

Opinionated Explanations for Recommendation Systems

Aonghus Lawlor, Khalil Muhammad, Rachael Rafter
and Barry Smyth

Abstract This paper describes a novel approach for generating explanations for recommender systems based on opinions in user-generated reviews. We show how these opinions can be used to construct helpful and compelling explanations at recommendation time. The explanation highlights how the pros and cons of a recommended item compares to alternative items. We propose a way to score these explanations based on their content. The scores help to identify compelling explanations, providing a strong reason why the item being explained is better or worse than the alternatives. We describe the results of offline experiments and a live-user study based on TripAdvisor data to demonstrate the usefulness of this approach.

1 Introduction

Recommender systems learn about our likes and dislikes to make suggestions to help us decide what to watch, read, and buy. But generating a list of suggestions is just part of the recommendation process. Explaining recommendations can make it easier for users to make good decisions, increasing conversion rates for businesses, and leading to more satisfied users; see [1, 2, 5, 6, 12, 14, 16]. For example, early work explored the utility of explanations in collaborative filtering with [6] reviewing various ways to explain movie suggestions using ratings, meta-data, neighbours and testing different presentation styles (histograms, confidence intervals, text).

A. Lawlor (✉) · K. Muhammad · R. Rafter · B. Smyth
Insight Centre for Data Analytics, University College Dublin, Belfield, Dublin 4, Ireland
e-mail: aonghus.lawlor@insight-centre.org

K. Muhammad
e-mail: khalil.muhammad@insight-centre.org

R. Rafter
e-mail: rachael.rafter@insight-centre.org

B. Smyth
e-mail: barry.smyth@insight-centre.org

Bilgic and Mooney [1] used keywords to justify items arguing that the goal of an explanation should not be to “sell” the user on the item but to help the user to make an informed judgment. They found users overestimated item quality when presented with similar-user style explanations. Keyword approaches were also developed by [14] to generate explanations in the style of “*Item A is suggested because it contains feature X & Y that are also included in items B, C, & D, which you also like.*”; see [16] for related ideas based on user tags. Explanations like this relate one item to others. Pu and Chen [12] build explanations that emphasise the tradeoffs between items, such as “*Here are laptops that are cheaper and lighter but with a slower processor*”; see also [13]. In Zhang et al. [17] a hybrid matrix factorisation framework for personalised recommendations was developed based on user-feature and item-feature relationships, and feature-level explanations are designed which highlight the features pushing the item into the top-K list.

In this work we generate *feature-based, personalised, opinionated* explanations (see also [15]). Like the work of [12, 13], our explanations relate items to alternative recommendations, to help the user to better understand the trade-offs and compromises that exist within a product-space; see also [9]. We also leverage the opinions in user-generated reviews as our primary source of item and recommendation knowledge. This paper pays particular attention to the clarity and helpfulness of these opinionated explanations, thereby complementing related work by [11], which focused on the role of opinionated explanations in ranking.

2 Opinionated Recommendation

This paper builds on recent work about mining opinions from user reviews to generate item descriptions for recommender systems. The work of [4] describes how shallow NLP, opinion mining, and sentiment analysis can be used to extract rich feature-based product cases. It is not possible to fully cover these techniques here and the interested reader is referred to [3, 4]. However, we will provide a brief summary based on Fig. 1. We will rely on a TripAdvisor dataset of hotels and reviews and so from time to time we will refer to this data without loss of generality.

2.1 Review Feature Extraction

As in [3, 4], for review r_i we mine *bi-gram* features and *single-noun* features; see [7, 8]; e.g. bi-grams which conform to a noun followed by a noun (e.g. *bath tub*) or an adjective followed by a noun (e.g. *double room*) are considered, excluding bi-grams whose adjective is a sentiment word (e.g. *excellent, terrible* etc.). Separately, single-noun features are validated by eliminating nouns that are rarely associated with sentiment words in reviews as per [7], since such nouns are unlikely to refer to item features. We refer to each of these extracted features, f_j 's as *review features*.

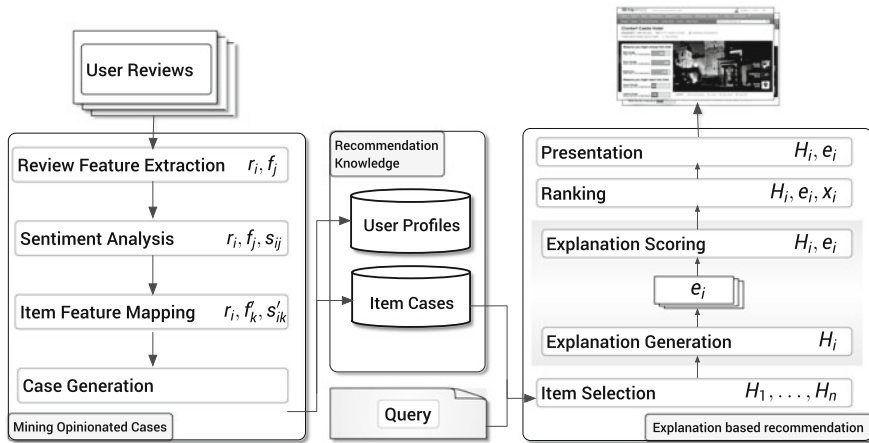


Fig. 1 An overview of the explanation-based recommendation architecture

2.2 Sentiment Analysis

For a review feature f_k we determine if there are any sentiment words in the sentence containing f_k . If not, f_k is marked *neutral*, otherwise we identify the sentiment word w_{min} with the minimum word-distance to f_k . Next we determine the part-of-speech (POS) tags for w_{min} , f_k and any words that occur between w_{min} and f_k . This POS sequence is an *opinion pattern*. We compute the frequency of all opinion patterns in all reviews; a pattern is *valid* if it occurs more than average. For valid patterns, we assign sentiment to f_k based on the sentiment of w_{min} and subject to whether the corresponding sentence contains any negation terms within 4 words of w_{min} . If there are no negation terms, then the sentiment assigned to f_k is that of the sentiment word in the sentiment lexicon; otherwise this sentiment is reversed. If an opinion pattern is not valid then we assign a *neutral* sentiment to each of its occurrences within the review set; see [10] for a fuller description. The end result of sentiment analysis is that we determine a sentiment label s_{ik} for each f_k in review r_i .

2.3 Item Feature Mapping

These review features often refer to very specific hotel details (e.g. the *orange juice* at breakfast) which are often too fine-grained for explanation purposes; although [3, 4] have shown their utility in recommendation. Therefore we map these low-level features to higher-level *item features* which correspond to features such as *bar/restaurant*, *room quality*, *breakfast*, etc. To automate this mapping, we apply a k-means clustering to the set of review sentences to find words that tend to co-occur frequently in review sentences. After some manual adjustment the resulting

clusters can be labelled with a known set of high-level features; we use TripAdvisor's *amenities*. Thus, each cluster contains a set of low-level *review features* and is mapped to a high-level *item feature*. Using this information we can automatically map each (r_i, f_k, s_{ik}) review feature tuple to a corresponding (r_i, f'_j, s'_{ij}) item feature tuple.

2.4 Case Generation: Constructing Item Cases

For each item/hotel H we have review features $\{f_1, \dots, f_m\}$ mined from $reviews(H)$. Each review feature is mapped to an item feature f'_j and we aggregate the review feature's mentions and sentiment scores to associate them with the corresponding f'_j . We can compute various properties of each f'_j : the fraction of times it is mentioned in reviews (its *importance* and the degree to which it is mentioned positively or negatively (its *sentiment* as in Eqs. 1 and 2; note, $pos(f'_j, H)$ and $neg(f'_j, H)$ denote the number of times that feature f'_j has positive or negative sentiment in reviews for H , respectively). Thus, each hotel can be represented as a *case*, $item(H)$, which aggregates item features, importance and sentiment data as in Eq. 3.

$$imp(f'_j, H) = \frac{count(f'_j, H)}{\sum_{f' \in H} count(f', H)} \quad (1)$$

$$sent(f'_j, H) = \frac{pos(f'_j, H)}{pos(f'_j, H) + neg(f'_j, H)} \quad (2)$$

$$item(H) = \{(f'_j, sent(f'_j, H), imp(f'_j, H)) : f'_j \in features(H)\} \quad (3)$$

2.5 Case Generation: Constructing User Profiles

User profiles are produced in a similar way: for user U review features are mined from U 's reviews; each review feature is mapped to an item feature; and we aggregate these item features and their popularity scores as a profile. Currently, we don't store sentiment user profiles as it is not required at the present time. However, in the future we intend to consider this option in more detail as such information may prove valuable when it comes to understanding a user's rating tendency.

$$user(U) = \{(f'_j, imp(f'_j, U)) : f'_j \in reviews(U)\} \quad (4)$$

3 From Opinions to Compelling Explanations

In what follows we assume the target user U_T is presented with a set of hotel recommendations $\{H_1 \dots H_n\}$. Our task is to generate an explanation for each of these recommendations in turn and we will refer to the current one as the *target* hotel or H_T and the other items as the *alternatives* or H' .

3.1 Generating a Basic Explanation Structure

We will describe the construction of a *basic* explanation structure, which begins with the data structure shown in Fig. 2.

Explanations come in two parts. The *pro* part is a set of positive features; reasons you might choose the hotel. The *con* part comprises the negative features; reasons to reject the hotel. These features are selected on the basis of three key components:

1. Sentiment Score—item feature $f'_j \in H_T$ is a *pro* if it has a majority of positive sentiments ($sent(f'_j, H_T) > 0.7$ in our TripAdvisor data) otherwise it is a *con*.
2. Relationship to Alternatives—To be a *pro* (or *con*) item feature f'_j must have a sentiment score that is *better than* (or *worse than*) at least one of the alternatives.
3. Importance to User—item feature f'_j must be contained within the user profile to ensure it has been mentioned by the user in his or her past reviews.

We generate explanations with these *pro* and *con* features using Eqs. 5–10. In Fig. 2 we can see an example of how this selects *pros* such as *Bar/Lounge* and *Free Breakfast*, which are important to the user, positive in sentiment, and better than some of the alternative recommendations. Likewise, we see *cons* such as *Leisure Centre*, which are also relevant to the user, but this time less favourably reviewed and worse than some of the alternatives.

		Feature	Importance	Sentiment	BetterThan/WorseThan
Hotel: Clontarf Castle User: Peter Parker	PROS	Bar/Lounge *	0.25	0.60	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%
		:	:	:	:

Fig. 2 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

$$betterThan(f'_j, H_T, H') = \frac{\sum_{H_c \in H'} 1[sent(f'_j, H_T) > sent(f'_j, H_c)]}{|H'|} \quad (5)$$

$$worseThan(f'_j, H_T, H') = \frac{\sum_{H_c \in H'} 1[sent(f'_j, H_T) < sent(f'_j, H_c)]}{|H'|} \quad (6)$$

$$pro(f'_j, U_T, H_T, H') \leftrightarrow sent(f'_j, H_T) > 0.7 \wedge betterThan(f'_j, H_T, H') > 0 \wedge imp(f'_j, U_T) > 0 \quad (7)$$

$$con(f'_j, U_T, H_T, H') \leftrightarrow sent(f'_j, H_T) < 0.7 \wedge worseThan(f'_j, H_T, H') > 0 \wedge imp(f'_j, U_T) > 0 \quad (8)$$

$$Pros(U_T, H_T, H') = \{(f'_j, v, m) : pro(f'_j, H_T, H') \wedge v = betterThan(f'_j, H_T, H') \wedge m = imp(f'_j, U_T)\} \quad (9)$$

$$Cons(U_T, H_T, H') = \{(f'_j, v, m) : con(f'_j, H_T, H') \wedge v = worseThan(f'_j, H_T, H') \wedge m = imp(f'_j, U_T)\} \quad (10)$$

3.2 Filtering Compelling Explanations

This basic explanation structure can be made up of many features, which may complicate the decision making if presented to the end user in this way. Moreover, many of the pros might be better than only a small fraction of the alternative recommendations, and conversely for cons, thereby limiting their usefulness as compelling reasons to choose or avoid the hotel in question. However, we can filter features based on how strong a reason they represent for choosing or rejecting the target hotel. To do this we define a *compelling* feature to be one that has a *betterThan* (pro) or *worseThan* (con) score of >50% instead of just >0. Thus, a compelling pro is *better* than a *majority* of alternative recommendations and a compelling con is *worse* than a *majority* of alternatives. A compelling pro may be a strong reason to choose the target hotel while a compelling con may be a strong reason to avoid it.

We define a *compelling explanation* as a non-empty explanation which contains only compelling pros and/or compelling cons. For instance, referring back to Fig. 2, we have marked compelling features with an asterisk after their name; so, the compelling explanation derived from this basic explanation structure includes *Bar/Lounge*, *Room Quality*, *Restaurant*, as pros, and *Airport Transport* and *Leisure Centre* as cons. These are all features that matter to the user and they distinguish the hotel as either better or worse than a majority of alternatives.

3.3 From Explanations to Ranking

As an aside it is worth highlighting another aspect of this work: the idea that explanations might also be used for the *ranking* of recommendations. We can estimate the quality of an explanation numerically and use this for ranking purposes. To do this we use a straightforward scoring function to measure the strength of an explanation as the weighted difference of its pros and cons as shown in Eq. 11; this can be applied to either basic or compelling explanation structures.

$$\begin{aligned}
 strength(U_T, H_T, H') = & \\
 & \sum_{f \in Pros(U_T, H_T, H')} betterThan(f, H_T, H') \times imp(f, U_T) - \\
 & \sum_{f \in Cons(U_T, H_T, H')} worseThan(f, H_T, H') \times imp(f, U_T) \quad (11)
 \end{aligned}$$

A further discussion of the role of explanations in recommendation ranking is beyond the scope of this paper. However, the interested reader is referred to the work of [11] for a more in-depth treatment of this idea.

4 The Explanation Interface

How can explanation information be presented in a helpful way to users? Fig. 3 shows three example treatments that could be presented alongside a given hotel description. In each, features that matter to the user are separated into pros and cons; we also only present compelling features. The features are ranked based on their importance to the user and each is associated with a sentiment bar to indicate the percentage of positive sentiments expressed by reviewers. Treatments Fig. 3b, c further enrich the explanation by relating each feature to the other recommendation alternatives at different levels of precision. Figure 4 shows an example of one of these explanation types in context in TripAdvisor.

5 Evaluation

Next, we describe a pair of evaluations designed to explore the form and function of our explanations in the context of a TripAdvisor dataset and user judgements.

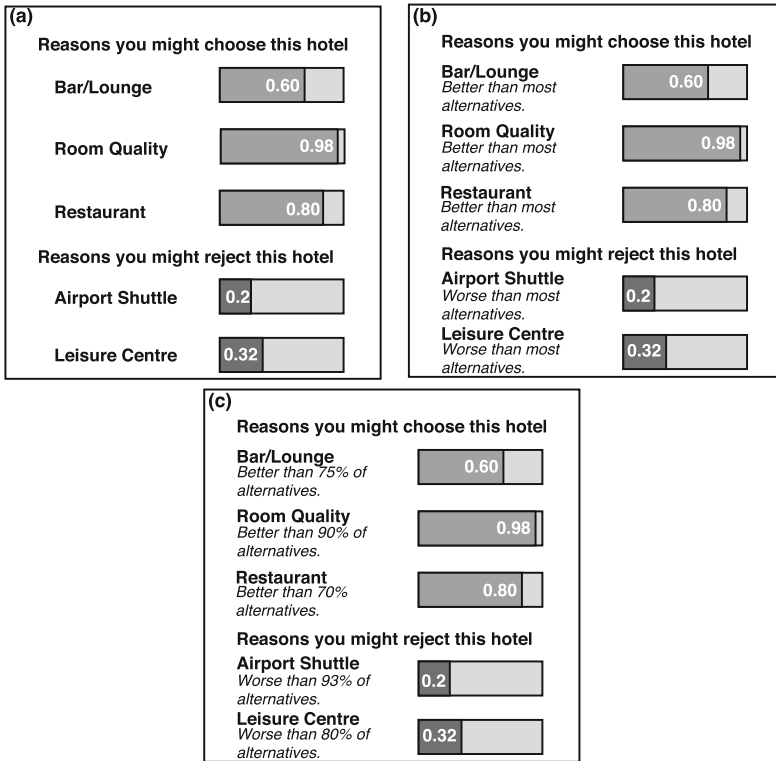


Fig. 3 Explanation styles: **a** sentiment only; **b** sentiment plus alternatives; **c** sentiment plus alternatives plus percentages

5.1 Offline Evaluation

For the first part of our evaluation, we use a large TripAdvisor dataset as a source of user profiles, reviews, and hotel cases. It contains 10,000 users who have each written at least 10 hotel reviews for 2,062 hotels. In addition, we had more than 220,000 reviews by almost 150,000 reviewers available for the hotel cases. For each target user U_T we know the hotel they booked, H_B , and the related hotels recommended by TripAdvisor; we understand that TripAdvisor generates these using a combination of location, similar users, and meta-data. Thus, we can generate approximately 100,000 *user sessions*, one for each user booking and containing the booked hotel and the related TripAdvisor suggestions. Next we generate basic and compelling explanations for each of the hotels in a user session—that’s approximately 1,000,000 explanations of each type—and analyse their form, focusing on the number and type of features that are commonplace in the resulting explanations.

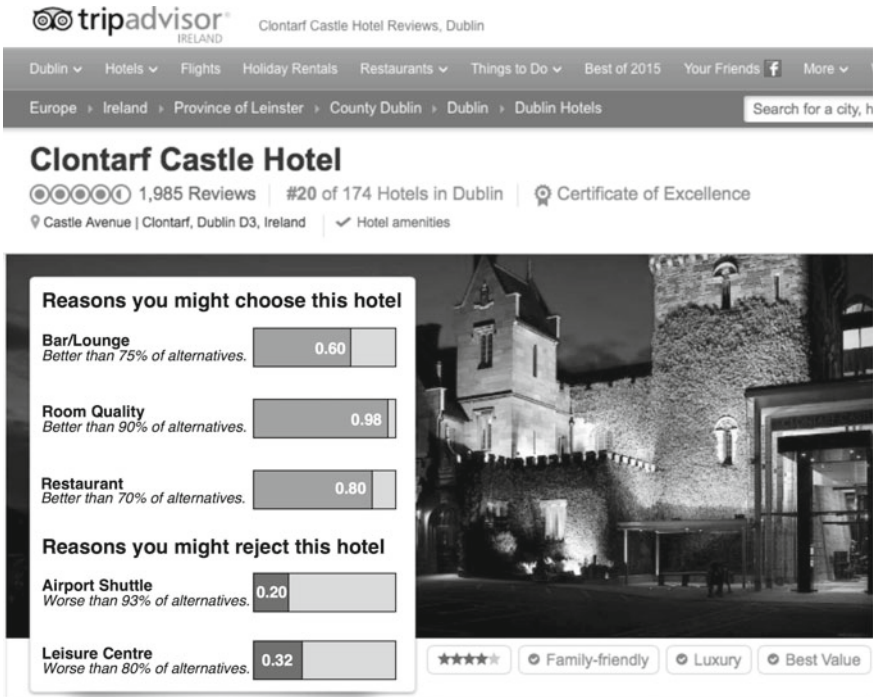


Fig. 4 An example explanation (sentiment plus alternatives plus percentages) in context. By mousing over the sentiment bars, the user sees a preview of relevant review fragments. It is also feasible to use the explanation as a navigation aid so that by clicking on the sentiment bars or explanation text the user can navigate to corresponding reviews or alternative candidates

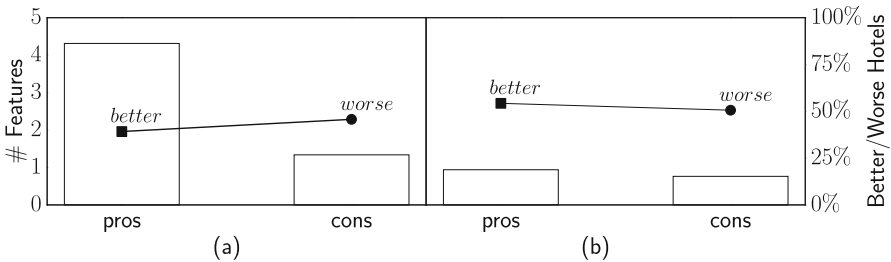


Fig. 5 The average number of pros and cons and the average *betterThan* and *worseThan* per explanation for basic and compelling explanations. **a** Basic. **b** Compelling

5.1.1 Pros versus Cons, Better versus Worse

Figure 5a, b shows the average number of pros and cons (left y-axis), and the average *betterThan/worseThan* scores (right y-axis), in basic and compelling explanations. We see that on average we are recommending about 4 pros versus only 2 cons in

basic explanations compared to 2 pros and 2 cons in compelling explanations. The extra pros in basic explanations reflect the positive bias in TripAdvisor reviews but it is interesting that approximately half of these pros are not compelling.

This bias is also suggested by the difference between the average *betterThan* score for pros (49 %) and the *worseThan* score for cons (69 %). For a typical hotel, its basic pros will typically be better than about 49 % of the alternatives in the recommendation session. In contrast, when it comes to the basic cons, it is usually the case that the hotel in question does worse than most of the alternatives in the recommendation session. A similar pattern is seen for compelling explanations, although the difference now is less pronounced; 70 % average *betterThan* scores for compelling pros versus approximately 75 % *worseThan* scores for compelling cons.

Overall we see that compelling explanations are simpler than basic explanations—they contain fewer pros and cons—and they are more compelling because their features are better or worse than a large majority of the alternative recommendations. Intuitively this combination of simplicity and compellingness should make compelling explanations particularly effective when it comes to helping users to decide whether to accept or reject a given recommendation.

5.1.2 On the Frequency of Explanation Features

Figure 6 shows the frequency distributions for the features contained in basic and compelling explanations. In each histogram, the individual bars refer to a specific item feature, and each bar shows the number of times that the feature occurs as a pro and as a con. The histograms also show the average *betterThan* and *worseThan* scores for these features, based on their pro and con occurrences, respectively.

We see that a handful of item features tend to dominate in explanations. Features like *free breakfast* and *bus service* appear very frequently compared to others such as *fitness centre*, *high-speed wifi*, and *kids activities*. We also see the strong positive review bias in TripAdvisor as a majority of features present mostly as pros. For example, in Fig. 6a we can see that *free breakfast* appears as a con in 47,263 explanations but as a pro in 238,577 explanations; it is worth noting the unusually high negative sentiment associated with the *fitness centre* and *high speed wifi* features, both of which appear more frequently as cons than pros. This also explains the relatively high value for the *worseThan* scores compared to *betterThan* scores mentioned previously. It is relatively unusual for a feature to be listed as a con (<20 % of the time in most cases) and so if a feature is a con it is likely that it is a pro in the alternative recommendation candidates and so it is likely to have a sentiment score that is worse than a majority of alternatives. In contrast if a feature is a pro in an explanation it is also likely to be a pro in the explanations of the alternatives and so it is less likely to have a sentiment score that is better than most alternatives.

This data tells us about the features that matter the most to users (based on their reviews) but it also indicates whether a particular feature is likely to appear as a compelling pro or a compelling con in an explanation. For example, *free breakfast* is the most common feature to appear in explanations and it appears as a pro over 85 % of

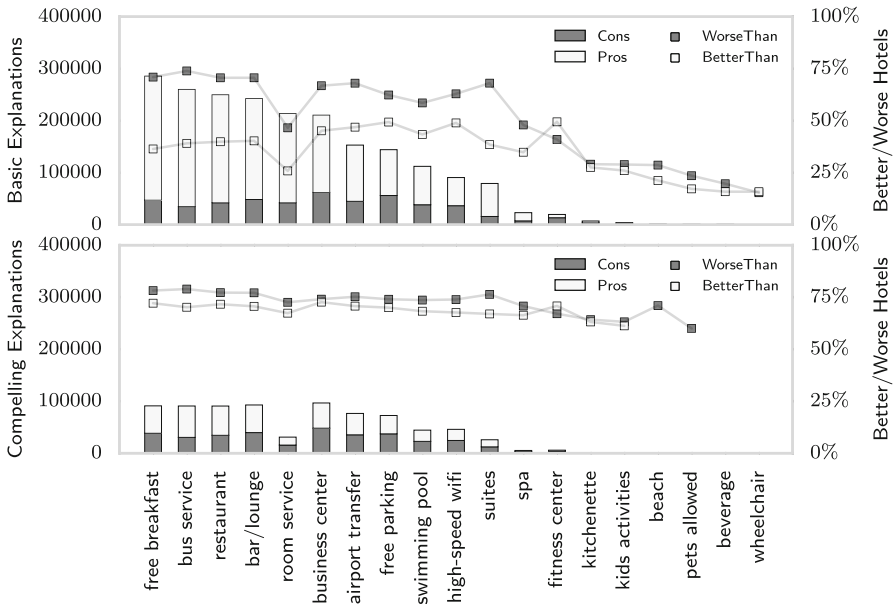


Fig. 6 An analysis of the relative frequency of features in the pros and cons of explanations and their corresponding *betterThan* and *worseThan* scores

the time and a con just under 15 % of the time. However, as a pro it has an average *betterThan* score of only about 30 %, whereas as a con it has a *worseThan* score of almost 80 %. Therefore, this feature is less likely to appear as a pro in a compelling explanation whereas it is very likely to appear as a con in a compelling explanation. This is evident in Fig. 6b which shows the corresponding data for compelling explanations. This time *free breakfast* appears as a con in 38,271 compelling explanations and as a pro in 52,763 explanations. As a hotel owner, if your hotel’s *free breakfast* is being negatively reviewed then there is a strong likelihood that this feature will be exposed as a compelling con in any explanation generated for your hotel. As a user who has a preference for *free breakfast*, you will likely be influenced by this feature as a con in compelling explanations.

5.2 Live-User Study

The true test of this approach will depend on the opinions of users in a live setting and whether or not the explanations help users make better decisions in the long-term. This is a challenging evaluation setting and it is beyond the scope of the present work to fully explore this broader issue. That being said, we have completed an initial user study to gather initial impressions of different explanation styles and types of information and we will summarise the results of this study in what follows.

5.2.1 Setup

Our user study took the form of an online questionnaire, which placed participants in a simple hotel booking setting, asking them to evaluate the 3 styles of explanation interface presented earlier (Fig. 3) in the context of TripAdvisor as per Fig. 4. In what follows we will refer to these 3 styles as *S* (pros and cons with sentiment only), *S + A* (pros and cons with sentiment and comparison to alternatives), and *S + A + %* (pros and cons with sentiment and percentage comparison to alternatives).

48 people participated in the user study, mostly Ph.D. students and researchers in our research centre. They were presented with each interface in turn—varying the presentation order—and they were asked to express their agreement on a scale of 1 (strongly disagree) to 10 (strongly agree) with each of the following two statements:

1. *Clarity*: The explanation is clear and easy to understand.
2. *Helpfulness*: The explanation will help me to make a choice about whether or not to choose or reject this hotel.

Finally, each participant was asked to rate the usefulness of the various explanation components used in these interfaces on a scale of 1 (not useful) to 10 (very useful) by responding to the following questions:

1. How useful was it to separate the amenities into groups of pros (positive sentiment) and cons (negative sentiment)?
2. How useful did you find the sentiment bars?
3. How useful did you find the explanations that compared the hotel to alternative recommendations?
4. When comparing the hotel to alternatives how useful did you find the precise percentage information?

5.2.2 Results

The results are presented in Fig. 7a for each of the 3 interfaces. We can see that overall participants found the interfaces clear and helpful with a preference for interfaces 2 and 3, which included extra information about how the hotel compared with alternative recommendations in addition to simple sentiment information.

Figure 7b shows the average utility ratings for each of the various explanation components. In general these ratings are high across all of the different explanation components with an average overall rating that is greater than 7. We can see that participants found the separation of amenities into pros and cons particularly useful (an average rating of 8.48) followed by the use of sentiment information in the explanations (7.60). There is little difference expressed between the purely text-based comparison to alternatives (e.g. "... better than most alternatives") and a more precise comparison (e.g. "... better than 93 % of alternatives") with both components scoring above 7 on average.

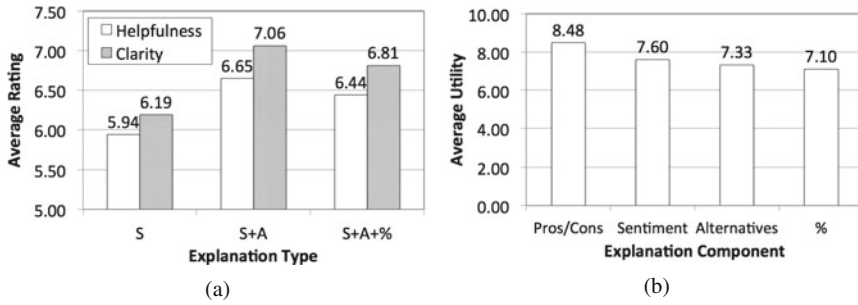


Fig. 7 In **a** we show the ratings for different explanation types. In **b** we show the average utility scores for each component in the explanations

The results, preliminary as they may be, do suggest that users are perceiving value in the type of explanations that we are generating. The combination of sentiment and a comparison to alternatives presents as the preferred interface type with users reporting high levels of clarity and helpfulness.

6 Conclusions

This work brings together ideas from case-based reasoning, opinion mining, and recommender systems [3, 4]. We have described an approach to generating explanations for recommender systems from user reviews. We have evaluated these explanations using a combination of offline and online evaluations using large-scale TripAdvisor data and live-users. As part of our future work, we plan to make progress on a more extensive live-user evaluation involving real-time recommendation sessions. It will also be interesting to incorporate additional information as part of our explanations. For example, in the work presented we compare recommendation candidates to alternative recommendations but we could also consider the relationship to a user’s previous bookings. In this way, our explanations could help the user to understand how a particular hotel/item relates to alternative recommendations but also to hotels they have booked in the past.

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

1. Bilgic, M., Mooney, R.J.: Explaining recommendations: satisfaction versus promotion. In: Proceedings of Beyond Personalization Workshop at the 2005 International Conference on Intelligent User Interfaces, vol. 5, pp. 13–18 (2005)

2. Coyle, M., Smyth, B.: Explaining search results. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence, pp. 1553–1555 (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 (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, IJCAI'13, pp. 1310–1316 (2013)
5. Friedrich, G., Zanker, M.: A taxonomy for generating explanations in recommender systems. *AI Mag.* **32**(3), 90–98 (2011)
6. Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, CSCW'00, pp. 241–250 (2000)
7. Hu, M., Liu, B.: Mining opinion features in customer reviews. In: Proceedings of the 19th National Conference on Artificial Intelligence, pp. 755–760 (2004)
8. Justeson, J.S., Katz, S.M.: Technical terminology: some linguistic properties and an algorithm for identification in text. *Nat. Lang. Eng.* **1**(1), 9–27 (1995)
9. McSherry, D.: Similarity and compromise. In: Proceedings of the 5th International Conference on Case-based Reasoning, pp. 291–305 (2003)
10. Moghaddam, S., Ester, M.: Opinion digger: an unsupervised opinion miner from unstructured product reviews. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 1825–1828 (2010)
11. Muhammad, K., Lawlor, A., Rafter, R., Smyth, B.: Great explanations: opinionated explanations for recommendation. In: Proceedings of the 23rd International Conference on Case-based Reasoning pp. 1–15 (2015)
12. Pu, P., Chen, L.: Trust-inspiring explanation interfaces for recommender systems. *Knowl.-Based Syst.* **20**(6), 542–556 (2007)
13. Reilly, J., McCarthy, K., McGinty, L., Smyth, B.: Explaining compound critiques. *Artif. Intell. Rev.* **24**(2), 199–220 (2005)
14. Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: Providing justifications in recommender systems. *IEEE Trans. Syst., Man, Cybern.—Part A: Syst. Hum.* **38**(6), 1262–1272 (2008)
15. 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 (2008)
16. 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 (2008)
17. Zhang, Y., Lai, G., Zhang, M., Zhang, Y., Liu, Y., Ma, S.: Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In: Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'14, pp. 83–92 (2014)