Explanations In Recommender Systems: Overview And Research Approaches

MOHAMMED Z. AL-TAIE
Computer Science Department, Al-Salam University College, Iraq
mza004@live.aul.edu.lb

Abstract: Recommender systems are software tools that supply users with suggestions for items to buy. However, it was found that many recommender systems functioned as black boxes and did not provide transparency or any information on how their internal parts work. Therefore, explanations were used to show why a specific recommendation was provided. The importance of explanations has been approved in a number of fields such as expert systems, decision support systems, intelligent tutoring systems and data explanation systems. It was found that not generating a suitable explanation might degrade the performance of recommender systems, their applicability and eventually their value for monetization. Our goal in this paper is to provide a comprehensive review on the main research fields of explanations in recommender systems along with suitable examples from literature. Open challenges in the field are also manifested. The results show that most of the work in the field focus on the set of characteristics that can be associated with explanations: transparency, validity, scrutability, trust, relevance, persuasiveness, comprehensibility, effectiveness, efficiency, satisfaction and education. All of these characteristics can increase the system's trustworthiness. Other research areas include explanation interfaces, over and underestimation and decision making.

Keywords — Recommender Systems; Explanations; Explanation Styles; Explanation Attributes; Decision Making; Research Approaches;

Received July 31, 2013; accepted September 19, 2013

1. Introduction

In recommender systems (RSs), an explanation can be defined as a bit of information that serves different goals such as why a specific recommendation was given, and how to maintain better communication styles in commercial transactions [15]. Another definition for an explanation is that it is a description that makes users better realize if the recommended item is relevant to their needs or not [17]. It can also be defined as an important piece of information that is used by both selling and buying agents through their communication process to increase their performance [7].

Explanations in recommender systems have gained an increasing importance in the last few years. Although they cannot completely compensate for poor recommendations, it was found that they can increase user acceptance of Collaborative Filtering (CF) recommender systems, help them make decisions more quickly, convince them to buy and develop users' trust in the system as a whole [4]. It was also found that providing explanations along with recommendations would lead to better understand the recommender system and establish a "sense of forgiveness" when users do not like the new recommended items [1].

Explanations can benefit from the advances that come from a number of disciplines such as intelligent systems, human-computer interaction, and information systems.

The purpose of this paper is to give an overview of the area of explanations in recommender systems in addition to survey existing research approaches. Even though previous studies such as [14] and [12] provided a review on the field (with a special focus on explanation attributes) they failed to comprehend all important research approaches.

The main contribution of this work is that it surveys the main strands of research in the field and put them in one complete work.

The remainder of this paper is organized as follows. The next subsections review the relevant literature on explanations in RSs, their history, their types and their styles. Section 2 addresses the main research approaches in explanations. Section 3 discusses open challenges and finally section 4 is for conclusions and future work.

1.1 Related Work

A number of studies, from different perspectives, have studied the effect of explanations in the performance of recommender systems or explored different variants of explanations.

For example, Tintarev and Masthoff [14], [16] discussed the influence of a number of explanation characteristics in the behavior of the system: transparency, scrutability, trust, persuasiveness, effectiveness, efficiency, satisfaction. All of these characteristics can help increase the system trustworthiness.

Herlocker et al. [6] compared the performance of 21 explanation interfaces and sought the best techniques that help support explanations in a collaborative-filtering recommendation environment. Gedikli et al. [4] evaluated a number of explanation attributes: satisfaction, efficiency, effectiveness, and trust in recommender systems through
The inputs to all influence the style of explanation used, even or has no control in decision systems.

- explanations that deal are the ones that are robust

1.2 History

Due to the decline of studies in Expert Systems (ES) in the 1990s, recommender systems borrowed the concepts of explanations. Therefore, expert systems, which were based on heuristics, case-based reasoning and model-based approaches, can be considered the predecessors of recommender systems. The importance of explanations in ES has been well identified, especially in the domain of medical decision making systems such as MYCIN - a medical expert system developed during 1970s that contained an explanation facility [14]. However, the use of explanations within recommender systems requires the use of an approach different from the rule-based reasoning used before [6].

Since the dawn of RSs, most of research has focused on finding the most accurate recommendation algorithms such as CF systems and Content-Based Filtering (CF) systems and with less care for supplying transparency that show how these systems work, which, if present, can contribute to users' trust in the system.

Today, RSs are becoming increasingly popular in diverse application domains such as music, videos, queries, friends on social networks, news and are being developed for either research purposes such as the work being done by GroupLens research group, or for commercial purposes such as Amazon.com 2 and Netflix.com 3 [2].

Since the seminal study by Herlocker et al. [6] in 2000 on explanation interfaces, which stated that recommender systems had worked as black boxes, the body of research in explanations began to grow [17].

Recommender system, now, is an active research area in data mining and machine learning. Some of the well-known conferences that discuss explanations in RSs amongst other field-related topics include RecSys (the ACM Conference Series on Recommender Systems), SIGIR (Special Interest Group on Information Retrieval) and KDD (Knowledge Discovery and Data Mining). Explanations-Aware Computing (ExACom) workshop series aims to draw on multiple perspectives on explanation and examine how explanation can be applied to further the development of robust and intelligent systems.

1.3 Types Of Explanations

A generic approach for producing explanations is still difficult due to the fact that there isn't an accepted unified theory that describes the concepts of explanations. In any case, depending on circumstances, explanations can be in any of the following forms [7]:

- Functional explanations: explanations that deal with the function of systems.
- Casual explanations: explanations that provide casual relationships between events.
- Intentional explanations: explanations that give reasons for human behavior.
- Scientific explanations: explanations that are used to express relations between the concepts of different scientific fields and are usually based on refutable theories.

Explanations, depending on the type of intermediary unit that relates the user to the recommended item, fall into three types: item-based: where items are the intermediary entities, user-based: where other users are the intermediary entities, and feature-based: where the characteristics of the recommended item are the intermediary entities. Tags were also used in this context as intermediary entities [19].

Depending on the provisioning mechanism of explanations, they can also be classified into the following three types [5]:

- User-invoked explanations: are the ones that are accessed upon a request of the user.
- Automatic explanations: are the ones that are provided at all times and the user has no control in this case.
- Intelligent explanations: the system, in this case, provides the user with the explanation, whenever it appears to be a need for it.

1.4 Explanation Styles

In essence, the underlying algorithm of a recommender engine will influence the style of explanation used, even though these explanations do not necessarily reflect the underlying algorithm (Fig. 1).

The most commonly used explanation styles, according to

![Figure 1: Explanation styles](http://www.groupLens.org/)

1. Collaborative-based style explanations: The inputs to the recommender engine are user's ratings of an item so that these ratings are used to find users with similar
2. **Content-based style explanations**: this type considers similarity between items, putting in mind their properties. An example for this type is: “... was recommended because you said you owned Book A”. Other types of explanation styles that come under this type include Keyword Style Explanation (KSE) and Tag Style Explanation (TSE) with its both types: Tag preference, which is the user's sentiment toward a tag, and Tag relevance, which is the extent to which a tag describes an item.

3. **Case-based reasoning style explanations**: the focus here is on the similar items that are used to make the recommendation. An example for this type is: “Because you have selected or highly rated: Movie A”. Sub-types here are such as the Influence Style Explanation (ISE) where the influence of an item on the recommendation is measured by the difference in the score of the recommendation with and without that item.

4. **Knowledge and utility-based style explanations**: the input to the system in this case is the description of user's needs and interests. The recommender engine then suggests recommendations based on these needs. Example “... because your profile has: You are single; you have a high budget”.

It is accepted that there is some overlap between three types of explanation styles: the knowledge-based style, the content-based style and the case-based style explanations.

5. **Demographic style explanations**: the input to the recommender engine is demographic information about a user while the algorithm identifies users that are demographically similar to the first one. This type of explanation style is the least one used by commercial recommender systems today due to the sensitivity of demographic information revealed. For example "We recommend you the movie ... because you are a male aged 16-32”.

2. **Research Approaches**

In this section, we address the fields that took much care from RSs researchers in regard to explanations. These can be either fields that are related to the aims that explanations in recommender systems are found for, or fields that discuss other issues related to explanations:

### 2.1 Explanation Attributes

The research here is focused on explanation attributes that represent the benefits explanations provide to recommender systems (Fig. 3). They can also be considered the criteria that measure to what extent an explanation is good [14]. These attributes may interrelate therefore; they require certain trade-offs. For example, personalized explanations may lead to greater satisfaction but not necessarily increase effectiveness. Also, transparency does not necessarily aid trust. On the other hand, some attributes of explanations may lead to achieve more than one goal. For example, by measuring understandability; transparency, trust and satisfaction are also determined.

These metrics (attributes) can be influenced not only by the underlying recommendation algorithm but also by the way recommendations are presented to customers and the method of interacting with the recommendations [17]. These attributes are:

1. **Transparency**: means that explanations show how recommendations were given [15]. While transparency gives an overview on how a recommendation was selected and how the overall system works, justification is a property of the recommender system that shows if the system, through explanations, is able to help users understand why an item was recommended. However, it may not reveal how the algorithm of the recommender system really works [19]. Certain cases require the use of justification instead of transparency such as in the difficulty of explaining some algorithms, for protection of business secrets, and for greater freedom designing the explanations. If a recommendation does not describe how the recommendation was selected, then it is not transparent [15].

2. **Validity**: explanations can be used to give the space to users to check the validity of recommendations by showing a comparison of the required and offered item features. Although similar, validity is not necessarily related to transparency [7].

3. **Scrutability**: means that users can tell if the system is wrong. Scrutability is closely related to transparency, follows it and allows users to correct reasoning made by the system. Users will trust the system that provides transparency into the working of that system by
providing explanations containing a mechanism to manage errors within recommendations. It was not likely that users try to correct system mistakes by their own therefore; an extra effort was needed to make this process easier to achieve [15].

4. **Trust (Confidence):** users' trust in a system can increase when using recommendations. In uncertain situations, trust building can be seen as a strategy to reduce the complexity of making decisions [7]. It was found that users intend to return to recommender systems which they find trustworthy. It is usually related to transparency since the possibility to interact with the system can increase users' trust and it is dependent on the accuracy of the recommendation algorithm [15]. Amazon's recommendation system, which recommends items with similar features, was found to increase users' trust and consequently increase sales rates. Pu and Chen [9] developed an organization-based explanation interfaces with concepts of diversity, tradeoffs reasoning and explanations. Their aim was to discover the benefits of using explanations for trust building and whether the system features can provide trust-related benefits.

5. **Relevance:** explanations can be used to explain why, in some cases, it is needed from the user to provide additional information. This is true in the case of conversational recommender systems, where customers keep interacting with the system in order to refine their profiles [7].

6. **Persuasiveness:** this implies convincing users to try to buy. Explanations may enhance user acceptance of the system or the recommendation. However, too much persuation may have reverse reactions if the item recommended by the system was found to be unacceptable [15]. Persuasiveness can be measured as the difference between a previous rating and a second re-rating using an explanation interface. It can also be measured as the difference between the people buying items using a system with explanation interface and the people using a system without that facility [14].

7. **Comprehensibility:** explanations can support recommender systems by relating the user's known knowledge to the knowledge used by the recommender systems since these systems can never be sure about the knowledge of their users until other methods that can provide help are present [7].

8. **Effectiveness:** means helping users make good decisions about what to buy, rent, watch, etc. Effectiveness is closely related to accuracy measures such Mean Average Error (MAE), precision and recall as it is highly dependent on the accuracy of the recommender algorithm. One of the benefits of effectiveness is that it can introduce a whole bunch of items to new users and thus, those users can understand the full range of available options [15]. Effectiveness can be measured as how much people still like an item they have bought after consuming it. Another way to measure effectiveness is to see if people who receive explanations end up with items more suited to their needs as compared to people not using systems with explanation facilities. It is also possible to use a metric from marketing to find the "best item" rather than just "good enough items." [14].

Bilgic and Mooney [2] evaluated the effectiveness property from two sides: *promotion*, which is the success in convincing users to buy items, and *satisfaction*, which means that the item is best fits users' needs. Satisfaction was considered more important to users since it develops users' trust in using the system. The evaluation process was done by studying the performance of three explanation systems: the KSE, the NSE and the ISE. Tintarev and Masthoff [16] evaluated the effectiveness property of explanations in the movie domain by studying what makes moviegoers like a movie. They conducted two focus groups to know if specific movie features make users see and like a movie. A number of movie features were found to be effective in this context: good in its genre, subject matter, director, originality, script complexity, initial expectations, visuals, mood, cast and realistic. In [17], Tintarev and Masthoff conducted a number of empirical studies in two domains: movies and cameras as the goal was to measure the effect of personalized feature-based explanations on both effectiveness and satisfaction.

9. **Efficiency:** means that explanations can make it faster for users to choose the right product. It measures how quickly a task can be performed. Efficiency is one of the most talked about criteria in recommender systems literature. It is often used in the evaluation of conversational recommender systems [15] and it can be measured as the total interaction time or the number of interactions needed to find good items. In other systems, efficiency is measured as the number of inspected explanations or the number of activations of repair actions when no good items are found [14].

10. **Satisfaction:** means that explanations can make users more satisfied with the overall system. It was found that the longer the description of an item, the better the perceived usefulness and ease of use of the recommender system [15]. Satisfaction can be measured either directly by asking users if they liked the system or indirectly by measuring user loyalty [14]. Symeonidis et al. [11] proposed a prototype called MovieExplain which is a movie recommender system that is supported by explanations. They tried to evaluate users' satisfaction in their explanation style, which was called (KISE) against the KSE and the ISE style explanations. Gedikli et al. [4] evaluated satisfaction along with efficiency, effectiveness, and trust of explanations in recommender systems by comparing the performance of Tag Clouds (TC) explanation method, a frequently used visualization and interaction technique on the web, against the performance of the KSE method and the Personalized Tag Clouds (PTC) Method.

11. **Education:** in order to make better purchasing decisions, explanations can be used to help users better understand the product domain by providing them with deep knowledge about that specific domain [7].
2.2 Other Research Approaches

Here, we address research fields that discuss other issues related to explanations.

1. Explanation Interfaces

Explanation interface is the technique mainly used to explain why a recommendation was suggested to a user. A good explanation presentation can better explain recommendations and even push them to make further requests.

When designing the recommender system interface, a number of factors must be present. In regard to explanations, these factors are (i) Level of detail: means how detailed the explanations are, (ii) System transparency: do explanations provide transparency? (iii) Modality: type of explanation modality used (e.g. text, graphs, tables, images etc.) and (iv) provisioning mechanism: are explanations automatically provided or they need to be invoked by the user? [1].

Explanations in RSS may need to be presented in a limited space and there should be a balance that takes into account the number of features to mention and the degree of detail for each feature. In all cases, there should be a limited number of item features otherwise; the system will suffer from severe space limitations [12].

Herlocker et al. [6] sought the best models and techniques that help support explanations in Automatic Collaborative Filtering (ACF) systems.

Xu [20] extensively discussed the various aspects of explanation interfaces in recommender systems and proposed a design framework for building the explanation interface for collaborative filtering recommender systems.

Among other people who discussed the importance of explanation interfaces to recommender systems with the aim of increasing the effectiveness of explanations are Friedrich and Zanker [3] and Tintarev and Masthoff [13], while Gonul et al. [5] discussed in detail the content, the provisioning mechanism and the format of presentation which collectively form an explanation.

2. Over and Underestimation

In recommender systems, overestimation means that users may try a product they do not end up liking. On the other hand, underestimation means that users miss products they might have appreciated. Therefore, overestimation may result in false positives while underestimation may result in false negatives [18].

Skews in valuation of recommendations come due to malfunction in any of the following factors: the quality of information used to form recommendations, the recommendation accuracy, the nature of the recommended object and the presentation of the recommended items [17], [15], [18].

Another source for skews in valuation of recommendation may result from additional information, such as explanations, which can either aid or hinder decision support since these explanations may contain both positive and negative information and thus may have a polarity.

Tintarev and Masthoff [18] investigated the phenomenon of over and underestimation in recommendations from the angle of explanations. Their results show that overestimation was considered more severely than underestimation in regard to both perceived effectiveness and high investment domains compared to low investment domains.

3. Decision Making

Decision making, which is an essential leadership skill, can be defined as the selection of a course of action from among two or more alternatives in order to arrive at a solution for a given problem. For effective decision making, a person must be able to forecast the outcome of each option, and determine which option is the best for that specific situation. Although studies of decision making have not been directly addressed by researchers of recommender systems, other scientific disciplines (e.g. psychology and sociology) have been examining this topic for some time.

Due to the huge amount of information that overwhelmed us since the Internet revolution, decision makers (users in our case) seek various forms of information assistance to help them make better decisions. What makes Decision Support Systems (DSS) (recommender systems in our case) successful is not only the accuracy of its given information but also how users receive information. It has been found that users are influenced by the format of the information given and thus make different decisions in different contexts.

Therefore, recommender systems are required to integrate a deep knowledge about human decision making with the recommendation algorithm [8]. Since not all types of explanations are accepted and that the underlying structural properties of explanations could influence their perspective effectiveness, Gonul et al. [5] examined the effects of structural properties of explanations within the DDS domain. Via an experimental study, they studied the effects of explanation length and the confidence level on the acceptance of explanations. They found that long and strongly confident explanations are more effective than short and weakly connected explanations in the context of explanation acceptance. Sharma and Cosley [10] developed a framework for realizing what effects social explanations have on decision making in a music recommendation context. Social explanations are based on the idea that people will follow other people's points of view according to the assumption that those people have reasons for doing these things. Facebook is an example for social websites that provide social explanations powered by the "Like" button. For instance, how many people in general, or how many of a person's own friends have liked the item. Their results show that social explanations can influence people's opinions about artists even though people's preferences varied quite a bit between artists.

3. Open Challenges

A number of challenges are still waiting to be investigated by people working in recommender systems field. One research line is to compare between explanation styles with regard to their performance on explanatory goals [19], [17], and [4]. As we have seen, these goals can interrelate which necessitates certain trade-offs, such as the relation between satisfaction and effectiveness. Also, one characteristic may achieve more than one goal such as when measuring understandability; transparency, trust and satisfaction can also be measured. Another research line is to devise more advanced explanation interfaces in order to increase system
effectiveness [3], better explain recommendations and push users to make further requests [15]. This trend can be found for example in [14].

4. CONCLUSIONS AND FUTURE WORK

A system’s ability to explain its recommendations is a crucial aspect of its usability and utility and can contribute significantly to users’ acceptance of its suggestions. Work related to explanations can be found in a number of disciplines such as cognitive science, psychology, and philosophy. They provide us with a mechanism for handling errors that come with a recommendation since users will trust a recommendation when they know the reason(s) behind it.

In this article, we reviewed some work on explanations in recommender systems. Work in this area has been motivated to a high degree by the aim of removing the black box from around the recommender system. In section 1, a list of previous works followed by the history, the types and the styles of explanations were given. In section 2, research approaches, supported by suitable examples from literature, were outlined while section 3 discussed the main lines of research to be followed by researches. The study shows that a set of explanation characteristics were found to increase the system’s trustworthiness: transparency, validity, scrutability, trust, relevance, persuasiveness, comprehensibility, effectiveness, efficiency, satisfaction and education. While previous studies provided an overview on explanations, they failed to comprehend all important research approaches in the field. So, the main advantage of this work is that it presents the main strands of research in the field and put them in one complete work.

One limitation of this work is that, since it is a survey, it is relying solely on theoretical evaluations. Future research might address a broader set of research approaches or investigate further possible benefits for explanations.

References


Mohammed Z. Taie is an assistant teacher at Al-Salam University College in Baghdad, Iraq. He holds master's degree in computer science and communication from the Arts, Science and Technology University (AUL) in Lebanon and bachelor's degree in computer science from Al-Mustansiria University in Iraq. His fields of interest lie in social networks, recommender systems and advanced web technologies. Mr. Taie has more than 10 years of experience in IT-related projects. Moreover, he published a number of studies in fields such as social networks, e-government and e-commerce. Currently, he teaches modules at BSc level in intelligent systems, Web technologies and software engineering.