Explanation-based Ranking in Opinionated Recommender Systems

Khalil Muhammad, Aonghus Lawlor, Barry Smyth

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

Abstract. Explanations can help people to make better choices, but their use in recommender systems has so far been limited to the annotation of recommendations after they have been ranked and suggested to the user. In this paper we argue that explanations can also be used to rank recommendations. We describe a technique that uses the strength of an item's explanation as a ranking signal – preferring items with *compelling* explanations – and demonstrate its efficacy on a real-world dataset.

1 Introduction

In recommender systems, an explanation is any additional information provided to help users understand why or how recommendations are made to them. Recently, explanations have been become an interesting area of research for recommender systems [6] where they have been shown to significantly increase how users perceive the utility of recommender systems [14]. The right explanation at the right time can help users to make better choices, promote trust and loyalty [9], increase user satisfaction [1], and improve the conversion rate of browsers into buyers [7]. Usually explanations take the form of recommendation annotations, such as the star rating of a movie or the number of friends who have also liked a book [2, 5].

Early work explored the utility of explanations in collaborative filtering systems. For instance, [7] evaluated a variety of explanation formats by leveraging different combinations of data and presentation styles using MovieLens data. Bilgic and Mooney [1] used keywords to justify items, finding that users tended to overestimate item quality when presented with this style of explanation. Keyword approaches were further developed by [13, 11] to justify 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."*. Explanations have also been used to relate one item to others [9, 10]. For example, Pu and Chen [9] built explanations that emphasise the tradeoffs between items, where 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"*.

In this paper, we present an approach to explanation for recommender systems that is novel in several important respects. First, the explanations are generated directly from the opinions expressed in user-generated reviews. They are personalised for the target user by highlighting item features (*pros* and *cons*) that are likely to matter to the user (see also [12]) and that distinguish the item from other recommendations. We argue that this makes for a more *compelling* form of explanation because it helps the user to understand the tradeoffs and compromises that exist within a product-space; see also [8–10]. Second, we show how the strength of explanations can be measured and used during recommendation ranking. Finally, we provide detailed evaluation of ranking performance against a suitable baseline. In what follows we will describe our approach to generating recommendations from user-generated reviews and demonstrate its effectiveness using a large-scale real-world dataset from TripAdvisor.

2 Opinion Mining for Recommendation

This paper builds on recent work about using opinions from user reviews to generate user profiles and item descriptions for use in recommendation. For example, [3, 4] describe how natural language processing and opinion mining can be used to extract rich feature-based product descriptions. While an in-depth description of this approach is beyond the scope of this paper, in the interest of what follows, we will briefly summarise how we use similar techniques to generate user profiles and item descriptions from TripAdvisor reviews.

2.1 Generating Item Descriptions

In our TripAdvisor data, an item/hotel (h_i) is associated with a set of reviews $reviews(h_i) = \{r_1, \ldots, r_n\}$. The opinion mining process extracts a set of features, $F = \{f_1, \ldots, f_m\}$, from these reviews, 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 as per Equations 2 and 3. Each item/hotel is represented by these features and scores (Equation 1).

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

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.

$$imp(f_j, h_i) = \frac{count(f_j, h_i)}{\sum_{f' \in reviews(h_i)} count(f', h_i)} \tag{2}$$

The sentiment score of f_j, $s(f_j, h_i)$, is the degree to which f_j is mentioned positively or negatively in $reviews(h_i)$. 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.

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

2.2 Generating User Profiles

We can generate a profile of a target user u_T in a similar manner, by extracting features and importance information from u_T's reviews as in Equation 4.

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

3 From Opinions to Explanations

Our aim 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.

3.1 Explanation Components

Consider an example (in Fig. 1) of our explanation for a particular hotel. There are a number of components worth highlighting. First, the explanation comprises of a number of features extracted from the reviews of this hotel *and* that are known to matter to the user (because the user has mentioned them in their own past reviews). Second, these features are divided into *pros* and *cons*, the

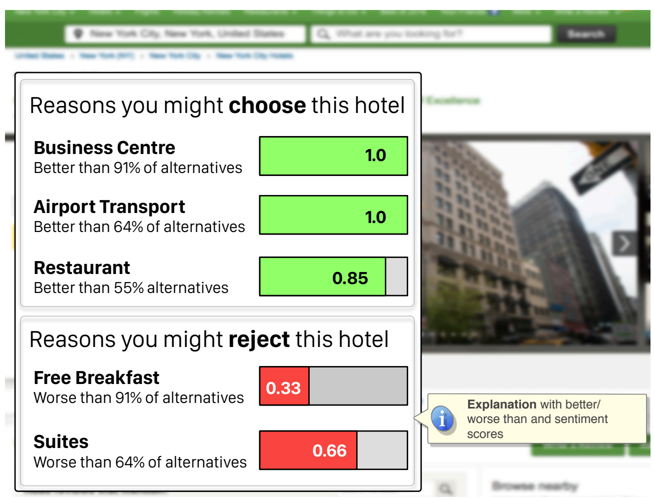

Fig. 1. An example explanation showing *pros* and *cons* that matter to the target user along with sentiment indicators (horizontal bars) and information about how this item fares with respect to alternatives.

former with a predominantly positive sentiment score ($s(f_j, h_i) > 0.7$) and the latter with a more negative sentiment score ($s(f_j, h_i) \leq 0.7$). Our choice to place the threshold at 0.7 may appear somewhat arbitrary, and reflects a desire to ensure a firm positive sentiment where there tends to be a bias towards the positive – however our results do not strongly depend on the precise value of this threshold. *Pros* may be reasons to choose the hotel whereas *cons* may be reasons to avoid it. Third, each feature is associated with a *sentiment bar* that shows the actual sentiment score for that feature. Finally, each feature is associated with additional 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. For example, we see that the Nomad Hotel in Fig. 1 has an excellent *Business Centre* (sentiment score: 1.0), which is better than 91% of the other recommendations. However, its *Free Breakfast* is not favourably reviewed (sentiment score: 0.33) making it worse than 91% of the current recommendation alternatives.

3.2 Generating a Basic Explanation Structure

To generate this type of explanation, we start with a basic explanation structure composed of the features of the item (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 & 6 with respect to some reference set; again, we will use the alternative recommendations as the reference set but other sets could be used, such as the user's past bookings. Thus, we calculate the percentage of alternative recommendations for which f_j in h_i has a better (for *pros*) or worse (for *cons*) sentiment score in h_i, respectively.

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

$$worse(f_j, h_i, H') = \frac{\sum_{h_a \in H'} 1[s(f_j, h_i) \leq s(f_j, h_a)]}{|H'|} \tag{6}$$

An example basic explanation structure is shown in Fig. 2. It corresponds to the hotel and explanation presented in Fig 1. There are 5 *pros* and 2 *cons*, and for each we can see its sentiment score and the corresponding *better*/*worse* scores with respect to the alternative recommendations made alongside this hotel (which are not shown here). For example, we see that the *Restaurant* feature, with a sentiment score of 0.85, is better than 55% of the alternative recommendations.

3.3 From Basic to Compelling Explanations

Not every *pro* or *con* in the above example makes for a compelling reason to choose or reject the hotel in question. For example, the *Shuttle Bus Service*, with an average sentiment of 75%, is only better than 18% of the alternative recommendations, and so if this feature is important to the user, then it may not constitute a strong reason to choose this particular hotel.

To simplify the explanations that are presented to users, and make them more compelling at the same time, we filter out *pros*/*cons* that have low *better*/*worse*

	Feature	Sentiment	better/worse
PROS	Business Centre	1.00	91%
	Airport Transport	1.00	64%
	Restaurant	0.85	55%
	* *Shuttle Bus Service*	0.75	18%
	* *Free Parking*	0.80	27%
CONS	Free Breakfast	0.33	91%
	Suites	0.66	64%

Fig. 2. An example of an explanation structure showing *pros* and *cons* that matter to the user along with associated sentiment, and *better/worse* than scores.

scores ($< 50\%$) so that only those features that are *better/worse* than a *majority* of alternatives remain; these features that are filtered out are indicated with an asterisk in Fig. 2, the remaining features are all compelling. They are all features that matter to the user (they are in the user's profile) and they distinguish the hotel as either significantly better or worse than a majority of alternative recommendations.

3.4 Using Explanations to Rank Recommendations

A key element of this work is the use of explanations for ranking as well as explaining. We can score explanations based on their *strength*, so that hotels with the strongest explanations 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*. Because these explanations are generated using features that matter to the individual user, this scoring metric is personalised to the target user.

$$\begin{aligned} Strength(u_T, h_i, H') = \\ & \sum_{f \in Pros(u_T, h_i, H')} better(f, h_i, H') - \\ & \sum_{f \in Cons(u_T, h_i, H')} worse(f, h_i, H') \end{aligned} \tag{7}$$

Thus, recommendations associated with explanations that are predominantly positive — more *pros* with high *better* scores and few *cons* with lower *worse* scores — will have a high *strength* score; they offer the user a better choice with fewer compromises with respect to the features that matter to them. By contrast, recommendations that are associated with a lower or even negative *strength* score will correspond to recommendations that demand far more compromise from the user in terms of the features that matter to them. Indeed, if the *strength* score is negative then it indicates that the recommendation may have more negatives (*cons*) than positives (*pros*) or that the *cons* are significant worse than the *pros*. Such recommendations then will be ranked lower than more positive ones.

This is just one way to score explanations for use in recommendation ranking; it can serve as a useful baseline against which to evaluate more sophisticated and effective strategies. For example, a user's feature importance weights could be used to weight the *better*/*worse* scores in the above scoring functions so that more important features have a greater influence as *pros* or *cons*. Alternatively, the actual *better*/*worse* scores could be incorporated so that features that are much better than the alternatives are more influential than features that are only marginally better, and vice versa for *cons* and *worse* scores. These adaptations will be considered as a matter for future work.

4 Evaluation

In this evaluation, we use a TripAdvisor[1] dataset which we collected from June 2013 to August 2013. This data comprises of $224,760$ reviews about $2,298$ hotels, from 6 major cities. From this, we select $10,000$ users who have written at least 4 reviews each. We build hotel descriptions and user profiles as described earlier.

4.1 Setup & Approach

For each user, we simulate the TripAdvisor session when they came to select each of their reviewed (booked) hotels. For a given user, u_T, and booked hotel h_B, we have the 10 best *related* hotels that TripAdvisor suggests for h_B and we use these hotels $(h_1, ..., h_{10})$ as recommendations for the u_T as they search for h_B; in this way h_B acts as a pseudo-query for the u_T. 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 which correspond to a recommendation session and our dataset allows us to produce $27,423$ (*users* $\times$ *reviews*) such sessions (we include only those sessions where we have profiles for all the related hotels in the session).

We have a number of options when ranking our 10 recommendations per sessions. Each of the following generates a *ranking score* per recommendation which is then used to sort the recommendations in descending order of ranking score.

1. *Basic Explanation Score, BES* – the ranking score of the recommendation is the strength of its basic explanation.
2. *Compelling Explanation Score, CES* – the ranking score of the recommendation is the strength of its compelling explanation.
3. *Cosine Similarity Score, CSS* – the ranking score of the recommendation is the cosine similarity of the booked hotel h_B to the target hotel. We personalise this score to the user by weighting the booked hotel h_B by the user

[1] This dataset is available from the authors on request.

importance score: $sim = (u_T \times h_B) \cdot h_T$. This will serve as a *classical* baseline against which to evaluate our explanation-based ranking as it corresponds to a fairly conventional personalised content-based recommendation approach.

4. *TripAdvisor Rating, TAR* – the ranking score of the recommendation is its TripAdvisor average review rating; we will use this as a plausible ground-truth and as an *ideal* baseline against which to evaluate our explanation-based approach as it is the ranking score that TripAdvisor appears to use when ranking its recommendations. Note the TripAdvisor rating is independent of the features and sentiments mined from customer reviews.

4.2 Ranking Quality

To begin with, we will compare the quality of the rankings produced by our explanation-based approach to the content-based approach. For each (u_T, h_B), we rank the 10 related hotels $\{h_1, ..., h_{10}\}$ by using BES, CES, and CSS. As a proxy for ranking quality we will use the average hotel ratings as our ground truth. As mentioned above, this appears to be used as the primary ranking signal by TripAdvisor itself, and the average rating is the collective rating score assigned by many independent users to each hotel and so should act as a reasonably unbiased ranking signal. In Fig. 3, we show the average rating for recommended

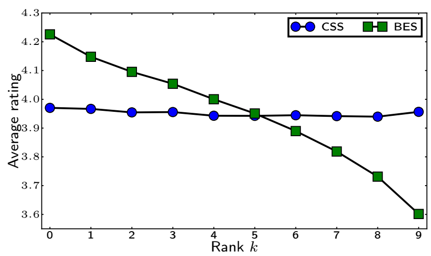

Fig. 3. Average ratings versus rank position for explanation-based (basic and compelling) and similarity-based rankings.

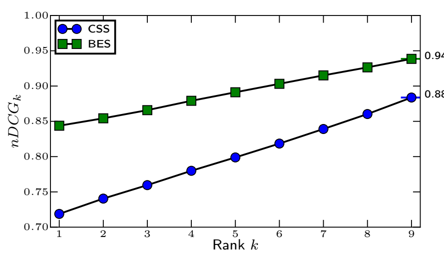

Fig. 4. Average $nDCG_k$ per session against rank position for both explanation-based and similarity-based ranking strategies.

hotels at at each rank position, using our different ranking approaches; each ranking approach re-ranks the same set of hotels so only the order of the hotels change. We find no difference between the rankings produced when using BES versus CES and so a single line (BES) is shown for the explanation-based ranking. Incidentally, this means we can deliver similar ranking performance using the strength of compelling explanations but with only a subset of the features in basic explanations. Thus we exploit the simpler compelling explanations (with fewer features) without any loss of ranking performance.

Evidently, the explanation-based ranking (BES) technique is better able to identify more highly rated hotels at higher rank positions. For example, the top-

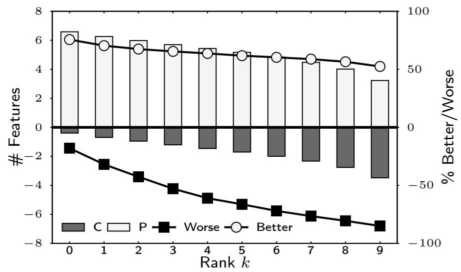

Fig. 5. The average number of *pros/cons* and *better/worse* scores for items ranked by the strength of basic explanations.

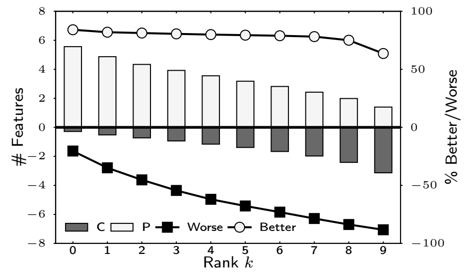

Fig. 6. The average number of *pros/cons* and *better/worse* scores for items ranked by the strength of compelling explanations.

ranked hotel according to explanation-based ranking has an average rating of just over 4.2 compared to just under 4 for the similarity-based ranking; note that it is quite normal for recommended hotels to have highly similar rating scores since TripAdvisor's related hotels usually all have very high average ratings.

The explanation-based ranking is better than the similarity-based ranking in sorting the recommendations in the way that correlates well with their average rating scores; its ratings are better at all ranks until the 5th position. However, similarity-based ranking fails to distinguish between hotels of different average ratings at different rank positions at all; its average rating results are almost flat across the rank positions.

We can directly compare the performance of these explanation-based and similarity-based rankings by computing a normalised discounted cumulative gain $nDCG$ – with the position discounted logarithmically. We use average ratings as relevance score, and the ideal ranking produced by the average ratings as the normalising factor, and calculate $nDCG_k$ for each recommendation session (u_T, h_B); see Fig. 4. Again, basic and compelling explanation scores produce very similar results and we show only BES. In this case we find that the explanation-based ranking performs much better than similarity-based ranking particularly at the top of the ranked list. Even so, we find explanation-based ranking has a higher $nDCG_k$ than similarity-based ranking across all ranks, reaching $nDCG_{10} = 0.94$ compared to $nDCG_{10} = 0.88$ for similarity-based ranking.

4.3 Explanations by Rank Position

In this section we take a closer look at the type of explanations that are associated with each rank position. To do this, we rank our recommendation lists using each of the 4 ranking strategies (BES, CES, CSS, TAR), and for each rank position we average the number of *pros/cons* and *better/worse* scores associated with this rank; thus we are averaging $27,423$ explanations for each rank position because there are this many separate sessions. The results are presented in Figures 5–7 for BES, CES, CSS and TAR, respectively. In each graph, the bar chart presents the average number of *pros* and *cons* at each rank, and the line-graphs indicate the corresponding average *better* and *worse* scores; note that for reasons of clarity we present *cons* and worse scores as 'negatives' below the zero-line.

The profile of explanations associated with the explanation-based ranking should be qualitatively different from those associated with CSS and TAR. In the explanation-based rankings (Figures 5 & 6), for example, we can see a clear difference in the makeup of the top-ranked explanations – or more correctly, the explanations associated with the top-ranked recommendations – than the bottom-ranked explanations. The top-ranked explanations are dominated by *pros* and have fewer *cons*. Moreover, their *pros* have high *better* scores, and their *cons*, if any, have relatively low *worse* score. In contrast, the bottom-ranked explanations have many more *cons* and fewer *pros*, and the *worse* scores of the *cons* are very high while the *worse* scores of the *pros* are relatively low.

This is as we might expect. It means that the explanations associated with the top-ranked recommendations (when explanation ranking is used) are much more positive than the explanations at the bottom of the rankings. For example, in the case of compelling explanations (see Fig. 6) the top-ranked explanations have about 5 *pros* and typically less than 1 *con*, on average. The *pros* tend to be better than about 80% of the alternatives and the *cons*, if they exist, are worse than only about 20% of the alternatives. By comparison the bottom-ranked recommendations are associated with explanations that have only 1 or 2 *pros* but 2 or 3 *cons*. And while the average *pro* is only better than a little more than half of the alternatives the average *con* is worse than up to 90% of alternatives.

This is consistent with how we might expect a recommendation ranking to look. The end-user's expectation will be for hotels at the top of the ranking to be more positively reviewed (and less negatively reviewed) than hotels at the bottom of the ranking. They should also involve fewer compromises (on features that matter to the user) than hotels at the bottom of the ranking. This is what both explanation-based rankings succeed in offering.

The picture for non-explanation-based ranking is very different. Both CSS and TAR rankings produce recommendation lists in which the items at the top of the ranking are distinguishable from the bottom-ranked items in very limited ways, if at all. For example, in Fig. 7, where the recommendations have been ranked by CSS we can see that there is virtually no difference between items at the top and bottom of the ranking. On average, independent of rank position, recommendations are associated with explanations that contain about 4-5 *pros* and 2 *cons*. And the *pros* and *cons* are relatively weak with *better*/*worse* scores of just over 50%. While the hotels chosen by TripAdvisor might be broadly relevant to the end-user, ranking them in this way may do little to help the user make their choice. And more likely than not, the user will be confused by what differentiates the hotel at the top of the list from the one at the bottom.

The picture is a little better when we rank by TAR as shown in Fig. 8. At least now there is difference between the items at the top of the ranking compared with those at the bottom. The former has about 4 *pros* and 1 *cons* compared with the latter which has 3 *pros* and 2 *cons*; there is not much difference in the *better/worse* scores across the rank positions however. Therefore, at least the top-ranked hotels appear to be more positively reviewed than those further down the ranking, but difference is far more subtle than that exposed by the explanation-based ranking. In summary, the explanation-based ranking

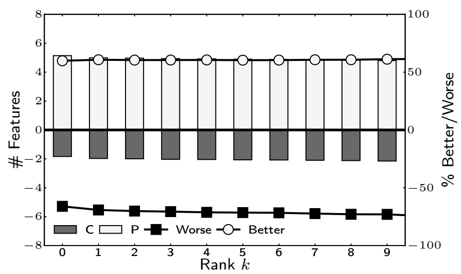

Fig. 7. The number of *pros/cons* and *better/worse* scores by rank position for similarity-based ranking.

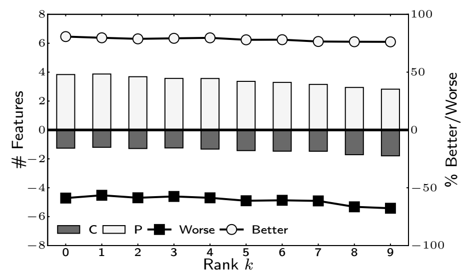

Fig. 8. Average number of *pros/cons* and *better/worse* scores for the TripAdvisor baseline ranking.

technique does a better job at differentiating the hotel recommendations in terms of the sentiment expressed in their reviews. Those at the top of the ranking are strongly positive, and those at the bottom are much more negative. We argue that this is as it should be because it will help the user to make a choice because of the strong differentiation between hotels as one moves through the ranking. And easier choices make for more frequent decisions and more satisfied users.

4.4 Discussion & Limitations

This section has attempted to evaluate the quality of the rankings produced by our explanation-based techniques on real-user data. While the results suggest a clear benefit for explanation-based ranking, they are of course limited, and in this section we will discuss some of these limitations.

Firstly, the rankings have obviously not been evaluated using live-user feedback. Ideally, we should present the recommendation rankings to real users booking real hotels in an effort to capture their perception of ranking quality. This is difficult to do outside of a live, large-scale systems (in this case, TripAdvisor). Nevertheless, it is hoped that the combination of reasonably large-scale real-user data and the inclusion of two plausible baseline ranking techniques is robust enough to afford a reasonable level of confidence in these results.

Secondly, one could question the use of content-based similarity and the average user rating as baseline approaches or metrics. Average user ratings serve as a reasonably plausible ground-truth (in the case of our *nDCG* analysis) and as a reasonable ranking strategy in the case of our rank-position analysis. The average ratings score has been crowdsourced from thousands of TripAdvisor users and their use in ranking is well-accepted in TripAdvisor itself, which uses average review rating as its default ranking metric. Moreover, the cosine similarity metric that we have implemented is typical of the type of content-based approaches that are often implemented in practice, and at least serves as a useful benchmark to judge the ranking effectiveness of our explanation-based approaches.

We have already mentioned a number of opportunities to further develop our explanation-based ranking approaches – by including user weights or relative differences in *better*/*worse* scores, for example – and likewise there are opportunities to encode further improvements in our baseline metrics, particularly the cosine-based content similarity approach. These will be left as a matter of future work without detracting from the value of the results presented herein.

Finally, to bring this evaluation section to a close, it is worth highlighting a further obvious limitation of the evaluation as presented, namely the lack of a real-time, live-user trial. As previously mentioned, such a trial amounts to a major evaluation challenge that is all but impossible to pursue in the absence of a supportive industry partner capable of delivering the scale of users needed for live-testing.

5 Conclusions

The right explanation at the right time can help a user to make a more informed decision. The increasingly important role of explanations in recommender systems reflects a growing trend to improve the overall recommendation experience, as opposed to a narrow focus on prediction accuracy or ranking quality. We describe an approach to explanation based on the opinions of users in their online reviews. These explanations are feature-based, personalised, and they are designed to clarify the tradeoffs that exist among recommendation alternatives.

Secondly, we described how to use these explanations for recommendation ranking by leveraging explanation *strength* as a novel ranking signal. This is motivated by the intuition that users are likely to be swayed by more compelling explanations. Thus, those items with the most compelling explanation should be prioritised in the rankings.

We included a detailed evaluation of this ranking method using real-world TripAdvisor data. This helped to answer key questions about the type of recommendations that are produced in a realistic setting. We have also looked at how these recommendations are ranked when using explanation strength compared to TripAdvisor default ranking (average user ratings) and an alternative content-based similarity approach. It is clear that our approach does a superior job of prioritising those items in a way that should appeal to users by simplifying their decision making. For example, we have shown how our explanation-based approaches are capable of ordering recommendations so that those that are most positive (and least negative) appear at the top of the rankings and those that are most negative (and least positive) appear at the bottom of the ranking. The same is not true for the alternative rankings tested as baselines.

In the future, we plan to explore different types of explanation scoring metrics. The current *strength* formula remains simple and straightforward in nature and there are obvious opportunities to explore when it comes to, for example, the differential weighting of *pros* and *cons*.

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

References

1. Bilgic, M., Mooney, R.J.: Explaining Recommendations: Satisfaction vs. Promotion. In: Proceedings of Beyond Personalization Workshop at the 2005 International Conference on Intelligent User Interfaces. pp. 13–18. San Diego, USA (Jan 2005)
2. Coyle, M., Smyth, B.: Explaining Search Results. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence. pp. 1553–1555. Edinburgh, UK (Jul 2005)
3. Dong, R., O'Mahony, M.P., Smyth, B.: Further Experiments in Opinionated Product Recommendation. In: Proceedings of the 22nd International Conference on Case-Based Reasoning. pp. 110–124. Cork, Ireland (Sep 2014)
4. Dong, R., Schaal, M., O'Mahony, M.P., Smyth, B.: Topic Extraction from Online Reviews for Classification and Recommendation. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence. pp. 1310–1316 (2013)
5. Friedrich, G., Zanker, M.: A Taxonomy for Generating Explanations in Recommender Systems. AI Magazine 32(3), 90–98 (Jun 2011)
6. Gedikli, F., Jannach, D., Ge, M.: How Should I Explain? A Comparison of Different Explanation Types for Recommender Systems. International Journal of Human-Computer Studies 72(4), 367–382 (2014)
7. Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining Collaborative Filtering Recommendations. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work. pp. 241–250. Philadelphia, USA (Dec 2000)
8. McSherry, D.: Similarity and Compromise. In: Proceedings of the 5th International Conference on Case-based Reasoning. pp. 291–305. Trondheim, Norway (Jun 2003)
9. Pu, P., Chen, L.: Trust-Inspiring Explanation Interfaces for Recommender Systems. Knowledge-Based Systems 20(6), 542–556 (Aug 2007)
10. Reilly, J., McCarthy, K., McGinty, L., Smyth, B.: Explaining Compound Critiques. Artificial Intelligence Review 24(2), 199–220 (Oct 2005)
11. Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: Providing Justifications in Recommender Systems. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 38(6), 1262–1272 (Nov 2008)
12. Tintarev, N., Masthoff, J.: 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. vol. 5149, pp. 204–213. Hannover, Germany (Jul 2008)
13. Vig, J., Sen, S., Riedl, J.: Tagsplanations: Explaining Recommendations using Tags. In: Proceedings of the 13th International Conference on Intelligent User Interfaces. pp. 47–56. Florida, USA (Feb 2008)
14. Zanker, M.: The Influence of Knowledgeable Explanations on Users' Perception of a Recommender System. In: Proceedings of the 6th ACM Conference on Recommender Systems. p. 269. New York, USA (2012)