<table>
<thead>
<tr>
<th>Title</th>
<th>A multi-criteria evaluation of a user generated content based recommender system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors(s)</td>
<td>Garcia Esparza, Sandra; O'Mahony, Michael P.; Smyth, Barry</td>
</tr>
<tr>
<td>Publication date</td>
<td>2011-10-23</td>
</tr>
<tr>
<td>Publication information</td>
<td>ëFreyne, J. et al. (eds.). Proceedings of the 3rd ACM RecSys 10 Workshop on Recommender Systems and the Social Web</td>
</tr>
<tr>
<td>Conference details</td>
<td>Presented at the 3rd Workshop on Recommender Systems and the Social Web (RSWEB-11), 5th ACM Conference on Recommender Systems, Chicago, IL, USA, 23-27 October 2011</td>
</tr>
<tr>
<td>Link to online version</td>
<td><a href="http://www.dcs.warwick.ac.uk/~ssanand/RSWeb11/rsweb2011proceedingsfinal.pdf">http://www.dcs.warwick.ac.uk/~ssanand/RSWeb11/rsweb2011proceedingsfinal.pdf</a></td>
</tr>
<tr>
<td>Item record/more information</td>
<td><a href="http://hdl.handle.net/10197/3509">http://hdl.handle.net/10197/3509</a></td>
</tr>
</tbody>
</table>
A Multi-Criteria Evaluation of a User-Generated Content Based Recommender System

Sandra Garcia Esparza, Michael P. O’Mahony, Barry Smyth
CLARITY: Centre for Sensor Web Technologies
School of Computer Science and Informatics
University College Dublin, Ireland
{sandra.garcia-esparza, michael.omahony, barry.smyth}@ucd.ie

ABSTRACT
The Social Web provides new and exciting sources of information that may be used by recommender systems as a complementary source of recommendation knowledge. For example, User-Generated Content, such as reviews, tags, comments, tweets etc. can provide a useful source of item information and user preference data, if a clear signal can be extracted from the inevitable noise that exists within these sources. In previous work we explored this idea, mining term-based recommendation knowledge from user reviews, to develop a recommender that compares favourably to conventional collaborative-filtering style techniques across a range of product types. However, this previous work focused solely on recommendation accuracy and it is now well accepted in the literature that accuracy alone tells just part of the recommendation story. For example, for many, the promise of recommender systems lies in their ability to surprise with novel recommendations for less popular items that users might otherwise miss. This makes for a riskier recommendation prospect, of course, but it could greatly enhance the practical value of recommender systems to end-users. In this paper we analyse our User-Generated Content (UGC) approach to recommendation using metrics such as novelty, diversity, and coverage and demonstrate superior performance, when compared to conventional user-based and item-based collaborative filtering techniques, while highlighting a number of interesting performance trade-offs.

Categories and Subject Descriptors
H.4 [Information Systems Applications]: Miscellaneous

General Terms
Algorithms, Experimentation

Keywords
Recommender Systems, User-Generated Content, Performance Metrics

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Copyright 20XX ACM X-XXXX-XX/XX/XX ...$10.00.

1. INTRODUCTION
Recommender systems allow users to discover information, products and services by predicting their needs based on the past behaviour of like-minded users. Typically these systems can be classified in three main categories: collaborative filtering (CF), content-based (CB) and hybrid approaches. In CF approaches [4, 7], users are recommended items that users with similar interests have liked in the past, whereas users interests in items are represented by ratings. In contrast, in CB approaches [14], users are recommended items that are similar to those items that the user liked in the past, where item descriptions (e.g. movies can be described using metadata such as actors, genres etc.) are used to measure the similarity between items. Finally, researchers have looked at the potential of combining CF and CB approaches as the basis for hybrid recommendation strategies [5]. However, one of the problems with these systems is that they need sufficient amounts of data in order to provide useful recommendations, and sometimes neither ratings nor metadata are available in such quantities. For this reason researchers have started looking into additional sources of recommendation data. In the last few years, the Social Web has experienced significant growth, with the emergence of new services, such as Twitter, Flixster and Foursquare, whose users collectively generate very large volumes of content in the form of micro-blogs, reviews, ratings and check-ins. These rich sources of information, namely User-Generated Content (UGC), which sometimes relate to products and services (such as movie reviews or restaurant check-ins), are becoming increasingly plentiful and researchers have already started to utilise this content for the purposes of recommendation.

Here we focus on UGC in the form of product reviews (see, for example, Figure 1). We believe this type of information offers important advantages in comparison to other sources of recommendation knowledge such as ratings and metadata. For instance, when reviewing products, users often discuss particular aspects that they like about them (e.g. “Tom Hanks and Tim Allen at their best”), as well as commenting on general interests (e.g. “I love animation”), which is not always reflected in other sources of recommendation knowledge. In this sense, in a collaborative filtering approach, two users that have rated movies similarly are treated as people with similar interests. However, it often happens that users may like the same movies for different reasons. For example, one user may have rated a movie with a high score because they loved the special effects while the other one rated the same movie highly because they loved the plot and the ac-
tors’ performances. In a similar way, a content-based approach usually relies on product descriptions to draw item similarities, but it does not consider how the user feels towards each of these descriptions. A solution to this problem is to consider multi-criteria ratings, where different aspects of a product (or service) are rated separately [2]. For example, when rating restaurants in TripAdvisor, users can rate along the dimensions of service, food, value and atmosphere. One of the disadvantages of this approach is that these multi-criteria ratings tend to be predefined and thus can restrict users from commenting on other aspects of the product. Social tagging provides a solution to this by allowing users to associate tags to content. In [18] these tags, which reflect users’ preferences, are introduced into recommender algorithms called tagommenders, and results show that these algorithms performed better than state of the art algorithms. [8] also comments on the benefits of using tag-based user profiles and they investigate different techniques to build condensed and optimised profiles. However, not many systems provide this social tagging functionality; instead users’ ratings and reviews are more common and abundant and for this reason we believe it is important to consider them for recommendation purposes.

The most common way to evaluate recommenders is to measure how accurate they are at predicting items that users like. However, one of the problems with evaluating the accuracy of top-N recommendation lists is that current metrics (such as precision and recall) reward algorithms that are able to accurately predict test set items (typically constructed by selecting a random subset of users’ known ratings), while failing to consider items that are not present in test sets, but which users may in fact like. Hence current recommendation accuracy evaluation techniques are limited, which may result in promising new algorithms being labelled as poor performers. Further, it has been shown that accuracy on its own is insufficient to fully measure the utility of recommendations [19]. Other performance criteria, such as novelty and diversity of recommended lists are now acknowledged as also being important when it comes to user satisfaction. In addition, the ability to recommend as many products as possible, known as coverage, is also a desirable system property.

While these performance criteria have been considered in the past [20, 24, 25], however they have been less explored in the context of UGC-based recommenders. In past work we have implemented a recommendation approach where UGC in the form of micro-reviews was used as the source of recommendation knowledge [9]. An evaluation performed on 4 different domains showed that UGC, while inherently noisy, provided a useful recommendation signal and outperformed a variation of a collaborative-filtering based approach. Here, we expand on this work by considering additional performance criteria, such as novelty, diversity and coverage, and we compare the performance of our approach with traditional user-based and item-based collaborative filtering approaches [4, 7].

2. RELATED WORK

UGC has been leveraged by recommender systems for different purposes such as enriching user profiles or extracting ratings from reviews by using sentiment analysis techniques. In addition, in the last few years it has been shown that accuracy on its own is insufficient to fully measure the utility of recommendations and new metrics have been proposed. Here, we provide an overview of some of the work that has been carried out in these areas.

2.1 Review-based Recommendations

Recent work has focused on leveraging UGC in the form of reviews for recommendations. For example, a methodology to build a recommender system which leverages user-generated content is described in [23]. Although an evaluation is not performed, they propose a hybrid of a collaborative filtering and a content-based approach to recommend hotels and attractions, where the collaborative filtering component utilises the review text to compute user similarities in place of traditional preference-based similarity computations. Moreover, they also comment on the advantages of using user-generated content for recommender systems; such as, for example, providing a better rationale for recommended products and increasing user trust in the system.

An early attempt to build a recommender system based on user-generated review data is described in [1]. Here, an ontology is used to extract concepts from camera reviews and recommendations are provided based on users’ requests about a product; for example, “I would like to know if Sony361 is a good camera, specifically its interface and battery consumption”. In this case, the features interface and battery are identified, and for each of them a score is computed according to the opinions (i.e. polarities) of other users and presented to the user.

Similar ideas are described in [3], which look at using user-generated movie reviews from IMDb in combination with movie meta-data (e.g. keywords, genres, plot outlines and synopses) as input for a movie recommender system. Their results show that user reviews provide the best source of information for movie recommendations, followed by movie genre data. In addition, in [12, 15], the number of ratings in a collaborative filtering system is increased by inferring new ratings from user reviews using sentiment analysis techniques. While [15] generate ratings for Flixster reviews by extracting the overall sentiment expressed in the review, [12] extract features and their associated opinions from IMDb reviews and a rating is created by averaging the opinion polarities (i.e. positive or negative) across the various features. Both approaches achieve better performance when using the ratings inferred from reviews when compared to using ratings predicted by traditional collaborative filtering approaches.

2.2 Beyond Accuracy Metrics

Typically, the performance evaluation of recommendation algorithms is done in terms of accuracy, which measures how well a recommender system can make recommendations. However, it has been shown that accuracy on its
own is insufficient to fully measure the utility of recommendations and over the last few years new metrics have been proposed [13,19]. Here we are interested in coverage, novelty and diversity.

Coverage is a measure of the domain of items in the system over which the recommender system can form predictions or make recommendations [10]. Sometimes algorithms can provide highly accurate recommendations but only for a small portion of the item space. Systems with poor coverage may be capable of just recommending well-known or popular products, while less mainstream products (belonging to the long tail) are rarely if ever recommended, resulting in a poor experience for recommendation consumers and providers alike. For this reason it is useful to examine accuracy and coverage in combination; a “good” recommender will achieve high performance along both dimensions.

Novelty measures how new or different recommendations are from a user’s perspective. For instance, a movie recommender that keeps suggesting movies that the user is already aware of is unlikely to be very useful to the user, although the recommendations may well have high accuracy. Diversity refers to how dissimilar the products in recommendation lists are. For example, a music recommendation list where most albums are from the same band will have a low diversity, and such a situation is not desirable. Indeed, in [20], it is argued that diversity can sometimes be as important as similarity.

Current research is focused on improving the novelty and diversity of the recommendations without sacrificing accuracy. For instance, in [24] the authors suggest an approach to recommend novel items by partitioning the user profile into clusters of similar items. Further, in [25], the authors introduce diversity in their recommendation lists and results show that although the new lists show a decrease in accuracy, users are more satisfied with the diversified lists. UGC in the form of tags has also been used to introduce diversity in recommendation lists and results showed that their method was also able to improve the accuracy [22].

The goal of this paper is to explore the advantages of using UGC in the form of reviews in a recommender system. Similar work has been discussed above; however, these approaches have been evaluated in terms of accuracy and other metrics, which have proven to be equally important, have not been taken into account. In this paper we provide an evaluation of a UGC-based recommendation approach which considers metrics that go beyond accuracy; in particular we are interested in the properties of novelty, diversity and coverage.

### 3. METHODOLOGY

The goal of this paper is to evaluate our UGC-based recommender by providing a multi-criteria evaluation. The recommender is similar to that described in our previous work [9], where UGC in the form of product micro-reviews are used as the source of recommendation knowledge. In particular, we propose an index-based approach where users and items are represented by the terms used in their associated reviews. This section describes two variants of our index-based approach and two variants of collaborative recommenders, which are used as benchmark techniques.

#### 3.1 Index-based Approaches

The approach to recommend products to users consists of an index-based approach where products and users are represented by the terms in their associated reviews. In this initial work we only consider terms from positive user reviews (i.e. reviews which are associated with a rating of greater than 3 on a 5 point scale). The reason for this is that in such reviews users tend to talk about the aspects they like in products. This poses some obvious limitations discussed in section 5 which will be addressed in future work.

Our recommendation approach involves the creation of a product index, where each product $P_i$ can be viewed as a document made up of the set of terms, $t_1, \ldots, t_n$, used in its associated user reviews, $r_1, \ldots, r_k$, as per Equation 1.

$$P_i = \{r_1, \ldots, r_k\} \equiv \{t_1, \ldots, t_n\}$$

(1)

Using techniques from the information retrieval community, we can apply weights to terms associated with a particular product according to how representative they are with respect of that product. In this work we have used the well known TFIDF approach [17] to term weighting (Equation 2).

Briefly, the weight of a term $t_j$ in a product $P$, with respect to some collection of products $P$, is proportional to the frequency of occurrence of $t_j$ in $P_i$ (denoted by $n_{t_j, P_i}$), but inversely proportional to the frequency of occurrence of $t_j$ in $P$ overall, thus giving preference to terms that help to discriminate $P_i$ from the other products in the collection.

We use Lucene\(^1\) to provide this indexing, term-weighting and retrieval functionality.

$$\text{TFIDF}(P, t_j, P) = \frac{n_{t_j, P_i}}{\sum_{t_k \in P_i} n_{t_k, P_i}} \times \log \left( \frac{|P|}{|\{P_k \in P : t_j \in P_k\}|} \right)$$

(2)

\(^1\)http://lucene.apache.org/
Similarly, we can create the profile of a user by using the terms in their associated (positive) reviews. To provide recommendations for this user we can use their profile as a query into the product index, and return the top-N list of products that best match the query as recommendations.

In addition to the vanilla index-based approach (IB) outlined above, we have also consider a variation where only nouns and adjectives from reviews\(^2\) (IB-\(a\)) are used to form the product index and user queries. We also considered extracting only from reviews, but better results were obtained when adjectives were included. Further, for both index-based approaches we applied stemming, stop-word removal and removed words that appeared in more than 60% of user profiles and products to exclude common domain-specific words (such as “movie”, “plot” and “story”).

This index-based approach is illustrated in Figure 2. One of the advantages of this approach is that user profiles can be independent from the product index (i.e. users may not have reviewed products from the product index), allowing us to use a product index from one particular source (e.g. a movie index created from Flixster reviews) with user profiles from another source (e.g. user interests extracted by analysing that user’s Twitter messages). This independence allows for cross-domain possibilities which in turn can be used to mitigate the cold start problem.

3.2 Collaborative Filtering Approaches

We study two variations of collaborative filtering (CF): user-based and item-based techniques. For these techniques, entries in the user-product ratings matrix consist of 1 (if the user likes a product, i.e. has assigned a rating of \(\geq 4\)) or the special symbol ⊥ (meaning that the user has not reviewed the product or has assigned a rating of \(\leq 3\)).

User-based CF (UBCF) [4]. In order to provide a top-N list of recommended items for a target user, the \(k\) most similar users (neighbours) to that user are selected using cosine similarity. Then, the union of the products rated by each of the neighbours (less those already present in the target user’s profile) is returned, ordered by descending frequency of occurrence. The top-N products are returned as recommendations.

Item-based CF (IBCF) [7]. In this case, recommended products are generated by first taking the union of the \(k\) most similar products (using cosine similarity) to each of the products in the target user’s profile, again ordered by descending frequency of occurrence. After removing the products already present in the target user’s profile, the top-N products are selected as recommendations.

4. EVALUATION

4.1 Metrics

We use four different metrics in order to evaluate our index-based recommendation approach. Accuracy measures the extent to which the system can predict the items that the users like. We measure this in terms of the \(F1\) metric (Equation 3), which is the harmonic mean of precision and recall [21].

\[
F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.
\]

Novelty measures how new or different recommended products are to a user. Typically, the most popular items in a system are the ones that users will be the most familiar with. Likewise, less mainstream items are more likely to be unknown by users. In related work novelty is often based on item popularity [6, 25]. Here we follow a similar approach and compute the novelty of a product as one minus its popularity; where the popularity (\(\text{popularity}(i)\)) of a product, \(i\), is given by the number of reviews submitted for the product divided by the maximum number of reviews submitted over all products. Hence, the novelty of a top-N list of recommended products is computed as the average of each product’s novelty as per Equation 4.

\[
\text{Novelty} = \frac{\sum_{i=1}^{N} (1 - \text{popularity}(i))}{N}.
\]

We define the diversity of a top-N list of recommended products as the average of their pairwise dissimilarities [20]. Let \(i\) and \(j\) denote two products and let \(\text{dissimilarity}(i, j) = 1 - \text{similarity}(i, j)\) and, assuming a symmetric similarity measure, diversity is given by:

\[
\text{Diversity} = \frac{2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (1 - \text{similarity}(i, j))}{N \times (N - 1)}.
\]

The similarity, \(\text{similarity}(i, j)\), between two products is computed using cosine similarity on the corresponding columns of the ratings matrix; for normalisation purposes, this is divided by the maximum similarity obtained between two products. Similar trends were obtained by computed similarity over the documents in the product index.

Finally, the ability to make recommendations for as many products as possible is also a desirable system property. This is reflected by the coverage metric, which for a given user is defined as the percentage of the unrated set of products for which the recommender is capable of making recommendations. Then the overall coverage provided by the system is given by the mean coverage over all users.

4.2 Dataset and Methodology

In this paper we consider Flixster\(^3\) as our source of data. Flixster is an online movie website where users can rate movies and also write reviews about them. We selected reviews authored in the English language only and performed some standard preprocessing on the reviews; such as removing stop-words, special symbols, digits and multiple character repetitions (e.g. we reduce coool to cool). Further, we selected users and movies with at least 10 associated positive reviews. This produced a total of 43179 reviews (and ratings) by 2157 users on 763 movies. The average number of reviews per user is 20 and per movie is 57.

To evaluate each algorithm, first we randomly split each users’ reviews and ratings into training (60%) and test (40%) sets. Second, we create the product index or ratings matrix using the training data. Then, for each user we produce a top-N list of recommendations using the approaches described in Section 3, and compute accuracy using the test set.

\(^3\)http://www.flixster.com
IB+ providing 87% diversity for top-10 recommended lists, index-based approaches performed significantly better, with est novelty (approximately 34%). In terms of diversity, the followed by IB and IBCF, with UBCF providing the worst performing approach in terms of coverage was IBCF approaches in terms of diversity, novelty and coverage. The first, our index-based approaches outperformed both CF ap-

4.3 Results

The results shown in Figure 3 indicate that the collaborative filtering approaches outperformed both index-based approaches in terms of accuracy; with user-based CF performing best overall. The index-based approach using nouns and adjectives (IB+) actually performed slightly worse than the standard bag-of-words approach (IB), which indicates a loss of information using only the terms selected.

Although the accuracy results may seem discouraging at first, our index-based approaches outperformed both CF approaches in terms of diversity, novelty and coverage. The worst performing approach in terms of coverage was IBCF (only 63.23%), while both IB and IB+ achieved in excess of 90% coverage. In terms of novelty, IB+ was the best approach, with 63% novelty for top-10 recommended lists, followed by IB and IBCF, with UBCF providing the poorest novelty (approximately 34%). In terms of diversity, the index-based approaches performed significantly better, with IB+ providing 87% diversity for top-10 recommended lists, compared to 77% for the best CF approach (IBCF).

It is interesting to note that in the above results, the neighbourhood size (k) for both CF approaches was tuned based on delivering the best accuracy performance. An interesting question that arises is how well the CF approaches would perform if neighbourhood sizes were tuned according to other performance criteria (e.g. novelty or diversity) — would the recommendation accuracy provided by CF still outperform the index-based approaches? To answer this question, we repeated the above analysis using different neighbourhood sizes for CF. In particular we computed accuracy versus novelty and diversity for the two CF approaches and compared them with the IB+ approach.

Results are presented in Figure 4 and show that by reducing the neighbourhood size, CF can achieve better diversity and novelty than when using a bigger neighbourhood size. However, this comes at the cost of reduced accuracy performance. In fact, for neighbourhood sizes of k = 10, the novelty and diversity performance of the CF approaches is closest to that achieved by the IB+ approach but at the cost of poorer (in the case of UBCF) or almost equivalent (in the case of IBCF) accuracy compared to IB+.

For comparison purposes, we also show novelty and diversity versus coverage in Figure 5. It can be seen that, for the CF approaches, a higher coverage is achieved when using larger neighbourhood sizes, although neither CF approach can beat the coverage achieved by the IB+ approach. For the CF approaches, the results in Figure 5 are also interesting in that they show a clear tradeoff exists between optimising for coverage performance on the one hand (larger neighbourhood sizes), or optimising for novelty and diversity performance on the other hand (smaller neighbourhood sizes).
Further, examining the results in Figures 4 and 5 together, although similar accuracy to IB+ was achieved by the CF approaches when using a small neighbourhood (k = 10), we can also see that the coverage achieved at k = 10 is only 9% and 11% for UBCF and IBCF, respectively. This scenario is obviously never desirable since 90% of the system’s items cannot be recommended, showing that if we want to maintain coverage while having high levels of accuracy, novelty and diversity, then the CF approaches are preferable to both CF approaches in this evaluation setting.

Hence we can conclude that the index-based approaches compare quite favourably to the collaborative filtering techniques, when the range of performance metrics evaluated in this work are taken into consideration. This is a noteworthy result, and underlines the potential of the UGC-based recommenders as described in this paper.

5. DISCUSSION AND FUTURE WORK

While past work has focused on improving recommendation accuracy, recent work has shown that other metrics need to be explored in order to improve the user experience. In fact, algorithms which recommend novel items are likely to be more useful to end-users than those recommending more popular items. Such algorithms may, for example, introduce users to an entirely new space of items, regardless of the rating that they might actually give to said novel items. Further, a live evaluation performed by [25] showed that users preferred more diverse lists instead of more accurate (and less diverse) lists.

In this paper we consider a multi-criteria performance evaluation and, although trade-offs exist for all evaluated approaches, we believe the findings indicate that the UGC-based approach offers the best trade-off between all metrics and algorithms considered. For example, in order to achieve similar levels of novelty and diversity using the CF approaches as achieved by the UGC-based approach, a significant loss in coverage must apply.

We believe a reason for the higher novelty and diversity performance achieved by the UGC-based approach is that profiles created using this technique often reflect particular aspects and topics that users are interested in, allowing for more diverse (and often novel) recommendation lists compared to using ratings alone. For example, even if a user rated only science fiction movies, there will be specific aspects that differentiate this user from another who is also a fan of this genre (e.g. one may prefer aliens and strange creatures while the other might prefer the more romantic elements in the storyline).

There is an interesting range of future work to be carried in order to improve our current approach and to explore other benefits of UGC content:

- Enhancing the Real-Time Web In this paper we used UGC in the form of long-form movie reviews. However we are also interested in UGC in the form of Real-Time Web data (e.g. Twitter messages) which captures users’ preferences in real-time. For instance, people often post messages about the movies they liked, their favourite football team or their dream vacation experience. This data facilitates the building of rich user profiles which in turn allows recommenders to better address users’ needs. In fact, in past work [9], we demonstrated that micro-blogging messages can provide a useful recommendation signal despite their short-form and inconsistent use of language. Further, as discussed above, in our index-based approach user profiles can be independent from the product index,
allowing us to use a product index from one particular source (e.g. Flixster) with user profiles from another source (e.g. Twitter). This allows for cross-domain possibilities which will be explored in future work.

- **The cold-start problem.** In future work we will also explore how UGC can help in solving (or at least in mitigating) the well-known cold-start problem, which is related to the new user and new item problems. One of the advantages of using UGC as recommendation knowledge is that it facilitates a cross-domain approach for users who have not reviewed any products from a particular product index. If data relating to such users can be sourced from other domains (e.g. Twitter or Facebook feeds), then they can still benefit from recommendations. Further, a system which does not have reviews for particular products could provide recommendations for these products by building an index based on reviews from another system.

- **Integrating UGC in traditional recommenders.** Collaborative filtering algorithms have proven to be effective when there is a sufficient amount of ratings available, but their performance decreases when the number of ratings is limited. Our work shows preliminary evidence that UGC-based approaches have the potential to complement recommendation knowledge in the form of ratings and to improve the response of recommender systems to data sparsity. In future work we will study the performance of a hybrid recommender that benefits from the strengths of multiple data sources.

- **Improved index-based approach.** One of the limitations of our approach is that it is based on positive reviews. In future we will also consider negative reviews which may be useful to avoid recommending certain products to users. Another limitation is that positive reviews may have negative aspects (e.g. ‘I don’t generally like romantic comedies but I loved this one’) or negative reviews may have positive aspects (e.g. ‘Susan Sarandon was the only good thing in this movie’). To address this problem, we will extend our approach by using feature extraction techniques together with sentiment analysis [11,16] in order to create richer user profiles and product indexes. Choosing the optimum terms to represent users and items is also a problem to be solved in order to reduce the sparsity of the term-based profiles. Evaluating the effect of these improvements on various performance metrics will also be carried out in future work.

6. **CONCLUSIONS**

In this paper we have considered an alternative source of recommendation knowledge based on user-generated product reviews. Our findings indicate that recommenders utilising this source of knowledge can deliver comparable recommendation performance — across a range of criteria — compared to traditional CF-based techniques. In future work we would like to study other properties of UGC for recommendation, such as the ability to address the well known cold-start problem. Further, despite the simplicity of the current approach our results are promising and further improvements can be incorporated in order to increase recommendation accuracy, while trying to maintain high levels of coverage, novelty and diversity.
7. ACKNOWLEDGMENTS

Based on work supported by Science Foundation Ireland, Grant No. 07/CE/I1147.

8. REFERENCES


