arXiv:1804.11192v3 [cs.IR] 13 May 2018

Foundations and Trends® in Information Retrieval
Vol. XX, No. XX (2018) 1–85
© 2018 now Publishers Inc.
DOI: 10.1561/XXXXXXXXXX

Explainable Recommendation: A Survey and New Perspectives

Yongfeng Zhang
Rutgers University
yongfeng.zhang@rutgers.edu

Xu Chen
Tsinghua University
xu-ch14@mails.tsinghua.edu.cn

Contents

Abstract

Explainable Recommendation refers to the personalized recommendation algorithms that address the problem of *why* - they not only provide the user with the recommendations, but also make the user aware why such items are recommended by generating recommendation explanations, which help to improve the effectiveness, efficiency, persuasiveness, and user satisfaction of recommender systems. In recent years, a large number of explainable recommendation approaches – especially model-based explainable recommendation algorithms – have been proposed and adopted in real-world systems.

In this survey, we review the work on explainable recommendation that has been published in or before the year of 2018. We first highlight the position of explainable recommendation in recommender system research by categorizing recommendation problems into the 5W, i.e., what, when, who, where, and why. We then conduct a comprehensive survey of explainable recommendation itself in terms of three aspects: 1) We provide a chronological research line of explanations in recommender systems, including the user study approaches in the early years, as well as the more recent model-based approaches. 2) We provide a taxonomy for explainable recommendation algorithms, including user-based, item-based, model-based, and post-model explanations. 3) We summarize the application of explainable recommendation in different recommendation tasks, including product recommendation, social recommendation, POI recommendation, etc. We devote a chapter to discuss the explanation perspectives in the broader IR and machine learning settings, as well as their relationship with explainable recommendation research. We end the survey by discussing potential future research directions to promote the explainable recommendation research area.

now Publishers Inc.. *Explainable Recommendation: A Survey and New Perspectives*. Foundations and Trends® in Information Retrieval, vol. XX, no. XX, pp. 1–85, 2018.

DOI: 10.1561/XXXXXXXXXX.

1

Introduction

1.1 Explainable Recommendation

In this section, we will introduce the definition of the explainable recommendation problem itself. We will highlight the position of explainable recommendation in the whole recommendation research area by providing a taxonomy to classify most of the personalized recommendation research works. Specifically, the many recommendation research tasks can be classified as addressing the 5W problems - when, where, who, what, why, and the five W's generally correspond to time-aware recommendation (when), location-based recommendation (where), social recommendation (who), application-aware recommendation (what), and explainable recommendation (why), respectively, where explainable recommendation aims to answer the question of *why*.

With this section, the readers will get a clear understanding on not only the explainable recommendation problem itself – which is the key topic of this survey – but also, they will get a big picture of the whole recommendation research area, so as to understand what is unique about explainable recommendation and why the research on explainable recommendation is important to this research field.

1.2 A Historical Overview

In this section we will provide a history overview of the explainable recommendation research. Though the term *explainable recommendation* was formally introduced in the recent years (Zhang et al. [2014a]), the basic concept, however, dates back to some of the most early works in personalized recommendation research. For example, Schafer et al. [1999] noted that recommender system would be used to explain to a user what type of thing a product is, such as "this product you are looking at is similar to these other products that you have liked in the past", which is the fundamental idea of item-based collaborative filtering; Herlocker et al. [2000] studied how to explain collaborative filtering algorithms in MovieLens based on user surveys; and Sinha and Swearingen [2002] highlighted the role of transparency in recommender systems. Besides, even before explainable recommendation has attracted serious research attention, the industry has been using semi-automatic or manually designed explanations in practical systems, such as the "people also viewed" explanation in Amazon e-commerce.

To help the readers understand the "pre-history" research of recommendation explanation and how explainable recommendation emerged as an important research task in the recent years, we provide a historical overview of the research line in this section.

Early approaches to personalized recommender systems mostly focused on content-based recommendation or collaborative-filtering (CF) based recommendation (Ricci et al. [2011]). Content-based recommender systems attempt to model user and/or item profiles with various available content information, such as the price, color, brand of the goods in e-commerce, or the genre, director, duration of the movies in review systems (Balabanović and Shoham [1997], Pazzani and Billsus [2007]). Because the item contents are usually easily understandable to the users, it is usually intuitive to explain to the users why an item is recommended out of other candidates in content-based recommendation. For example, one straight forward way is to let the users know about the certain content features he/she might be interested in on the recommended item. Ferwerda et al. [2012] provided a comprehensive study of possible protocols to provide explanations for content-based recommendations.

However, collecting content information in different application scenarios for content-based recommendation is a time consuming task. Collaborative filtering (CF) based (Ekstrand et al. [2011]) approaches, on the other hand, attempts to avoid this difficulty by leveraging “wisdom of the crowds”. One of the most early CF-based recommendation algorithms is the User-based CF in the GroupLens news recommendation system introduced by Resnick et al. [1994], which represents each user as a vector of ratings, and predicts the missing ratings of a user on a news message based on weighted average of other users’ ratings on this message. Symmetrically, Sarwar et al. [2001] introduced the item-based CF method, and Linden et al. [2003] further described its application in Amazon product recommendation system. Item-based CF takes each item as a vector of ratings, and predicts a missing rating by weighted average of the ratings from similar items.

Though the predicted ratings would be relatively difficult to understand for normal users of the system, user- and item-based CF are somewhat explainable based on the philosophy of the their algorithm design. For example, the items recommended by user-based CF can be explained as “users that are similar to you loved this item”, while item-based CF can be explained as “the item is similar to your previously loved items”. However, although the idea of collaborative filtering has achieved significant improvement on recommendation accuracy, it is less intuitive to explain compared with many content-based algorithms, and research pioneers in very early stages also noticed the importance of the problem (Herlocker et al. [2000], Herlocker and Konstan [2000], Sinha and Swearingen [2002]).

The idea of collaborative filtering achieved further success when integrated with Latent Factor Models (LFM) introduced by Koren [2008] in late 2000’s, among which Matrix Factorization (MF)-based CF and its variants were especially successful in rating predictions (Koren et al. [2009]). Latent factor models led the research and application of recommender systems for the years to come. Though successful in recommendation performance, the “latent factors” in latent factor models such as matrix factorization do not possess intuitive meanings, which makes it difficult to understand why an item achieved better predictions and got recommended out of the others. This lack of model explainability also makes it difficulty to provide intuitive recommendation explanations to the users, and it would be hardly acceptable to tell the users that an item is recommended just because it gets higher predicted scores by the model.

To make personalized recommendation models intuitively understandable, researchers have more and more turned to the study of *model-based explainable recommendation*, where the recommendation algorithm not only provides a recommendation list as output, but also naturally works in an explainable way and provides explanations to accompany the recommendations. For example, McAuley and Leskovec [2013] aligned the latent dimensions with latent topics from latent dirichlet allocation (LDA) for recommendation, and Zhang et al. [2014a] formally defined the *explainable recommendation* problem, and proposed the Explicit Factor Model (EFM) by aligning the latent dimensions with explicit product features for explainable recommendation. A lot of other approaches were also proposed to address the problem of explainability, which will be introduced in detail in the following parts of the paper. It is worthwhile to note that the application of deep learning models to personalized recommendation has further improved recommendation performances in the recent years, but the black box nature of deep models also brings about the difficulty of model explainability. In this survey, we will also review the research efforts for explainable recommendation with deep models.

In a broader sense, researchers in the Greater-AI community have also realized the importance of *Explainable AI* in the recent years, which aims to address a wide range of AI explainability problems in deep learning, computer vision, automatic driving systems, natural language processing tasks, etc. As an important branch of AI research, this also highlights the importance for the recommendation system community to address the explainability issues of various recommendation systems, and the research on explainable recommendation has also been a suitable setting to develop and investigate new *Explainable Machine Learning* algorithms and theories.

1.3 Classification of the Methods

In this survey, we provide a classification paradigm of existing explainable recommendation methods, which can help the readers to better understand the state-of-the-art of explainable recommendation research.

We classify existing explainable recommendation research based on two dimensions: 1) The type of the generated explanations (e.g., textual, visual, etc), and 2) the model or algorithm used to generate such an explanation, which can be broadly classified into matrix factorization, topic modeling, graph-based, deep learning, knowledge-graph, association rules, and post-hoc models, etc. Based on this classification taxonomy, we devise two chapters to introduced the research on explainable recommendation, corresponding to Chapter 2 and Chapter 3, respectively.

For example, the Explicit Factor Model (EFM) for explainable recommendation proposed in Zhang et al. [2014a] falls into the category of shallow model-based method with textual explanations; the Interpretable Convolutional Neural Network approach to explainable recommendation in Seo et al. [2017] falls into the category of deep model-based method with textual explanations; while the visually explainable recommendation approach proposed in Chen et al. [2018c] belongs to deep model-based methods with visual explanations. We also classify other research works according to this taxonomy so that the readers can get a clear understanding of the existing explainable recommendation methods, and these methods are also going to be analyzed in detail in the following sections.

1.4 Explainability and Effectiveness

It was long believed that explainability and effectiveness of recommendation models are two conflicting goals that can not be achieved at the same time, i.e., you can either choose a simple method for better explainability, or you can choose to design a complex recommendation model while sacrificing the explainability.

However, researchers have more and more realized recently that these two goals may not necessarily conflict with each other when designing recommendation models (Bilgic et al. [2004], Zhang et al. [2014a]), and state-of-the-art techniques – especially the prospering deep representation learning approaches – can help us to design recommendation models that are both highly effective and explainable, and this is also one of the most attractive research direction of explanation recommendation in the recent years, which not only leads into progress in personalized recommendation, but also helps us more generally in the research of Explainable Machine Learning.

When introducing each explainable recommendation model in detail in the following sections, we will also discuss the relationships of explainability and effectiveness of the models in terms of providing personalized recommendations.

2

Different Forms of Explanations

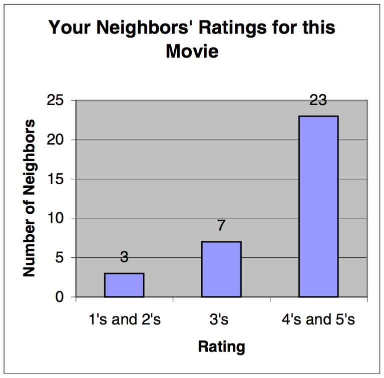
Recommendation explanation can come in different forms in practice (Tintarev and Masthoff [2015]). For example, Zhang et al. [2014a], He et al. [2015], Vig et al. [2009], etc., generated a (personalized) textual sentence as recommendation explanation to help users understand each particular result; McAuley and Leskovec [2013], Zhang [2015], Al-Taie and Kadry [2014], etc., provided topical word cloud as explanations; Chen et al. [2018c], etc., highlighted particular regions of the product image as a visual explanation to the users; Sharma and Cosley [2013], Quijano-Sanchez et al. [2017]), etc., generated a list of social friends who liked the product as social explanation for each target user; in the early research stages, Herlocker et al. [2000], Bilgic and Mooney [2005], Tintarev and Masthoff [2007], McSherry [2005], etc., adopted statistical histogram or pie chart to help users understand the rating distribution and pros/cons of each product as intuitive explanations. Figure 2.1 shows several representative recommendation explanations as examples.

In this section, we provide a summary of the different types of recommendation explanations to help the readers understand what an explanation can look like in practice. We also categorize the related wok into these different types for easy reference. Specifically, the following subsections presents an overview of several frequently seen explanations in practice.

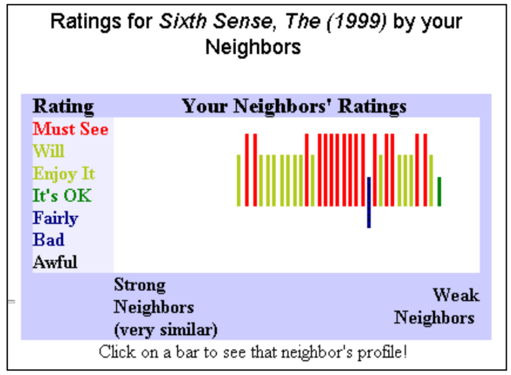
Figure 2.1: Different forms of explanations. On the left panel, there are four example users together with their purchased items and the corresponding reviews and ratings. On the right panel, we review the different types of explanation investigated in the previous research.

2.1 User-based and Item-based Explanations

We start from the very early stages of recommendation explanation research. In this section we introduce explainable recommendation with user-based and item-based collaborative filtering (Resnick et al. [1994], Sarwar et al. [2001], Zanker and Ninaus [2010], Cleger-Tamayo et al.



(a)



(b)

Figure 2.2: Example of user-based explanations (a) A histogram of neighbors' ratings for the recommended item as explanation, with the "good" ratings clustered together and the "bad" ratings clustered together, and the ambivalent ratings separated out. The result is that the user has to do only a single binary visual comparison to understand the consequence of the explanation. (b) A screen explaining the recommendation for the movie "The Sixth Sense". Each bar represents a rating of a neighbor. Upwardly trending bars are positive ratings, while downward trending ones are negative. The x-axis represents similarity to the user. Courtesy images from Herlocker et al. [2000].

[2012]), which are two of the most fundamental methods for personalized recommendation. Research works that are extensions of the two basic methods will also be introduced in this section.

User-based and item-based explanations are usually provided based on users' implicit or explicit feedbacks. In user-based collaborative filtering (Resnick et al. [1994]), user-based explanation can be generated by letting the user know that a recommendation is provided because his/her ratings are similar to a group of "neighbor" users, and these neighbor users made good ratings on the recommended item. For example, Herlocker et al. [2000] compared the effectiveness of different display forms of explanation for user-based collaborative filtering, where the explanation can be shown as an aggregated histogram of the neighbors' ratings, or be shown as the detailed ratings of the neighbors, as shown in Figure 2.2. State-of-the-art model-based explanations can generate more personalized and meticulously designed explanations

Your ratings for similar movies

Movie	Your Rating out of 5
L.A. Confidential	4
Air Force One	5
The Game	5
12 Angry Men	3
Carrie	4

Your neighbors' ratings for this movie

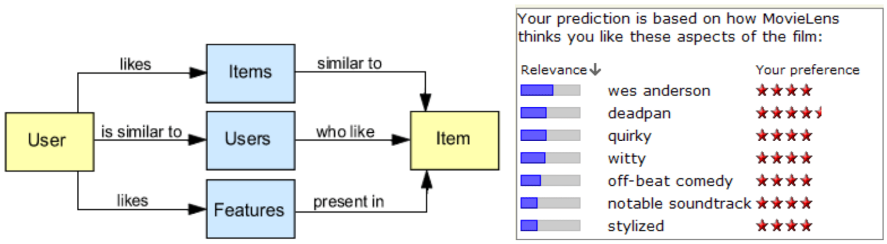
Rating	Number of Neighbors
★	0
★★	0
★★★	3
★★★★	4
★★★★★	2

Figure 2.3: A comparison between item-based explanation and user-based explanation. Courtesy image from Abdollahi and Nasraoui [2017].

than this, but this research illustrated the based ideas of providing explanations for users.

In item-based collaborative filtering (Sarwar et al. [2001]), explanations can be provided by letting the user know that an item is recommended because it is similar to some other items that the user liked before, as shown in the left subfigure of Figure 2.3, where several similar movies that the user made high ratings (4 or 5 stars) before are displayed as explanations. More intuitively, as shown in Figure 2.1, for the recommended item (i.e., the camera lens), user-based explanation tells Bob that similar users William and Fred bought this item, while item-based explanation persuades Bob by pointing out that the lens is relevant to his previously purchased camera.

Item-based explanations are usually more intuitive for users to understand, because users are usually familiar with those products that he/she has purchased before, as a result, providing these products will be acceptable explanations to most users. In user-based explanation, however, providing other users that made similar choices with the current user as explanation may be less convincing, because the target user may not know about the explanation users at all. Besides, disclosing the explanation based on users' purchasing history information may also cause privacy problems in practical systems. This drives user-based explanation to a new direction, which is to leverage social friend



(a) (b)

Figure 2.4: Tags explanation: generating explanations based on content information of items from tags Herlocker et al. [2000]. (a) The basic idea of tags-based recommendation is to find the tags that a user likes and then recommend items with these tags. (b) The tags-explanation provided for a recommended movie *Rushmore*, where the relevant tags (aspects) are displayed as explanations.

information and to provide a user with his/her social friends public behavior as recommendation explanations. We will review this research direction in the following section of social explanation.

2.2 Content-based Explanation

Content-based recommendation provides personalized recommendations by matching user preference with the available item content features (Pazzani and Billsus [2007], Ferwerda et al. [2012], Cramer et al. [2008]). Based on these content features, content-based recommendation is usually intuitive to explain in practice.

Content-based recommendation can be realized based on many different item contents in different application scenarios. For example, movie recommendation can be generated based on movie genres, actor, director, etc.; and book recommendation can be provided based on book types, price, author, etc. A common paradigm for content-based explanation is to provide users with the item features that match with the target user's interest profile.

Vig et al. [2009] adopted tags as item content features to generate recommendations and the corresponding explanations, as shown in Figure 2.4. To explain the recommended movie, the system displays the

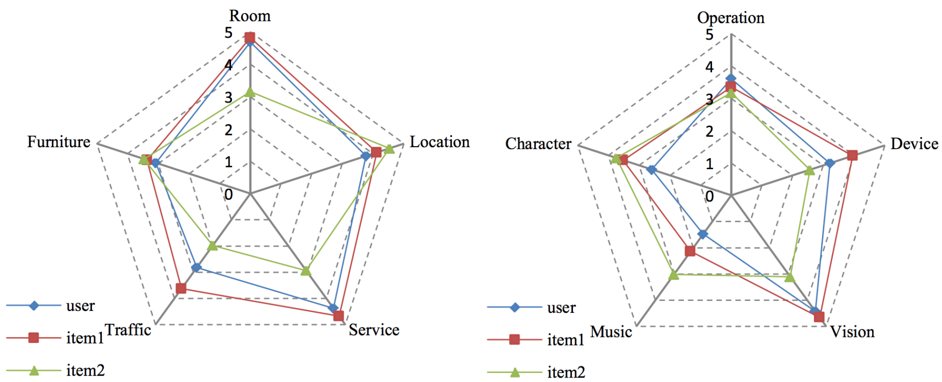


Figure 2.5: Using radar charts to explained why an item is recommended. The left figure shows hotel recommendation for a user, where item 1 is recommended because it satisfies user preferences on nearly all aspects. Similarly, the right figure shows video game recommendation and also, item 1 is recommended to the user. Courtesy image from Hou et al. [2018].

content features (aspects) of the movie, and also tells the user about the relevance of each aspect. The authors also designed user study experiments to analyze the effectiveness of the content-based explanations. Tintarev [2007] studied different types of contents for explanations, and also investigated the effectiveness, trustworthiness, and different ways to present explanations to users; Ferwerda et al. [2012], Wang and Benbasat [2007] adopted the user study approach based on subject surveys to investigate the trust and understandability of content-based explanations.

The content features can also be displayed in intuitive forms for users to understand. For example, Hou et al. [2018] used radar charts to explain why an item is recommended to a user and why others are not recommended, as shown in Figure 2.5, where a recommended item can be explained that most of its aspects satisfy the preference of the target user.

User demographic information describes the content of users, and the demographic contents can also be used to generate content-based explanations. Demographic-based recommendation (Pazzani [1999]) is one of the most early approaches to personalized recommendation, and in the recent years, researchers have also integrated demographic-based method with social media to provide product recommendations in social environments (Zhao et al. [2014, 2016]).

Basically, demographic-based recommendation makes recommendations based on the user's demographic information such as age, gender, location, etc. Intuitively, a recommended item based on demographic information can be explained as being appropriate for a particular type of user, e.g., by letting the user know that "80% of customers in an age group will buy a particular product". Zhao et al. [2014] represented products and users in the same demographic feature space, and used the weights of the demographic features learned by a ranking function to intuitively explain the results; Zhao et al. [2016] further explored demographic information in social media for product recommendation with explanation.

2.3 Textual Explanations

With the ever prospering of Web 2.0, more and more user generated contents have been accumulated on the web, such as user reviews in e-commerce and posts in social networks. Such information is of great value for accessing more comprehensive user preference, and can be utilized to provide finer-grained and more reliable recommendation explanations to persuade and help costumers. Motivated by this intuition, recently, many models have been designed to explain recommendations leveraging various types of text information, and they usually generate a piece of textual sentence as explanation.

The related methods can be generally classified into feature-level and sentence-level approaches according to how the textual explanations are displayed to users. See Figure 2.1 for example, the feature-level models present product features (such as color, quality, etc) as well as the possible sentiments to the customers for recommendation explanations, and the sentence-level methods directly present a sentence to Bob to tell him why the camera lens is recommended.

Feature-level explanation is similar to content-based explanation except that the features are usually not directly available content pro-

(a) (b)

Figure 2.6: Word cloud explanation for hotel recommendation generated based on latent topic modeling with textual reviews. (a) Word cloud about the *Location* of the recommended hotel, and (b) Word cloud about the *Service* of the recommended hotel. Courtesy image from Wu and Ester [2015].

files, instead, they are extracted or learned as part of the recommendation model from – e.g., the textual reviews – and the features can be paired up with consumer opinions to express a clear sentiment on the feature.

To extract product features and user sentiments from large-scale textual reviews, Zhang et al. developed and publicized a phrase-level sentiment analysis toolkit called *Sentires*[1] (Zhang et al. [2014b]), which can extract "feature–opinion–sentiment" triplets from textual reviews of a product domain. For example, given large-scale user reviews about mobile phones, the toolkit can extract triplets such as "noise–high–negative", "screen–clear–positive", and "battery_life–long–positive", etc. The toolkit also has the ability to detect contextual sentiments of opinion words given different feature words, for example, though "noise" paired with "high" usually represents a negative sentiment, when "quality" is paired with "high", it instead shows a positive sentiment. Based on the dictionary of feature–opinion–sentiment triplets that the program constructed, it can further detect which triplets are contained in given review sentence automatically.

Based on this toolkit, Zhang et al. [2014a] and Zhang [2015] developed an explicit factor model for explainable recommendation, and presented word clouds of the feature-opinion pairs as explanation, for

[1] http://yongfeng.me/software/

example, "bathroom-clean", which not only indicates the available features of an item, but also the consumers' aggregated opinions on the features, so as to highlight the performance of a recommended item on these features. These sentiment-enhanced modeling approach was also leveraged in point-of-interest recommendation Zhao et al. [2015] and social recommendation Ren et al. [2017].

Without using the opinion words and sentiment scores, the explanations can also be shown as a word cloud of product features that the user may be interested in. For example, Wu and Ester [2015] developed a topic modeling approach for explainable recommendation of hotels on TripAdvisor, which generates topical word clouds on three perspectives of hotels (Location, Service, and Room) as explanations, as shown in Figure 2.6, where the size of words in the word cloud reflects the importance of the corresponding feature.

`You might be interested in [feature], on which this product performs well.`
`You might be interested in [feature], on which this product performs poorly.`

Figure 2.7: Generating sentence explanations with template-based methods.

Dataset	(User, Item)	Explanation
Amazon Books	(9163, 11021)	i love this series. i can't wait for the next book. i love the characters and the story line. i was so glad that the story was a little longer. i would recommend this book to anyone who enjoy a good mystery.
BeerAdvocate	(shivtim, 2023)	poured from a bottle into a pint glass. a: pours a dark brown with a small head. s - smells of caramel and chocolate. t - a bit of a caramel malt and a little bit of coffee. m- medium body with a solid carbonation. d - medium bodied with a smooth mouthfeel. i can taste the sweetness and a bit of caramel and a little bit of a bit of alcohol.

Figure 2.8: Generating sentence explanations directly based on natural language generation models such as LSTM. Courtesy image from Costa et al. [2017].

In sentence-level approaches, the explanations are provided as a complete and semantically coherent sentence. The sentence can be constructed based on templates, for example, Zhang et al. [2014a] attempts to construct an explanation by telling the user that "*You might be interested in* `feature`*, on which this product performs well*", and in this

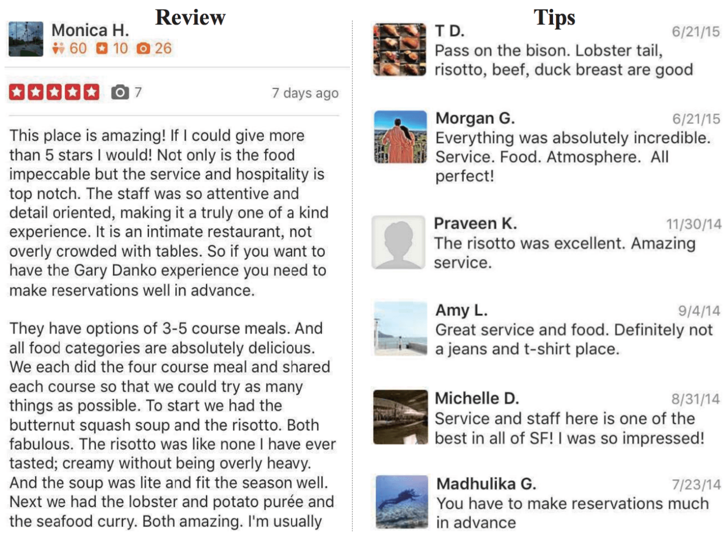


Figure 2.9: Example of reviews and tips selected from the restaurant "Gary Danko" on Yelp. Tips are more concise than reviews and can reveal user experience, feelings, and suggestions with only a few words. Courtesy image from Li et al. [2017].

template, the `feature` will be selected based on personalization algorithms so as to construct a personalized explanation, as shown in Figure 2.7. Based on the templates, the model can also provide "dis-recommendations", and let the user know why an item is not a good fit for the user by telling the user that "*You might be interested in* `feature`*, on which this product does not perform very well*". As shown in Zhang et al. [2014a] based on real-world user studies, providing both recommendations and dis-recommendations as well as their explanations improves the persuasiveness, conversion rate, and trustworthiness of recommender systems.

The textual explanation sentence can also be generated without templates, for example, Costa et al. [2017] attempted to generate an item's review explanations using long-short term memory (LSTM), and by learning based large-scale user review data, the model can generate reasonable review sentence as explanations automatically, as shown in Figure 2.8. Inspired by how people explain word-of-mouth recommendations, Chang et al. [2016] proposed a process that combines crowdsourcing and computation to generate personalized natural language explanations, and the authors also evaluated the generated natural language explanations in terms of efficiency, effectiveness, trust, and satisfaction. Li et al. [2017] leveraged gated recurrent units (GRU) to summarize the massive reviews of an item and generate tips for an item, as shown in Figure 2.9. Although tips generation does not directly come from an explainable recommendation model, the tips are still very helpful for users to understand the key features of the recommended item quickly and accurately.

2.4 Visual Explanations

To take advantage of the intuition of visual images, there has been new attempts to leverage product images for explainable recommendation recently. In Figure 2.1 for example, to tell Bob that the lens is recommended because of the collar appearance, the system highlights the image region corresponding to the necklet of the lens.

In particular, Chen et al. [2018c] proposed visually explainable recommendation to highlight the image regions that a user may be interested in, as shown in Figure 2.10. The basic intuitive is that different users may be attracted by different regions of the product image, as shown in the example in Figure 2.11, even for the same shirt, some users may care about the collar while others may pay more attention to the pocket. As a result, they adopted neural attention mechanism integrated with both image and textual reviews information to learn the importance of each region in an image, and highlight the important areas as visual explanations.

2.5 Social Explanation

As discussed in the previous subsections, a problem with user-based explanation is trustworthiness and privacy concerns, because the target user may have no idea about the explanation users who have "similar interests" at all. Usually, it will be more acceptable if we tell the user that his/her friends have similar interests on the recommended item. As a result, researchers proposed to generate social explanations with

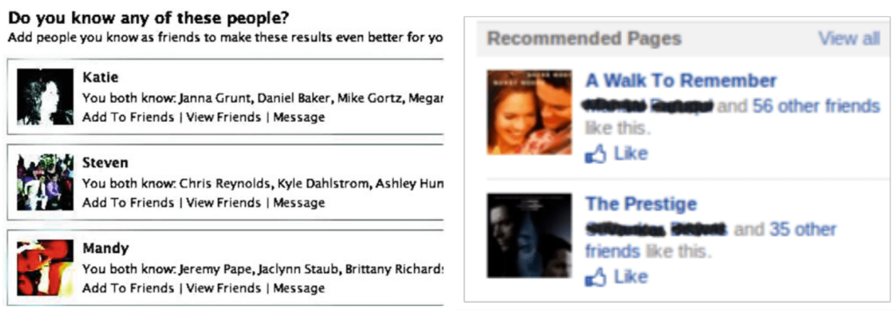
#	Target Item	Historical Records	Textual Review	Visual Explanation	
				VECF	Re-VECF
1			this is a large watch... nearly as large as my suunto but due to ***its articulated strap it fits on the wrist very well.***		
2			***this is a really comfortable v-neck. i found that the size and location of the v are just right for me.*** i'm 5'8 & #34, but 200 lbs (and dropping :))		
3			***Great leggings. perfect for fly fishing or hunting or running.*** just perfect anytime you are cold!		
4			The socks on the shoes are a perfect fit for me. ***first time with a shoe with the speed laces and i like them a lot***		
5			Really like these socks! they are really thick woolen socks and are good for cold days. ***they cover a good portion of your feet as they go a little (halfway) above the calf muscle area.***		
6			***I like the front pocket~!*** Very cool!		

Figure 2.10: Examples of the visual explanations in Chen et al. [2018c], where each row represents the target item of a user. The first column lists the image of the target item, and the second column lists two most similar products to the target item that the user purchased before. The third column shows the user's review on the target item, and the last two columns compare the highlighted regions provided by two visually explainable recommendation models for the target item. In the review column, the bolded italic texts highlight the part of user review that the generated visual explanations correspond to.

Figure 2.11: According to user reviews, different users may care about different regions of the image even for the same product.

the help of social information.

Papadimitriou et al. [2012] studied human-style, item-style, feature-style and hybrid-style explanations in social recommender systems, and they also studied geo-social explanations that combine geographical



(a) (b)

Figure 2.12: Social explanations in Facebook. (a) Facebook provides the common friends as explanation when recommending a new friend to a user (Papadimitriou et al. [2012]). (b) Providing friends who liked the same item when recommending items to a user (Sharma and Cosley [2013]).

with social data. For example, Facebook provides common friends as explanation when recommending a new friend to a user (Figure 2.12(b)). Sharma and Cosley [2013] studied the effects of social explanations in music recommendation context by providing the target user with the number of friends that liked the recommended item (Figure 2.12(b)), and found that the explanations do have influence on the likelihood of user checking out the recommended artists, but there could be little correlation between the likelihood and the actual rating for the same artist. Chaney et al. [2015] presented social Poisson factorization, a Bayesian model that incorporates a user's latent preferences for items with the latent influences of her friends, which provides a source of explainable serendipity (i.e., pleasant surprise due to novelty) to users.

Social explanations can also be provided in other scenarios except for friend recommendation in social networks, for example, Park et al. [2017] proposed the UniWalk algorithm to exploit both rating data and social network to generate explainable and accurate product recommendations. In this framework, a recommendation can be explained based on similar users that are friends with the target user, as shown in Figure 2.13. Quijano-Sanchez et al. [2017] introduced a social explanation system applied to group recommendation, which significantly

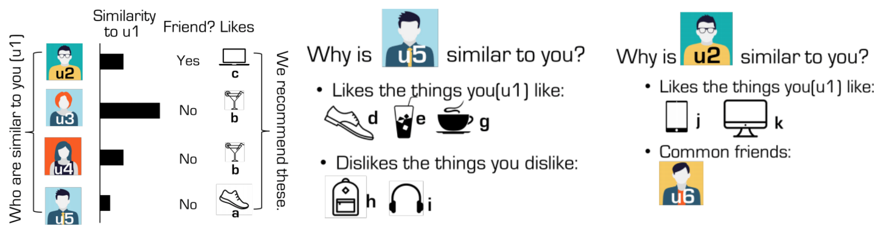


Figure 2.13: Explanations based on similar users, where the similar users can be social friends or users that have the same preference on the same subset of products (Park et al. [2017]).

increased the user intent (likelihood) to follow the recommendations, the user satisfaction, and the system efficiency to help users make decisions. Wang et al. [2014] generates social explanations such as "*A and B also like the item*". They proposed to generate the most persuasive social explanation by recommending the optimal set of users to be put in the explanation. Specifically, a two-phase ranking algorithm is proposed, which predicts the persuasiveness of a set users, taking factors like marginal utility of persuasiveness, credibility of explanation, and reading cost into consideration.

2.6 Hybrid Explanations

In many practical systems, multiple types of explanations can be integrated as hybrid explanation to users, so that the users can understand the recommendation from multiple perspectives. On the other hand, many practical systems rely on more than one recommendation algorithms to build a hybrid recommendation systems, as a result, it is important and sometimes challenging to make explanations for hybrid recommendation outputs consistently.

Kouki et al. [2017] described a hybrid recommender system built on a probabilistic programming language, and discuss the benefits and challenges of explaining its recommendations to users. To achieve this goal, the authors performed a mixed model statistical analysis of user preferences for explanations in the system. Through an online user

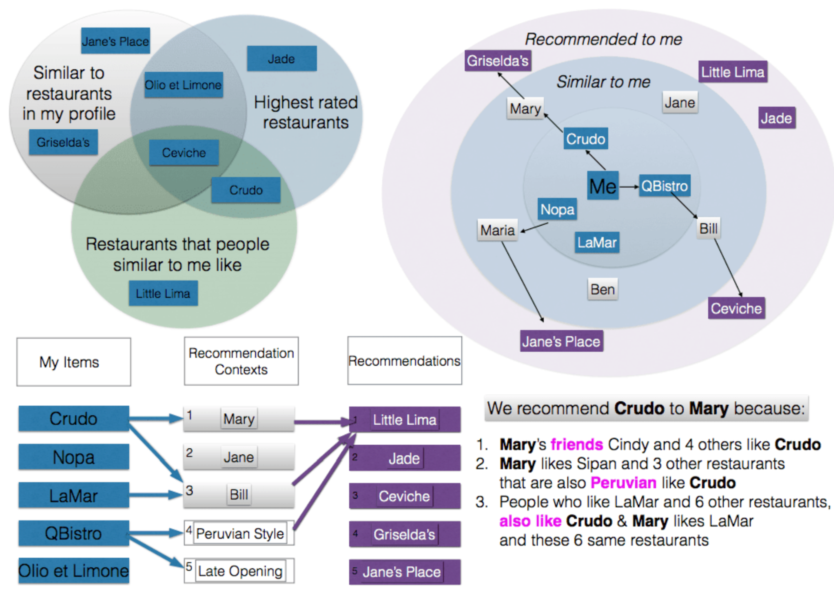


Figure 2.14: A subset of visualizations presented in the user study of hybrid explanations in Kouki et al. [2017].

survey, the authors evaluated explanations for hybrid algorithms in a variety of text and visual, graph-based formats, as shown in Figure 2.14, that are either novel designs or derived from existing hybrid recommender systems.

Ben-Elazar and Koenigstein [2014] introduced a hybrid explanation framework based on item tags. In particular, the authors presented a flexible explanations framework utilizing item tags to automatically generate personalized explanations in a natural language format. Given a specific user and a recommended item, the algorithm utilizes the user's personal information as well as global information (e.g., item similarities, metadata) in order to rank item tags based on their "explanatory power".

3

Explainable Recommendation Models

Recently, popular explainable recommendation approaches are mostly model-based methods, i.e., the recommendation is provided by a model such as matrix/tensor factorization, factorization machines, topic modeling, and deep recommendation models, an meanwhile, the models and recommendation results are explainable. In this section, we provide a survey of model-based explainable recommendation methods.

3.1 Overview of Machine Learning for Recommendation

Model-based explainable recommendation is closely related to machine learning methods for recommendation research. We first provide a brief overview of machine learning for recommendation in this section.

The most classical model-based approach to recommendation could be Latent Factor Models (LFM) based on Matrix Factorization (MF) techniques Koren et al. [2009], which attempts to learn latent factors to predict the missing ratings in a user-item rating matrix. Representative matrix factorization methods include Singular Value Decomposition (SVD) (Koren et al. [2009], Koren [2008], Srebro and Jaakkola [2003]), Non-negative Matrix Factorization (NMF) (Lee and

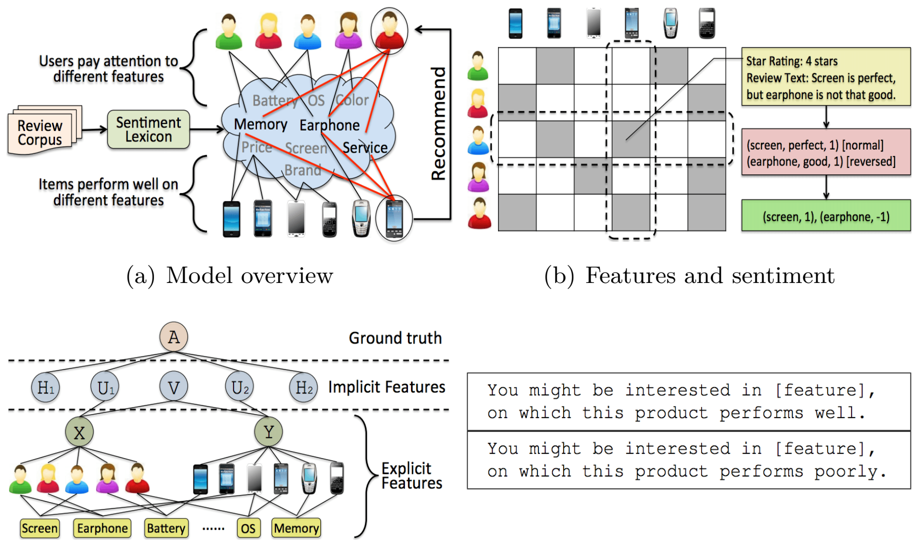
Seung [1999, 2001]), Max-Margin Matrix Factorization (MMMF) (Srebro et al. [2005], Rennie and Srebro [2005]), Probabilistic Matrix Factorization (PMF) (Mnih and Salakhutdinov [2008], Salakhutdinov and Mnih [2008]), Localized Matrix Factorization (LMF) (Zhang et al. [2013b,a, 2014c]), etc. Matrix factorization methods are also commonly referred to as point-wise prediction methods, and they are frequently used for prediction with explicit feedbacks such as numerical star ratings in e-commerce or movie review website, etc.

To learn for the rankings of items with implicit feedbacks, pair-wise learning to rank methods are also frequently used for recommendation. For example, Rendle et al. [2009] proposed Bayesian Personalized Ranking (BPR) to learn the relative ranking of purchased items (positive item) against unpurchased items (negative items). Rendle and Schmidt-Thieme [2010] further extended the idea to tensor factorization to model pairwise interactions. Except for pair-wise learning to rank methods, Shi et al. [2010] adopted list-wise learning to rank with matrix factorization for collaborative filtering.

More recently, deep learning and representation learning approaches have gained much attention in recommendation research. For example, Wang et al. [2015] proposed collaborative deep learning for recommendation systems, He et al. [2017] proposed neural collaborative filtering for recommendation, Zhang et al. [2017] proposed joint representation learning for recommendation. Besides, researchers have also investigated various types of deep neural networks for recommendation, such as convolutional neural networks (Zheng et al. [2017]), recurrent neural network and its variations (LSTM, GRU, etc.) (Hidasi et al. [2015], Donkers et al. [2017], Devooght and Bersini [2017]), auto-encoders (Wu et al. [2016]), and memory networks (Chen et al. [2018b]), etc. A lot of the deep methods have also been used for explainable recommendation, which will be introduced in the following subsections.

3.2 Matrix Factorization for Explainable Recommendation

In this section, we introduce how matrix/tensor-factorization and factorization machines are used for explainable recommendation.



(a) Model overview

(b) Features and sentiment

(c) Multi-matrix factorization

(d) Recommendation explanations

Figure 3.1: Overview of the Explicit Factor Model. (a) The basic idea is to recommend products that performs well on the features that a user cares about. (b) Each review (shaded block) is transformed to a set of product features accompanied with the sentiment that the user expressed on the feature. (c) Users' different attention on features is constructed as the user-feature attention matrix X, item qualities on features are constructed as the item-quality matrix Y, these two matrices are collaborated to predict the rating matrix A. (d) The explicit product features can be used to generate personalized explanations.

A lot of explainable recommendation models have been proposed based on matrix factorization methods. One problem of matrix factorization methods – or more generally, latent factor models – is that the user/item embedding dimensions are latent. Usually, we assume that the user and item representation vectors are embedded in a low-dimensional space where each dimension represents a particular factor that affects user decisions, but we do not explicitly know the exact meaning of each factor, which makes the predictions or recommendations provided by latent factor models difficult to be explained.

To alleviate the problem, Zhang et al. [2014a] proposed Explicit Factor Models (EFM), where the basic idea is to recommend prod-

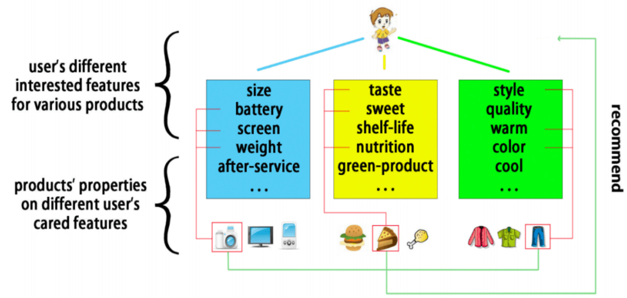
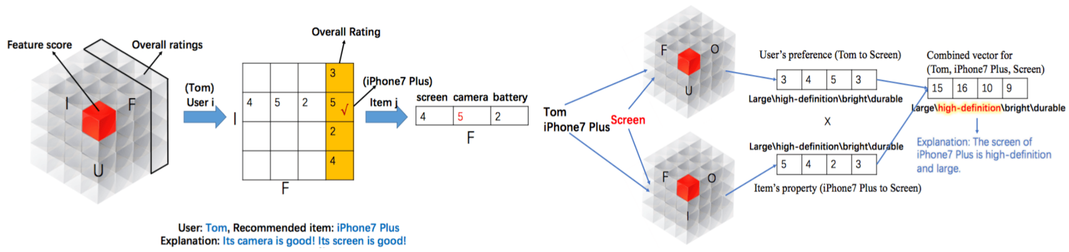


Figure 3.2: Learning to rank features for explainable recommendation over multiple categories based on tensor factorization.

ucts that performs well on the features that a user cares about, as shown in Figure 3.1. Specifically, the proposed approach extracts explicit product features from textual user reviews, and align each latent dimension in matrix factorization with a particular explicit feature, so that the factorization/prediction procedure can be trackable to provide explicit explanations for the recommendations. The proposed approach can provide personalized explanations accompanying the recommendations leveraging the explicit features, e.g., "*The product is recommended because you are interested in a particular feature, and this product performs well on the feature*". The model can even provide disrecommendations by telling the user that "*The product does not perform very well on a feature that you care about*", which can help to improve the trustworthiness of recommendation systems. Because user preferences on item features are dynamic and may change over time, Zhang et al. [2015b] extended the idea by modeling the features that a user cares about in a dynamic manner on daily resolution.

Chen et al. [2016] further extended the EFM model to tensor factorization. In particular, the authors extracted product features from textual reviews and constructed the user-item-feature cube. Based on this cube, the authored conducted pair-wise learning to rank to predict user preferences on features and items, and to provide personalized recommendations based on these predictions. The model was further extended to consider multiple categories of products simultaneously, which can help to alleviate the data sparsity problem in recommenda-



(a) (b)

Figure 3.3: (a) A user-item-feature tensor is used to predict user/item preference on features and the overall ratings. The last dimension of the feature slice is the overall rating matrix. (b) User-feature-opinion tensor and item-feature-opinion tensor. Feature-level text-based explanation is generated by integrating these two tensors for a given tuple of user, item and feature. Courtesy image from Wang et al. [2018b].

tion systems, as shown in Figure 3.2.

Wang et al. [2018b] further generalized previous MF-based explainable recommendation models by multi-task learning with tensor factorization. In particular, two companion learning tasks of "user preference modeling for recommendation" and "opinionated content modeling for explanation" are integrated via a joint tensor factorization solution for explainable recommendation, as shown in Figure 3.3. As a result, the algorithm predicts not only a user's preference over a list of items, i.e., recommendation, but also how the user would appreciate a particular item at the feature level, i.e., opinionated textual explanation.

Cheng et al. [2018] proposed aspect-aware latent factor model for rating prediction with ratings and reviews. Specially, the authors first apply an aspect-aware topic model on the review text to model user preferences and item features from different aspects, and then estimate the aspect importance of a user towards an item. The aspect importance is then integrated into an aspect-aware latent factor model to learn user's and item's latent factors based on ratings. In this way, a user's preference on an item is decomposed into the user's preference on different aspects and the importance of those aspects, and the recommendation can be explained by the target user's preference on each of the aspects.

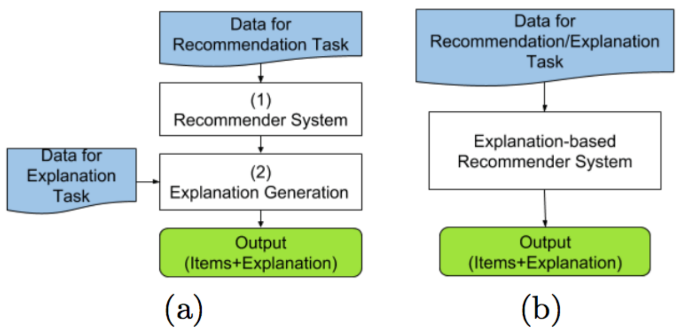


Figure 3.4: (a) Explainable recommendation with external data. (b) Explainable recommendation without external data support. Courtesy image from Abdollahi and Nasraoui [2016b].

The features themselves extracted from reviews can be recommended to users as a type of explanation. Bauman et al. [2017] proposed the Sentiment Utility Logistic Model (SULM), which extracts features (i.e., aspects) and the user sentiment on these features. The features and sentiments are integrated into a matrix factorization model to fit the unknown sentiments and ratings, which are finally used to generate recommendations. The proposed method not only provides recommended items to users, but also provides the recommended features for an item, and the features serve as the explanations for a recommendation. Furthermore, the method can recommend restaurants together with those most important aspects over which the user has control and can potentially select them, such as the time to go to a restaurant, e.g. lunch vs. dinner, and what to order there, e.g., seafood. Qiu et al. [2016] and Hou et al. [2018] also investigated aspect-based latent factor models by integrating ratings and reviews for recommendation.

Researchers have also investigated model-based approaches for user- and/or item-based explanation, which can provide explainable recommendation based only on the user-item rating matrix (see Figure 3.4). Specifically, Abdollahi and Nasraoui [2016b, 2017] described Explainable Matrix Factorization (EMF) for explainable recommendation. In this model, the authors considered neighborhood style explanations, where a recommended item is to be explained as “a lot of users similar to you purchased this item". To achieve this goal, the authors added an "explainability regularizer" into objective function of matrix factorization, and the explainability regularizer will force the a user latent vector and an item latent vector to be close to each other if a lot of the user's neighbors also purchased the item. In this way, the model will naturally select those commonly purchased items from a user's neighbors as recommendation, and at the same time maintain high rating prediction accuracy.

Abdollahi and Nasraoui [2016a] also applied the idea to restricted Boltzmann machines, and proposed explainable restricted Boltzmann machines for collaborative filtering and recommendation. In this paper, the authors defined "explainability scores" for each user-item pair based on the percentage of ratings that the user's neighbors rated on the item. This explainability score is integrated into restricted Boltzmann machine by adding an additional visible layer to define a user-item probability conditioned on the explainability scores. Similar to the above explainable matrix factorization method, this approach also provides user-based neighborhood style explanations.

3.3 Topic Modeling for Explainable Recommendation

Based on available text information – especially the widely available textual reviews in e-commerce – topic modeling approach has also been widely adopted for explanations in recommender systems. In these approaches, users can usually be provided with intuitive explanations in the form of topical word clouds (McAuley and Leskovec [2013], Wu and Ester [2015], Zhao et al. [2015]). In this section, we review the related work that can be categories into this approach.

McAuley and Leskovec [2013] proposed to understand the hidden factors in latent factor models based on the hidden topics extracted from textual reviews. To achieve this goal, the authors proposed the Hidden Factor and Topic (HFT) model, which bridges latent factor models and Latent Dirichlet Allocation (LDA) by linking each dimension of the item (or user) latent vector to each dimension of the topic distribution in LDA using a softmax function. By considering review

Beer (Beeradvocate)

pale ales	lambics	dark beers	spices	wheat beer
ipa	funk	chocolate	pumpkin	wheat
pine	brett	coffee	nutmeg	yellow
grapefruit	saison	black	corn	straw
citrus	vinegar	dark	cinnamon	pilsner
ipas	raspberry	roasted	pie	summer
piney	lambic	stout	cheap	pale
citrusy	barnyard	bourbon	bud	lager
floral	funky	tan	water	banana
hoppy	tart	porter	macro	coriander
dipa	raspberries	vanilla	adjunct	pils

Musical instruments (Amazon)

drums	strings	wind	microphones	software
cartridge	guitar	reeds	mic	software
sticks	violin	harmonica	microphone	interface
strings	strap	cream	stand	midi
snare	neck	reed	mics	windows
stylus	capo	harp	wireless	drivers
cymbals	tune	fog	microphones	inputs
mute	guitars	mouthpiece	condenser	usb
heads	picks	bruce	battery	computer
these	bridge	harmonicas	filter	mp3
daddario	tuner	harps	stands	program

Yelp Phoenix

theaters	spas	mexican	vietnamese	snacks	italian	medical	donuts	coffee	seafood
theater	massage	mexican	pho	cupcakes	pizza	dr	donuts	coffee	sushi
movie	spa	salsa	vietnamese	cupcake	crust	stadium	donut	starbucks	dish
harkins	yoga	tacos	yogurt	hotel	pizzas	dentist	museum	books	restaurant
theaters	classes	chicken	brisket	resort	italian	doctor	target	latte	rolls
theatre	pedicure	burrito	beer	rooms	bianco	insurance	subs	bowling	server
movies	trail	beans	peaks	dog	pizzeria	doctors	sub	lux	shrimp
dance	studio	taco	mojo	dogs	wings	dental	dunkin	library	dishes
popcorn	gym	burger	shoes	frosting	pasta	appointment	frys	espresso	menu
tickets	hike	carne	froyo	bagel	mozzarella	exam	tour	stores	waiter
flight	nails	food	zoo	bagels	pepperoni	prescription	bike	gelato	crab

Figure 3.5: Example hidden topics extracted textual reviews on different datasets, each topic is aligned to a latent factor in matrix factorization to make the latent dimensions understandable. Courtesy image from McAuley and Leskovec [2013].

information for recommendation, the proposed method improves rating prediction accuracy. Besides, by projecting each user latent vector onto the learned topics from LDA (Figure 3.5), it helps us to understand why a user made a particular rating on a target item by detecting the important topics that a user cares about.

Following this idea, Tan et al. [2016] proposed to model item recommendability and user preference in a unified semantic space based on review information. In the modeling process, an item is embedded as a topical recommendability distribution, and the topics in those reviews of higher ratings are repeated to enhance the importances. Similarly, a user is embedded in the same space, which is determined by his/her historical rating behaviors. The recommendability and preference distributions are, at last, integrated into the latent factorization framework to fit the ground truth, and the explanations for the recommended items are derived based on the learned latent topics.

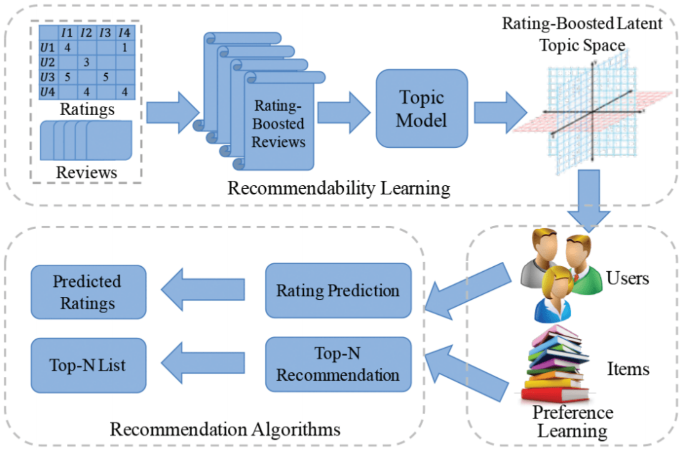
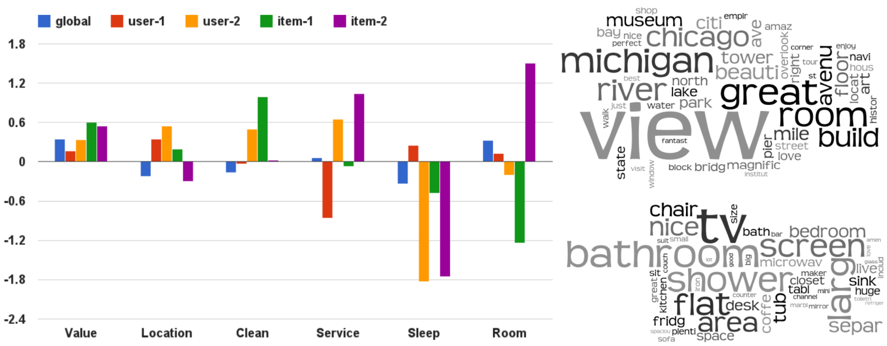


Figure 3.6: Recommendation framework for understanding users and items with ratings and reviews in "Rating-Boosted Latent Topics" model. Users and items are embedded as preference and recommendability vectors in the same space, which are later used by latent factorization model for both rating prediction and top-n recommendation. Courtesy image from Tan et al. [2016].

In a more general sense, researchers have also investigated using probabilistic graphic models beyond LDA for explainable recommendation. Wu and Ester [2015] studied the problem of estimating personalized sentiment polarities on different aspects of the items. In particular, the authors proposed the FLAME model (Factorized Latent Aspect ModEl), which combines the advantages of collaborative filtering and aspect based opinion mining. It learns users' personalized preferences on different aspects from their past reviews, and predicts users' aspect ratings on new items by collective intelligence, as shown in Figure 3.7(a). The proposed method showed improved performance for hotel recommendation based on TripAdvisor, besides, for the recommended hotels, it can provide the aspects of the hotel as a word cloud for explanation, as shown in Figure 3.7(b), where the size of each aspect is proportional to the sentiment on the aspect.

Zhao et al. [2015] designed a probabilistic graphical model to integrate sentiment, aspect, and region information in a unified framework for improving the performance as well as the explainability of point-of-interest (POI) recommendation. The explanations are determined by



(a) (b)

Figure 3.7: (a) The FLAME model learns users' different sentiments on different aspects of the items. (b) Displaying the aspects proportional to its sentiment in a word cloud for explanation. Courtesy image from Wu and Ester [2015].

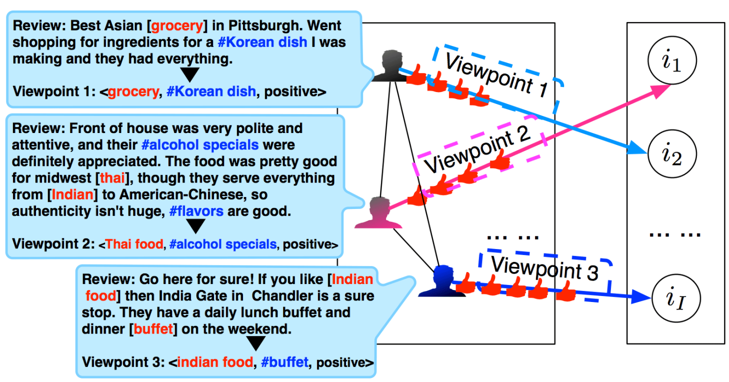
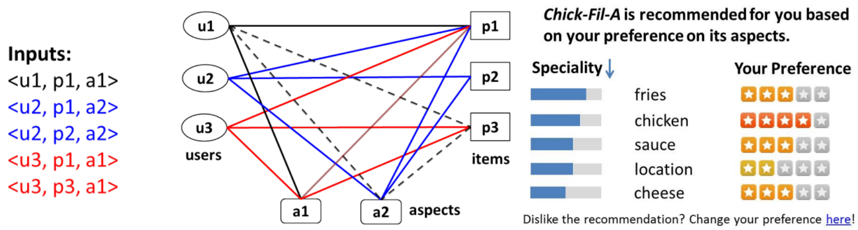


Figure 3.8: An example of trusted social relations, user reviews and ratings in a recommender system. Black arrows connect users with trusted social relations. "ThumpUp" symbols reflect the ratings of items. Concepts and topics have been highlighted in red and blue, respectively. Three viewpoints are represented in three different colors. A viewpoint is a mixture over a concept, a topic, and a sentiment. Courtesy image from Ren et al. [2017].

the learned topical-aspect preferences.

Ren et al. [2017] introduced topic modeling approach to explainable recommendation for social recommendation. Specifically, the authors proposed social collaborative viewpoint regression (sCVR), where



(a) (b)

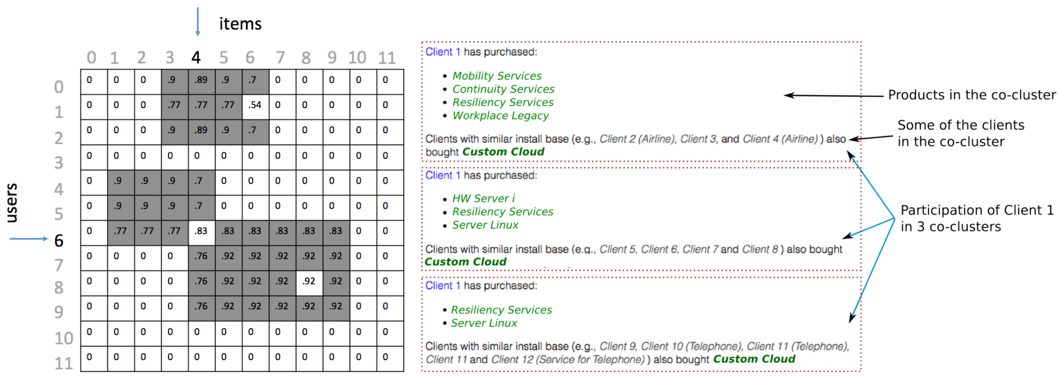
Figure 3.9: (a) An example tripartite structure of the given inputs used by the TriRank algorithm. (b) A mock user interface for showing the explanation for recommending *Chick-Fil-A* to a user. Courtesy image from He et al. [2015].

a viewpoint is defined as a tuple of concept, topic, and a sentiment label from both user reviews and trusted social relations, as shown in Figure 3.8, and the viewpoints are used as explanations. A probabilistic graphical model based on the viewpoints were proposed to improve the rating prediction by leveraging user reviews and trusted social relations simultaneously. Similar to previous work, the explanations are generated based on the discovered user favorite topics embedded in the viewpoints. For example, to explain a recommended "Steakhouse", the model provides its top rated topic (potato, appetizer, good, place, pork, rib, bread, rib-eye, filet, beef) for explanation.

3.4 Graph-based Models for Explainable Recommendation

Many user-user or user-item relationships can be represented as graphs, especially in social network related application scenarios. In this section, we introduce how explainable recommendation can be generated based on graph learning approaches such as graph-based propagation and graph clustering.

He et al. [2015] introduced a tripartite graph to model the user-item-aspect ternary relation for top-N recommendation, as shown in Figure 3.9, where an aspect is an item feature extracted from user reviews. The authors proposed TriRank, a generic algorithm for ranking the vertices of tripartite graph by regularizing the smoothness and fit-



(a) (b)

Figure 3.10: (a) Example of overlapping user-item co-clusters identified by the OCuLaR algorithm in Heckel et al. [2017]. Dark squares correspond to positive examples, and the white squares within the clusters correspond to recommendations. (b) Based on the clustering results, the algorithm can provide user-based and item-based explanations, for example, customers with similar install base also purchased the recommended item.

ting constraints. The graph-based ranking algorithm is used for review-aware recommendation, where the ranking constraints directly model the collaborative and aspect filtering, and also personalization. In this paper, the explanations are attributed to the top-ranked aspects matching the target user and the recommended item.

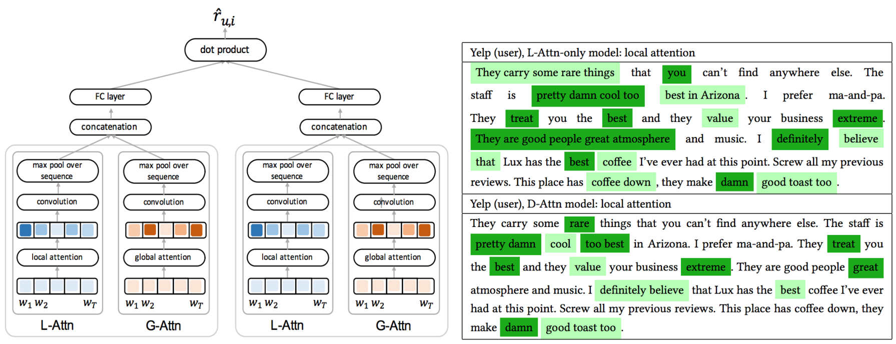
Without using external information such as aspects, Heckel et al. [2017] proposed to conduct over-lapping co-clustering based on user-item bipartite graph for explainable recommendation, and in each co-cluster the users have similar interests and the items are of similar properties, as shown in Figure 3.10. The explanations are generated by leveraging user collaborative information, for example, in the form of "Item A is recommended to Client X with confidence α because: Client X has purchased Item B, C and D, and clients with similar purchase history (e.g., Clients Y and Z) also bought Item A". If a user-item pair falls into multiple co-clusters, we can thus generate multiple user-based and item-based explanations based on each of the co-cluster.

As a special case of graph, tree structures can also be leveraged for explainable recommendation. Wang et al. [2018c] proposed the tree-enhanced embedding model for explainable recommendation, which attempts to combine the generalization ability of embedding-based models with the explainability of tree-based models. In this model, the authors first employed a tree-based model to learn explicit decision rules based on cross features from rich side information, and then designed an embedding model that incorporates explicit cross features and generalize to unseen cross features on users and items. For explanation, an attention network is used to make the recommendation process transparent and explainable.

3.5 Deep Learning for Explainable Recommendation

Recently, deep learning and representation learning have attracted much attention in the recommendation research community, and they have also be widely applied for explainable recommendations. By now, the related explainable recommendation models cover a wide range of deep learning techniques, including CNN (Seo et al. [2017], Tang and Wang [2018]), RNN/LSTM (Donkers et al. [2017]), attention mechanism (Chen et al. [2018c]), memory networks (Chen et al. [2018b]), etc., and they have also been applied to different recommendation tasks regarding explainability, such as top-n recommendation, sequential recommendation, etc, and based on LSTM models, the system can even automatically generate explanation sentences instead of using explanation templates (Seo et al. [2017]). In this section, we will review deep learning approaches to explainable recommendation, and analyze their advantages and shortcomings.

Seo et al. [2017] proposed to model user preferences and item properties using convolutional neural networks (CNNs) upon textual reviews with dual local and global attention, as shown in Figure 3.11. When predicting the user-item rating, the model selectively chooses the words from reviews with different attention weights, and with the learned attention weights, the model can indicate which part of a review is more important for the current prediction. Besides, the model can also highlight the relevant words in reviews as explanation to help



(a) (b)

Figure 3.11: (a) Architecture of dual-attention model to extract latent representations of users and items. A user document and an item document are fed into (Left) the user network and (Right) the item network, respectively. (b) The model generates attention scores for each review and highlight the words with high attention scores as explanations. Courtesy image from Seo et al. [2017].

users understand the recommendations.

Similarly, Wu et al. [2017] combined the user-item interaction and review information in a unified framework. The user reviews are attentively summarized as content features integrated with the user/item embedding to predict the final ratings. Lu et al. [2018] presented a deep learning recommendation model which co-learns user and item information from ratings and customer reviews by optimizing matrix factorization and an attention-based GRU network. In both of the models, the attention weights over review words are leveraged to explain the predictions and recommendations.

Different from highlighting the words in existing reviews as explanation, Costa et al. [2017] proposed a method for automatically generating natural language explanations based on character-level RNN structure, where the review rating is concatenated into the input component as an auxiliary information, so that the model can generate reviews according to the expected rating (sentiment). Different from many explainable recommendation models where the explanation is generated based on a predefined template, the learned model can automatically generate explanation in a natural language manner, and by choosing different

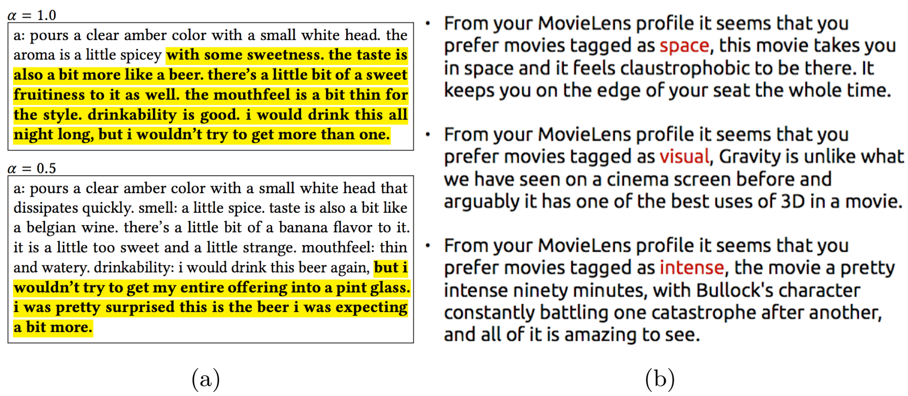


Figure 3.12: (a) The automatically generated textual reviews (explanations) based on natural language generation. Setting different model parameters will generate different explanations. Courtesy image from Costa et al. [2017]. (b) Example natural language explanations for the movie "Gravity". Depending on the model of a user's interest, the system selects one of the three crowd-sourced explanations for the user. Courtesy image from Chang et al. [2016].

parameters, the model can generate different explanations, as shown in Figure 3.12(a), which can be attractive for web users.

Also for generating natural language explanations, Chang et al. [2016] attempted another approach based on human users and crowd sourcing. Inspired by how people explain word-of-mouth recommendations, the authors designed a process combining crowdsourcing and computation, that generates personalized natural language explanations. They modeled key topical aspects of movies, asked crowd workers to write explanations based on quotes from online movie reviews, and personalized the explanations presented to users based on their rating history, as shown in Figure 3.12(b). Controlled experiments with 220 MovieLens users were conducted to evaluate the efficiency, effectiveness, trust, and satisfaction of the personalized natural language explanations compared with personalized tag based explanations.

Chen et al. [2018a] leveraged attention mechanisms upon the user/item reviews for rating prediction. Specifically, the authors introduced an attention mechanism to explore the usefulness of reviews, and proposed a neural attentional regression model with review-level expla-

Image	True Review	Re-VECF	Re-CF	NRT
	It's an excellent poplin solid color long ***sleeved*** shirt	Much like the ***sleeve***	Not bad for the price	Very good choice
	Very ***good-looking*** sturdy belt with a good ribbed weave and strong ***buckle***	I like this ***good looking buckle***	Great for the price	Makes a great price

Figure 3.13: The visually explainable recommendation model can also generate natural language explanations for the highlighted image regions. In the example, the generated review can properly describe the highlighted image regions. Courtesy image from Chen et al. [2018c].

nations for recommendation. The proposed approach not only predicts the ratings, but also learns the usefulness of each review simultaneously. Therefore, the predicted highly-useful reviews are adopted as review-level explanations to help users understand the recommendations for better and faster decisions.

Chen et al. [2018c] proposed Visually Explainable Recommendation based on joint neural modeling of visual images and textual reviews, which highlights the image regions that a user may be interested in as explanations, as shown in Figure 2.10. By jointly modeling images and reviews, the proposed model can also generate natural language explanations, more particularly, the generated natural language explanations are supposed to describe the highlighted regions, as shown in Figure 3.13, so that the users can understand why the particular regions of a recommended product image are highlighted and why the particular item is recommended.

Recently, natural language generation-based explainable recommendation has even been applied to commercial e-commerce systems. For example, Alibaba e-commerce recommendation system generates explanations for the recommended items based on data-to-sequence natural language generation, as shown in Figure 3.14, which generates easily readable natural language recommendation explanations, as shown in Figure 3.15.

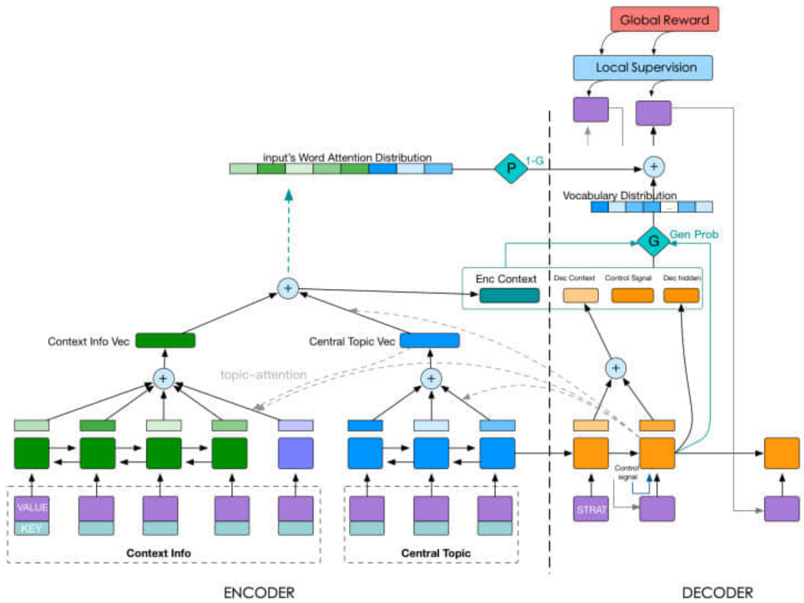
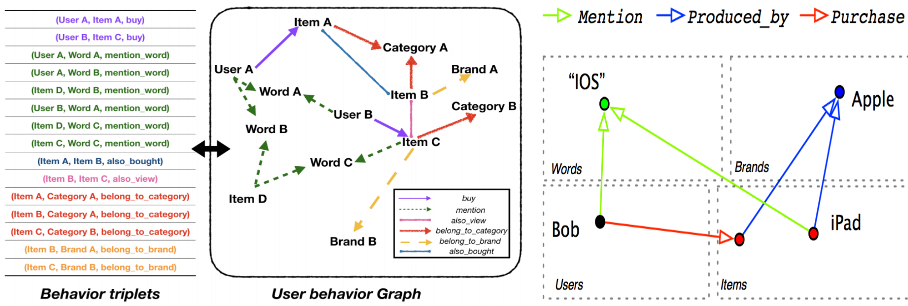


Figure 3.14: A data-to-sequence model for explanation generation in Alibaba e-commerce recommendation systems, where the encoder encodes the input contextual and topical data based on RNN and embeddings, and the decoder generates explanation sentences based on attention mechanism. Courtesy image from `https://zhuanlan.zhihu.com/p/33956907`.

Product	Popularity Trend	Recommendation Explanation
	街头风	今年大热的街头风的连帽卫衣，采用了纯棉的面料，手感柔软细腻，上身亲肤透气，经典的袋鼠兜口袋设计，方便实用
Translation	Street style	Street style hoodie which is very popular this year, pure cotton fabric, feels soft and smooth, skin-friendly and breathable, classical kangaroo-style pocket, convenient and easy to use.

Figure 3.15: An example of the natural language recommendation explanations provided in Alibaba e-commerce recommendation system. The original explanations are in Chinese and the corresponding English translation is provided accordingly. The highlighted sentence is the sentence that contains the pre-specified keyword (in this case, style) to generate the explanation.



(a) (b)

Figure 3.16: (a) The user-item knowledge graph constructed for Amazon product domain. In the left is a set of triplets of user behaviors and item properties, and in the right is the corresponding graph structure. The knowledge graph contains various different types of relations such as purchase, mention, also bought, also view, category, brand, etc. (b) When a user is recommended with an item based on knowledge-base embeddings, recommendation can be explained by *explanation paths* from the user to the item in the graph, in this example, there are two paths from *Bob* to the recommendation *iPad*.

3.6 Knowledge-base Embedding for Explainable Recommendation

Knowledge base contain rich information of the users and items, which can help to generate intuitive and more tailored explanations for the recommended items. Recently, there has been some work on leveraging knowledge bases for explainable recommendation. In a broader sense, after years of prospering of emperism

Catherine et al. [2017] illustrated how explanations can be generated by leveraging external knowledge in the form of knowledge graphs. The proposed method jointly ranks items and knowledge graph entities using a Personalized PageRank procedure to produce recommendations together with their explanations. The paper works on movie recommendation scenario, and it produces a ranked list of entities as explanations by jointly ranking them with the corresponding movies.

Different from Catherine et al. [2017] that adopts rules and programming on knowledge graph for explainable recommendation, Ai et al. [2018] proposed to adopt knowledge graph embeddings for ex-

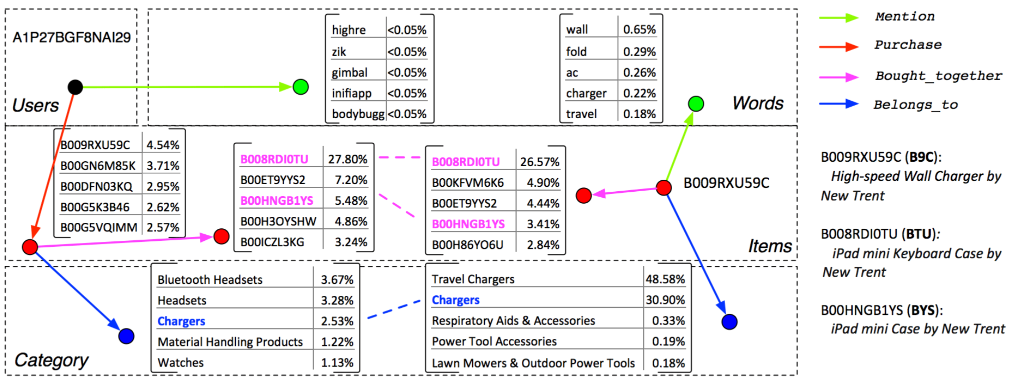


Figure 3.17: Example explanation paths between the user `A1P27BGF8NAI29` and the item `B009RXU59C` in Cell Phones. There are three explanation paths in the example, two of them are based on the **`bought together`** relationship through products `B008RDI0TU` and `B00HNGB1YS`, respectively, and the third explanation is based on the `category` relationship.

plainable recommendation, as shown in Figure 3.16. In this research, the authors constructed user-item knowledge graph, which contains various user, item, and entity relations, such as user purchasing item, item belonging to category, and item are co-purchased together, etc. Knowledge base embeddings are learned over the graph to obtain the embeddings of each user, item, entity, and relation, and recommendations are provided for a user by finding the most similar items under the 'purchase' relation. Besides, explanations can be provided by finding an "explanation path" from the user to the recommended item through the knowledge graph (Figure 3.16(b)), and each path can be described as a natural language explanation. More specially, the authors proposed an entity soft-matching algorithm to extract explanation paths with knowledge-graph embeddings, as shown by the example in Figure 3.17. In this example, there are three explanation paths from the target user to the recommended item, and thus three explanations can be constructed by describing the paths with simple natural language templates, for example, "Product `B009RXU59C` is recommended because the user often purchases items that are bought together with `B008RDI0TU`, and `B009RXU59C` is also frequently bought together with `B008RDI0TU`",

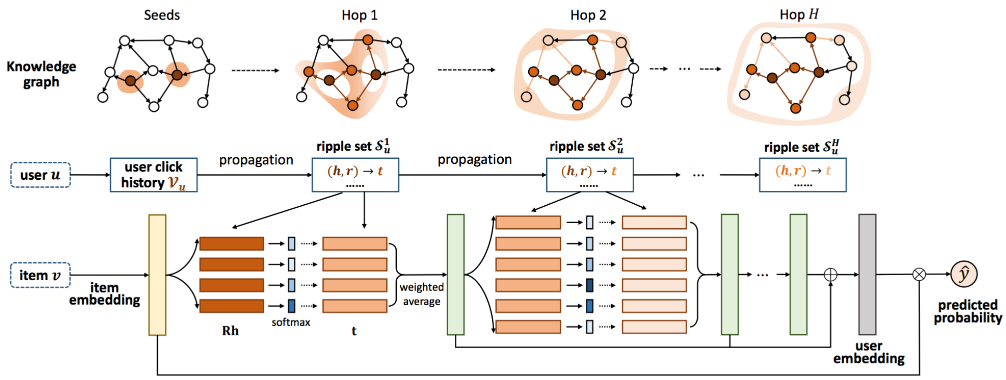


Figure 3.18: Propagating user preferences on the knowledge graph through ripple network for recommendation. Courtesy image from Wang et al. [2018a].

and similar for the other two explanations.

Wang et al. [2018a] proposed Ripple Network, an end-to-end framework to incorporate the knowledge graph into recommender systems, as shown in Figure 3.18. Similar to actual ripples propagating on the surface of water, Ripple Network stimulates the propagation of user preferences over the set of knowledge entities by automatically and iteratively extending a user's potential interests along links in the knowledge graph. The multiple "ripples" activated by a user's historically clicked items are thus superposed to form the preference distribution of the user with respect to a candidate item, which could be used for predicting the final clicking probability. Explanations can also be provided by finding a path from the user and the recommended item over the knowledge graph, and the paths can also be specified to certain number of hops from the user to the recommended item.

3.7 Data Mining for Explainable Recommendation

Data mining approaches to recommendation are important for recommendation research, and data mining approaches usually have particular advantages for explainable recommendation, because they are capable of generating very straight forward explanations that are easy to understand for users. The most frequently used data mining technique for explainable recommendation is associate rule mining (Agrawal et al. [1993], Agarwal et al. [1994]), and a very classical example is the "beer-diaper" recommendation originated from data mining research.

For example, Mobasher et al. [2001] leveraged association rule mining for efficient web page recommendation at large-scale; Cho et al. [2002] combined decision trees and association rule mining for a web-based shop recommender system; Smyth et al. [2005] adopted a-priori association rule mining to help calculate better item-item similarity, and applied association rule mining for conversational recommendation; Sandvig et al. [2007] studied the robustness of collaborative recommendation algorithms based on association rule mining; Zhang et al. [2015a] defined a sequence of user demands as a task in web browsing, and leveraged frequent pattern mining and association rule mining for task-based recommendation by analyzing user browsing logs. More comprehensively, Amatriain and Pujol [2015] provided a survey of data mining methods for recommendation, including association rules-based approaches.

In terms of explainable recommendation, Lin et al. [2000, 2002] investigated association rules for recommendation systems. In particular, the authors proposed a "personalized" association rule mining technique, which mines association rules for a specific target user, and the associations between users as well as associations between items are employed to make recommendations. Recommendation results generated by association rule mining are self-explainable, for example, "90% of articles liked by user A and user B are also liked by user C".

Davidson et al. [2010] introduced the YouTube Video Recommendation System. The authors considered the sessions of user watch activities on the site. For a given time period (usually 24 hours), the authors adopted association rule mining to count for each pair of videos (v_i, v_j) how often they were co-watched within sessions, which helps to calculate the relatedness score for each pair of videos. To compute personalized recommendations, the authors consider a seed set of videos for each user, which can include both videos that were watched by the user, as well as videos that were explicitly favorited, "liked", rated, or added to playlists. The related videos of these seed videos are taken

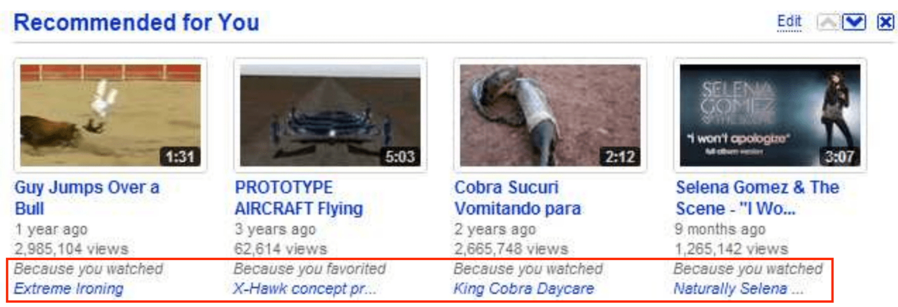


Figure 3.19: A screenshot of the recommendations module on the YouTube home page. The boxed explanations are generated based on association rule mining. Courtesy image from Davidson et al. [2010].

as candidate items for recommendation, and the seed video as well as the association rule that triggered the recommendation will be taken as explanations, as shown in Figure 3.19.

Lawlor et al. [2015] and Muhammad et al. [2015] leveraged opinion mining to generate explanations. Based on opinions mined from user-generated reviews, the explanation highlights how the pros and cons of a recommended item compares to alternative items. Muhammad et al. [2016a,b] further studied explanation-based ranking, which used the strength of an item's explanation as a ranking signal, where items with compelling explanations are preferred when generating the recommendation list.

3.8 Post Hoc Explanation

Very frequently in practice, the recommendation explanation is not generated from the recommendation model itself. Instead, we do not care about explainability when designing the recommendation model that generates the recommendations, but provide a post-hoc explanation for the items after they have been recommended by the algorithm.

For example, in many e-commerce applications the items are recommended based on very complex hybrid recommendation methods, but after an item is recommended, we provide some statistical information to the user as explanation, for example, "70% of your friends also bought this item". Usually, we pre-define a candidate set of possible explanations based on data mining techniques such as frequently item set mining and association rule mining, and decide which explanation(s) to display based on a post-hoc strategy such as maximum confidence.

4

Evaluation of Explainable Recommendation

In this chapter, we provide a review for the evaluation methods of explainable recommendation. Though sometimes it is reasonable to sacrifice some recommendation performance for better explainability, it is desirable if the explainable recommendation model can achieve comparable or even better performance than conventional "non-explainable" methods. As a result, explainable recommendation algorithms primarily evaluate the recommendation performance in terms of rating prediction or top-n recommendation, and it is also encouraged to evaluate the explanation performance in terms of persuasiveness and effectiveness, etc. In this chapter, we introduce commonly used evaluation measures and protocols for both perspectives.

4.1 Evaluation of Recommendation Performance

The evaluation of recommendation performance for explainable recommendation models is similar to that evaluating other personalized recommendation models. We can do both offline evaluation based on training/testing dataset and online evaluation by analyzing the behaviors of real users.

4.1.1 Offline Evaluation

Usually, we can evaluate based on two tasks, including rating prediction and top-n recommendation.

For rating prediction, a model is trained based on training dataset and is later used to predict the user-item ratings in the test dataset. Normally, we can use mean absolute error (MAE) and root mean square error (RMSE) to evaluate the performance of rating prediction.

$$MAE = \frac{1}{|\mathcal{T}|} \sum_{(u,i)\in\mathcal{T}} |r_{ui} - \hat{r}_{ui}|, \quad RMSE = \sqrt{\frac{1}{|\mathcal{T}|} \sum_{(u,i)\in\mathcal{T}} (r_{ui} - \hat{r}_{ui})^2}$$

where $\mathcal{T}$ is the test dataset, r_{ui} is the rating that user u made on item i in the testing set, and $\hat{r}_{ui}$ is the predicted rating. Usually, RMSE is more sensitive to large errors in predictions.

For top-n recommendation, a lot of ranking measures can be used for evaluation. The most frequently used measures could be precision, recall, F_1-measure, normalized discounted cumulative gain (NDCG), etc. Suppose the length of recommendation list is n, the set of all users is $\mathcal{U}$, for each user $u \in \mathcal{U}$, the set of recommended items for u is $\mathcal{S}_u$, and the set of truly purchased items for user u in testing dataset is $\mathcal{T}_u$, then the precision, recall, and F_1-measure can be defined as:

$$P@n = \frac{1}{|\mathcal{U}|} \sum_{u\in\mathcal{U}} \frac{|\mathcal{S}_u \cap \mathcal{T}_u|}{\mathcal{S}_u}, \; R@n = \frac{1}{|\mathcal{U}|} \sum_{u\in\mathcal{U}} \frac{|\mathcal{S}_u \cap \mathcal{T}_u|}{\mathcal{T}_u}, \; F_1@n = \frac{2 \cdot P \cdot R}{P + R}$$

When we take the position of the correctly recommended items into consideration, NDCG is frequently used among the many position-sensitive measures:

$$DCG@n = \frac{1}{|\mathcal{U}|} \sum_{u\in\mathcal{U}} \sum_{i=1}^{n} \frac{2^{rel_i} - 1}{\log_2(i+1)}, \quad NDCG@n = \frac{1}{|\mathcal{U}|} \sum_{u\in\mathcal{U}} \frac{DCG@n}{IDCG@n}$$

where $rel_i = 1$ if the i-th element is a positive item (i.e., purchased by the corresponding user), and 0 otherwise, and IDCG is the idealized discounted cumulative gain where all the positive items are ranked at the top positions in the recommendation list.

A lot of other measures are available for the evaluation of both rating prediction and top-n recommendation tasks in the offline, for example, mean average precision (MAP), Mean Reciprocal Rank (MRR), Hit Ratio (HR), Area Under the Curve (AUC), etc. Readers can refer to Shani and Gunawardana [2011] and Karypis [2001] for a comprehensive survey of offline evaluation methods for recommendation systems.

4.1.2 Online Evaluation

When possible, we can deploy the recommendation algorithm online and evaluate by doing A/B test with real users to study the recommendation performance of the algorithm. Usually, the experiment involves two different groups of users, where users in group A (experimental group) receive recommendations based on the designed algorithm to be evaluated, and users in group B (comparison group) receive recommendations from a baseline algorithm that is used for comparison. By comparing the behavior of users in two groups, we verify the recommendation performance of the target algorithm.

Frequently used online evaluation measures include click through rate (CTR), conversion rate (CR), and other business related measures such as average revenue. Click through rate is measured as the percentage of clicked items among the total number of recommended items by the algorithm, and conversion rate is the percentage of purchased items among the total number of clicked items.

Online evaluation is commonly used by commercial companies in practice where sufficient number of real users can be accessed and assigned for experiment. Because this survey focuses on explainable recommendation, readers are are suggested to refer Gunawardana and Shani [2009] and Beel et al. [2013] for a comprehensive analysis of online evaluation methods under the background of conventional "non-explainable" recommendation.

4.2 Evaluation of Recommendation Explanations

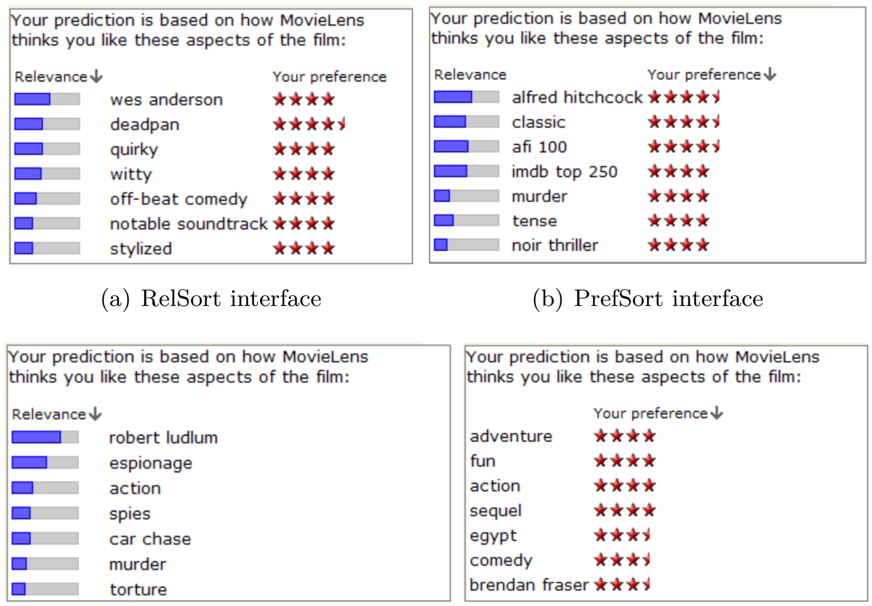
We introduce approaches to evaluate the recommendation explanations in explainable recommendation systems. Similarly, explanations can also be evaluated with both offline and online protocols, and researchers can use either or both of the two settings to evaluate the explanations. Usually, offline evaluation is easier to implement, while online evaluation and user studies depend on the availability of data and users in real-world systems, which are not always accessible to researchers, as a result, online evaluation is encouraged but not always required for explainable recommendation research.

4.2.1 Offline Evaluation

There are generally two approaches to evaluating the recommendation explanations. One is to evaluate the percentage of recommended items that can be explained by the explainable recommendation model, regardless of the quality of the explanations; and the second approach is to evaluate the quality of the explanations exactly. However, more offline evaluation measures and protocols are to be proposed for more comprehensive evaluation of recommendation explanations.

For the first approach, Abdollahi and Nasraoui [2017] adopted mean explainability precision (MEP) and mean explainability recall (MER) for evaluation. More specifically, explainability precision (EP) is defined as the proportion of explainable items in the top-n recommendation list relative to the number of recommended (top-n) items for each user, and explainability recall (ER) is the proportion of explainable items in the top-n recommendation list relative to the number of all explainable items for a given user. Finally, mean explainability precision (MEP) and mean explainability recall (MER) are EP and ER averaged across all testing users, respectively.

In the second approach, evaluating the quality of the explanations usually depends on the particular type of the explanation. One commonly used form of explanation is a piece of textual sentence, and in this case, offline evaluation can be conducted with text-based measures. For example, in many websites such as e-commerce we can consider the review that a user wrote for an item as the ground-truth explanation that the user purchased the item. If our recommendation explanation is automatically generated as a piece of free text, we can take frequently used text generation measures for evaluation, such as BLEU (bilingual evaluation understudy) score (Papineni et al. [2002]) and ROUGE (recall-oriented understudy for gisting evaluation) score (Lin [2004]). The quality of the free-text explanations can also be evaluated



(a) RelSort interface (b) PrefSort interface

(c) RelOnly interface (d) PrefOnly interface

Figure 4.1: Four explanation interfaces for online evaluation in Vig et al. [2009]. (a) RelSort: Shows relevance and preference, sorts tags by relevance. (b) PrefSort: Shows relevance and preference, sorts tags by preference. (c) RelOnly: Shows relevance only, sorts tags by relevance. (d) PrefOnly: Shows preference only, sorts tags by preference.

in terms of readability based on frequently used readability measures, such as Gunning Fog Index (Gunning [1952]), Flesch Reading Ease (Flesch [1948]), Flesch Kincaid Grade Level (Kincaid et al. [1975]), Automated Readability Index (Senter and Smith [1967]), and Smog Index (Mc Laughlin [1969]).

4.2.2 Online Evaluation

Another approach to evaluate explainable recommendation is through online experiments, also based on online measures such as conversion rate (CR) and click through rate (CTR), similar to online evaluation of recommendation performance.

There could be several different perspectives to consider when eval-

Figure 4.2: Top-4 recommended items are presented by the browser at right hand side when the user is browsing an online product, and the feature-opinion word pair cloud is displayed to assist explanations, where the positive pairs are green and negative pairs are blue, and the size is proportional to the term frequency in the reviews of the recommended item. For example, the largest pair in the right figure means "PictureClarity-High". The explanations will be displayed only when the user hovers the mouse on a recommended item, so that the system knows that the explanation has indeed been examined by the user.

uating explanations online, including persuasiveness, effectiveness, efficiency, and satisfaction of the explanations. Due to the limited type of information that one can collect in online systems, it is usually easier to evaluate the persuasiveness of the explanations, i.e., to see if the explanations can help to make users take the recommendations.

Vig et al. [2009] conducted a within-subjects study for four explanation interfaces based on MovieLens website, where the four interfaces are RelSort, PrefSort, RelOnly, and PrefOnly, as shown in Figure 4.1. Subjects completed an online survey in which they evaluated each interfaces on how well it helped them (1) understand why an item was recommended (justification), (2) decide if they would like the recommended item (effectiveness), and (3) determine if the recommended item matched their mood (mood compatibility). Based on survey responses, the authors draw conclusions about the role of tag preference and tag relevance in promoting justification, effectiveness, and mood compatibility.

Zhang et al. [2014a] conducted online experiments focusing on how the explains affect users' acceptance of the recommendation (i.e. persuasiveness). The authors conducted A/B-tests based on a popular commercial web browser which has more than 100 million users with 26% monthly active users, and the experiments attempted to recommend relevant phones when a user is browsing mobile phones in an online shopping website, as shown in Figure 4.2. To evaluate the persuasiveness of the explanations, the authors designed three groups of users, including an experimental group that receives the testing explanations, a comparison group that receives the baseline ‘People also viewed’ explanations, and a control group that receives no explanation. Click through rate is adopted to compare the groups so as to evaluate the effect of providing personalized explanations to users.

Besides, the authors also calculated the percentage of recommendations that are added to cart by users to evaluate the conversation rate, and calculated the percentage of agreement to evaluate the effectiveness of the explanations, where a recommendation (or a dis-recommendation) is considered as an agreement if it was (or was not) added to cart, respectively.

4.2.3 Simulated Online Evaluation based on User Study

Online evaluation needs a deployed system with sufficiently many users, which usually requires extensive efforts or collaboration with commercial company. A comparably easier approach to simulate online evaluation is through user study based on volunteers or paid experiment subjects. The volunteers or paid subjects can either be hired by the researchers directly, or hired based on various online crowdsourcing platforms such as Amazon Mechanical Turk[1].

For example, Chen et al. [2018c] generated visual explanations by highlighting certain areas of a product image to users, and leveraged Amazon MTurk to hire freelancers to label the ground-truth areas of images for evaluation; Chen et al. [2018a] used the high-quality reviews as explanations for the recommended items, and made crowdsourcing evaluation via CrowdFlower[2] platform to generate usefulness annotations for reviews, so as to evaluate the usefulness of the explanations.

Except for large-scale online workers, we can also conduct user

[1] https://www.mturk.com/
[2] https://www.crowdflower.com

Figure 4.3: Three sampled explanation interfaces designed in Muhammad et al. [2016c] (there are a total of nine interfaces), based on which user study was conducted to evaluate their performance in terms of explainability. Courtesy image from Muhammad et al. [2016c].

study with relative small-scale volunteers, paid subjects, or manually labeling the explanations. For example, Wang and Benbasat [2007] adopted the user study based on subject surveys to investigate the trust and understandability of content-based explanations; Muhammad et al. [2016c] designed nine different types of explanations based on the pros and cons of item features (e.g., three sampled explanation interfaces are shown in Figure 4.3), and conducted user study to evaluate different explanations based on 181 participants who provided rating for the overall clarity and utility of the explanations; Ren et al. [2017] took a random sample of 100 recommendations and manually evaluated the corresponding explanations for accuracy of topic and sentiment label, which helped to verify that the proposed viewpoint-based explanations are more informative than topic labels in prior work.

4.2.4 Qualitative Evaluation by Case Study

Case study as qualitative analysis is also frequently used for explainable recommendation system research. Providing case studies can help to understand the intuition behind the explainable recommendation model and the effectiveness of the explanations, and providing case studies as qualitative analysis also helps readers to understand when the proposed approach works and when it does not work.

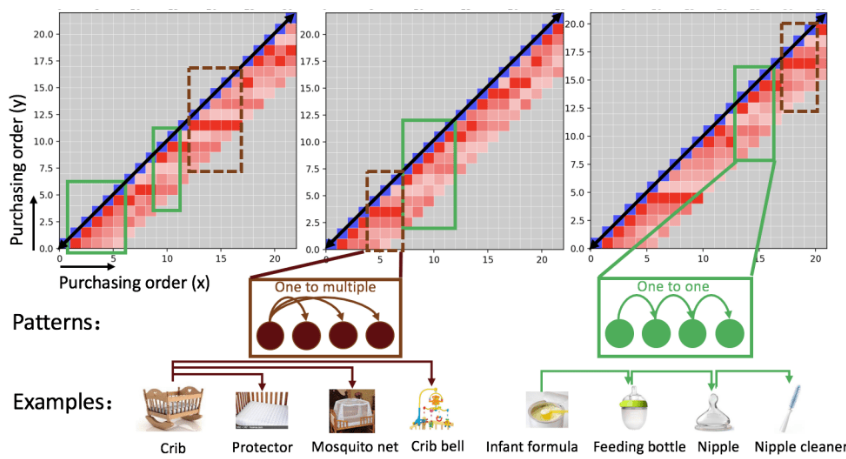
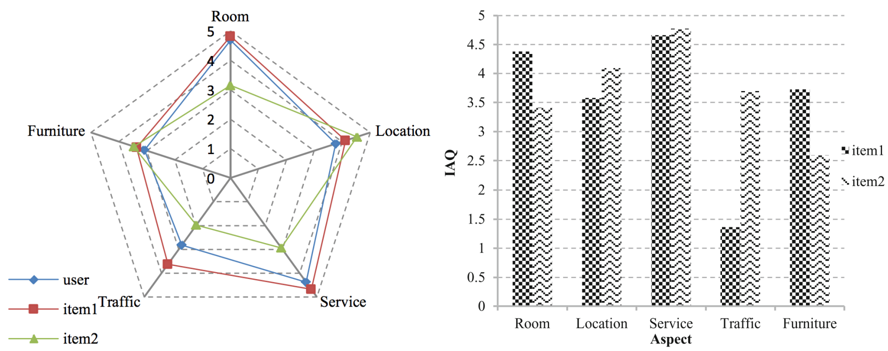


Figure 4.4: Case study of user sequential behaviors in e-commerce for explanation of sequential recommendations. Recommendations can be explained with "one-to-multiple" behaviors, for example, a mom bought a crib for her baby, after that she bought a water-proof mattress protector and a mosquito net for the crib, and further bought a bed bell to decorate the crib; in other cases the recommendation may be explained with "one-to-one" behaviors, for example, a user purchased some infant formula, and then bought a feeding bottle, which caused her to buy some nipples, and these nipples further made her buy a nipple cleaner.

Chen et al. [2018b] provided case study to explain the sequential recommendations, as shown in Figure 4.4. Though the case studies, the authors found that many of the sequential recommendations can be explained based on either "one-to-multiple" or "one-to-one" behavior patter of users, where "one-to-multiple" means that a series of subsequent purchased are triggered by the same already bought product, and "one-to-one" means that a subsequent purchase is triggered by its own preceding purchase. These explanations can help users to clearly understand why an item is recommended, and how the recommended item matches with his/her already purchased items.

He et al. [2015] adopted case study to analyze the explainability and scrutability of the proposed model, and showed that the model can provide explainable restaurant recommendations according to the restaurant aspects mentioned in user reviews, as in Figure 4.5. From the reviews, the authors found that the user is interested in "chicken",

Figure 4.5: Case study of the reviews of a user. This case study shows that the explainable recommendation model is able to provide recommendations based on the aspects that a user liked in his/her reviews. Courtesy image from He et al. [2015].



(a) Radar chart of a user (b) Item aspect quality of two items

Figure 4.6: Case study of explainable hotel recommendation for a target user. For each item, the algorithm calculates its quality on each aspect, and for each user, the algorithm calculates user preference on each aspect. The system then draws the radar chart and bar chart of user preference/item quality for explanation.

although she gives low ratings to the two businesses. Based on this feature, the recommendation model was able to rank a relevant restaurant "Chicken-Fil-A" at high positions in the recommendation list.

Hou et al. [2018] adopted case study to analyze the user preference, item quality, and explainability of the hotel recommendations. The authors first proposed a metric Satisfaction Degree on Aspects (SDA) to measure user satisfaction on aspects of the recommended items, and then conducted case studies to show how the proposed model explains the recommendation results by SDA. As shown in Figure 4.6, for the selected user, item 1 is recommended instead of item 2, and by examining the user preference and item quality, this recommendation can be explained by the fact that item 1 satisfies user preferences on most aspects while item 2 is less effective on these aspects.

5

Explainable Recommendation in Different Applications

The research and application of explainable recommendation span across many different scenarios, such as explainable e-commerce recommendation, explainable social recommendation, explainable multimedia recommendation, etc.

In this section, we provide a review of explainable recommendation methods in different applications. Most of the papers to be reviewed in this section have already been introduced in previous sections, instead, we organize them based on their application scenario to help readers better understand the current scope of explainable recommendation research and how it can be helpful in different applications.

5.1 Explainable E-commerce Recommendation

E-commerce product recommendation is one of the most widely adopted scenarios for explainable recommendation, and it has been a standard test setting for explainable recommendation research.

McAuley and Leskovec [2013] leveraged topic modeling to help understand the latent factors for rating prediction in Amazon e-commerce, and Zhang et al. [2014a] proposed explainable recommendation based on the explicit factor model, and conducted online experiments to evaluate the explainable recommendation model based on real users in a commercial e-commerce website (JD.com). Later, based on the public Amazon e-commerce dataset[1] provided by McAuley et al. [2015], many explainable recommendation models were proposed for e-commerce recommendation.

For instance, He et al. [2015] introduced a tripartite graph ranking-based algorithm to conduct explainable recommendation for electronics products; Chen et al. [2016] proposed a learning to rank approach based on tensor factorization to provide cross-category explainable recommendation of the products; Seo et al. [2017] and Wu et al. [2017] conducted explainable recommendation for multiple product categories (individually) in Amazon, and automatically highlighted important words in user reviews based on attention mechanism; Heckel et al. [2017] adopted overlapping co-clustering to provide scalable and interpretable product recommendations; Chen et al. [2018c] proposed the visually explainable collaborative filtering model to conduct visually explainable recommendation for fashion products; Hou et al. [2018] took advantages of product aspects to conduct explainable video game recommendation in Amazon; Chen et al. [2018a] leveraged neural attention regression based on reviews to conduct rating prediction on three Amazon product categories; Chen et al. [2018b] adopted memory networks to provide sequential recommendation while providing explanations based on what the user previously purchased in Amazon; Wang et al. [2018b] leveraged multi-task learning with tensor factorization to learn recommendations and textual explanations automatically for Amazon product recommendation.

5.2 Explainable Point-of-Interest Recommendation

Point-of-Interest (POI) recommendation, or more generally location recommendation, tries to recommend users with potential locations of interest, such as hotel, restaurant, museum, etc. Explainable POI recommendation has also gained great interest in recent years. Most of

[1]jmcauley.ucsd.edu/data/amazon

the explainable recommendation research are based on location review website datasets, such as Yelp[2] and TripAdvisor[3].

Wu and Ester [2015] conducted restaurant recommendation in Yelp and hotel recommendation in TripAdvisor, and the authors proposed a probabilistic model combining aspect based opinion mining and collaborative filtering to provide explainable recommendations, where the recommended locations are explained based on a word cloud of its aspects; Bauman et al. [2017] developed models to extract the most valuable aspects from reviews for restaurant, hotel and beauty&spa recommendation in Yelp; Seo et al. [2017] also conducted explainable restaurant recommendation in Yelp, and the authors proposed interpretable convolutional neural networks to highlight informative words in reviews for explanation; Zhao et al. [2015] conducted POI recommendation based on Yelp data in Phoenix city and Singapore, respectively, and the authors proposed a joint sentiment-aspect-region modeling approach for recommendation; Wang et al. [2018c] proposed a tree-enhanced embedding model for explainable tourist attraction and restaurant recommendation based on TripAdvisor data in London and New York City.

5.3 Explainable Social Recommendation

Explainable recommendation has also been used for social recommendation, such as friend recommendation, news feeding recommendation, as well as the recommendation of blogs, news, music, travel plans, web pages, images, tags, etc. in a social environment.

For example, Ren et al. [2017] proposed the social collaborative viewpoint regression model for predicting item ratings based on user opinions and social relations, where the social relations not only help to improve the recommendation performance but also the explainability of the recommendations; Quijano-Sanchez et al. [2017] developed a social explanation system applied to group recommendation, which integrates explanations about the system's group recommendation and

[2]http://www.yelp.com
[3]http://www.tripadvisor.com

explanations about the group's social reality for better perception of the group recommendations; Tsai and Brusilovsky [2018] studied how to design explanation interfaces for causal (non-expert) users to achieve different explanatory goals, and in particular, the authors conducted an international (across 13 countries) online survey of 14 active users of a social recommender system to capture user feedback and frame it in terms of design principles of explainable social recommender systems.

5.4 Explainable Multi-media Recommendation

Explainable multi-media recommendation broadly includes the explainable recommendation of books (Wang et al. [2018a]), news/articles (Kraus [2016]), music (Celma [2010]), movie (Tintarev and Masthoff [2008], Nanou et al. [2010]), video (Toderici et al. [2010]), etc, such as Youtube recommendation engine. We review the related work on explainable multi-media recommendation in this section.

The MovieLens dataset[4] is one of the most frequently used dataset for movie recommendation. Based on this dataset, Abdollahi and Nasraoui [2016b] proposed explainable matrix factorization by learning the rating distribution within the active user's neighborhood; in Abdollahi and Nasraoui [2017], the authors further extended the idea for explainability of constrained matrix factorization; Chang et al. [2016] adopted crowd-sourcing to generate crowd-based natural language explanations for movie recommendations in MovieLens; Lee and Jung [2018] attempted to provide story-based explanations for movie recommendation systems, achieved by a multi-aspect explanation and narrative analysis method.

Based on a knowledge-base of the movies such as genre, type, actor, director, etc, recent research has been trying to provide explainable recommendation with knowledge of the users and items, for example, Catherine et al. [2017] proposed explainable entity-based recommendation with knowledge graphs for movie recommendation, which can provide explanations by reasoning over the knowledge graph entities about the movies; Wang et al. [2018a] proposed ripple network structure to

[4]https://grouplens.org/datasets/movielens/

propagate user preferences on knowledge graphs for recommendation, which can also provide explanations based on network hops from the user to the recommended movie, book, or news.

Davidson et al. [2010] introduced the YouTube Video Recommendation System, and leveraged association rule mining to find the related videos as explanation of a recommended video. Online media frequently provide news article recommendations to users, and more and more, such functionally has been integrated into independent news feeding mobile applications, and Kraus [2016] studied how news feedings can be explainable based on political topics.

5.5 Other Explainable Recommendation Applications

Explainable recommendation is also important to a lot of other application scenarios, such as academic recommendation, citation recommendation, healthcare recommendation, etc. Though direct explainable recommendation work on these topics are still limited, researchers have begun to consider the explainability issues within these systems.

For example, Gao et al. [2017] studied the explainability of text classification in online healthcare forums, where a sentence is classified into three classes: medication, symptom, and background, and an interpretation method is also developed, where the decision rules can be explicitly extracted to gain an insight of useful information in texts; and Liu et al. [2018] further studied interpretable outlier detection for health monitoring. Because of the importance for doctors and patients to understand why a treatment is recommended by data-driven systems, it would be important to study the explainability of healthcare recommendation systems.

Similar issues exist in other explainable recommendation scenarios. For example, financial advising/recommendation systems are required to provide explanations for the financial advices, so as to help the investors understand why a certain financial decision is reasonable, such as buying or dropping stocks. The ability to provide explanations is extremely important in some critical financial decisions where millions or even billions of investments are involved. In some cases the explanations can also be time sensitive because of the rapidly changing market condition and the high frequency of modern transactions. Another example is legal advice/recommendation systems oriented either to the users or to the legal practitioners. In this case, the users or legal practitioners not only require a piece of advice aggregated from articles or legal cases, but also helpful evidences, logical reasonings, or supporting documents to explain why such an advice/decision is reasonable.

In a more general sense, the fundamental problem that explainable recommendation strives to solve is to answer the question of 'why' in decision systems, as a result, it is important to any decision system where knowing the 'why' is needed or required, no matter the decision is a recommendation, an action, or a piece of advice.

6

Open Directions and New Perspectives

We discuss some open research directions and new research perspectives of explainable recommendation in this chapter.

6.1 Explainable Deep Learning for Recommendation

The research community has been leveraging deep learning techniques for explainable recommendation. Current approaches focus on designing deep models to generate explanations to accompany the recommendation results, where the explanation can come from attention weights over text, image, video frames, etc. However, the research of leveraging deep models for explainable recommendation is still in its initial stage, and there is much more to be explored in the future. Except for using designing deep models for explainable recommendation, the explainability of the deep model itself also needs further exploration. In most cases, the recommendation/explanation model is still a black box and we do not fully understand how an item is recommended out of the other alternatives. This is mostly due to the fact the the hidden layers in most deep neural networks do not possess certain understandable meanings. As a result, an important task is to make the deep model itself explainable for recommendation, and this will not only benefit the personalized recommendation research, but also many other research areas such as personalized healthcare, personalized online education, chatbots, and self-autonomous systems, etc.

Recent advances in machine learning have shed light on this problem, for example, Koh and Liang [2017] provided a framework to analyze deep neural networks based on influence analyses, while Pei et al. [2017] proposed a whitebox testing mechanism to help understand the nature of deep learning systems. Regarding explainable recommendation, this will help us to understand what are the meanings of each latent component in a neural network and how they interact with each other to generate the final results.

6.2 Knowledge-enhanced Explainable Recommendation

Most of the research on explainable recommendation are based on various unstructured data, such as textual reviews, visual images, etc. However, if the recommendation system possesses certain knowledge about the recommendation domain like human-beings, it will help to generate more tailored recommendations and explanations. For example, with knowledge graph about movies, actors, and directors, the system can explain to the user a movie is recommended because he has watched a lot of movies starred by an actor. Previous work based on this idea dates back to content-based recommendation, which is effective but lacks serendipity and requires extensive manual efforts to match the user interests to content profiles.

With the fast progress of knowledge graph embedding techniques recently, it has been possible for us to integrate the learning of graph embeddings and recommendation models for explainable recommendation, so that the system can make recommendations with certain knowledge about the domain, and tell the user why such items are recommended based on knowledge reasoning, similar to what humans do when asked to make recommendations. This will also help to construct conversational recommendation systems that can communicate with users to provide explainable recommendations based on knowledge. And in a more general sense, this represents one of the future directions for the research of intelligent systems, which is to integrate the philosophy of empiricism and rationalism again for intelligent agent modeling.

6.3 Heterogenous Information Modeling

Modern information retrieval and recommendation systems work on a lot of heterogeneous multi-modal information sources. For example, web search engines have access to documents, images, videos, audios as candidate results for queries; e-commerce recommendation system works on user numerical ratings, textual reviews, product images, demographic information, etc. for user personalization and recommendation; social networks leverage user social relations, and contextual information such as time and location for search and recommendation. Current systems mostly leverage heterogeneous information sources to improve search and recommendation performance, while a lot of research efforts are needed regarding how to jointly leveraging heterogeneous information sources for explainable recommendation and search. These include research tasks such as multi-modal explanation by aligning two or more different information sources, transfer learning over heterogeneous information sources for explainable recommendation, cross-domain explanation in information retrieval and recommendation systems, etc.

6.4 Natural Language Generation for Explanation

Most existing explainable recommendation models are designed to generate explanations of predefined forms, which could be sentence templates, certain association rules, or word clouds, etc. A more natural explanation form could be a piece of free-text that explains to the user with natural language.

Recently, there has been some related work trying to generate natural language explanations, and the basic idea is to train sequence-to-sequence models based on user reviews, and to generate “review-like” sentences as explanations of the recommendation, such as Costa et al.

[2017] and Chen et al. [2018c]. The research of generating natural language explanation is still in initial stage, and there is still a lot of work to do. For example, not all of the review content are of explanation purpose, and it is challenging to decide which part of the review is informative for generating explanations. Beyond textual review, we can also integrate visual images, knowledge base, sentiments and other external information to generate more informed natural language explanation, such as explanation with certain sentiment orientations.

6.5 Explanation beyond Persuasiveness

Existing explainable recommendation mostly focus on generating explanations to persuade the users to accept the explanations (Nanou et al. [2010]). However, explanations can also help to improve the trustworthiness (Cramer et al. [2008]), efficiency (Tintarev and Masthoff [2011]), diversity (Yu et al. [2009]), satisfaction (Bilgic and Mooney [2005]), and scrutability of the system (Knijnenburg et al. [2012]). For example, by letting the user know why not to buy a certain product, the system can help to save time for the users and to win user's trust in the system (Zhang et al. [2014a]).

As a result, it would be important to investigate how explainable recommendation can help to benefit recommendation systems in other aspects beyond persuading the users to accept the recommended items.

6.6 Evaluation of Explainable Recommendations

Evaluation of explainable recommendation systems remains an important problem. Explainable recommendation systems can be easily evaluated with traditional rating prediction or top-n ranking measures to test its recommendation performance, and to evaluate the explanation performance, a currently reliable protocol is to test explainable vs non-explainable recommendation models based on real-world user study, such as A/B testing in practical systems or evaluation with online workers in M-Turk. However, there is still a lack of easily usable offline measure to evaluate the explanation performance. Evaluation of explanations is related to multiple perspectives of information systems, including not only persuasiveness, but also effectiveness, efficiency, transparency, trustworthiness, and user satisfaction, etc. Developing reliable and easily usable evaluation measures for different evaluation perspectives will save a lot of efforts for offline evaluation of explainable recommendation systems.

6.7 Dynamic Explainable Recommendation

User preferences or item profiles may change over time, as a result, the personalized recommendations should be dynamic in accordance to the latest preferences of the users, which leads to the important research direction of dynamic/time-aware recommendation. The same idea applies to explainable recommendation. Because user preferences may change over time, the explanations should also be dynamic, so that a recommendation can be explained according to most recent interests of the users. Most of the current explainable recommendation models are static, i.e., users are profiled based on a training dataset and explanations are generated accordingly, while the explanations are not time-aware or context-aware.

Dynamic explainable recommendation can be investigated both as an extension of time/context-aware recommendation, or as an extension of sequential recommendation, or even other dynamic recommendation settings. For example, Zhang et al. [2015b] attempted to leverage users' time-sensitive interests on product aspects to explain the user purchasing behaviors, while based on memory networks, Chen et al. [2018b] learned how user's previously purchased items contribute to the recommended item as an explanation in sequential recommendation. However, more work is needed towards dynamic explainable recommendation so that the system can provide time-dependent explanations for users.

6.8 Aggregation of Different Explanations

Different explainable recommendation models may generate different explanations, and the explanation may highly rely on the recommendation model. As a result, a common problem is that, we usually have to design different explainable models to generate different explanations for different purposes, and the explanations may not be logically consistent. When the system generates a lot of candidate explanations for a search or recommendation result, a great challenge is how to select the best combination of the explanations to display in a limited space, and how to aggregate different explanations into a logically consistent unified explanation. Solving this problem may require extensive efforts to integrate statistical and logical approaches to machine learning, so that the decision making system is equipped with certain ability of logical inference to explain the results.

6.9 Answering the "Why" in Conversations

The research of recommendation system has extended itself to multiple perspectives, including *what* to recommend (user/item profiling), *when* to recommend (time-aware), *where* to recommend (location-based), *who* to recommend (social recommendation), etc. Beyond these, explainable recommendation aims at answering the question of *why* to recommend, which attempts to solve the problem regarding users' inherent curiosity of why a recommended item is suitable for him/her. Demonstrating why an item is recommended not only helps users to understand the rationality of the recommendations, but also helps to improve the system efficiency, transparency, and trustworthiness.

Based on different application scenario, users can receive recommendation explanations either passively or actively. In conventional web-based systems such as online e-commerce, the explanations can be displayed together with the recommended item, so that the users passively receive the explanations for each recommendation. In the emerging environment of conversational recommendation based on smart agent devices, users can ask "why-related" questions so as to actively seek for explanations when a recommendation is not intuitive to understand. In this case, explainable recommendation will significantly increase the scope of queries that intelligent systems can process.

7

Conclusions

For a long period of time, recommendation systems – or more generally, information retrieval systems – mostly focused on developing extensive models to find the most relevant results as efficiently and effectively as possible, however, the explainability of the IR/recommendation models were largely neglected for a long time. The lack of explainability mainly exists in terms of two perspectives, 1) the outputs of the recommendation system (i.e., recommendation results) are hardly explainable to system users, and 2) the mechanism of the recommendation model (i.e., recommendation algorithm) is hardly explainable to system designers. This lack of explainability for IR/recommendation systems and algorithms leads to problems in practice. Without making the users aware of why certain results are provided, the system may be less effective in persuading the users to accept the results, and may decrease the trustworthiness of the system. More importantly, many IR/recommendation systems nowadays are not only useful for information seeking, but also useful for complicated decision making by providing supportive information and evidences. For example, medical workers may need comprehensive healthcare document recommendation/retrieval to make medical diagnosis. In these decision making tasks, explainability of the results and systems are extremely important, so that system users can understand why a particular result is provided and how to take advantage of the result to take actions.

Recently, deep neural models have been widely used in many IR/recommendation systems. Though researchers have achieved important success in promoting the performance of IR and recommendation, the complexity and inexplainability of many neural models have further highlighted the importance of the research of explainable recommendation and search, and there is a wide range of research topics for the community to address in the coming years.

In this survey, we provided the history of explainable recommendation research ever since the early stage of recommendation system towards the very recent research achievements. We introduced some different forms of recommendation explanations, including user/item-based, content-based, textual, visual, and social explanations, etc. We also introduced different explainable recommendation models, including MF-based, topic-based, graph-based, deep learning-based, knowledge-based, mining-based, and post-hoc explainable recommendation models. We further summarized representative evaluation methods for explainable recommendation, as well as different explainable recommendation applications, including explainable e-commerce/POI/social/multi-media recommendation as well as many other application scenarios. As an outlook to the future, we summarized several possible new research perspectives on explainable recommendation, and we expect that knowledge-base techniques, deep representation learning, natural language generation, dynamic modeling, model aggregation, and conversational systems to gain more achievements in terms of explainable recommendation, and the goal of explainable recommendation systems will also go beyond persuasiveness to further benefit the system users/designers in many other aspects.

In a broader sense, researchers in the Greater-AI community have also realized the importance of Explainable AI, which aims to address a wide range of AI explainability problems in deep learning, computer vision, automatic driving systems, and natural language processing tasks. As an important branch of AI research, this highlights the importance for our IR/RecSys community to address the explainability issues of various recommendation and search systems.

Acknowledgements

We sincerely thank the reviewers for providing the valuable reviews and constructive suggestions.

References

Behnoush Abdollahi and Olfa Nasraoui. Explainable restricted boltzmann machines for collaborative filtering. *2016 ICML Workshop on Human Interpretability in Machine Learning (WHI)*, 2016a.

Behnoush Abdollahi and Olfa Nasraoui. Explainable matrix factorization for collaborative filtering. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 5–6. International World Wide Web Conferences Steering Committee, 2016b.

Behnoush Abdollahi and Olfa Nasraoui. Using explainability for constrained matrix factorization. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pages 79–83. ACM, 2017.

Rakesh Agarwal, Ramakrishnan Srikant, et al. Fast algorithms for mining association rules. In *Proc. of the 20th VLDB Conference*, pages 487–499, 1994.

Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. *Acm sigmod record*, 22(2): 207–216, 1993.

Qingyao Ai, Vahid Azizi, Xu Chen, and Yongfeng Zhang. Learning heterogeneous knowledge base embeddings for explainable recommendation. *arXiv preprint arXiv:1805.03352*, 2018.

Mohammed Z Al-Taie and Seifedine Kadry. Visualization of explanations in recommender systems. *The Journal of Advanced Management Science*, 2 (2), 2014.

Xavier Amatriain and Josep M Pujol. Data mining methods for recommender systems. In *Recommender systems handbook*, pages 227–262. Springer, 2015.

Marko Balabanović and Yoav Shoham. Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66–72, 1997.

Konstantin Bauman, Bing Liu, and Alexander Tuzhilin. Aspect based recommendations: Recommending items with the most valuable aspects based on user reviews. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 717–725. ACM, 2017.

Joeran Beel, Marcel Genzmehr, Stefan Langer, Andreas Nürnberger, and Bela Gipp. A comparative analysis of offline and online evaluations and discussion of research paper recommender system evaluation. In *Proceedings of the international workshop on reproducibility and replication in recommender systems evaluation*, pages 7–14. ACM, 2013.

Shay Ben-Elazar and Noam Koenigstein. A hybrid explanations framework for collaborative filtering recommender systems. In *RecSys Posters*. Citeseer, 2014.

Mustafa Bilgic and Raymond J Mooney. Explaining recommendations: Satisfaction vs. promotion. In *Beyond Personalization Workshop, IUI*, volume 5, page 153, 2005.

Mustafa Bilgic, R Mooney, and E Rich. Explanation for recommender systems: satisfaction vs. promotion. *Computer Sciences Austin, University of Texas. Undergraduate Honors*, 27, 2004.

Rose Catherine, Kathryn Mazaitis, Maxine Eskenazi, and William Cohen. Explainable entity-based recommendations with knowledge graphs. *RecSys 2017 Poster Proceedings*, 2017.

Oscar Celma. Music recommendation. In *Music Recommendation and Discovery*, pages 43–85. Springer, 2010.

Allison JB Chaney, David M Blei, and Tina Eliassi-Rad. A probabilistic model for using social networks in personalized item recommendation. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 43–50. ACM, 2015.

Shuo Chang, F Maxwell Harper, and Loren Gilbert Terveen. Crowd-based personalized natural language explanations for recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 175–182. ACM, 2016.

Chong Chen, Min Zhang, Yiqun Liu, and Shaoping Ma. Neural attentional rating regression with review-level explanations. *WWW*, 2018a.

Xu Chen, Zheng Qin, Yongfeng Zhang, and Tao Xu. Learning to rank features for recommendation over multiple categories. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 305–314. ACM, 2016.

Xu Chen, Hongteng Xu, Yongfeng Zhang, Yixin Cao, Hongyuan Zha, Zheng Qin, and Jiaxi Tang. Sequential recommendation with user memory networks. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 2018b.

Xu Chen, Yongfeng Zhang, Hongteng Xu, Yixin Cao, Zheng Qin, and Hongyuan Zha. Visually explainable recommendation. *arXiv preprint arXiv:1801.10288*, 2018c.

Zhiyong Cheng, Ying Ding, Lei Zhu, and Mohan Kankanhalli. Aspect-aware latent factor model: Rating prediction with ratings and reviews. *WWW*, 2018.

Yoon Ho Cho, Jae Kyeong Kim, and Soung Hie Kim. A personalized recommender system based on web usage mining and decision tree induction. *Expert systems with Applications*, 23(3):329–342, 2002.

Sergio Cleger-Tamayo, Juan M Fernandez-Luna, and Juan F Huete. Explaining neighborhood-based recommendations. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 1063–1064. ACM, 2012.

Felipe Costa, Sixun Ouyang, Peter Dolog, and Aonghus Lawlor. Automatic generation of natural language explanations. *arXiv preprint arXiv:1707.01561*, 2017.

Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, 18(5):455, 2008.

James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296. ACM, 2010.

Robin Devooght and Hugues Bersini. Long and short-term recommendations with recurrent neural networks. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, pages 13–21. ACM, 2017.

Tim Donkers, Benedikt Loepp, and Jürgen Ziegler. Sequential user-based recurrent neural network recommendations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pages 152–160. ACM, 2017.

Michael D Ekstrand, John T Riedl, Joseph A Konstan, et al. Collaborative filtering recommender systems. *Foundations and Trends® in Human–Computer Interaction*, 4(2):81–173, 2011.

Bruce Ferwerda, Kevin Swelsen, and Emily Yang. Explaining content-based recommendations. *regular paper*, 2012.

Rudolph Flesch. A new readability yardstick. *Journal of applied psychology*, 32(3):221, 1948.

Jun Gao, Ninghao Liu, Mark Lawley, and Xia Hu. An interpretable classification framework for information extraction from online healthcare forums. *Journal of healthcare engineering*, 2017, 2017.

Asela Gunawardana and Guy Shani. A survey of accuracy evaluation metrics of recommendation tasks. *Journal of Machine Learning Research*, 10(Dec): 2935–2962, 2009.

Robert Gunning. *The technique of clear writing.* McGraw-Hill, New York, 1952.

Xiangnan He, Tao Chen, Min-Yen Kan, and Xiao Chen. Trirank: Review-aware explainable recommendation by modeling aspects. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 1661–1670. ACM, 2015.

Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, pages 173–182. International World Wide Web Conferences Steering Committee, 2017.

Reinhard Heckel, Michail Vlachos, Thomas Parnell, and Celestine Duenner. Scalable and interpretable product recommendations via overlapping co-clustering. In *Data Engineering (ICDE), 2017 IEEE 33rd International Conference on*, pages 1033–1044. IEEE, 2017.

Jonathan L Herlocker, Joseph A Konstan, and John Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 241–250. ACM, 2000.

Jonathan Lee Herlocker and Joseph A Konstan. *Understanding and improving automated collaborative filtering systems.* University of Minnesota Minnesota, 2000.

Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939*, 2015.

Yunfeng Hou, Ning Yang, Yi Wu, and S Yu Philip. Explainable recommendation with fusion of aspect information. *World Wide Web*, pages 1–20, 2018.

George Karypis. Evaluation of item-based top-n recommendation algorithms. In *Proceedings of the tenth international conference on Information and knowledge management*, pages 247–254. ACM, 2001.

J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, Naval Technical Training Command Millington TN Research Branch, 1975.

Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5):441–504, 2012.

Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. *arXiv preprint arXiv:1703.04730*, 2017.

Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008.

Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8), 2009.

Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. User preferences for hybrid explanations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pages 84–88. ACM, 2017.

Christina Luisa Kraus. A news recommendation engine for a multi-perspective understanding of political topics. *Master Thesis, Technical University of Berlin*, 2016.

Aonghus Lawlor, Khalil Muhammad, Rachael Rafter, and Barry Smyth. Opinionated explanations for recommendation systems. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pages 331–344. Springer, 2015.

Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.

Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.

O-Joun Lee and Jason J Jung. Explainable movie recommendation systems by using story-based similarity. *regular paper*, 2018.

Piji Li, Zihao Wang, Zhaochun Ren, Lidong Bing, and Wai Lam. Neural rating regression with abstractive tips generation for recommendation. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 345–354. ACM, 2017.

Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. *Text Summarization Branches Out*, 2004.

Weiyang Lin, Sergio A Alvarez, and Carolina Ruiz. Collaborative recommendation via adaptive association rule mining. *Data Mining and Knowledge Discovery*, 6:83–105, 2000.

Weiyang Lin, Sergio A Alvarez, and Carolina Ruiz. Efficient adaptive-support association rule mining for recommender systems. *Data mining and knowledge discovery*, 6(1):83–105, 2002.

Greg Linden, Brent Smith, and Jeremy York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1):76–80, 2003.

Ninghao Liu, Donghwa Shin, and Xia Hu. Contextual outlier interpretation. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018.

Yichao Lu, Ruihai Dong, and Barry Smyth. Coevolutionary recommendation model: Mutual learning between ratings and reviews. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pages 773–782. International World Wide Web Conferences Steering Committee, 2018.

G Harry Mc Laughlin. Smog grading-a new readability formula. *Journal of reading*, 12(8):639–646, 1969.

Julian McAuley and Jure Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 165–172. ACM, 2013.

Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 43–52. ACM, 2015.

David McSherry. Explanation in recommender systems. *Artificial Intelligence Review*, 24(2):179–197, 2005.

Andriy Mnih and Ruslan R Salakhutdinov. Probabilistic matrix factorization. In *Advances in neural information processing systems*, pages 1257–1264, 2008.

Bamshad Mobasher, Honghua Dai, Tao Luo, and Miki Nakagawa. Effective personalization based on association rule discovery from web usage data. In *Proceedings of the 3rd international workshop on Web information and data management*, pages 9–15. ACM, 2001.

Khalil Muhammad, Aonghus Lawlor, Rachael Rafter, and Barry Smyth. Great explanations: Opinionated explanations for recommendations. In *International Conference on Case-Based Reasoning*, pages 244–258. Springer, 2015.

Khalil Muhammad, Aonghus Lawlor, and Barry Smyth. Explanation-based ranking in opinionated recommender systems. In *AICS*, pages 128–139, 2016a.

Khalil Muhammad, Aonghus Lawlor, and Barry Smyth. On the use of opinionated explanations to rank and justify recommendations. In *FLAIRS Conference*, pages 554–559, 2016b.

Khalil Ibrahim Muhammad, Aonghus Lawlor, and Barry Smyth. A live-user study of opinionated explanations for recommender systems. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pages 256–260. ACM, 2016c.

Theodora Nanou, George Lekakos, and Konstantinos Fouskas. The effects of recommendations? presentation on persuasion and satisfaction in a movie recommender system. *Multimedia systems*, 16(4-5):219–230, 2010.

Alexis Papadimitriou, Panagiotis Symeonidis, and Yannis Manolopoulos. A generalized taxonomy of explanations styles for traditional and social recommender systems. *Data Mining and Knowledge Discovery*, 24(3):555–583, 2012.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics, 2002.

Haekyu Park, Hyunsik Jeon, Junghwan Kim, Beunguk Ahn, and U Kang. Uniwalk: Explainable and accurate recommendation for rating and network data. *arXiv preprint arXiv:1710.07134*, 2017.

Michael J Pazzani. A framework for collaborative, content-based and demographic filtering. *Artificial intelligence review*, 13(5-6):393–408, 1999.

Michael J Pazzani and Daniel Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007.

Kexin Pei, Yinzhi Cao, Junfeng Yang, and Suman Jana. Deepxplore: Automated whitebox testing of deep learning systems. In *Proceedings of the 26th Symposium on Operating Systems Principles*, pages 1–18. ACM, 2017.

Lin Qiu, Sheng Gao, Wenlong Cheng, and Jun Guo. Aspect-based latent factor model by integrating ratings and reviews for recommender system. *Knowledge-Based Systems*, 110:233–243, 2016.

Lara Quijano-Sanchez, Christian Sauer, Juan A Recio-Garcia, and Belen Diaz-Agudo. Make it personal: A social explanation system applied to group recommendations. *Expert Systems with Applications*, 76:36–48, 2017.

Zhaochun Ren, Shangsong Liang, Piji Li, Shuaiqiang Wang, and Maarten de Rijke. Social collaborative viewpoint regression with explainable recommendations. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 485–494. ACM, 2017.

Steffen Rendle and Lars Schmidt-Thieme. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 81–90. ACM, 2010.

Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 452–461. AUAI Press, 2009.

Jasson DM Rennie and Nathan Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22nd international conference on Machine learning*, pages 713–719. ACM, 2005.

Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175–186. ACM, 1994.

Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer, 2011.

Ruslan Salakhutdinov and Andriy Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *Proceedings of the 25th international conference on Machine learning*, pages 880–887. ACM, 2008.

Jeff J Sandvig, Bamshad Mobasher, and Robin Burke. Robustness of collaborative recommendation based on association rule mining. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 105–112. ACM, 2007.

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

J Ben Schafer, Joseph Konstan, and John Riedl. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 158–166. ACM, 1999.

RJ Senter and Edgar A Smith. Automated readability index. Technical report, CINCINNATI UNIV OH, 1967.

Sungyong Seo, Jing Huang, Hao Yang, and Yan Liu. Interpretable convolutional neural networks with dual local and global attention for review rating prediction. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pages 297–305. ACM, 2017.

Guy Shani and Asela Gunawardana. Evaluating recommendation systems. *Recommender systems handbook*, pages 257–297, 2011.

Amit Sharma and Dan Cosley. Do social explanations work?: studying and modeling the effects of social explanations in recommender systems. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1133–1144. ACM, 2013.

Yue Shi, Martha Larson, and Alan Hanjalic. List-wise learning to rank with matrix factorization for collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 269–272. ACM, 2010.

Rashmi Sinha and Kirsten Swearingen. The role of transparency in recommender systems. In *CHI'02 extended abstracts on Human factors in computing systems*, pages 830–831. ACM, 2002.

Barry Smyth, Kevin McCarthy, James Reilly, Derry O'Sullivan, Lorraine McGinty, and David C Wilson. Case studies in association rule mining for recommender systems. In *IC-AI*, pages 809–815, 2005.

Nathan Srebro and Tommi Jaakkola. Weighted low-rank approximations. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 720–727, 2003.

Nathan Srebro, Jason Rennie, and Tommi S Jaakkola. Maximum-margin matrix factorization. In *Advances in neural information processing systems*, pages 1329–1336, 2005.

Yunzhi Tan, Min Zhang, Yiqun Liu, and Shaoping Ma. Rating-boosted latent topics: Understanding users and items with ratings and reviews. In *IJCAI*, pages 2640–2646, 2016.

Jiaxi Tang and Ke Wang. Personalized top-n sequential recommendation via convolutional sequence embedding. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 2018.

Nava Tintarev. Explanations of recommendations. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 203–206. ACM, 2007.

Nava Tintarev and Judith Masthoff. Effective explanations of recommendations: user-centered design. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 153–156. ACM, 2007.

Nava Tintarev and Judith Masthoff. The effectiveness of personalized movie explanations: An experiment using commercial meta-data. In *Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 204–213. Springer, 2008.

Nava Tintarev and Judith Masthoff. Designing and evaluating explanations for recommender systems. *Recommender Systems Handbook*, pages 479–510, 2011.

Nava Tintarev and Judith Masthoff. Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook*, pages 353–382. Springer, 2015.

George Toderici, Hrishikesh Aradhye, Marius Pasca, Luciano Sbaiz, and Jay Yagnik. Finding meaning on youtube: Tag recommendation and category discovery. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3447–3454. IEEE, 2010.

Chun-Hua Tsai and Peter Brusilovsky. Explaining social recommendations to casual users: Design principles and opportunities. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, page 59. ACM, 2018.

Jesse Vig, Shilad Sen, and John Riedl. Tagsplanations: explaining recommendations using tags. In *Proceedings of the 14th international conference on Intelligent user interfaces*, pages 47–56. ACM, 2009.

Beidou Wang, Martin Ester, Jiajun Bu, and Deng Cai. Who also likes it? generating the most persuasive social explanations in recommender systems. In *AAAI*, pages 173–179, 2014.

Hao Wang, Naiyan Wang, and Dit-Yan Yeung. Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1235–1244. ACM, 2015.

Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. Ripple network: Propagating user preferences on the knowledge graph for recommender systems. *arXiv preprint arXiv:1803.03467*, 2018a.

Nan Wang, Hongning Wang, Yiling Jia, and Yue Yin. Explainable recommendation via multi-task learning in opinionated text data. In *Proceedings of the 41st international ACM SIGIR conference on Research & development in information retrieval*. ACM, 2018b.

Weiquan Wang and Izak Benbasat. Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems*, 23(4):217–246, 2007.

Xiang Wang, Xiangnan He, Fuli Feng, Liqiang Nie, and Tat-Seng Chua. Tem: Tree-enhanced embedding model for explainable recommendation. In *Proceedings of the 27th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee*, 2018c.

Libing Wu, Cong Quan, Chenliang Li, Qian Wang, and Bolong Zheng. A context-aware user-item representation learning for item recommendation. *arXiv preprint arXiv:1712.02342*, 2017.

Yao Wu and Martin Ester. Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pages 199–208. ACM, 2015.

Yao Wu, Christopher DuBois, Alice X Zheng, and Martin Ester. Collaborative denoising auto-encoders for top-n recommender systems. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 153–162. ACM, 2016.

Cong Yu, Laks VS Lakshmanan, and Sihem Amer-Yahia. Recommendation diversification using explanations. In *Data Engineering, 2009. ICDE'09. IEEE 25th International Conference on*, pages 1299–1302. IEEE, 2009.

Markus Zanker and Daniel Ninaus. Knowledgeable explanations for recommender systems. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 1, pages 657–660. IEEE, 2010.

Yongfeng Zhang. Incorporating phrase-level sentiment analysis on textual reviews for personalized recommendation. In *Proceedings of the eighth ACM international conference on web search and data mining*, pages 435–440. ACM, 2015.

Yongfeng Zhang, Min Zhang, Yiqun Liu, and Shaoping Ma. Improve collaborative filtering through bordered block diagonal form matrices. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 313–322. ACM, 2013a.

Yongfeng Zhang, Min Zhang, Yiqun Liu, Shaoping Ma, and Shi Feng. Localized matrix factorization for recommendation based on matrix block diagonal forms. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1511–1520. ACM, 2013b.

Yongfeng Zhang, Guokun Lai, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 83–92. ACM, 2014a.

Yongfeng Zhang, Haochen Zhang, Min Zhang, Yiqun Liu, and Shaoping Ma. Do users rate or review?: Boost phrase-level sentiment labeling with review-level sentiment classification. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 1027–1030. ACM, 2014b.

Yongfeng Zhang, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. Understanding the sparsity: Augmented matrix factorization with sampled constraints on unobservables. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1189–1198. ACM, 2014c.

Yongfeng Zhang, Min Zhang, Yiqun Liu, Chua Tat-Seng, Yi Zhang, and Shaoping Ma. Task-based recommendation on a web-scale. In *Big Data (Big Data), 2015 IEEE International Conference on*, pages 827–836. IEEE, 2015a.

Yongfeng Zhang, Min Zhang, Yi Zhang, Guokun Lai, Yiqun Liu, Honghui Zhang, and Shaoping Ma. Daily-aware personalized recommendation based on feature-level time series analysis. In *Proceedings of the 24th international conference on world wide web*, pages 1373–1383. International World Wide Web Conferences Steering Committee, 2015b.

Yongfeng Zhang, Qingyao Ai, Xu Chen, and W Bruce Croft. Joint representation learning for top-n recommendation with heterogeneous information sources. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1449–1458. ACM, 2017.

Kaiqi Zhao, Gao Cong, Quan Yuan, and Kenny Q Zhu. Sar: A sentiment-aspect-region model for user preference analysis in geo-tagged reviews. In *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*, pages 675–686. IEEE, 2015.

Wayne Xin Zhao, Sui Li, Yulan He, Liwei Wang, Ji-Rong Wen, and Xiaoming Li. Exploring demographic information in social media for product recommendation. *Knowledge and Information Systems*, 49(1):61–89, 2016.

Xin Wayne Zhao, Yanwei Guo, Yulan He, Han Jiang, Yuexin Wu, and Xiaoming Li. We know what you want to buy: a demographic-based system for product recommendation on microblogs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1935–1944. ACM, 2014.

Lei Zheng, Vahid Noroozi, and Philip S Yu. Joint deep modeling of users and items using reviews for recommendation. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 425–434. ACM, 2017.