Proceedings of the Twenty-Ninth International
Florida Artificial Intelligence Research Society Conference

On the Use of Opinionated Explanations to Rank and Justify Recommendations

Khalil Muhammad, Aonghus Lawlor, Barry Smyth

Insight Centre for Data Analytics
University College Dublin
Belfield, Dublin 4, Ireland
{khalil.muhammad, aonghus.lawlor, barry.smyth}@insight-centre.org

Abstract

Explanations are an important part of modern recommender systems. They help users to make better decisions, improve the conversion rate of browsers into buyers, and lead to greater user satisfaction in the long-run. In this paper, we extend recent work on generating explanations by mining user reviews. We show how this leads to a novel explanation format that can be tailored for the needs of the individual user. Moreover, we demonstrate how the explanations themselves can be used to rank recommendations so that items which can be associated with a more compelling explanation are ranked ahead of items that have a less compelling explanation. We evaluate our approach using a large-scale, real-world TripAdvisor dataset.

Introduction

Explanations are an important part of recommender systems designed to help users make better choices by providing them with justifications to support the recommendations made. Very often, these explanations take the form of simple annotations such as the average star-rating of a recommended movie or the number of friends who have also liked a given book. Nearly always, these explanations are generated after the recommendations have been selected and ranked.

The work presented in this paper aims to go further and is novel in two important respects. First, we aim to generate richer explanations, which convey more meaningful reasons why a recommendation might be accepted or rejected by a user. Second, we use these explanations in recommendation ranking. Regarding the form of our explanations, our approach emphasises the features of an item or product that are positive (*pros*) and negative (*cons*). Moreover, they tell us whether the features of a given item are *better* or *worse* than the alternative recommendations. Regarding how explanations are used during ranking, we use the strength of an explanation as the primary ranking signal for ordering recommendations, instead of more conventional ranking measures such as relevance or similarity. Thus, items that are associated with stronger or more *compelling* explanations are ranked ahead of items with less compelling explanations.

Related Work

Explanations have been used to support intelligent reasoning systems for some time; for instance, there are examples based on heuristics (Buchanan and Shortliffe 1984), case-based reasoning (Sørmo, Cassens, and Aamodt 2005; McSherry 2004; Doyle et al. 2004), and model-based techniques (Druzdzel 1996). More recently, in the recommender systems community, explanations have been used to support the recommendation process (Herlocker, Konstan, and Riedl 2000; Symeonidis, Nanopoulos, and Manolopoulos 2008; Pu and Chen 2007; Coyle and Smyth 2005; Friedrich and Zanker 2011) by providing additional evidence to support a recommendation. Good explanations promote trust and loyalty (Pu and Chen 2007). They have the potential to increase user satisfaction (Bilgic and Mooney 2005), and make it easier for users to find what they want, increasing the conversion rate of browsers into buyers (Herlocker, Konstan, and Riedl 2000).

Some of the earliest recommender systems work explored the utility of explanations in collaborative filtering systems with (Herlocker, Konstan, and Riedl 2000) reviewing different explanation approaches using MovieLens data. They considered a variety of explanation types by leveraging different combinations of data (ratings, meta-data, neighbours, confidence scores etc.) and presentation styles (histograms, confidence intervals, text etc.) concluding that most users recognised the value of explanations.

Bilgic and Mooney (Bilgic and Mooney 2005) used keywords to justify items rather than disclosing the behaviour of similar users. They argued that the goal of an explanation should not be to "sell" the user on the item but rather to help the user to make an informed judgment. They found, for example, that users tended to overestimate item quality when presented with similar-user style explanations. Elsewhere, keyword approaches were further developed by (Symeonidis, Nanopoulos, and Manolopoulos 2008) in a content-based, collaborative hybrid capable of justifying recommendations as: *"Item A is suggested because it contains feature X and Y that are also included in items B, C, and D, which you have also liked."*; see also the work of (Vig, Sen, and Riedl 2008) for related ideas based on user-generated tags instead of keywords. Note, this style of explanation justifies the item with reference to other items, in this case, items that the user had previously liked. These keyword-based approaches share some similarities with the work presented in this paper: they highlight key features of items and harness this information as part of the explanation process. We too focus on item features albeit features that are mined directly from user-generated reviews. Moreover, our features are associated with sentiment information (positive and negative) rather than simply reflecting the existence of an item characteristic.

Explanations can also relate one item to others. For example, Pu and Chen (Pu and Chen 2007) build explanations that emphasise the tradeoffs between items. For example, a recommended item can be augmented by an explanation that highlights alternatives with different tradeoffs such as *"Here are laptops that are cheaper and lighter but with a slower processor"* for instance; see also related work by (Reilly et al. 2005).

Here we focus on generating explanations that are feature-based and personalised (see also (Tintarev and Masthoff 2008)), highlighting features that are likely to matter most to the user. But, like the work of (Pu and Chen 2007; Reilly et al. 2005; Pu and Chen 2007), our explanations also relate items to other recommendation alternatives to help the user to better understand the tradeoffs and compromises that exist within a product-space; see also (McSherry 2003).

The above approaches reflect a larger body of work on the role of explanations in recommender systems. They are directly relevant to the work in this paper but different in both the form of explanations and the source of explanation information used. We generate explanations from features mined from user-generated reviews and leverage the sentiment of these features in order to draw attention to the pros and cons of a particular item, as we shall see. Moreover, a unique feature of our approach is that explanations are not generated purely to justify recommendations but also to influence their ranking in the recommendation set.

Opinion Mining for Recommendation

This paper builds on recent work about mining opinions from user reviews to generate user profiles and item descriptions for use in recommendation. The work of (Dong et al. 2013) is particularly relevant and describes how shallow natural language processing, opinion mining, and sentiment analysis can be used to extract rich feature-based product descriptions (product *cases*) based on the features that users refer to in their reviews and the polarity of their opinions. An in-depth description of this approach is beyond the scope of this short paper and the interested reader is referred to (Dong et al. 2013; Dong, O'Mahony, and Smyth 2014) for further details. However, in the interest of what follows we will briefly summarise the type of opinion data that is produced for the purpose of recommendation using TripAdvisor hotel and review data.

Item, User, and Compelling Features

To begin, in order to frame what follows, it is worthwhile describing the context of the work in this paper. We wish to construct explanations for a particular user u_T about items h_i (TripAdvisor hotels in this instance), from a set of recommendations H' generated with that user in mind.

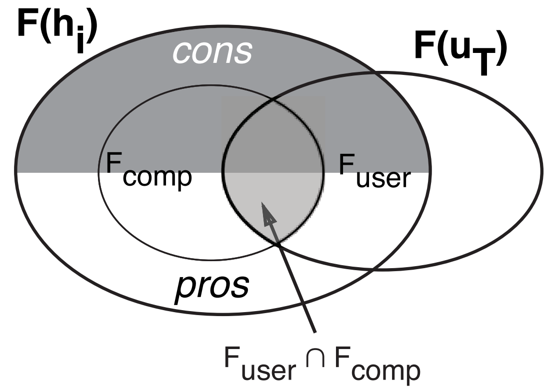

Figure 1: The relationship between item features, those that matter to the user F_{user}, and those that are compelling F_{comp}.

A given item, h_i, is represented by a set of features $F(h_i)$ mined from the reviews of h_i. These include a set of positive features (*pros*) and a set of negative features (*cons*), as we shall describe in the following sections; see Figure1. Separately, each user is also represented by a set of features mined from the reviews that they have written, $F(u_T)$; note that user features are not segregated into *pros* and *cons*. In general, some of these user features will overlap with the *pros* and *cons* of the item features ($F_{user} = F(h_i) \cap F(u_T)$) but some will not, as shown in Figure1.

Moreover, as we will describe later, some features of h_i will be *compelling* in the sense that they help to distinguish h_i from other recommendations in H'. For now we refer to these as F_{comp} (see Figure 1) and some of these compelling features will also matter to the user $F_{user} \cap F_{comp}$. When generating an explanation about h_i for u_T, we will focus on $F_{comp} \cap F_{user}$ or those compelling item features that are shared by the user; see Figure1. We will construct explanations from these features and use these explanations not only to justify an item to the user but also to rank items for recommendation. But first we will summarise how we mine item and user features from item reviews and describe how we identify compelling features during recommendation to build our explanations.

Generating Item Descriptions

Each item/hotel (h_i) is associated with a set of reviews $R(h_i) = \{r_1, \ldots, r_n\}$. The opinion mining process extracts a set of features, $F = \{f_1, \ldots, f_m\}$, from these reviews, based on the techniques described in (Dong et al. 2013; Dong, O'Mahony, and Smyth 2014), by looking for frequently occurring patterns of sentiment rich words and phrases such as "a great location" or a "disappointing restaurant". Each feature, f_j (e.g. "location" or "restaurant") is associated with an *importance* score and a *sentiment* score (in the range -1 to +1) as per Equations 2 and 3. And the item/hotel is represented by these features and scores as per

Equation 1.

$$item(h_i) = \{(f_j, s(f_j, h_i), imp(f_j, h_i)) : f_j \in R(h_i)\} \quad (1)$$

The importance score of f_j, $imp(f_j, h_i)$, is the relative number of times that f_j is mentioned in the reviews of hotel h_i.

$$imp(f_j, h_i) = \frac{count(f_j, h_i)}{\sum_{\forall f' \in R(h_i)} count(f', h_i)} \quad (2)$$

The sentiment score of f_j, $s(f_j, h_i)$, is the degree to which f_j is mentioned positively or negatively in $R(h_i)$. Note, $pos(f_j, h_i)$ and $neg(f_j, h_i)$ denote the number of mentions of f_j labeled as positive or negative during the sentiment analysis phase.

$$s(f_j, h_i) = \frac{pos(f_j, h_i)}{pos(f_j, h_i) + neg(f_j, h_i)} \quad (3)$$

Generating User Profiles

In much the same way we can generate a profile of a user u_T based on the reviews that they have written by extracting features and importance information from these reviews as in Equation 4.

$$user(u_T) = \{(f_j, imp(f_j, u_T)) : f_j \in R(u_T)\} \quad (4)$$

Note that we give more meaning to the frequency with which the user reviews a particular feature, as opposed to the average sentiment of the user‘s review opinions. This is because the frequency of mentions is a better indication of which features matter most to a user, whereas the sentiment of these features is a property of the hotels themselves and will typically vary.

From Opinions to Explanations

Our aim in this work is to describe an approach for generating explanations for each item, h_i, in a set of recommendations $H = \{h_1...h_k\}$ generated for some user u_T. The novelty of our approach stems from how we leverage opinions mined from user-generated reviews in two ways: (1) to highlight those important features (*pros* and *cons*) of an item that likely matter to u_T; (2) to emphasise those feature that distinguish the recommendation relative to other items, such as alternative recommendations or past bookings.

Explanation Components

To begin with, we present an example explanation for one particular hotel in Figure 2. There are a number of components worth highlighting. First, the explanation is made up of a number of features that have been extracted from reviews of this hotel and that are known to matter to the user; that is, they are features that the user has mentioned in their own past reviews. Second, these features are divided into *pros* and *cons*, the former with a positive sentiment score ($s(f_j, h_i) < 0.7$) and the latter with a more negative sentiment score ($s(f_j, h_i) < 0.7$). *Pros* might be reasons to choose the hotel whereas *cons* might be reasons to avoid it. Third, each feature is associated with a *sentiment bar* that shows the actual sentiment score for that feature. And finally, each feature is associated with an additional piece of

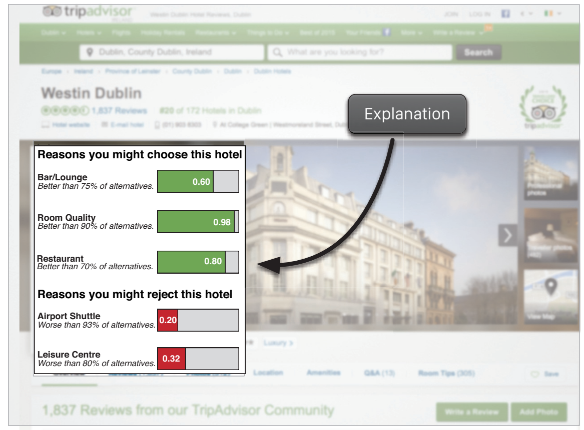

Figure 2: An example explanation showing *pros* and *cons* that matter to the target user along with sentiment indicators (horizontal bars) and information about how this item fares with respect to alternatives.

explanatory text that highlights how the hotel compares to other relevant items called a *reference set* (such as alternative recommendations as in this example) in terms of this feature.

Generating a Basic Explanation Structure

To generate an explanation like the one shown in the previous section, we start with a basic explanation structure that is made up of the features of the item in question (h_i) that are also present in the user‘s profile (u_T). These features are divided into *pros* and *cons* based on their sentiment score $s(f_j, h_i)$ and ranked in order of importance $imp(f_j, u_T)$. We also compute so-called *better* and *worse* scores as in Equations 5 and 6 with respect to some reference set. These calculate the percentage of items in the reference set for which f_j has a better sentiment score (for *pros*) or worse sentiment score (for *cons*) in h_i, respectively.

$$better(f_j, h_i, H') = \frac{\sum_{h_a \in H'} 1[s(f'_j, h_i) > s(f'_j, h_a)]}{|H'|} \quad (5)$$

$$worse(f'_j, h_i, H') = \frac{\sum_{h_a \in H'} 1[s(f'_j, h_i) < s(f'_j, h_a)]}{|H'|} \quad (6)$$

An example basic explanation structure is shown in Figure 3. It shows a set of 5 *pros* and 4 *cons* and for each we can see its importance to the user, its sentiment score for the hotel’s reviews, and the corresponding *better/worse* scores. In this case the reference set is the alternative recommendations made alongside this hotel (which are not shown here). And so, for example, we see that the *Bar/Lounge* feature, with a sentiment score of 0.71, is better than 75% of the alternative recommendations.

From Basic to Compelling Explanations

Not every pro and con in the above example makes for a very compelling reason to choose or reject the hotel in question.

	Feature	Importance	Sentiment	better/worse
PROS	Bar/Lounge *	0.25	0.71	75%
	Free Breakfast	0.22	0.79	10%
	Room Quality *	0.18	0.98	90%
	Restaurant *	0.15	0.80	80%
	Shuttle Bus	0.06	0.75	10%
CONS	Airport Shuttle *	0.21	0.20	90%
	Leisure Centre *	0.11	0.32	80%
	Swimming Pool	0.10	0.45	33%
	Room Service	0.5	0.46	20%
	:	:	:	:

Figure 3: An example of an explanation structure showing *pros* and *cons* that matter to the user along with associated importance, sentiment, and *better/worse* than scores.

For example, the *Free Breakfast*, while positively reviewed, is only better than 10% of the alternative recommendations and so if this feature is important to the user then there are better alternatives to chose from; the *Free Breakfast* is not something that distinguishes this hotel from the alternatives. In contrast , this hotel's *Room Quality* beats 90% of the alternatives and so it does distinguish this hotel and may make for a strong reason to prefer this hotel.

To simplify the explanations that are presented to users, and make them more compelling at the same time, we filter out *pros/cons* than have low *better/worse* scores ($< 50\%$) so that only those features that are better/worse than a *majority* of alternatives remain; these features are indicated with an asterisk in Figure 3. They are all features that matter to the user (they are in the user's profile) and they distinguish the hotel as either better or worse than a majority of alternative recommendations; that is, they are in the set of compelling features of the hotel.

Using Explanations to Rank Recommendations

A unique element of this work is our proposal to use explanations to rank recommendations as well as justify them. To do this we score explanations based on their strength or *compellingness*; hotels with the strongest explanations should appear at the top of the ranking. We use the scoring function shown in Equation 7 which calculates the difference between the *better* scores of the *pros* and the *worse* scores of the *cons*.

$$Compellingness(u_T, h_i, H') = \sum_{f \in Pros(u_T, h_i, H')} better(f, h_i, H') - \sum_{f \in Cons(u_T, h_i, H')} worse(f, h_i, H') \tag{7}$$

In this way, recommendations that are associated with explanations that are predominantly positive (more *pros* with high *better* scores and few *cons* with lower *worse* scores) are preferred. They offer the user a better choice with fewer compromises with respect to the features that matter to them.

Evaluation

For the purpose of this evaluation we will use data collected from TripAdvisor during the period June 2013 to August 2013. This data includes $2,800$ target users. For each user we have 10 reviews that they have written for 10 hotels (*booked hotels*) that they have stayed in. In addition we gathered $227,125$ hotel reviews for $2,379$ hotels, including the reviews of these target users. This data allow us to build descriptions of hotels and user profiles as described earlier.

Setup & Approach

For each target user, u_T, we simulate the TripAdvisor session when they came to select each of their booked hotels. Specifically, for a given session for u_T and a booked hotel h_B, we determine the 9 best *related* hotels that TripAdvisor suggests for h_B and we use these hotels $(h_B, h_1, ..., h_9)$ as a reasonable set of recommendations for the user as they search for h_B. Of course, strictly speaking it is unlikely that they will have seen these exact related hotels but it is reasonable to assume that they would have been presented with similar at the time. Thus, for every u_T-h_B pair we can produce a set of 10 recommendations $\{h_B\} \cup \{h_1, ..., h_9\}$ which correspond to the user's recommendation session. We select 5 booked hotels from each user profile to seed our sessions, thus yielding a total of $14,000$ sessions.

For each session, we generate a compelling explanation for each of the 10 hotels (including the booked hotel) with reference to the alternatives in the session as described previously. Then the 10 hotels are ranked based on the compellingness of these sessions as described above. In what follows we analyse the explanations produced and critique the recommendation ranking that results.

Explanation Types & Compellingness

First of all, we graph the average compellingness score of the explanations in each session by rank position; see Figure 4(a). As expected, there is a steady decrease in compellingness but the extent of the decline is somewhat striking. We can see that the top 5 recommendations are associated with explanations that have positive compellingness scores, but from position 6 onwards we see the explanations turn to be increasingly negative. In other words, the top 5 or so recommendations are net-positive with respect to the features that matter to the user, but after that, recommendations are associated with explanations where negative features tend to dominate. Figure 4(a) also shows the mix of different *explanation types* at each of the rank positions. There are 3 possible types of explanation: *pros only (P)*, *cons only (C)*, and *pros and cons (PC)* and Figure 4 shows the number of each type by rank position as stacked bars. About 70% of the explanations at the top ranked position contain *pros* only; in other words, users are offered recommendations with little or no compromises on features that matter to them. As we move through the rankings the number of recommendations with only *pros* declines and as the number of *cons* begins to

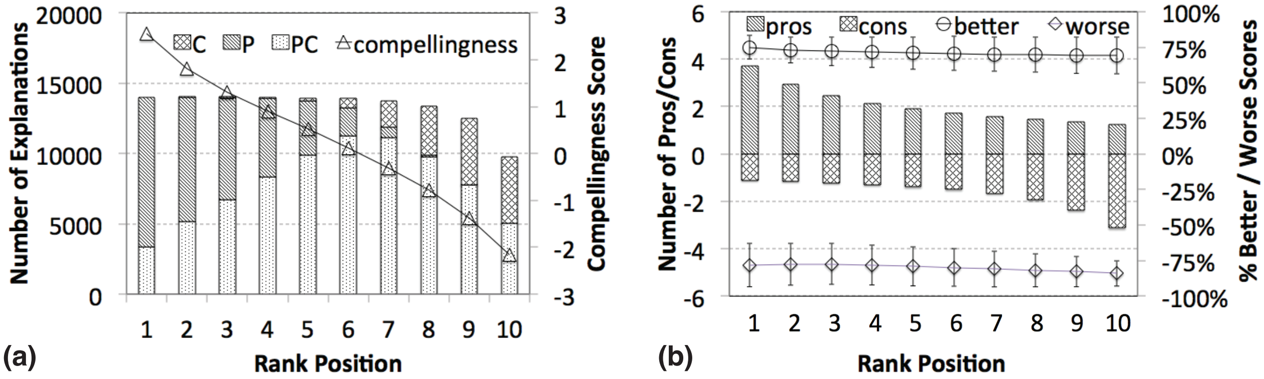

Figure 4: (a) The number of different types of compelling explanation and average compellingness score per rank; (b) The average number of *pros* and *cons* per explanation at each rank position and their average better and worse scores.

increase. For instance, the bottom half of the rankings see an increasing number of explanations with only *cons*, and by rank position 10 a full 50% of the explanations are made up only of *cons*.

Overall then, we can see that this approach to ranking recommendations by compellingness is capable of ordering and separating recommendations based on the comparative sentiment of the features that matter to the user. The resulting ranking emphasises the degree of compromise facing the user which choosing a hotel among a set of alternatives.

Pros/Cons & Better/Worse Scores

To help us understand the precise make-up of the explanations produced at different rank positions, Figure 4(b) plots the average number of *pros* and *cons* and their average *better/worse* scores (plus standard deviations); for visual reasons the *cons* and *worse* scores have been plotted on a negative axis to reflect their negative connotations. We can see that the top ranking recommendations are associated with explanations that have 3-4 *pros* and 0 or 1 *cons*. And as we move through the ranking, the number of *pros* per explanation declines, only to be replaced with an increasing number of *cons*. By the bottom of the ranking, a typical explanation has 3-4 *cons* and 0 or 1 *pros*. The *better/worse* scores are somewhat similar to each other; remember that compelling explanations are, by definition, made up of features with *better/worse* scores of at least 50%. It is worth noting that the *better* scores tend to decrease and become more variable as we move down the ranking whereas the *worse* scores tend to increase and become less variable.

This makes the degree of compromise facing the user clear. The top ranked recommendations have explanations that have a majority of *pros* and usually only one con, if any. But the bottom ranked recommendations are associated with a clear majority of *cons*.

Summary Discussion

The above helps us understand the type of explanations that we obtain for recommendations at different ranks. But what can we say about this approach to ranking and its ability to improve the recommendation experience for the user? After all in a simulated (offline) study such as this we are limited by the lack of live-user feedback. That being said, for each session we do know the booked hotel and its position in the explanation-based ranking. The user chose this hotel based on some earlier ranking by TripAdvisor. And the position of this hotel appears in our explanation-based ranking will help us understand whether our ranking might be able to improve the recommendation experience for users in practice.

In fact, we find that on average the booked hotel appears in our explanation-based rankings at approximately position 5. In other words, our explanation-based approach ranks 4 other hotels ahead of the one that the user booked. And when we examine the *pros* and *cons* of these hotels, compared to the booked hotel, it is clear that they offer some considerable benefits. For example, the average booked hotel has about 1.7 *pros* (with average better scores of 0.6) and about 1.3 *cons* (with average worse scores of 0.7). Compare this to a typical top-ranked recommendation based on our approach, which has almost 4 *pros* and less than one *con* on average (with *better/worse* scores of approximately 75%).

Clearly our approach is able to identify hotels that have many more strong *pros* and far fewer *cons* than the hotel the user booked. Whether the user would prefer these in practice remains to be seen. Perhaps these hotels were not available at the time of booking by the user and so they never saw them when making their choice. This is certainly possible but it seems unlikely, at least not on the scale that these results would suggest. More likely, the default TripAdvisor ranking is based on some global assessment of how well hotels match a user's needs. For example, the evidence clearly suggests that the average review rating of a hotel plays a key role in the TripAdvisor rankings. But of course average rating is unlikely to be a good fit for an individual user, who likely has a specific set of features that they prioritise over others when making their choice.

All of this remains a matter for future work, probably involving live-user trials. For now it is at least extremely promising to see that our approach is able to identify recommendations that appear to offer a number of advantages over the hotel booked by the user. And, as such, we remain confident that our approach — the combination of opinionated hotel descriptions, user-tailored explanations, and the use of these explanations for ranking — has significant potential.

Conclusions

In this paper we have presented a novel form of explanation for recommender systems based on the opinions of users mined from product reviews. Our approach is unique in two important ways: first, the explanations themselves are derived from user opinions, they are personalised based on the things that matter to the target user, and they not only highlight the *pros* and *cons* of a particular item but also relate this item to alternative recommendations. Second, these explanations are the basis for a novel approach to recommendation ranking in which the strength of compellingness of the explanation is used as the primary ranking signal.

Having described the details of our approach to explanation generation, filtering, and ranking the main contribution of this paper has been to evaluate these ideas using large-scale, real-world data. We did this by using more than $227,125$ hotel reviews from $150,961$ users from TripAdvisor. And we were able to show how ranking hotels based on the compellingness of their explanations produced recommendation lists that prioritised items with a preponderance of positive features that were better than a large majority of alternative recommendations, and with few if any negative features. We also demonstrated how this approach tended to identify a number of items that seemed to offer improved tradeoffs for the user than the actual item that was eventually booked by the user on TripAdvisor. We propose that this result provides strong evidence in support of this approach to explanation and ranking.

Acknowledgments

This work is supported by Science Foundation Ireland through through the Insight Centre for Data Analytics under grant number SFI/12/RC/2289.

References

Bilgic, M., and Mooney, R. J. 2005. Explaining Recommendations: Satisfaction vs. Promotion. In *Proceedings of Beyond Personalization Workshop at the 2005 International Conference on Intelligent User Interfaces*, 13–18.

Buchanan, B. G., and Shortliffe, E. H. 1984. *Rule Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project (The Addison-Wesley Series in Artificial Intelligence)*, volume 3. New York, NY, USA: Addison-Wesley Longman Publishing Co., Inc.

Coyle, M., and Smyth, B. 2005. Explaining Search Results. In *Proceedings of The 19th International Joint Conference on Artificial iIntelligence*, 1553–1555. Edinburgh, UK: Morgan Kaufmann Publishers Inc.

Dong, R.; Schaal, M.; O'Mahony, M. P.; and Smyth, B. 2013. Topic Extraction from Online Reviews for Classification and Recommendation. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, IJCAI '13, 1310–1316. AAAI Press.

Dong, R.; O'Mahony, M. P.; and Smyth, B. 2014. Further Experiments in Opinionated Product Recommendation. In *Proceedings of The 22nd International Conference on Case-Based Reasoning*, 110–124.

Doyle, D.; Cunningham, P.; Bridge, D.; and Raham, Y. 2004. Explanation Oriented Retrieval. In *Proceedings of The 7th European Conference on Case-Based Reasoning, (ECCBR 04)*, volume 3155, 157 – 168. Madrid, Spain: Springer Berlin Heidelberg.

Druzdzel, M. J. 1996. Qualitative Verbal Explanations in Bayesian Belief Networks. *Artificial Intelligence and Simulation of Behaviour Quarterly, Special Issue on Bayesian Networks* 94:43–54.

Friedrich, G., and Zanker, M. 2011. A Taxonomy for Generating Explanations in Recommender Systems. *AI Magazine* 32(3):90–98.

Herlocker, J. L.; Konstan, J. A.; and Riedl, J. 2000. Explaining Collaborative Filtering Recommendations. In *Proceedings of The 2000 ACM Conference on Computer Supported Cooperative Work*, CSCW '00, 241–250. Philadelphia, USA: ACM.

McSherry, D. 2003. Similarity and Compromise. In *Proceedings of The 5th International Conference on Case-based Reasoning (ICCBR '03)*, 291–305. Trondheim, Norway: Springer-Verlag.

McSherry, D. 2004. Explaining the pros and cons of conclusions in CBR. In *Proceedings of The 7th European Conference on Advances in Case-Based Reasoning*, 317–330.

Pu, P., and Chen, L. 2007. Trust-Inspiring Explanation Interfaces for Recommender Systems. *Knowledge-Based Systems* 20(6):542–556.

Reilly, J.; McCarthy, K.; McGinty, L.; and Smyth, B. 2005. Explaining Compound Critiques. *Artificial Intelligence Review* 24(2):199–220.

Sørmo, F.; Cassens, J.; and Aamodt, A. 2005. Explanation in Case-Based ReasoningPerspectives and Goals. *Artificial Intelligence Review* 24(2):109–143.

Symeonidis, P.; Nanopoulos, A.; and Manolopoulos, Y. 2008. Providing Justifications in Recommender Systems. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 38(6):1262–1272.

Tintarev, N., and Masthoff, J. 2008. The Effectiveness of Personalized Movie Explanations: An Experiment Using Commercial Meta-data. In *Proceedings of The 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH '08)*, volume 5149, 204–213. Hannover, Germany: Springer Berlin Heidelberg.

Vig, J.; Sen, S.; and Riedl, J. 2008. Tagsplanations: Explaining Recommendations using Tags. In *Proceedings of The 13th International Conference on Intelligent User Interfaces*, 47–56. Florida, USA: ACM Press.