valued as it promotes clarity, efficiency, and usability
Stability	It measures the ability of a system to maintain its performance and reliability over time, even under changing conditions or in the face of failures or errors
Transparency	It refers to the ability of a software system to provide clear and understandable information to users, which can increase their trust and satisfaction in using the system
Usefulness	It measures the extent to which a system is capable of satisfying user needs and delivering value to its intended users
Validity	It refers to the accuracy and correctness of the data and information processed by a system, including recommendation systems

Acknowledgements We extend our sincere gratitude to the domain experts who actively participated in and contributed to this research project. Their valuable insights and expertise have significantly enriched the quality of this study. We would like to express our appreciation to Sjaak Brinkkemper, Fabiano Dalpiaz, Gerard Wagenaar, Fernando Castor de Lima Filho, and Sergio Espana Cubillo for their invaluable feedback, which has helped us in presenting the results of this study more effectively. We are also deeply thankful to all the participants of the case studies for their cooperation and willingness to share their valuable publications, which served as essential resources in evaluating and validating the proposed decision model. Their contributions have been pivotal in ensuring the practical applicability and effectiveness of the decision model in real-world scenarios. Finally, we extend our appreciation to the journal editors and reviewers for their meticulous review of this manuscript and their constructive feedback. Their efforts have played a crucial role in enhancing the quality and clarity of this research, making it a more valuable contribution to the scientific community.

References

Agarwal, N., Sikka, G., Awasthi, L.K.: Evaluation of web service clustering using Dirichlet multinomial mixture model based approach for dimensionality reduction in service representation. Inf. Process. Manag. **57**(4), 102238 (2020)

Allamanis, M., Barr, E.T., Devanbu, P., Sutton, C.: A survey of machine learning for big code and naturalness. ACM Comput. Surv. (CSUR) **51**(4), 1–37 (2018)

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., Zimmermann, T.: Software engineering for machine learning: A case study. In: 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), pp. 291–300. IEEE (2019)

Ashkan, A., Clarke, C.L., Agichtein, E., Guo, Q.: Classifying and characterizing query intent. In: Advances in Information Retrieval: 31th European Conference on IR Research, ECIR 2009, Toulouse, France, April 6–9, 2009. Proceedings 31, pp. 578–586. Springer (2009)

Baykan, E., Henzinger, M., Marian, L., Weber, I.: A comprehensive study of features and algorithms for URL-based topic classification. ACM Trans. Web (TWEB) **5**(3), 1–29 (2011)

Beel, J., Gipp, B., Langer, S., Breitinger, C.: Paper recommender systems: a literature survey. Int. J. Digit. Libr. **17**, 305–338 (2016)

Beemer, J., Spoon, K., He, L., Fan, J., Levine, R.A.: Ensemble learning for estimating individualized treatment effects in student success studies. Int. J. Artif. Intell. Educ. **28**, 315–335 (2018)

Bhaskaran, S., Santhi, B.: An efficient personalized trust based hybrid recommendation (TBHR) strategy for e-learning system in cloud computing. Clust. Comput. **22**, 1137–1149 (2019)

Bifet, A., Holmes, G., Pfahringer, B., Kirkby, R., Gavalda, R.: New ensemble methods for evolving data streams. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 139–148 (2009)

Bigi, B.: Using Kullback–Leibler distance for text categorization. In: European Conference on Information Retrieval, pp. 305–319. Springer (2003)

Binkley, D., Lawrie, D., Morrell, C.: The need for software specific natural language techniques. Empir. Softw. Eng. **23**, 2398–2425 (2018)

Cai, Y., Lau, R.Y., Liao, S.S., Li, C., Leung, H.-F., Ma, L.C.: Object typicality for effective web of things recommendations. Decis. Support Syst. **63**, 52–63 (2014)

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P.S., Sun, L.: A comprehensive survey of AI-generated content (AIGC): a history of generative AI from GAN to chatGPT. arXiv preprint arXiv:2303.04226 (2023)

Caprara, A., Toth, P., Fischetti, M.: Algorithms for the set covering problem. Ann. Oper. Res. **98**(1–4), 353–371 (2000)
Carmel, D., Chang, Y., Deng, H., Nie, J.-Y.: Future directions of query understanding. Query Understanding for Search Engines, pp. 205–224 (2020)
Carvallo, A., Parra, D., Lobel, H., Soto, A.: Automatic document screening of medical literature using word and text embeddings in an active learning setting. Scientometrics **125**, 3047–3084 (2020)
Chen, Y., Liu, Z., Li, J., McAuley, J., Xiong, C.: Intent contrastive learning for sequential recommendation. In: Proceedings of the ACM Web Conference 2022, pp. 2172–2182 (2022)
Chen, L., Wang, Y., Yu, Q., Zheng, Z., Wu, J.: WT-LDA: user tagging augmented LDA for web service clustering. In: Service-Oriented Computing: 11th International Conference, ICSOC 2013, Berlin, Germany, December 2–5, 2013, Proceedings 11, pp. 162–176. Springer (2013)
Chen, T., Wong, R.C.-W.: Handling information loss of graph neural networks for session-based recommendation. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1172–1180 (2020)
Chen, L., Chen, G., Wang, F.: Recommender systems based on user reviews: the state of the art. User Model. User-Adap. Inter. **25**, 99–154 (2015)
Colace, F., De Santo, M., Greco, L., Moscato, V., Picariello, A.: A collaborative user-centered framework for recommending items in online social networks. Comput. Hum. Behav. **51**, 694–704 (2015)
Cremonesi, P., Garzotto, F., Negro, S., Papadopoulos, A.V., Turrin, R.: Looking for "good" recommendations: a comparative evaluation of recommender systems. In: Human-Computer Interaction–INTERACT 2011: 13th IFIP TC 13 International Conference, Lisbon, Portugal, September 5–9, 2011, Proceedings, Part III 13, pp. 152–168. Springer (2011)
Da'u, A., Salim, N.: Sentiment-aware deep recommender system with neural attention networks. IEEE Access **7**, 45472–45484 (2019). https://doi.org/10.1109/ACCESS.2019.2907729
de Barcelos Silva, A., Gomes, M.M., da Costa, C.A., da Rosa Righi, R., Barbosa, J.L.V., Pessin, G., De Doncker, G., Federizzi, G.: Intelligent personal assistants: a systematic literature review. Expert Syst. Appl. **147**, 113193 (2020)
Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
Ding, X., Liu, T., Duan, J., Nie, J.-Y.: Mining user consumption intention from social media using domain adaptive convolutional neural network. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29 (2015)
Ding, H., Liu, Q., Hu, G.: TDTMF: a recommendation model based on user temporal interest drift and latent review topic evolution with regularization factor. Inf. Process. Manag. **59**(5), 103037 (2022)
Dodeja, L., Tambwekar, P., Hedlund-Botti, E., Gombolay, M.: Towards the design of user-centric strategy recommendation systems for collaborative human-AI tasks. Int. J. Hum Comput Stud. **184**, 103216 (2024)
Doumpos, M., Grigoroudis, E.: Multicriteria Decision Aid and Artificial Intelligence. Whiley, UK (2013)
Elfaik, H., et al.: Leveraging feature-level fusion representations and attentional bidirectional RNN-CNN deep models for Arabic affect analysis on Twitter. J. King Saud Univ. Comput. Inf. Sci. **35**(1), 462–482 (2023)
Fan, L., Li, Q., Liu, B., Wu, X.-M., Zhang, X., Lv, F., Lin, G., Li, S., Jin, T., Yang, K.: Modeling user behavior with graph convolution for personalized product search. In: Proceedings of the ACM Web Conference 2022, pp. 203–212 (2022)
Farshidi, S., Kwantes, I.B., Jansen, S.: Business process modeling language selection for research modelers. Softw Syst Model 1–26 (2023)
Farshidi, S.: Multi-criteria decision-making in software production. PhD thesis, Utrecht University (2020)
Farshidi, S.: Understanding user intent: a systematic literature review of modeling techniques. Mendeley Data (2024). https://doi.org/10.17632/zcbh9r37rc.1
Farshidi, S., Jansen, S., van der Werf, J.M.: Capturing software architecture knowledge for pattern-driven design. J. Syst. Softw. **169**, 110714 (2020)
Fitzgerald, B., Stol, K.-J.: Continuous software engineering and beyond: trends and challenges. In: Proceedings of the 1st International Workshop on Rapid Continuous Software Engineering, pp. 1–9 (2014)
Fitzgerald, D.R., Mohammed, S., Kremer, G.O.: Differences in the way we decide: the effect of decision style diversity on process conflict in design teams. Pers. Individ. Differ. **104**, 339–344 (2017)
Gao, C., Lam, W.: Search clarification selection via query-intent-clarification graph attention. In: European Conference on Information Retrieval, pp. 230–243. Springer (2022)

Garcia, K., Berton, L.: Topic detection and sentiment analysis in twitter content related to COVID-19 from Brazil and the USA. Appl. Soft Comput. **101**, 107057 (2021)
Garg, R.: Parametric selection of software reliability growth models using multi-criteria decision-making approach. Int. J. Reliab. Saf. **13**(4), 291–309 (2019)
Garg, R.: MCDM-based parametric selection of cloud deployment models for an academic organization. IEEE Trans. Cloud Comput. **10**, 863–871 (2020)
Garg, R., Sharma, R., Sharma, K.: MCDM based evaluation and ranking of commercial off-the-shelf using fuzzy based matrix method. Decis. Sci. Lett. **6**(2), 117–136 (2017)
Garg, R., Kumar, R., Garg, S.: MADM-based parametric selection and ranking of E-learning websites using fuzzy COPRAS. IEEE Trans. Educ. **62**(1), 11–18 (2018)
Gozuacik, N., Sakar, C.O., Ozcan, S.: Technological forecasting based on estimation of word embedding matrix using LSTM networks. Technol. Forecast. Soc. Change **191**, 122520 (2023)
Gu, Y., Zhao, B., Hardtke, D., Sun, Y.: Learning global term weights for content-based recommender systems. In: Proceedings of the 25th International Conference on World Wide Web. WWW '16, pp. 391–400. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2016). https://doi.org/10.1145/2872427.2883069
Gunawan, D., Sembiring, C., Budiman, M.A.: The implementation of cosine similarity to calculate text relevance between two documents. In: Journal of Physics: Conference Series, vol. 978, p. 012120. IOP Publishing (2018)
Guo, L., Hua, L., Jia, R., Fang, F., Zhao, B., Cui, B.: EdgeDIPN: a unified deep intent prediction network deployed at the edge. Proc. VLDB Endowm. **14**(3), 320–328 (2020)
Haefliger, S., Von Krogh, G., Spaeth, S.: Code reuse in open source software. Manag. Sci. **54**(1), 180–193 (2008)
Hashemi, S.H., Williams, K., El Kholy, A., Zitouni, I., Crook, P.A.: Measuring user satisfaction on smart speaker intelligent assistants using intent sensitive query embeddings. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 1183–1192 (2018)
Hashemi, H., Zamani, H., Croft, W.B.: Guided transformer: leveraging multiple external sources for representation learning in conversational search. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1131–1140 (2020)
Hernández-Rubio, M., Cantador, I., Bellogín, A.: A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews. User Model. User-Adap. Inter. **29**(2), 381–441 (2019)
Hidasi, B., Karatzoglou, A.: Recurrent neural networks with top-k gains for session-based recommendations. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 843–852 (2018)
Hill, C., Bellamy, R., Erickson, T., Burnett, M.: Trials and tribulations of developers of intelligent systems: a field study. In: 2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), pp. 162–170. IEEE (2016)
Hu, Y., Da, Q., Zeng, A., Yu, Y., Xu, Y.: Reinforcement learning to rank in e-commerce search engine: formalization, analysis, and application. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '18, pp. 368–377. Association for Computing Machinery, New York (2018). https://doi.org/10.1145/3219819.3219846
Hu, Z., Zhang, Z., Yang, H., Chen, Q., Zuo, D.: A deep learning approach for predicting the quality of online health expert question-answering services. J. Biomed. Inform. **71**, 241–253 (2017)
Huang, Q., Xia, X., Lo, D., Murphy, G.C.: Automating intention mining. IEEE Trans. Softw. Eng. **46**(10), 1098–1119 (2018)
Iovine, A., Narducci, F., Musto, C., de Gemmis, M., Semeraro, G.: Virtual customer assistants in finance: from state of the art and practices to design guidelines. Comput. Sci. Rev. **47**, 100534 (2023)
ISO: IEC/IEEE systems and software engineering: Architecture description. ISO/IEC/IEEE 42010: 2011 (E)(Revision of ISO/IEC 42010: 2007 and IEEE Std 1471-2000) (2011)
Ittoo, A., van den Bosch, A., et al.: Text analytics in industry: challenges, desiderata and trends. Comput. Ind. **78**, 96–107 (2016)
Izadi, M., Akbari, K., Heydarnoori, A.: Predicting the objective and priority of issue reports in software repositories. Empir. Softw. Eng. **27**(2), 50 (2022)
Jain, S., Grover, A., Thakur, P.S., Choudhary, S.K.: Trends, problems and solutions of recommender system. In: International Conference on Computing, Communication & Automation, pp. 955–958 (2015)

Jansen, S.: Applied multi-case research in a mixed-method research project: customer configuration updating improvement. In: Information Systems Research Methods, Epistemology, and Applications, pp. 120–139. IGI Global (2009)
Jiang, D., Pei, J., Li, H.: Mining search and browse logs for web search: a survey. ACM Trans. Intell. Syst. Technol. (TIST) **4**(4), 1–37 (2013)
Jindal, V., Bawa, S., Batra, S.: A review of ranking approaches for semantic search on web. Inform. Process. Manag. **50**(2), 416–425 (2014)
Johnson, R.B., Onwuegbuzie, A.J.: Mixed methods research: a research paradigm whose time has come. Educ. Res. **33**(7), 14–26 (2004)
Jordan, M.I., Mitchell, T.M.: Machine learning: trends, perspectives, and prospects. Science **349**(6245), 255–260 (2015)
Kaptein, R., Kamps, J.: Exploiting the category structure of wikipedia for entity ranking. Artif. Intell. **194**, 111–129 (2013)
Kaufmann, L., Kreft, S., Ehrgott, M., Reimann, F.: Rationality in supplier selection decisions: the effect of the buyer's national task environment. J. Purch. Supply Manag. **18**(2), 76–91 (2012)
Keyvan, K., Huang, J.X.: How to approach ambiguous queries in conversational search: a survey of techniques, approaches, tools, and challenges. ACM Comput. Surv. **55**(6), 1–40 (2022)
Khilji, A.F.U.R., Sinha, U., Singh, P., Ali, A., Dadure, P., Manna, R., Pakray, P.: Multimodal recipe recommendation system using deep learning and rule-based approach. SN Comput. Sci. **4**(4), 421 (2023)
Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlỳ, P., Suchomel, V.: The sketch engine: ten years on. Lexicography **1**(1), 7–36 (2014)
Kim, D., Park, C., Oh, J., Yu, H.: Deep hybrid recommender systems via exploiting document context and statistics of items. Inf. Sci. **417**, 72–87 (2017)
Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J., Linkman, S.: Systematic literature reviews in software engineering-a systematic literature review. Inf. Softw. Technol. **51**(1), 7–15 (2009)
Konishi, T., Ohwa, T., Fujita, S., Ikeda, K., Hayashi, K.: Extracting search query patterns via the pairwise coupled topic model. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, pp. 655–664 (2016)
Kuwajima, H., Yasuoka, H., Nakae, T.: Engineering problems in machine learning systems. Mach. Learn. **109**(5), 1103–1126 (2020)
Larson, S., Mahendran, A., Peper, J.J., Clarke, C., Lee, A., Hill, P., Kummerfeld, J.K., Leach, K., Laurenzano, M.A., Tang, L. et al.: An evaluation dataset for intent classification and out-of-scope prediction. arXiv preprint arXiv:1909.02027 (2019)
Latifi, S., Mauro, N., Jannach, D.: Session-aware recommendation: a surprising quest for the state-of-the-art. Inf. Sci. **573**, 291–315 (2021)
Li, L., Deng, H., Dong, A., Chang, Y., Zha, H.: Identifying and labeling search tasks via query-based hawkes processes. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 731–740 (2014)
Lin, H., Liu, G., Li, F., Zuo, Y.: Where to go? predicting next location in IoT environment. Front. Comput. Sci. **15**, 1–13 (2021)
Liu, Z., Chen, H., Sun, F., Xie, X., Gao, J., Ding, B., Shen, Y.: Intent preference decoupling for user representation on online recommender system. In: Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, pp. 2575–2582 (2021)
Liu, J., Dou, Z., Zhu, Q., Wen, J.-R.: A category-aware multi-interest model for personalized product search. In: Proceedings of the ACM Web Conference 2022, pp. 360–368 (2022)
Liu, P., Liao, D., Wang, J., Wu, Y., Li, G., Xia, S.-T., Xu, J.: Multi-task ranking with user behaviors for text-video search. In: Companion Proceedings of the Web Conference 2022, pp. 126–130 (2022)
Liu, P., Zhang, L., Gulla, J.A.: Dynamic attention-based explainable recommendation with textual and visual fusion. Inf. Process. Manag. **57**(6), 102099 (2020)
Liu, T., Wu, Q., Chang, L., Gu, T.: A review of deep learning-based recommender system in e-learning environments. Artif. Intell. Rev. **55**(8), 5953–5980 (2022)
Ludewig, M., Jannach, D.: Evaluation of session-based recommendation algorithms. User Model. User-Adap. Inter. **28**, 331–390 (2018)
Majumder, M.: Multi criteria decision making. In: Impact of Urbanization on Water Shortage in Face of Climatic Aberrations, pp. 35–47. Springer (2015)
Mandayam Comar, P., Sengamedu, S.H.: Intent based relevance estimation from click logs. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 59–66 (2017)

Manzoor, A., Jannach, D.: Towards retrieval-based conversational recommendation. Inf. Syst. **109**, 102083 (2022). https://doi.org/10.1016/j.is.2022.102083
Mao, M., Lu, J., Han, J., Zhang, G.: Multiobjective e-commerce recommendations based on hypergraph ranking. Inf. Sci. **471**, 269–287 (2019)
Musto, C., Narducci, F., Lops, P., de Gemmis, M., Semeraro, G.: Linked open data-based explanations for transparent recommender systems. Int. J. Hum. Comput. Stud. **121**, 93–107 (2019)
Ni, X., Lu, Y., Quan, X., Wenyin, L., Hua, B.: User interest modeling and its application for question recommendation in user-interactive question answering systems. Inf. Process. Manag. **48**(2), 218–233 (2012)
Okoli, C., Schabram, K.: A guide to conducting a systematic literature review of information systems research (2015)
Oulasvirta, A., Blom, J.: Motivations in personalisation behaviour. Interact. Comput. **20**(1), 1–16 (2008)
Pan, R., Bagherzadeh, M., Ghaleb, T.A., Briand, L.: Test case selection and prioritization using machine learning: a systematic literature review. Empir. Softw. Eng. **27**(2), 29 (2022)
Papadimitriou, A., Symeonidis, P., Manolopoulos, Y.: A generalized taxonomy of explanations styles for traditional and social recommender systems. Data Min. Knowl. Discov. **24**, 555–583 (2012)
Park, C., Kim, D., Yang, M.-C., Lee, J.-T., Yu, H.: Click-aware purchase prediction with push at the top. Inf. Sci. **521**, 350–364 (2020)
Paul, H., Nikolaev, A.: Fake review detection on online e-commerce platforms: a systematic literature review. Data Min. Knowl. Discov. **35**(5), 1830–1881 (2021)
Penha, G., Hauff, C.: What does BERT know about books, movies and music? probing BERT for conversational recommendation. In: Proceedings of the 14th ACM Conference on Recommender Systems, pp. 388–397 (2020)
Phan, X.-H., Nguyen, C.-T., Le, D.-T., Nguyen, L.-M., Horiguchi, S., Ha, Q.-T.: A hidden topic-based framework toward building applications with short web documents. IEEE Trans. Knowl. Data Eng. **23**(7), 961–976 (2010)
Pi, Q., Bian, W., Zhou, G., Zhu, X., Gai, K.: Practice on long sequential user behavior modeling for click-through rate prediction. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2671–2679 (2019)
Portugal, I., Alencar, P., Cowan, D.: The use of machine learning algorithms in recommender systems: a systematic review. Expert Syst. Appl. **97**, 205–227 (2018)
Pradhan, T., Kumar, P., Pal, S.: CLAVER: an integrated framework of convolutional layer, bidirectional LSTM with attention mechanism based scholarly venue recommendation. Inf. Sci. **559**, 212–235 (2021)
Pu, P., Chen, L., Hu, R.: Evaluating recommender systems from the user's perspective: survey of the state of the art. User Model. User-Adap. Inter. **22**(4), 317–355 (2012)
Qu, Y., Cai, H., Ren, K., Zhang, W., Yu, Y., Wen, Y., Wang, J.: Product-based neural networks for user response prediction. In: 2016 IEEE 16th International Conference on Data Mining (ICDM), pp. 1149–1154 (2016). https://doi.org/10.1109/ICDM.2016.0151
Qu, Y., Cai, H., Ren, K., Zhang, W., Yu, Y., Wen, Y., Wang, J.: Product-based neural networks for user response prediction. In: 2016 IEEE 16th International Conference on Data Mining (ICDM), pp. 1149–1154. IEEE (2016)
Qu, C., Yang, L., Croft, W.B., Zhang, Y., Trippas, J.R., Qiu, M.: User intent prediction in information-seeking conversations. In: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval, pp. 25–33 (2019)
Rapp, A., Curti, L., Boldi, A.: The human side of human-chatbot interaction: a systematic literature review of ten years of research on text-based chatbots. Int. J. Hum. Comput. Stud. **151**, 102630 (2021)
Ribeiro, M.T., Singh, S., Guestrin, C.: "why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144 (2016)
Ricci, F., Rokach, L., Shapira, B.: Recommender systems: introduction and challenges. Recom. Syst. Handb. 1–34 (2015)
Robertson, S., Zaragoza, H., Taylor, M.: Simple bm25 extension to multiple weighted fields. In: Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management, pp. 42–49 (2004)
Rus, I., Halling, M., Biffl, S.: Supporting decision-making in software engineering with process simulation and empirical studies. Int. J. Softw. Eng. Knowl. Eng. **13**(05), 531–545 (2003)

Sagi, O., Rokach, L.: Ensemble learning: a survey. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. **8**(4), 1249 (2018)
Saka, A.B., Oyedele, L.O., Akanbi, L.A., Ganiyu, S.A., Chan, D.W., Bello, S.A.: Conversational artificial intelligence in the AEC industry: a review of present status, challenges and opportunities. Adv. Eng. Inform. **55**, 101869 (2023)
Salle, A., Malmasi, S., Rokhlenko, O., Agichtein, E.: COSEARCHER: studying the effectiveness of conversational search refinement and clarification through user simulation. Inf. Retr. J. **25**(2), 209–238 (2022)
Sandhya, Garg, R., Kumar, R.: Computational MADM evaluation and ranking of cloud service providers using distance-based approach. Int J Inf Decis. Sci. **10**(3), 222–234 (2018)
Sarker, I.H.: Machine learning: algorithms, real-world applications and research directions. SN Comput. Sci. **2**(3), 160 (2021)
Schlaefer, N., Chu-Carroll, J., Nyberg, E., Fan, J., Zadrozny, W., Ferrucci, D.: Statistical source expansion for question answering. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 345–354 (2011)
Singh, A., Thakur, N., Sharma, A.: A review of supervised machine learning algorithms. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1310–1315. IEEE (2016)
Srivastava, M., Nushi, B., Kamar, E., Shah, S., Horvitz, E.: An empirical analysis of backward compatibility in machine learning systems. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 3272–3280 (2020)
Sun, C., Gan, C., Nevatia, R.: Automatic concept discovery from parallel text and visual corpora. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2596–2604 (2015)
Takács, G., Tikk, D.: Alternating least squares for personalized ranking. In: Proceedings of the Sixth ACM Conference on Recommender Systems, pp. 83–90 (2012)
Tamine-Lechani, L., Boughanem, M., Daoud, M.: Evaluation of contextual information retrieval effectiveness: overview of issues and research. Knowl. Inf. Syst. **24**, 1–34 (2010)
Tang, J., Yao, L., Zhang, D., Zhang, J.: A combination approach to web user profiling. ACM Trans. Knowl. Discov. Data (TKDD) **5**(1), 1–44 (2010)
Tanjim, M.M., Su, C., Benjamin, E., Hu, D., Hong, L., McAuley, J.: Attentive sequential models of latent intent for next item recommendation. In: Proceedings of The Web Conference 2020, pp. 2528–2534 (2020)
Tanjim, M.M., Su, C., Benjamin, E., Hu, D., Hong, L., McAuley, J.: Attentive sequential models of latent intent for next item recommendation. In: Proceedings of The Web Conference 2020. WWW '20, pp. 2528–2534. Association for Computing Machinery, New York (2020). https://doi.org/10.1145/3366423.3380002
Teevan, J., Dumais, S.T., Liebling, D.J.: To personalize or not to personalize: modeling queries with variation in user intent. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 163–170 (2008)
Telikani, A., Tahmassebi, A., Banzhaf, W., Gandomi, A.H.: Evolutionary machine learning: a survey. ACM Comput. Surv. (CSUR) **54**(8), 1–35 (2021)
Vayansky, I., Kumar, S.A.: A review of topic modeling methods. Inf. Syst. **94**, 101582 (2020)
Venkateswara Rao, P., Kumar, A.S.: The societal communication of the Q&A community on topic modeling. J. Supercomput. **78**(1), 1117–1143 (2022)
von Rueden, L., Mayer, S., Sifa, R., Bauckhage, C., Garcke, J.: Combining machine learning and simulation to a hybrid modelling approach: current and future directions. In: Advances in Intelligent Data Analysis XVIII: 18th International Symposium on Intelligent Data Analysis, IDA 2020, Konstanz, Germany, April 27–29, 2020, Proceedings 18, pp. 548–560. Springer (2020)
Wang, J., Ding, K., Hong, L., Liu, H., Caverlee, J.: Next-item recommendation with sequential hypergraphs. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1101–1110 (2020)
Wang, W., Hosseini, S., Awadallah, A.H., Bennett, P.N., Quirk, C.: Context-aware intent identification in email conversations. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 585–594 (2019)
Wang, Y., Wang, S., Li, Y., Dou, D.: Recognizing medical search query intent by few-shot learning. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 502–512 (2022)

Wang, T., Lin, Q.: Hybrid predictive models: when an interpretable model collaborates with a black-box model. J. Mach. Learn. Res. **22**(1), 6085–6122 (2021)
Wang, H.-C., Jhou, H.-T., Tsai, Y.-S.: Adapting topic map and social influence to the personalized hybrid recommender system. Inf. Sci. **575**, 762–778 (2021)
Wang, X., Li, Q., Yu, D., Cui, P., Wang, Z., Xu, G.: Causal disentanglement for semantics-aware intent learning in recommendation. IEEE Trans. Knowl. Data Eng. (2022). https://doi.org/10.1109/TKDE.2022.3159802
Weismayer, C., Pezenka, I.: Identifying emerging research fields: a longitudinal latent semantic keyword analysis. Scientometrics **113**(3), 1757–1785 (2017)
White, R.W., Chu, W., Hassan, A., He, X., Song, Y., Wang, H.: Enhancing personalized search by mining and modeling task behavior. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 1411–1420 (2013)
Wu, L., Quan, C., Li, C., Wang, Q., Zheng, B., Luo, X.: A context-aware user-item representation learning for item recommendation. ACM Trans. Inf. Syst. (TOIS) **37**(2), 1–29 (2019)
Wu, Z., Liang, J., Zhang, Z., Lei, J.: Exploration of text matching methods in Chinese disease Q&A systems: a method using ensemble based on BERT and boosted tree models. J. Biomed. Inform. **115**, 103683 (2021)
Xia, C., Zhang, C., Yan, X., Chang, Y., Yu, P.S.: Zero-shot user intent detection via capsule neural networks. arXiv preprint arXiv:1809.00385 (2018)
Xiao, Y., Watson, M.: Guidance on conducting a systematic literature review. J. Plan. Educ. Res. **39**(1), 93–112 (2019)
Xu, P., Sugano, Y., Bulling, A.: Spatio-temporal modeling and prediction of visual attention in graphical user interfaces. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 3299–3310 (2016)
Xu, L., Brinkkemper, S.: Concepts of product software. Eur. J. Inf. Syst. **16**(5), 531–541 (2007)
Xu, Z., Chen, L., Chen, G.: Topic based context-aware travel recommendation method exploiting geotagged photos. Neurocomputing **155**, 99–107 (2015)
Xu, H., Ding, W., Shen, W., Wang, J., Yang, Z.: Deep convolutional recurrent model for region recommendation with spatial and temporal contexts. Ad Hoc Netw. **129**, 102545 (2022)
Xu, H., Ding, W., Shen, W., Wang, J., Yang, Z.: Deep convolutional recurrent model for region recommendation with spatial and temporal contexts. Ad Hoc Netw. **129**, 102545 (2022). https://doi.org/10.1016/j.adhoc.2021.102545
Yadav, N., Pal, S., Singh, A.K., Singh, K.: Clus-DR: cluster-based pre-trained model for diverse recommendation generation. J. King Saud Univ. Comput. Inf. Sci. **34**(8), 6385–6399 (2022)
Yao, S., Tan, J., Chen, X., Zhang, J., Zeng, X., Yang, K.: ReprBERT: distilling BERT to an efficient representation-based relevance model for e-commerce. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 4363–4371 (2022)
Yao, Y., Zhao, W.X., Wang, Y., Tong, H., Xu, F., Lu, J.: Version-aware rating prediction for mobile app recommendation. ACM Trans. Inf. Syst. (TOIS) **35**(4), 1–33 (2017)
Ye, Q., Wang, F., Li, B.: Starrysky: A practical system to track millions of high-precision query intents. In: Proceedings of the 25th International Conference Companion on World Wide Web. WWW '16 Companion, pp. 961–966. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2016). https://doi.org/10.1145/2872518.2890588
Yengikand, A.K., Meghdadi, M., Ahmadian, S.: DHSIRS: a novel deep hybrid side information-based recommender system. Multimed Tools Appl 1–27 (2023)
Yin, R.K.: Case Study Research: Design and Methods, vol. 5. Sage, Thousand Oaks (2009)
Yin, R.K.: Case Study Research and Applications: Design and Methods. Sage publications, New York (2017)
Yu, Z., Lian, J., Mahmoody, A., Liu, G., Xie, X.: Adaptive user modeling with long and short-term preferences for personalized recommendation. In: IJCAI, pp. 4213–4219 (2019)
Yu, J., Zhu, T.: Combining long-term and short-term user interest for personalized hashtag recommendation. Front. Comput. Sci. **9**, 608–622 (2015)
Yu, S., Liu, J., Yang, Z., Chen, Z., Jiang, H., Tolba, A., Xia, F.: Pave: Personalized academic venue recommendation exploiting co-publication networks. J. Netw. Comput. Appl. **104**, 38–47 (2018)
Yu, B., Zhang, R., Chen, W., Fang, J.: Graph neural network based model for multi-behavior session-based recommendation. GeoInformatica **26**(2), 429–447 (2022)

Yuan, S., Zhang, Y., Tang, J., Hall, W., Cabotà, J.B.: Expert finding in community question answering: a review. Artif. Intell. Rev. **53**, 843–874 (2020)
Zaib, M., Zhang, W.E., Sheng, Q.Z., Mahmood, A., Zhang, Y.: Conversational question answering: a survey. Knowl. Inf. Syst. **64**(12), 3151–3195 (2022)
Zhang, Y., Chen, X., Ai, Q., Yang, L., Croft, W.B.: Towards conversational search and recommendation: System ask, user respond. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 177–186 (2018)
Zhang, C., Fan, W., Du, N., Yu, P.S.: Mining user intentions from medical queries: A neural network based heterogeneous jointly modeling approach. In: Proceedings of the 25th International Conference on World Wide Web, pp. 1373–1384 (2016)
Zhang, H., Xu, H., Lin, T.-E., Lyu, R.: Discovering new intents with deep aligned clustering. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 14365–14373 (2021)
Zhang, Y., Yin, H., Huang, Z., Du, X., Yang, G., Lian, D.: Discrete deep learning for fast content-aware recommendation. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pp. 717–726 (2018)
Zhang, H., Zhong, G.: Improving short text classification by learning vector representations of both words and hidden topics. Knowl.-Based Syst. **102**, 76–86 (2016)
Zhang, H., Babar, M.A., Tell, P.: Identifying relevant studies in software engineering. Inf. Softw. Technol. **53**(6), 625–637 (2011)
Zhang, S., Yao, L., Sun, A., Tay, Y.: Deep learning based recommender system: a survey and new perspectives. ACM Comput Surv. (CSUR) **52**(1), 1–38 (2019)
Zhang, C., Huang, X., An, J., Zou, S.: Improving conversational recommender systems via multi-preference modeling and knowledge-enhanced. Knowl. Based Syst. **286**, 111361 (2024)
Zhou, X., Jin, Y., Zhang, H., Li, S., Huang, X.: A map of threats to validity of systematic literature reviews in software engineering. In: 2016 23rd Asia-Pacific Software Engineering Conference (APSEC), pp. 153–160. IEEE (2016)
Zhou, K., Zhao, W.X., Wang, H., Wang, S., Zhang, F., Wang, Z., Wen, J.-R.: Leveraging historical interaction data for improving conversational recommender system. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 2349–2352 (2020)
Zhou, X., Qin, D., Chen, L., Zhang, Y.: Real-time context-aware social media recommendation. VLDB J. **28**, 197–219 (2019)
Zou, J., Kanoulas, E., Ren, P., Ren, Z., Sun, A., Long, C.: Improving conversational recommender systems via transformer-based sequential modelling. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2319–2324 (2022)

Siamak Farshidi is a research fellow at the University of Amsterdam in the Multiscale Networked Systems (MNS) research group. His work focuses on decision-making in software engineering, integrating advanced AI models and decision-making theories into software engineering practices. Throughout his academic career, he has tackled challenges in automated decision-making within software production and has conducted extensive research on multi-criteria decision-making problems, particularly in technology selection.

Kiyan Rezaee is computer science students at the University of Guilan in Rasht, Iran. He serve as research assistants and collaborate closely with Utrecht University. Their primary research interests lie in AI models and their applications in decision-making within software engineering.

Sara Mazaheri is computer science students at the University of Guilan in Rasht, Iran. He serve as research assistants and collaborate closely with Utrecht University. Their primary research interests lie in AI models and their applications in decision-making within software engineering.

Amir Hossein Rahimi is computer science students at the University of Guilan in Rasht, Iran. He serve as research assistants and collaborate closely with Utrecht University. Their primary research interests lie in AI models and their applications in decision-making within software engineering.

Ali Dadashzadeh is computer science students at the University of Guilan in Rasht, Iran. He serve as research assistants and collaborate closely with Utrecht University. Their primary research interests lie in AI models and their applications in decision-making within software engineering.

Morteza Ziabakhsh is computer science students at the University of Guilan in Rasht, Iran. He serve as research assistants and collaborate closely with Utrecht University. Their primary research interests lie in AI models and their applications in decision-making within software engineering.

Sadegh Eskandari is an Assistant Professor in the Department of Computer Science at the University of Guilan in Rasht, Iran. He holds a Ph.D. in Applied Mathematics and an MSc in Computer Science, both from Shahid Bahonar University of Kerman. His research focuses on feature selection, machine learning, and evolutionary computation.

Slinger Jansen is an Associate Professor at Utrecht University, where he leads the Software Ecosystems Security research group. A prominent researcher in the field of software ecosystems, he co-founded the International Conference on Software Business and the International Workshop on Software Ecosystems. He has authored and edited several books, including "Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry." Additionally, Slinger Jansen serves as an associate editor for the Empirical Software Engineering journal and actively supports startups by serving on advisory boards for multiple ventures.