Foundations and Trends® in Information Retrieval
Vol. XX, No. XX (2018) 1–87

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 users with the recommendations, but also provide explanations to make the user or system designer aware of why such items are recommended. In this way, it helps to improve the effectiveness, efficiency, persuasiveness, and user satisfaction of recommendation 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 section 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. 1–87, 2018.

DOI: 10.1561/XXXXXXXXXX.

1

Introduction

1.1 Explainable Recommendation

Explainable Recommendation refers to the personalized recommendation algorithms that address the problem of why - they not only provide users with the recommendations, but also provide explanations to make the user or system designer aware of why such items are recommended. In this way, it helps to improve the effectiveness, efficiency, persuasiveness, and user satisfaction of recommendation systems.

To highlight the position of explainable recommendation in the whole recommendation system research, we classify most of the existing personalized recommendation research with a broad conceptual taxonomy. 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*.

From the perspective of process-product distinction in AI research, explainable recommendation can consider the explainability of the rec-

ommendation method (i.e., process). In this way, explainable recommendation aims to devise interpretable models that work in a human way, and such models usually also lead to the explainability of recommendation results (i.e., product). Most of the current explainable recommendation research fall into this category, which aim to understand how the process works, and they are usually referred to as model-based explainable recommendation. Another approach to explainable recommendation is that we only focus on the product. In this way, we treat the recommendation model as a complex blackbox and ignore its explainability, but instead develop separate methods to explain the recommendation results produced by this blackbox. Methods that fall into this category are usually called post-hoc explainable recommendation. In this survey, we will introduce both of the two types of explainable recommendation methods.

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 recom-

mender 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 was usually intuitive to explain to the users why an item is recommended out of other candidates in content-based recommendation. For example, one straightforward 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 earliest 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, CF 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 *Explainable Recommendation Models*, 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, the explainability of AI systems was already a core discussion in the "old" or logical AI age in the 1980s, where early knowledge-based systems predicted (or diagnosed) well but could not explain why. For example, the work of Clancy showed that being able to explain predictions required far more knowledge than just making correct predictions (Clancey [1982]). Recent booming of big data and computational power have brought AI research and performance onto a new level, but researchers in the broader AI community have again realized the importance of *Explainable AI* in recent years, 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 also highlights the importance the IR and recommendation system community to address the explainability issues of various search and 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 in explainable recommendation research.

In particular, we classify existing explainable recommendation research based on two orthogonal dimensions: 1) The display style of the generated explanations (e.g., a piece of textual sentence, or visual image-based explanation, etc.), and 2) the model used to generate such an explanation, including matrix factorization, topic modeling, graph-based models, deep learning, knowledge-graph embedding, association rule mining, etc.

With this classification taxonomy, each combination of the two dimensions refers to a particular type of research on explainable recommendation. It should be noted that there could be particular conceptual differences between "how the explanations are presented" and "the type of information used for explanation". In the context of explainable recommendation, however, these two principles are conceptually aligned with each other because the type of information usually determines how the explanations can be displayed, and we believe that the display style is a better dimension to categorize the research in this area. However, we should point out that among the possibly many classification taxonomies, this is only one of them that we think would be appropriate to better organize the related research on explainable recommendation.

Table 1.1 shows how representative explainable recommendation research are classified into different categories. For example, the Explicit Factor Model (EFM) for explainable recommendation proposed in Zhang et al. [2014a] developed a multi-matrix factorization model for explainable recommendation, and based on this model, it provides a piece of explanation sentence to the user as recommendation explanation, as a result, it falls into the category of matrix factorization method with textual explanations. Similarity, the Interpretable Convolutional Neural Network approach to explainable recommendation in Seo et al. [2017] developed a deep convolutional neural network-based model and display the item features to users as explanations, which falls into the category of deep learning-based method with user/item feature explanation; while the visually explainable recommendation approach proposed in Chen et al. [2018c] developed a deep neural network that can generate image regional highlights as explanations for target users, which belongs to deep learning-based models with visual explanations.

We also classify other research 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.

It should be noted that due to the limited space of the table, Table 1.1 is an incomplete enumeration of existing explainable recommendation methods, and for each "model – display style" combination we only presented one representative work in the corresponding table cell. However, in the following parts of the paper, we will devise two section to introduced the details of the many research on explainable recommendation based on this classification taxonomy, corresponding to Section 2 and Section 3, respectively.

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, recent evidence suggests that these two goals may not necessarily conflict with each other when designing recommendation models (Bilgic et al. [2004], Zhang et al. [2014a]). For example, 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. Developing both accurate and explainable deep neural models is also an attractive research direction in the recent years, which not only leads to progress in the explainable recommendation research, but also helps us to gain progress in the research of Explainable Machine Learning.

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

Table 1.1: A classification of existing explainable recommendation methods. The classification is based on two dimensions, i.e., the type of model for explainable recommendation (e.g., matrix factorization, topic modeling, deep learning, etc.) and the display type of the generated explanation (e.g., textual sentence explanation, visual image-based explanation, etc.). Note that due to the table space this is an incomplete enumeration of the existing explainable recommendation methods, and more methods are introduced in detail in the following parts of the paper. Besides, some of the table cells are empty because to the best of our knowledge there has not been a work falling into the corresponding combination.

	Methods for Explainable Recommendation						
	neighbor -based	matrix factorization	topic modeling	graph -based	deep learning	knowledge -based	association analysis
Display Styles of the Recommendation Explanations							
relevant user or item	Herlocker et al. [2000]	Abdollahi and Nasraoui [2017]		Heckel et al. [2017]	Chen et al. [2018b]	Catherine et al. [2017]	Peake and Wang [2018]
user or item features	Vig et al. [2009]		McAuley and Leskovec [2013]	He et al. [2015]	Seo et al. [2017]	Huang et al. [2018]	Davidson et al. [2010]
textual sentence explanation		Zhang et al. [2014a]			Costa et al. [2017]	Ai et al. [2018]	
visual image explanation					Chen et al. [2018c]		
social explanation	Sharma and Cosley [2013]		Ren et al. [2017]	Park et al. [2017]			
word cluster		Zhang [2015]	Wu and Ester [2015]				

2

Different Display Styles of Explanations

An explanation is a piece of information displayed to the target user to explain why a particular item is recommended. In practice, recommendation explanations can be presented in very different display styles (Tintarev and Masthoff [2015]), which could be a relevant user or item, a radar chart, a sentence, an image, etc. Besides, explanation may not be unique, namely, there could be many different explanations for the same recommendation.

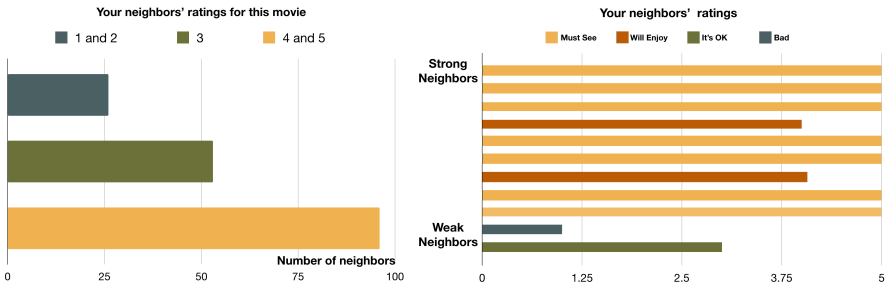
For example, Zhang et al. [2014a] generated (personalized) textual sentences as recommendation explanation to help users understand each recommendation result; McAuley and Leskovec [2013], Zhang [2015], Al-Taie and Kadry [2014] provided topical word cloud or clusters to highlight the key features of a recommended item as explanation; Chen et al. [2018c] proposed visually explainable recommendation where particular regions of a recommended image are highlighted as the visual explanations for users; Sharma and Cosley [2013] and Quijano-Sanchez et al. [2017]) generated a list of social friends who also liked the recommended product as social explanations for target user.

In the early research stages, Herlocker et al. [2000], Bilgic and Mooney [2005], Tintarev and Masthoff [2007] and McSherry [2005]

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

adopted statistical histograms or pie charts to help users understand the rating distribution and the pros/cons of a recommended product as intuitive explanations. As an aggregation, Figure 2.1 shows several representative recommendation explanations.

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 real-world settings. We also categorize the related work into different display styles of explanation for easy reference. More specifically, the following subsections present an overview



(a) (b)

Figure 2.2: An example for explanation based on relevant users (a) A histogram of the neighbors' ratings of the target user are displayed as an explanation for the recommended item, where the positive and negative ratings are clustered respectively, and the neutral ratings are displayed separately. Based on this explanation, it would be easy for the users to understand that the item is recommended because his/her neighbors made high ratings on the item. (b) An explanation for the recommended movie "The Sixth Sense", where each bar represents the rating of a neighbor, and the x-axis represents a neighbor's similarity to the user. From this explanation it would easy to understand how a user's most similar neighbors rated the target movie (Herlocker et al. [2000]).

of several frequently seen explanations in existing systems.

2.1 Explanation based on Relevant Users or Items

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. [2012]), which are two 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]), explanations based on similar users (i.e., neighbors) can be generated by letting the user know that a recommendation is provided because their ratings are similar to a group of "neighborhood" users, and these neighborhood users made good ratings on

Figure 2.3: A comparison between explanations with relevant items (left) and relevant users (right) (Abdollahi and Nasraoui [2017]).

the recommended item. For example, Herlocker et al. [2000] compared the effectiveness of different display styles of explanation for user-based collaborative filtering, where the explanation can be displayed as an aggregated histogram of the neighbors' ratings, or be displayed as the detailed ratings of the neighbors, as shown in Figure 2.2. State-of-the-art model-based explainable recommendation methods can generate more personalized and meticulously designed explanations than this, but this research illustrated the basic ideas of providing explanations to users.

In item-based collaborative filtering (Sarwar et al. [2001]), explanations can be provided by letting the users 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), explanation based on relevant users tells Bob that similar users William and Fred also bought this item, while explanation based on relevant items persuades Bob by pointing out that the lens is relevant to his previously purchased camera.

To study how explanations could make a difference in recommendation systems, Tintarev [2007] proposed to develop a prototype system to study the different types of explanations, especially the relevant-user or relevant-item-based explanations. In particular, the authors proposed seven aims of providing explanations in recommender systems, including transparency, scrutability, trustworthiness, effectiveness, persuasiveness, efficiency, and satisfaction, and based on user study, the

authors showed that providing appropriate explanations can be helpful to benefit the recommender system in terms of these aspects.

Explanation based on relevant items 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. For explanation based on relevant users, however, it could be less convincing by providing the target user with some other users that made similar choices, because the target user may know nothing about the other users at all, which decreases the trustworthiness of the explanation. Besides, disclosing other users' purchasing history information may also cause privacy problems in practical systems. This drives relevant-user-based explanation to a new direction, which is to leverage social friend information to provide social explanations, which provides a user with his/her social friends' public interests as recommendation explanations. We will review this research direction in the social explanation section in the following part of the paper.

2.2 Feature-based Explanation

Feature-based explanation is closely related to content-based recommendation methods, where the system provides personalized recommendations by matching the user preference with the available content features of items (Pazzani and Billsus [2007], Ferwerda et al. [2012], Cramer et al. [2008]). Based on these content features, content-based recommendation is usually intuitive to be explained by generating feature-based explanations.

Content-based recommendations can be based on various item features/characteristics depending on each application scenario. For example, movie recommendations can be generated based on movie genres, actor, director, etc; and book recommendations can be provided based on book types, price, author, etc. A common paradigm for feature-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 features to generate recom-

(a) (b)

Figure 2.4: Tags explanation: generating explanations based on content-based item features such as 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 that match with these tags. (b) The tags-explanation provided for a recommended movie *Rushmore*, where the relevant tags (features) are displayed as explanations.

mendations and the corresponding explanations, as shown in Figure 2.4. To explain the recommended movie, the system displays the content features (aspects) of the movie, and also tells the user why each feature is relevant to the user. The authors also designed user study experiments, and showed that providing the feature-based explanations can help to improve the effectiveness the personalized recommendations. Furthermore, Ferwerda et al. [2012] conducted user study experiments and the results supported the idea that user trust and choice satisfaction were highly correlated with the explanations for recommendations.

The content features can be displayed in many different intuitive styles 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 features of users, and the demographic features can also be used to generate feature-based explanations. Demographic-based recommendation (Pazzani [1999]) is one of the earliest approaches to personalized recommendation, and in the recent years, researchers have also integrated

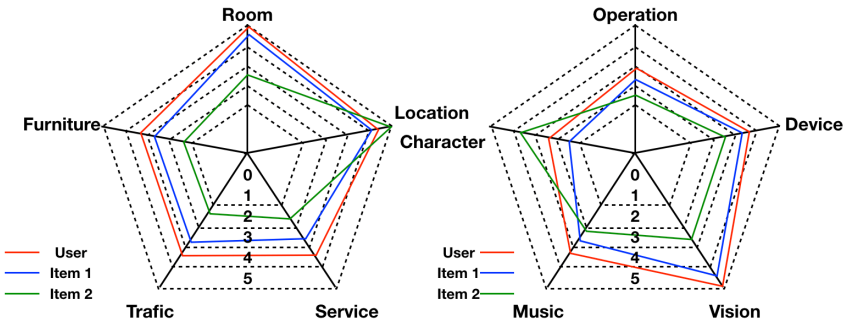


Figure 2.5: Using radar charts to explain a recommendation. The left figure shows hotel recommendations for a user, where item 1 is recommended because it satisfies the user preferences on nearly all aspects. Similarly, the right figure shows video game recommendations and also, item 1 is recommended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

demographic-based method with social media to provide product recommendations in social environments (Zhao et al. [2014, 2016]).

Basically, demographic-based recommendation make 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 environment for product recommendation with feature-based explanations.

2.3 Textual Sentence Explanations

More and more user generated contents such as user reviews in e-commerce and user posts in social networks have been accumulated on the web. Such information is of great value for estimating more comprehensive user preference, and can be utilized to provide finer-grained and more reliable recommendation explanations to persuade the costumers or to help the consumers make more informed decisions.

(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].

Motivated by this intuition, many models have been designed recently to explain recommendations leveraging various types of text information, and they usually generate a piece of textual sentence as recommendation explanation.

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

Aspect-level explanation is similar to feature-based explanations, except that aspects are usually not directly available within an item or user profile, instead, they are extracted or learned as part of the recommendation model from – e.g., the textual reviews – and the aspects can be paired up with consumer opinions to express a clear sentiment on the aspect.

To extract product aspects 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 "aspect–opinion–sentiment" triplets from textual reviews

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

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 the contextual sentiment of the opinion words given different aspect words, for example, though "noise" paired with "high" usually represents a negative sentiment, when "quality" is paired with "high", however, it instead shows a positive sentiment. Based on the dictionary of aspect–opinion–sentiment triplets that the program constructed, it can further detect which triplets are contained in a given piece of 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 aspect-opinion pairs as explanation, for example, "bathroom-clean", which not only indicates the available aspects of an item, but also the consumers' aggregated opinions on the aspects, so as to highlight the performance of a recommended item on these aspects. 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 aspects 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 reflect the important of the corresponding word.

```
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.

In sentence-level approaches, the explanations are provided as a

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

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 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 performs poorly*". 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 over 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,

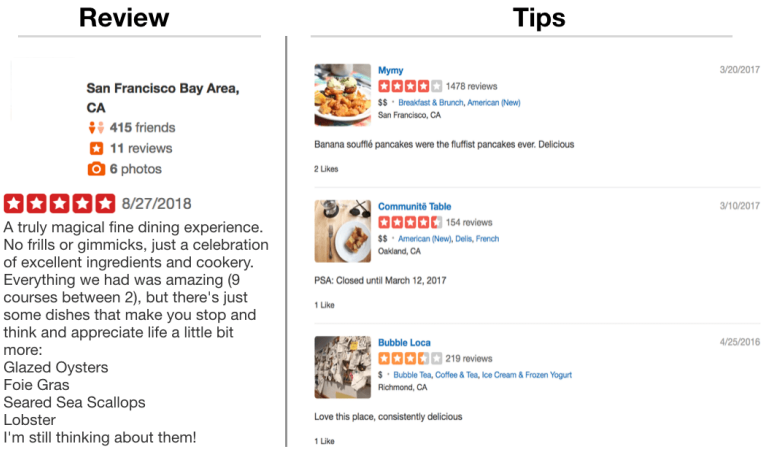


Figure 2.9: Example of the reviews and tips on Yelp. Tips are more concise than reviews and can reveal user experience, feelings, and suggestions with only a few words (Li et al. [2017]).

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 Image 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 intuition is that different users may be attracted by different regions of the product image, for example, even for the same shirt, some users may care about the collar design while others may pay more attention to the pocket. As a result, the authors adopted neural attention mechanism that integrates

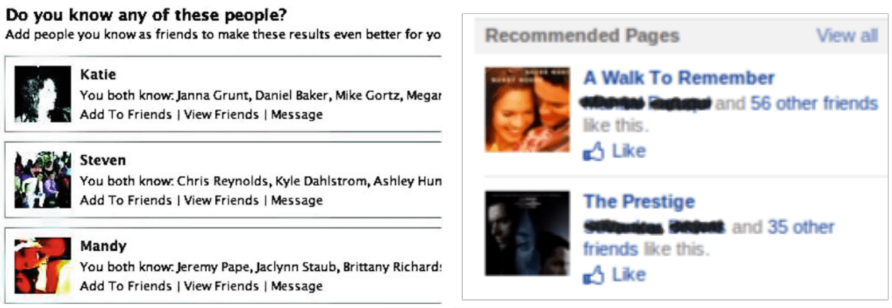
#	Target Item	Historical Records	Textual Review	Visual Explanation	
				VECF	Re-VECF
1			this is a large watch... nearly as large as my suunto but due to ***its articulated strap it fits on the wrist very well.***		
2			***this is a really comfortable v-neck. i found that the size and location of the 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.

both image and textual reviews information to learn the importance of each region in an image, and the important regions of an image are highlighted in a personalized manner as visual explanations for users.

Lin et al. [2018] studied the problem of explainable outfit recommendation, for example, given a top (i.e., upper garment), how to recommend a short list of bottoms (e.g., trousers or skirts) from a large collection that best match the top, and meanwhile generate explanations for each recommendation so as to explain why the top and the bottom match. Technically, this work proposed a convolutional neural network with a mutual attention mechanism to extract visual features of the outfits, and the visual features are fed into a neural prediction network to predict the rating scores for recommendation. During the prediction procedure, the attention mechanism will learn the importance of different regions of the product image as explanations, and the regional importance scores tell us which regions of the image are taking effect when generating the recommendations.

Generally speaking, the research on visually explainable recommen-



(a) (b)

Figure 2.11: 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]).

dation is still at its initial stage, and there has not been many research work on this topic. With the continuous development of deep image processing techniques, we expect that images will be better integrated into recommender systems for both better recommendation performance and explainability.

2.5 Social Explanation

As discussed in the previous subsections, a problem with explanation based on relevant users is trustworthiness and privacy concerns, because the target user may have no idea about other users that the explanation highlights who have "similar interests". 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 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 with social data. For example, Facebook provides common friends as explanation when recommending a new friend to a user (Figure 2.11(b)). Sharma and Cosley [2013] studied the effects of social expla-

nations in music recommendation context by providing the target user with the number of friends that liked the recommended item (Figure 2.11(b)), and founfound that explanations influence 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.

Figure 2.12: 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]).

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 2.12. Quijano-Sanchez et al. [2017] introduced a social explanation system applied to group recommendation, which significantly 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 of users,

taking factors such as marginal utility of persuasiveness, credibility of explanation, and reading cost into consideration.

2.6 Summary

In this section, we introduced several frequently seen display styles of recommendation explanations, including 1) Explanation based on relevant users or items, which presents nearest-neighbor users or items as explanation, and the relevant users or items are provided by user-based or item-based collaborative filtering methods. 2) Feature-based explanation, which provides users with the item features that match with the target user's interest profile as explanation, and this approach is closely related to content-based recommendation methods. 3) Textual sentence explanation, which provides the target use with an explanation sentence, and the sentence could be constructed based on pre-defined templates or generated directly based on natural language generation models. 4) Visual image explanation, which provides the user with a visual image as explanation, and the explanation can be reveal by the whole image or particular visual highlights in the image. 5) Social explanation, which provides explanations based on the target user's social relations that help to improve the user trust in recommendations and explanations.

3

Explainable Recommendation Models

Popular explainable recommendation approaches are 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, and 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

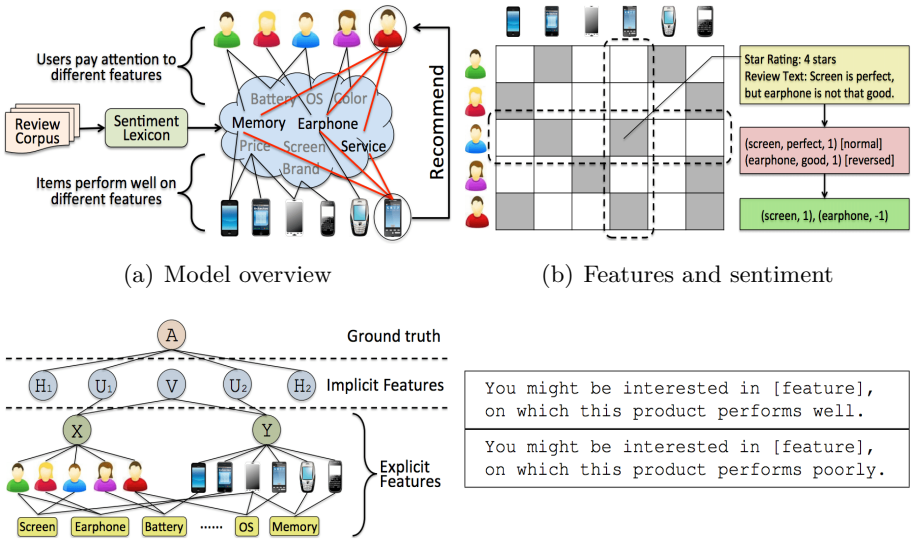
Seung [1999, 2001]), Max-Margin Matrix Factorization (MMMF) (Srebro et al. [2005], Rennie and Srebro [2005]), Probabilistic Matrix Factorization (PMF) (Mnih and Salakhutdinov [2008], Salakhutdinov and Mnih [2008]), and Localized Matrix Factorization (LMF) (Zhang et al. [2013b,a, 2014c]). Matrix factorization 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.

To learn for the rankings of items with implicit feedback, 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]). 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-

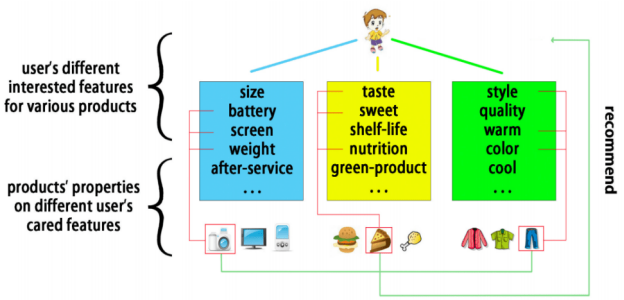
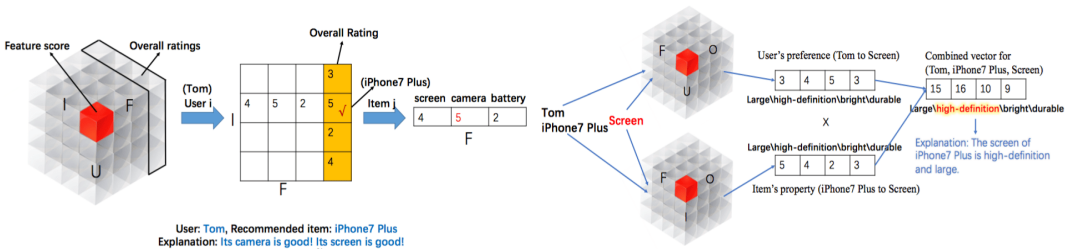


Figure 3.2: Learning to rank features for explainable recommendation over multiple categories based on tensor factorization.

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 (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.

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. For example, 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.

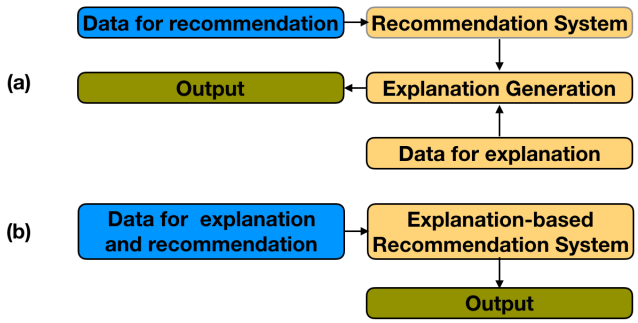


Figure 3.4: (a) Explainable recommendation with external data. (b) Explainable recommendation without external data support (Abdollahi and Nasraoui [2016b]).

Researchers have also investigated model-based approaches to generate explanations based on relevant users and/or items, 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 with naturally select those commonly purchased items from a user's neighbors as recommendation, and at the same time maintain high rating prediction accuracy.

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 information for recommendation, the proposed method improves rating prediction accuracy. Besides, by projecting each user latent vector onto the learned topics from LDA, 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.

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

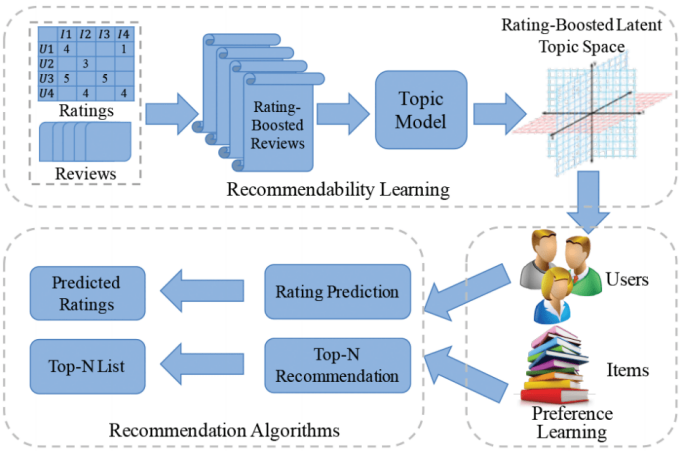
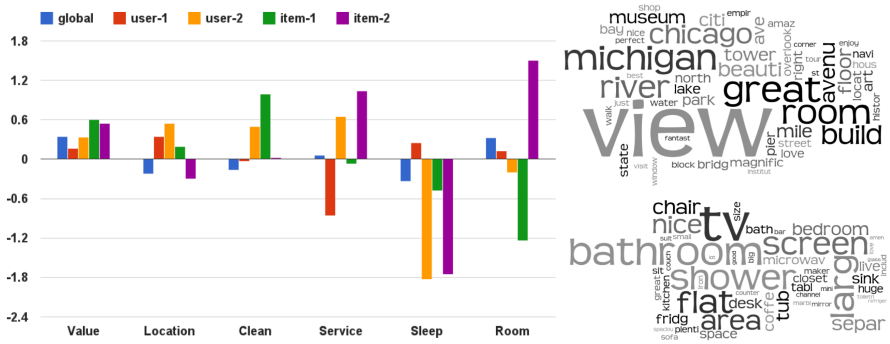


Figure 3.5: Recommendation framework for understanding users and items with ratings and reviews in "Rating-Boosted Latent Topics" model. Users and items are embedded as preference and recommendability vectors in the same space, which are later used by latent factorization model for both rating prediction and top-n recommendation. Courtesy image from Tan et al. [2016].

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 3.6(a). The proposed method showed improved performance for hotel recommendation based on TripAdvisor. Further, for the recommended hotels, it can provide the aspects of the hotel as a word cloud for explanation, as shown in Figure 3.6(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 the learned topical-aspect preferences of each user, which is similar to the topical clusters in McAuley and Leskovec [2013] but the difference is that the model can provide sentiment within each cluster as explanation.

Ren et al. [2017] introduced topic modeling approach to explain-



(a) (b)

Figure 3.6: (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].

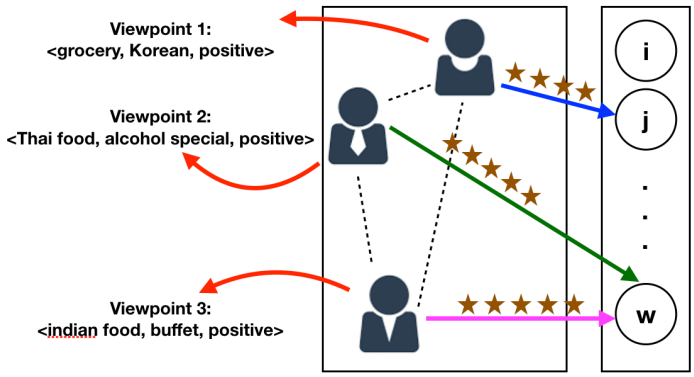
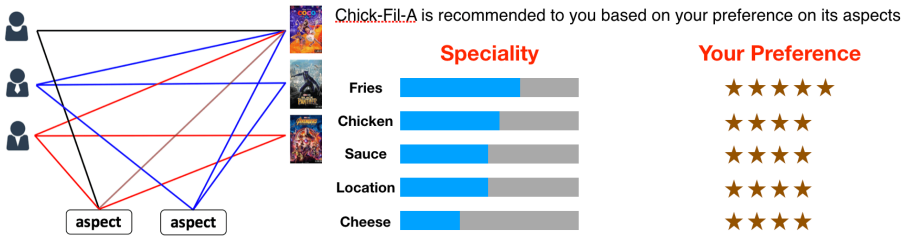


Figure 3.7: An example of trusted social relations, user reviews and ratings in a recommender system. Black arrows connect users with trusted social relations. "ThumpUp" symbols reflect the ratings of items. Concepts and topics have been highlighted in red and blue, respectively. Three viewpoints are represented in three different colors. A viewpoint is a mixture over a concept, a topic, and a sentiment (Ren et al. [2017]).

able recommendation for social recommendation. Specifically, the authors proposed social collaborative viewpoint regression (sCVR), where 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



(a) (b)

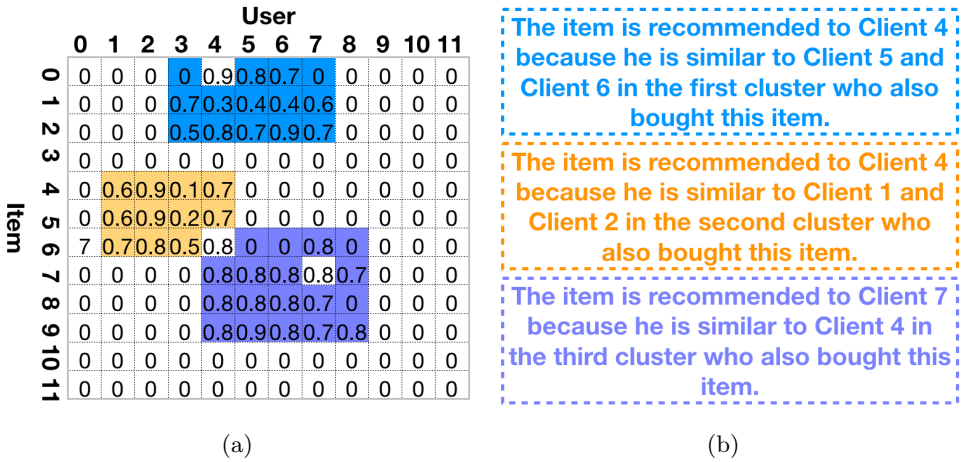
Figure 3.8: (a) An example tripartite structure of the TriRank algorithm. (b) A mock user interface for showing the explanation for recommending *Chick-Fil-A* to a user (He et al. [2015]).

Figure 3.7, 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.

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 3.8, 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 fitting 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 match-



(a) (b)

Figure 3.9: (a) Example of overlapping user-item co-clusters identified by the OCuLaR algorithm in (Heckel et al. [2017]). Colored 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.

ing 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 3.9. 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 mod-

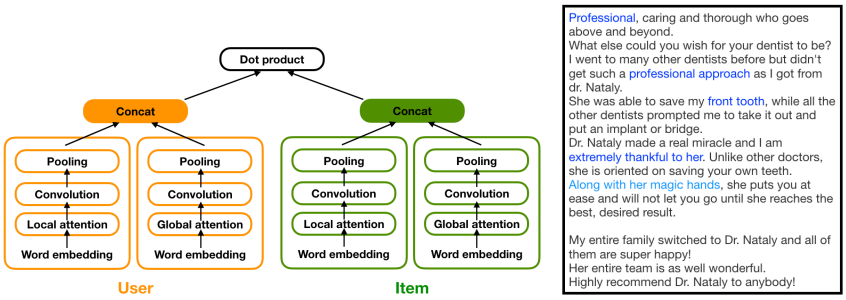
els with the explainability of tree-based models. In this model, the authors first employed a tree-based model to learn explicit decision rules based on cross features from rich side information, and then designed an embedding model that incorporates explicit cross features and generalize to unseen cross 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 been 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]), and many others, and they have also been applied to different recommendation tasks regarding explainability, such as top-n recommendation and sequential recommendation. 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 3.10. 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 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 atten-



(a) (b)

Figure 3.10: (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 (Seo et al. [2017]).

tively 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 parameters, the model can generate different explanations, as shown in Figure 3.11(a), which can be attractive for web users.

Also for generating natural language explanations, Chang et al. [2016] proposed another approach to generate natural language expla-

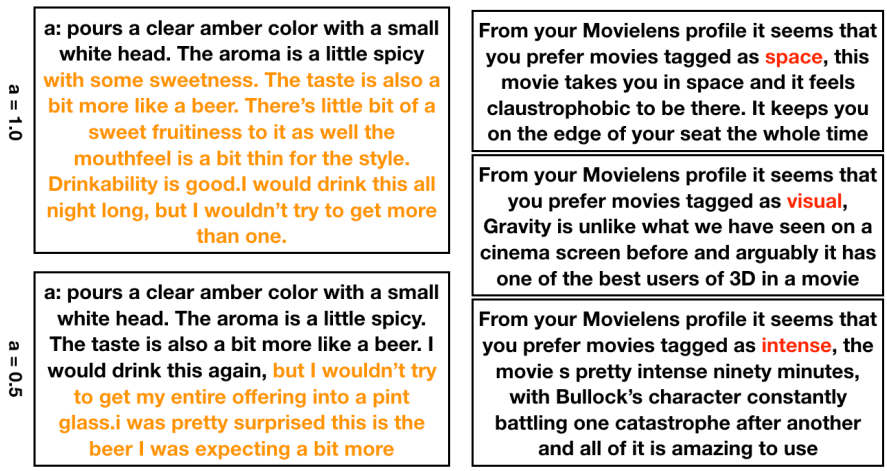


Figure 3.11: (a) The automatically generated textual reviews (explanations) based on natural language generation. Setting different model parameters will generate different explanations (Costa et al. [2017]). (b) Example natural language explanations for the movie "Gravity". Depending on the model of a user's interest, the system selects one of the three crowd-sourced explanations for the user (Chang et al. [2016]).

nations 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 3.11(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] designed an attention mechanisms over the user/item reviews for rating prediction. In this research, the authors claim that reviews written by others are critical reference information for users to make decision in e-commerce, however, the huge amount

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 3.12: The visually explainable recommendation model can also generate natural language explanations for the highlighted image regions (Chen et al. [2018c]).

of reviews for each product makes it difficult for consumers to examine all the reviews to evaluate a product. As a result, selecting and providing high-quality reviews for each product is an important approach to generate explanations. Specifically, the authors introduced an attention mechanism to learn the usefulness of reviews. Therefore, the highly-useful reviews can be adopted to provide review-level explanations, which help users to make 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 3.12, so that the users can understand why the particular images 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, which generates very readable natural language recommendation explanations, as shown in Figure 3.13.

As a type of (not necessarily very deep) neural network, restricted

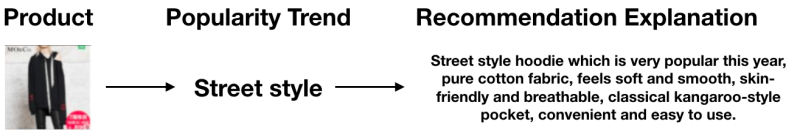


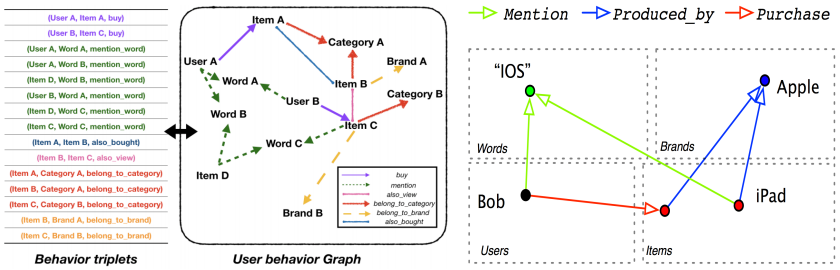
Figure 3.13: An example of the natural language recommendation explanations provided in Alibaba e-commerce recommendation system. The original explanations are in Chinese and the corresponding English translation is provided accordingly. The highlighted sentence is the sentence that contains the pre-specified keyword (in this case, style) to generate the explanation.

Boltzmann machines have also been used for recommendation systems Georgiev and Nakov [2013]. Abdollahi and Nasraoui [2016a] applied the idea of explainability 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.

3.6 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.

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



(a) A knowledge graph of users and items (b) Extracting explanation path

Figure 3.14: (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) Example explanation paths between a user Bob and a recommended item iPad in the product knowledge graph. Bob and iPad can be connected through a commonly mentioned word 'iOS' or the common brand 'Apple' from one of Bob's already purchased products.

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 explainable recommendation, as shown in Figure 3.14. 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. 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 item under the 'purchase' relation. Besides, explanations can be provided by finding the shortest path from the user to the recommended item through the knowledge graph.

To address the limitations of embedding-based and path-based methods for knowledge-graph aware recommendation, Wang et al. [2018a] proposed Ripple Network, an end-to-end framework to incorporate the knowledge graph into recommender systems. 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.

Huang et al. [2018] leveraged knowledge bases for recommendation with better explainability in a sequential recommendation setting. In particular, the authors bridged Recurrent Neural Network (RNN) with Key-Value Memory Network (KV-MN) for sequential recommendation, where the RNN component is used to capture the user's sequential preference on items, and the memory network component is enhanced with the knowledge of items to capture the users' attribute-based preferences. Finally, the sequential preferences together with the attribute-level preferences are combined as the final representation of user preference for recommendation. For explanation of the recommendations, the model can detect the attribute that is taking effect when predicting the recommended item. For example, it can detect whether the attribute of *album* is more important or the attribute of *singer* is more important for a particular music recommendation, where the attributes come from an external knowledge base. The model further enhances the explainability by providing value-level interpretability, i.e., suppose it is already known that some attribute (e.g., album) plays the key role in determining the recommendation, the model further predicts how the user will select among a set of entities for that attribute as explanations.

3.7 Association Rule Mining for Explainable Recommendation

Data mining approaches to recommendation is 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 association 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

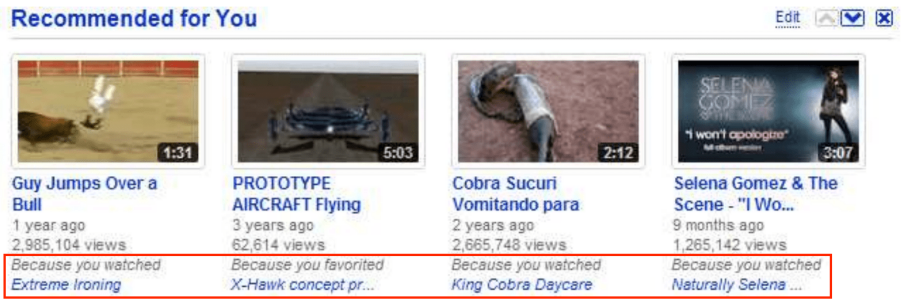


Figure 3.15: A screenshot of the recommendations module on the YouTube home page. The boxed explanations are generated based on association rule mining. Courtesy image from Davidson et al. [2010].

added to playlists. The related videos of these seed videos are taken 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 3.15.

3.8 Post Hoc Explanation

Sometimes in practice, the recommendation explanation is not generated from the recommendation model itself. Instead, it is generated by a post-hoc explanation model after an item has been recommended by the recommendation model.

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.

Peake and Wang [2018] provided a theoretical study on the post-hoc explanation strategy for recommendation systems. In particular, the authors treated the recommendation model – in this paper, a matrix

factorization model – as a black box, and attempted to prove that approximating a black-box model with an interpretable model can maintain the high predictive accuracy of the recommendation model whilst improving interpretability by extracting explanations that can be used to understand the model behavior. Technically, the authors proposed an approach to extracting explanations for latent factor recommendation systems by training association rules on the output of a matrix factorization black-box model. More specially, the completed rating prediction matrix of a latent factor model is taken as the input of an association rule-based explanation model, and the top D predicted items for each user is considered as a transaction, and transactions of all users constitute the transaction corpus, which is used to mine association rules for explanation. The proposed approach is able to extract second order explanation rules such as "$\{X \Rightarrow Y\}$: Because you watched X, we recommend Y".

Singh and Anand [2018] studied the post hoc explanations of learning-to-rank algorithms in term of web search. In this work, the authors focused on how best we can understand the decisions made by a ranker in a post-hoc model agnostic manner, in particular, the explainability of rankings is based on an interpretable feature space. Technically, the authors first train a blackbox ranker, and then use the ranking labels produced by the blackbox ranker model as secondary training data to train an explainable tree-based model, where the tree-based model is the post-hoc explanation model that generates explanations for the ranking list.

In this sense, Peake and Wang [2018] can be considered as a point-wise post-hoc explanation model, while Singh and Anand [2018] can be considered as a pair-wise or list-wise post-hoc explanation model.

3.9 Summary

In this section, we have aggregated and introduced key approaches to explainable recommendation. We first provided a broad overview of machine learning techniques for personalized recommendation, and then, we introduced key techniques for explainable recommendation,

including matrix/tensor-factorization approaches, topic modeling approaches, graph-based models, deep learning approaches, knowledge-base embedding approaches, association rule mining approaches, and finally post-hoc explanation approaches.

As in section 1.1, explainable recommendation can consider both the explainability of recommendation methods (i.e., process) and the explainability of recommendation results (i.e. product). When considering the explainability of methods, explainable recommendation aims to devise interpretable models that work in a human way, and such models usually also lead to the explainability of recommendation results (i.e., product). In this section, the matrix/tensor-factorization, topic modeling, graph-based, deep learning, knowledge-enhanced, and associate rule mining approaches adopt this philosophy – they aim to understand how the process works.

Another philosophy for explainable recommendation is that we only focus on the explainability of recommendation results. In this way, we treat the recommendation model as a complex blackbox and ignore its explainability, but instead develop separate methods to explain the recommendation results produced by this blackbox. In this section, the post-hoc explainable recommendation approach falls into this category.

Overall, explainable recommendation research covers a wide range of techniques and algorithms, and can be realized in many different ways in real-world systems.

4

Evaluation of Explainable Recommendation

In this section, we provide a review of the evaluation methods for explainable recommendation. It would be desirable if an explainable recommendation model can achieve comparable or even better recommendation performance than conventional "non-explainable" methods, and meanwhile achieve better explainability.

As a result, explainable recommendation algorithms should primarily evaluate the recommendation performance in terms of rating prediction or top-n recommendation. On top of this, it is also encouraged to evaluate the explanation performance in terms of persuasiveness, effectiveness, and other perspectives. Sometimes, there could be more than one explanations for the same recommendation, so we may also need to evaluate the quality of different explanations.

In this section, we introduce commonly used evaluation measures and protocols for both recommendation performance evaluation and explanation evaluation.

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, and normalized discounted cumulative gain (NDCG). 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), and Area Under the Curve (AUC). 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. Peake and Wang [2018] further generalized the idea and proposed *Model Fidelity* as a measure to evaluate explainable recommendation algorithms, which is defined as the percentage of explainable items in the recommended items:

$$Model\ Fidelity = \frac{|explainable\ items \cap recommended\ items|}{|recommended\ items|}$$

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 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 evaluating 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 accept the recommendations.

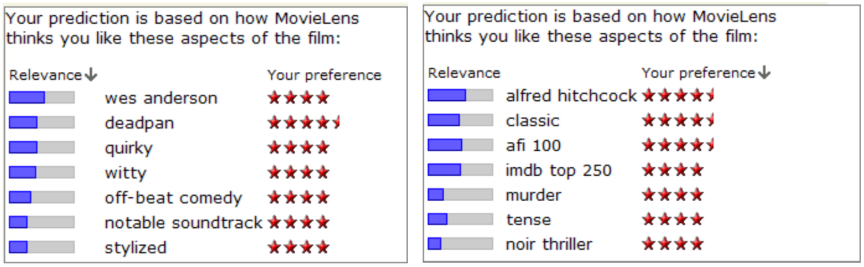
Zhang et al. [2014a] conducted online experiments focusing on how the explains affect users' acceptance of the recommendation (i.e. persuasiveness). The authors conducted A/B-tests based on a popular commercial web browser which has more than 100 million users with 26% monthly active users, and the experiments attempted to recommend relevant phones when a user is browsing mobile phones in an

Figure 4.1: 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. 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.

online shopping website, as shown in Figure 4.1. 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.

It should be noted that the available evaluation measure in online evaluation could vary depending on the available resources in the testing environment. For example, one may evaluate based on click-through-rate (CTR) when user click information is available, or calculate the purchase rate if user purchase action can be tracked, or even calculate the gross profit if product price information is available.



(a) RelSort interface (b) PrefSort interface

Figure 4.2: Some explanation interfaces for online user study 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.

4.2.3 Online 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.

Vig et al. [2009] conducted an online study for four explanation interfaces based on MovieLens website, where the four interfaces are RelSort, PrefSort, RelOnly, and PrefOnly, as shown in Figure 4.2. Subjects completed an online survey in which they evaluated each interfaces on how well it helped them (1) understand why an item was recommended

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

(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.

To evaluate the explanations, Wang et al. [2018b] recruited study participants through Amazon Mechanical Turk to perform the evaluation in a diverse population of users. The study is based on the review data in Amazon and Yelp datasets. For each participant, the authors randomly selected an existing user from the review datasets, and presented this user's previous reviews to the participant to read. Participants are then asked to infer the selected user's preference from these reviews. Then they will be asked to judge the provided recommendations and explanations by answering several survey questions from this assigned user's perspective.

Except for large-scale online workers, we can also conduct user study with relative small-scale volunteers, paid subjects, or manually labeling the explanations. For example, Wang and Benbasat [2007] adopted a user study approach based on surveys to investigate the trust and understandability of content-based explanations. They examined the effects of three types of explanations about a recommendation system and its use – how, why, and trade-off explanations – on consumers' trusting beliefs in competence, benevolence, and integrity. The authors built a recommendation system as the experimental platform and conducted laboratory experiments, where the results confirmed the important role of explanation facilities in enhancing consumers' initial trusting beliefs and indicated that consumers' use of different types of explanations enhances different trusting beliefs. 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.

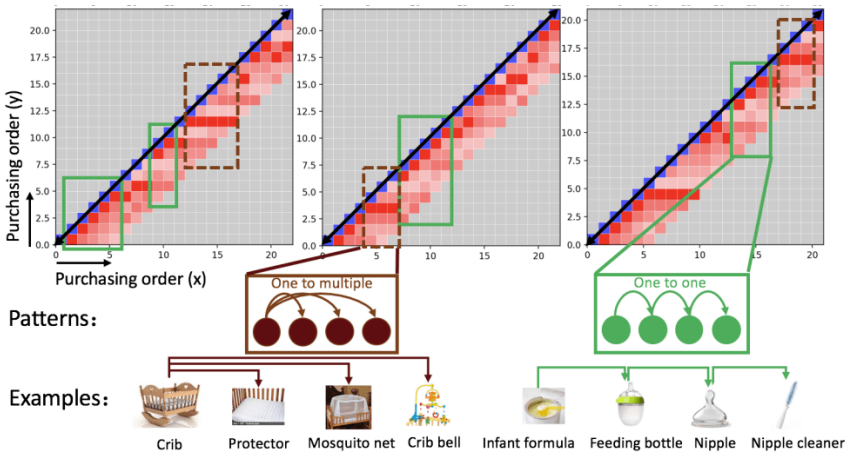


Figure 4.3: Case study of user sequential behaviors in e-commerce for explanation of sequential recommendations. Recommendations can be explained with "one-to-multiple" behaviors, for example, a mom bought a crib for her baby, after that she bought a water-proof mattress protector and a mosquito net 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.

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.

Chen et al. [2018b] provided case study to explain the sequential recommendations, as shown in Figure 4.3. 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

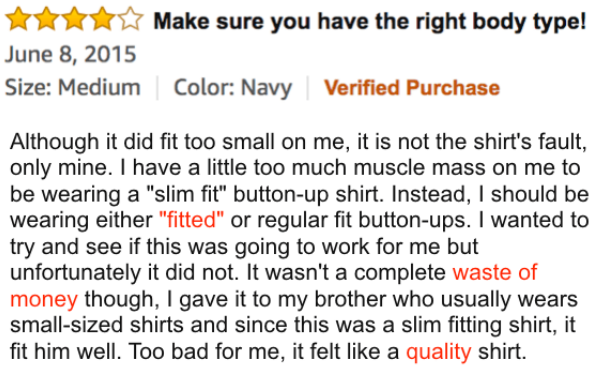
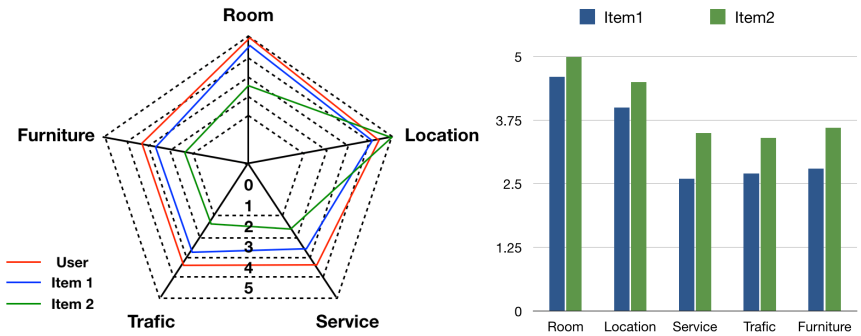


Figure 4.4: Case study of the reviews of a user. This case study shows that the explainable recommendation model is able to provide recommendations based on the aspects that a user liked in his/her reviews (He et al. [2015]).

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 4.4. From the reviews, the authors found that the user is interested in "chicken", although she gives low ratings to the two businesses, and the recommendation model ranks restaurant "Chick-Fil-A" highly 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 4.5, for the selected user, item 1 is recommended instead of item 2, and by examining the user preference and item quality, this recommendation is explained by the fact that item 1 satisfies user preferences on most aspects.



(a) Radar chart of a user

(b) Item aspect quality of two items

Figure 4.5: 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.

4.3 Summary

In this section, we introduced the evaluation methods for explainable recommendation. A desirable explainable recommendation method would not only be able to provide high-quality recommendations but also high-quality explanations, as a result, an explainable recommendation method should be evaluated in terms of both perspectives.

In this section, we first presented frequently used evaluation methods for recommendation evaluation, including both online and offline approaches. We further introduced methods for explanation evaluation, including both quantitative methods and qualitative methods. More specifically, quantitative methods include online approaches, offline approaches, and user study approaches, while qualitative evaluation is usually realized by conducting case study of the generated explanations.

5

Explainable Recommendation in Different Applications

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

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. It has been shown that by providing appropriate explanations in e-commerce, it can help to improve the transparency, scrutability, trustworthiness, effectiveness, persuasiveness, efficiency, and satisfaction of the product

recommendations (Tintarev [2007]).

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 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] takes 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, or museum. Explainable POI recommendation has also gained great interest in recent years. Most of the explainable recommendation research are based on location review website datasets, such as Yelp[1] and TripAdvisor[2].

By providing appropriate explanations in POI recommendation systems, it helps users to save time and to minimize the potential cost of making wrong decisions, because traveling from one place to another usually means extensive efforts in terms of time and money. Besides, by providing explanations in some travel-related POI applications such as TripAdvisor, it will help users better understand the relationship between different places, which could help the user to plan better routes of the trip in advance.

In terms of the research on explainable POI recommendation, 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.

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

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, and tags in a social environment.

Explainability of social recommender systems is vitally important to increase the trustworthiness of the recommender systems, and trustworthiness is a fundamental factor to maintain the sustainability of social networks. For example, by providing the common friends as explanations for friend recommendation in Facebook, it helps the user to understand why an unknown person is related to him/her and why the recommended friend would be trusted; and by telling the user which of his/her friends have twitted a news as explanation in Twitter, it helps the user to understand why the recommended news could be important or attractive.

In terms of research on social explainable recommendation, 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 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 Multimedia Recommendation

Explainable multimedia 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. Providing explanations for multimedia recommendations could help users to make informed decisions more efficiently with less trail and errors, which can help to save time by reducing unnecessary data-intensive media transmission through networks. We review the related work on explainable multimedia recommendation in this section.

The MovieLens dataset[3] 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, and director, 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 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

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

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, and healthcare recommendation. 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, and similarly for explainable recommendation in other scenarios.

5.6 Summary

In this section, we introduced several application scenarios of explainable recommendation to help readers understand how the idea of explanation is applied to different recommendation scenarios. In particular, we introduced explainable e-commerce, POI, social, and multimedia recommendations, as well as the related research for each application. We also briefly touched some new explainable recommendation tasks such as explainable academic, citation, and healthcare recommendations, which have been attracting attention in recent years.

It is also beneficial to discuss potential limitations of explainable recommendation, i.e., although explanations can be helpful to a lot of recommendation scenarios, there could exist scenarios where explanations are not needed or could even hurt, these include time critical

cases where decisions should be made in a real-time way, and users are not expected to spend time evaluating the decision. For example, while driving on highways users may want to know the correct decision about which exit to take without spending time listening to the explanations. More critical scenarios include emergency medical decisions or battlefield decisions where wasting time for evaluation is not permitted. Depending on the particular scenario, explainable recommendation systems may also need to avoid providing too much explanation, and to avoid repetitive explanations, explaining the obvious, or explaining in too much details, which may hurt in rather than improve the user experience.

6

Open Directions and New Perspectives

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

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, and 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 it-

self explainable for recommendation, and this will not only benefit the personalized recommendation research, but also many other research area such as personalized healthcare, personalized online education, chatbots, and self-autonomous systems.

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 or visual images. 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 knowl-

edge. And in a more general sense, this represents one of the future directions for the research of intelligent systems, which is to integrate rational and empirical approaches to 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, and other information. 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 a wide range of research tasks such as multi-modal explanation by aligning two or more different information sources, transfer learning over heterogeneous information sources for explainable recommendation, and cross-domain explanation in information retrieval and recommendation systems.

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. 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 lan-

guage 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 real-world 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, effi-

ciency, transparency, trustworthiness, and user satisfaction. 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 explana-

tions 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), and *who* to recommend (social recommendation). 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

Early IR and recommendation models – such as exact match Boolean search and user/item-based collaborative filtering – were very transparent and explainable, and the lack of transparency of best match ranking models has been known and studied as a serious downside from the start. Recent advances on more complex models – such as latent factor models or deep representation learning – have helped a lot to improve the performance of search and recommendation systems, but they also bring about the difficulty on transparency and explainability.

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 com-

plicated decision making by providing supportive information and evidence. 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. 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/multimedia 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 broader 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. In *Acm sigmod record*, volume 22, pages 207–216. ACM, 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.

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.

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.

William John Clancey. The epistemology of a rule-based expert system: A framework for explanation. Technical report, STANFORD UNIV CA DEPT OF COMPUTER SCIENCE, 1982.

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.

Kostadin Georgiev and Preslav Nakov. A non-iid framework for collaborative filtering with restricted boltzmann machines. In *International conference on machine learning*, pages 1148–1156, 2013.

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. 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.

Jin Huang, Wayne Xin Zhao, Hongjian Dou, Ji-Rong Wen, and Edward Y Chang. Improving sequential recommendation with knowledge-enhanced memory networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 505–514. ACM, 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.

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

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.

Yujie Lin, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Jun Ma, and Maarten de Rijke. Explainable fashion recommendation with joint outfit matching and comment generation. *arXiv preprint arXiv:1806.08977*, 2018.

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.

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.

Georgina Peake and Jun Wang. Explanation mining: Post hoc interpretability of latent factor models for recommendation systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2060–2069. ACM, 2018.

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.

Jaspreet Singh and Avishek Anand. Posthoc interpretability of learning to rank models using secondary training data. *Proceedings of the SIGIR 2018 International Workshop on ExplainAble Recommendation and Search (EARS)*, 2018.

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.