CrossMark

Trust-based Modelling of Multi-criteria Crowdsourced Data

Fátima Leal[1,2] · Benedita Malheiro[2,3] · Horacio González-Vélez[4] ·
Juan Carlos Burguillo[1]

Abstract As a recommendation technique based on historical user information, collaborative filtering typically predicts the classification of items using a single criterion for a given user. However, many application domains can benefit from the analysis of multiple criteria, e.g. tourists usually rate attractions (hotels, attractions, restaurants, etc.) using multiple criteria. In this paper, we argue that the personalised combination of multi-criteria data together with the creation and application of trust models should not only refine the tourist profile, but also improve the quality of the collaborative recommendations. The main contributions of this work are: (1) a novel profiling approach which takes advantage of the multi-criteria crowdsourced data and builds pairwise trust models and (2) the k-NN prediction of user ratings using trust-based neighbour selection. Significant experimental work has been performed using crowdsourced datasets from the Expedia and TripAdvisor platforms.

✉ Fátima Leal
  fatimaleal2@gmail.com

  Benedita Malheiro
  mbm@isep.ipp.pt

  Horacio González-Vélez
  horacio@ncirl.ie

  Juan Carlos Burguillo
  J.C.Burguillo@uvigo.es

[1] EET/Uvigo – School of Telecommunications Engineering, University of Vigo, Vigo, Spain

[2] INESC TEC, Porto, Portugal

[3] ISEP/IPP – School of Engineering, Polytechnic Institute of Porto, Porto, Portugal

[4] CCC/NCI – Cloud Competency Centre, National College of Ireland, Dublin, Ireland

1 Introduction

Coupled with information and communications technologies, *tourism crowdsourcing* has significantly revolutionised tourist behaviour over the past decade. Mobile technologies provide tourists with permanent access to endless web services which influence their decisions using crowdsourced information. Such information is shared collaboratively by tourists in well-known tourism business-to-customer online platforms (e.g. TripAdvisor, Expedia and Airbnb). They enable a tourist to actively share mementos, comments, reviews and, most importantly, rate their overall travel experience. By gathering voluntarily shared feedback, these online platforms have essentially become *crowdsourcing platforms* [9].

The value of crowdsourced tourism information is crucial to businesses and clients alike. However, the voluntary information sharing and the openness of crowdsourcing systems raise reliability and integrity questions. Therefore, when using crowdsourced data, it is necessary to take trustworthiness into account in order to ensure the accuracy and validity of the final results. Trust mechanisms must arguably underpin crowdsourcing platforms in order to validate both the quality level of the crowdsourced information and indeed the users.

In this work, we have modelled tourists and tourism attractions ('resources') employing multi-criteria tourism information from crowdsourcing platforms coupled with trust mechanisms in order to produce personalised recommendations.

🖄 Springer

Personalised recommendations are often based on the prediction of user classifications. Typically, the crowdsourced classification of hotels involves multi-criteria ratings, e.g. hotels are classified in the Expedia or TripAdvisor platforms in terms of *cleanliness*, *hotel condition*, *service and staff*, *room comfort* or *overall* opinion. The personalised combination of multi-criteria crowdsourced ratings together with trust modelling arguably improves the tourist profile and, consequently, the accuracy of the collaborative predictions.

Collaborative filtering is a classification-based technique, i.e. depends on the classification each user gave to the items he/she was exposed to [5]. Typically, this classification corresponds to a unique rating. Whenever the crowdsourced data hold multiple ratings per user and item, first, it is necessary to decide which user classification to use in order to apply collaborative filtering. This work explores both profiling approaches: single criterion (SC)—choosing the most representative of the crowdsourced user ratings [22, 23]—and multi-criteria (MC)—combining the different crowdsourced user ratings per item, using the non-null rating average (NNRA) or the personalised weighted rating average (PWRA), i.e. based on the individual user rating profile.

This work proposes a new approach to provide online tourism recommendations using collaborative filtering via k-nearest neighbours (k-NN) algorithm. Additionally, we apply Pearson correlation to determine the correlation among users, and, then, build a decentralised trust model depending on the selected recommendations regarding the current data stream event. Therefore, this research contributes to guest and hotel profiling—based on multi-criteria ratings incorporating trust modelling—and to the prediction of hotel guest ratings—based on the k-NN algorithm via data streams—ultimately providing reliable online recommendations. Our experiments with crowdsourced Expedia and TripAdvisor datasets show that the proposed profiling approach significantly improves the k-NN prediction accuracy of unknown hotel ratings. The results also show the relevance of multi-criteria and trustworthiness in prediction and recommendation accuracy.

This paper is organised as follows. Section 2 reviews previous approaches to personalisation via crowdsourced ratings and presents a critical comparison between our approach and the surveyed works. Section 3 describes our methodological approach and algorithms used. Section 4 describes our implementation and the processing details. The experiments and tests on the different datasets are reported in Sect. 5. Finally, Sect. 6 summarises and discusses the outcomes of this work.

2 Related Work

Technology plays an important role in the hotel and tourism industry. Both tourists and businesses benefit from technology advances regarding communication, reservation and guest feedback services. Individuals create a digital footprint while using web services to organise trips, i.e. to search, book and share their opinions in the form of ratings, textual reviews, photographs, *etc.*. This pervasive interaction between individuals, web services and mobile applications continually generates large volumes of useful data. Based on individual digital footprints, tourist profiles are used by recommender systems to personalise suggestions. Refined tourist profiles increase the quality of the suggestions and, ultimately, the tourist experience.

Collaborative filtering is a popular recommendation technique in the tourism domain. It often relies on rating information voluntarily provided by tourists, i.e. crowdsourced ratings, to recommend unknown resources to other tourists. Well-known tourism crowdsourcing platforms, e.g. TripAdvisor or Expedia, allow users to classify tourism resources using multi-criteria, e.g. *overall*, *service* and *cleanliness*.

Moreover, crowdsourced information influences the tourist decision making process. Therefore, trust modelling of the users of crowdsourcing platforms helps to provide more helpful and accurate recommendations. According to Josang et al. [19], trust is based on direct experiences between stakeholders. The trustworthiness is established over time, i.e. interaction by interaction. It involves an online scenario, i.e. incremental updating, where the model is built and updated whenever a new event occurs. The streaming approach is used to learn models and predict the behaviours in near real time, e.g. to learn the user behaviour and provide online recommendations. Therefore, we perform a data streaming based on Amatriain [3], Gama [13] and Sayed [30] research. Their works address the problems of modelling, prediction, classification, data understanding and processing in unpredictable environments by exploiting data stream processing techniques.

Online profiling and prediction together with trust-based modelling of tourism crowdsourced multi-criteria ratings are a relevant research topic for the actual tourism industry due to the impact of crowdsourced information in tourist behaviour. This related work contemplates: (1) multi-criteria tourism crowdsourced ratings in hotel recommendation systems; (2) collaborative filtering; and (3) trust-based modelling. Adomavicius and Kwon [2], Bilge and Kaleli [4], Lee and Teng [24], Jhalani et al. [17], Liu et al. [26], Manouselis and Costopoulou [27] and Shambour et al. [32] have explored the integration of multi-criteria ratings in the user profile, mainly using multimedia datasets to validate

their proposals. Davoudi et al. [7], Jia et al. [18] and Zhang et al. [37] have explored the trust modelling for rating prediction presenting trust models together with matrix factorisation algorithms or similarity metrics. However, scant research considers trust-based modelling of multi-criteria crowdsourced ratings for profiling and rating prediction applied to the tourism domain in order to obtain more accurate tourism recommendations.

Jannach et al. [16] apply the Adomavicius and Kwon [1] methods to incorporate multi-criteria ratings in the tourist profile based on support vector regression (SVR). It combines a user and item models, using a weighted approach, to provide better recommendations. The evaluation was performed with a dataset provided by HRS.com.

Fuchs and Zanker [12] perform multi-criteria rating analysis based on a TripAdvisor dataset. First, they use multiple linear regression (MLR) to identify correlations, patterns and trends among the TripAdvisor dataset parameters. Then, the authors apply the Penalty-Reward-Contrast analysis proposed by Randall Brant [29] to establish tourist satisfaction levels based on multi-criteria ratings. This work proposes a methodology for MC rating analysis.

Nilashi et al. [28] propose a SC profiling approach together with a hybrid hotel recommendation model for multi-criteria recommendation. They employed: (1) principal component analysis (PCA) for the selection of the most representative rating (dimensionality reduction); (2) expectation maximisation (EM) and adaptive neuro-fuzzy inference system (ANFIS) as prediction techniques; and (3) TripAdvisor data for evaluation.

Farokhi et al. [11] explore SC profiling together with collaborative filtering. First, the authors selected the *overall* as the most representative rating after determining the correlation between the multiple ratings, then applied data clustering (fuzzy *c*-means and *k*-means) to find the nearest neighbours and, finally, predicted the unknown hotel ratings using the Pearson correlation coefficient. The evaluation was performed with TripAdvisor data.

Finally, Ebadi and Krzyzak [8] developed an intelligent hybrid multi-criteria hotel recommender system. The system uses both textual reviews and ratings from TripAdvisor. Regarding the ratings, it adopts SC profiling to learn the guest preferences and singular value decomposition (SVD) matrix factorisation to predict unknown ratings.

2.1 Contributions

This paper explores a trust-based profiling and prediction methodology using crowdsourced multi-criteria ratings in the tourism domain. The main goal is to refine guest and hotel profiling by reusing the multiple hotel ratings each guest shares over time, using data streams and computing the trustworthiness. According to Nilashi et al. [28] and

Table 1 Comparison of tourism multi-criteria research approaches

Approach	Evaluation	Profiling	Prediction	Trust
[16]	HRS	MC	SVR	
[12]	TA	MC	–	
[28]	TA	SC	ANFIS	
[11]	TA	SC	*k*-means	
[8]	TA	SC	SVD	
Our proposal	TA & E	SC & MC	*k*-NN	✔

Adomavicius and Kwon [2], collaborative filtering with multi-criteria item ratings has been unexplored when compared with its single criterion item rating counterpart.

When compared with other research found in the literature regarding trust-based modelling of tourism multi-criteria crowdsourced data, our work: (1) contributes with single and multiple rating profiling; (2) employs *k*-NN as predictive technique together with trust modelling; and (3) uses Expedia (E) and TripAdvisor (TA) crowdsourced data for evaluation. Table 1 depicts a comparison of the surveyed tourism multi-criteria tourism approaches. We can verify that the trust modelling in tourism ratings predictions has not been explored yet. Therefore, this paper contributes mainly to improve the accuracy of predictions and enhance tourism recommendations.

3 Method

The proposed method includes: (1) profiling; (2) online rating prediction; (3) trust modelling; and (4) evaluation metrics. The profiling explores the multi-criteria ratings in order to obtain the most refined profile. Therefore, we analyse the most representative rating (MRR) using the Leal et al. approach [22] which relies on multiple linear regression. Additionally, we combine the multi-criteria ratings using a non-null rating average (NNRA) and personalised weighted rating average (PWRA). For rating prediction, we utilise *k*-NN algorithm with data streams to obtain online user-based recommendations. The trust model is based on Pearson correlation and items selected items by users. Our method uses data streams, i.e. the model is updated as soon as the user introduces a new rating, providing, thus, online recommendations. Finally, we experiment and assess our method with Expedia and TripAdvisor data using root mean squared error (RMSE), Target Recall (TRecall) and Recall as evaluation metrics.

3.1 Profiling

The profiling module addresses the activity of user modelling using the crowdsourced multi-criteria ratings. First,

we apply a multiple linear regression (MLR) to identify the most representative rating (MRR). Then, we combine the crowdsourced multi-criteria user ratings into a single rating using NNRA and PWRA.

Most Representative Rating (MRR) is the most meaningful of the multiple ratings available. We perform a multiple linear regression (MLR) to identify the MRR from the multi-criteria crowdsourced ratings as proposed by Leal et al. [22]. The MLR is typically applied to multivariate scenarios in order to predict one or more continuous variables based on other dataset attributes, i.e. by identifying existing dependencies among variables [33]. First, we do a correlation analysis to identify the relation between the different crowdsourced ratings and, then, perform the MLR to validate the correlation results and analyse the dependency of the MRR.

Non-Null Rating Average (NNRA) models the user using a standard average of the positive multi-criteria ratings. This profiling approach is defined by Eq. 1 ($r_{u,i}$) where $r_{u,i,c}$ is the non-null rating of criterion c given by user u to the item i and n is the number of non-null multi-criteria ratings given by user u to item i. The multiple criteria ratings c of tourism crowdsourcing platforms are *cleanliness, room service, overall, etc.*

$$r_{u,i} = \frac{\sum_{c=1}^{n} r_{u,i,c}}{n} \qquad (1)$$

Personalised Weighted Rating Average (PWRA) explores a personalised combination of the multi-criteria ratings in order to lead to a profile refinement. Platforms contain multi-criteria ratings, e.g. *cleanliness, hotel condition, service and staff*. The PWRA combines these multi-criteria ratings. Equation 2 displays the personalised weighted rating average—$r_{u,i}$—where $r_{u,i,c}$ is the non-null criterion c given by user u to item i, n_c presents the number of times user u has rated the criterion c, $n_{u,c}$ the number of non-null ratings of criterion c given by user u and n is the total number of non-null multi-criteria ratings given by user u.

$$r_{u,i} = \frac{\sum_{c=1}^{n} n_c r_{u,i,c}}{\sum_{c=1}^{n} n_{u,c}} \qquad (2)$$

While MRR uses just one rating for profiling, the PWRA and NNRA combine the multiple types of ratings into a single rating. Both profiling approaches—PWRA and NNRA—introduce the multi-criteria concept in recommendations. According to our data, the users are guests and the items are hotels.

3.2 Online Rating Prediction

The rating prediction module addresses the prediction of ratings regarding hotels not yet rated by the active user. This was implemented using a user-based collaborative recommendation filter.

We employ k-NN using data streams, i.e. the method predicts and updates the model in near real time whenever a new rating event occurs, providing online recommendations. The correlation among users is calculated via Pearson correlation which identifies the nearest neighbours. The set of k-nearest neighbours holds the k users with higher Pearson correlation with the active user. We use Pearson correlation not only to implement the k-NN, but also for trust-based modelling of multi-criteria crowdsourced data. The trust engine computes the trustworthiness analysing both (1) the number of times which a user k was identified as a neighbour of a user u and (2) the number of the items selected by u due to neighbour k. Therefore, the final rating prediction is based on k-NN algorithm together with a trust modelling built over time according to the user's selections.

3.2.1 Trust Modelling

The explosive growth of tourism crowdsourcing platforms has promoted the indirect interaction among tourists. The trust factor plays an important role in these interactions as well as in building higher-quality relationships among users. Therefore, we employ a trust modelling which involves: (1) Pearson correlation and (2) user interaction evaluation. The trust values among users are updated upon each user event. On the one hand, the Pearson correlation determines: (1) the nearest neighbours to be used for trust modelling and (2) the correlation values to be used as predictions. On the other hand, the trustworthiness takes into account the number of times the current user selects one of the top 10 recommendations provided by his/her neighbours.

Pearson Correlation Coefficient provides a measure of linear correlation between two users [31]. The similarity $PC_{u,k}$ computes the degree of linearity among the ratings of user u and user k. Equation 3 expresses the Pearson correlation between a user u and a neighbour k ($PC_{u,k}$), where $r_{u,i}$ is the rating given by user u to the item i, $\bar{r}_u$ and $\bar{r}_k$ are the average of the co-rated items given by user u and k, respectively, $r_{k,i}$ is the rating given by neighbour k to the item i and m is the total number of items [10].

$$PC_{u,k} = \frac{\sum_{i=1}^{m}[(r_{u,i} - \bar{r}_u)(r_{k,i} - \bar{r}_k)]}{\sqrt{\sum_{i=1}^{m}(r_{u,i} - \bar{r}_u)^2 \sum_{i=1}^{m}(r_{k,i} - \bar{r}_k)^2}} \qquad (3)$$

Different similarity metrics have been proposed [15, 25] to determine the similarity between two entities. However, according to Lathia et al. [20], these different metrics share the same features, i.e. they rely on a non-empty intersection between two user's profiles in order to find a similarity point using co-rated items. In this context, we propose a model which not only selects the neighbours using the Pearson correlation, but determines the influence of each neighbour in the final recommendations by computing their trustworthiness. This trustworthiness will be applied upon final predictions in order to improve the recommendations accuracy.

Trustworthiness quantifies the closeness between a user and his/her neighbours in terms of the resulting recommendations. Once the nearest neighbours are determined, it is important to analyse the user behaviour over time. If a given user u selects many recommendations due to a neighbour k, the trustworthiness between u in k increases. This approach relies on data streams, i.e. an ordered sequence of events allowing to add new ratings and update existing models, providing new and more accurate recommendations. In this context, Eq. 4 displays the trust $T_{u,k}$ between user u and k where $n_{u,k}$ represents the number of items actually recommended to u due to k and $N_{u,k}$ the number of times k was chosen as a neighbour of u.

$$T_{u,k} = \frac{n_{u,k}}{N_{u,k}} \qquad (4)$$

In fact, $n_{u,k}$ corresponds to the number of top 10 items which were previously rated by k and recommended while k was a neighbour of u. The idea is to emphasise the neighbours who can provide trustworthy recommendations and downgrade neighbours who provide uninteresting recommendations.

3.2.2 k-Nearest Neighbours

The k-NN is widely used in collaborative filtering. In the user-based version, the k-NN determines the k user neighbours (neighbourhood) to generate recommendations for a current user u. This memory-based approach uses similarity, correlation or distance metrics to extract the predictions and select the best recommendations.

This user-based collaborative filtering uses Pearson correlation to compute the neighbourhood (k) of the user u. Once the k has been computed, our method combines the ratings of the users to generate predictions. This is calculated by the weighted average of the neighbouring users' ratings regarding an item i using $PC_{u,k}$ as weights. To improve the accuracy of the final predictions, we employ the trust $T_{u,k}$ of the n neighbours. Equation 5 displays the prediction $\hat{r}_{u,i}$ of item i for a user u.

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{k=1}^{n}[(r_{k,i} - \bar{r}_k) * PC_{u,k}]}{\sum_{k=1}^{n}PC_{u,k}} * \frac{\sum_{k=1}^{n} T_{u,k}}{n} \qquad (5)$$

3.3 Evaluation Metrics

The evaluation of recommendation systems involves predictive accuracy and classification metrics.

On the one hand, the predictive accuracy metrics measure the error between the predicted rating and the real user rating. It is the case of RMSE which quantifies the prediction error. Using this evaluation metric, we can evaluate the systems in terms of predictive accuracy. Equation 6 represents the RMSE where $\hat{r}_{u,i}$ represents the rating predicted for user u and item i, $r_{u,i}$ the rating given by user u to item i, m the total number of users and n the total number of items. We calculate the global prediction RMSE adopted by Takács et al. [34], which is calculated incrementally after each incoming rating event.

$$RMSE = \frac{1}{u} \times \sum_{u=1}^{m} \left(\sqrt{\frac{1}{n} \times \sum_{i=1}^{n}(\hat{r}_{u,i} - r_{u,i})^2} \right) \qquad (6)$$

On the other hand, the classification metrics evaluate the recommendations accuracy. In this context, we determine for each new rating event the Recall of the top 10 recommendations proposed by Cremonesi et al. [6] and the Target Recall of the top 10 recommendations presented by Veloso et al. [35]. Cremonesi et al. [6] metric includes: (1) the prediction of the ratings of all items unseen by the user, including the newly rated item; (2) the selection of 1000 unrated items plus the newly rated item; and (3) the sorting in descending order of the predictions. If the newly rated item belongs to the list of the top N user predicted items, it is considered as a hit. The latter case, Veloso et al. [35] use all rated items instead of just the top-rated items. The Target Recall@N (TRecall@N) evaluates the recommendations accuracy using all user ratings. This metric verifies whether the recommendation is close to the target rating, i.e. within a radius of $\frac{N}{2}$ of the user actual rating. Therefore, to evaluate our online method we calculate RMSE, the Recall@N and the TRecall@N, making $N = 10$.

4 Implementation

Our recommendation engine, which is implemented in Java, runs on an OpenStack cloud instance with 16 GiB RAM, 8 CPU and 16 GiB of hard disk space In terms of architecture, our collaborative filter, which is depicted in

Fig. 1 Recommendation engine
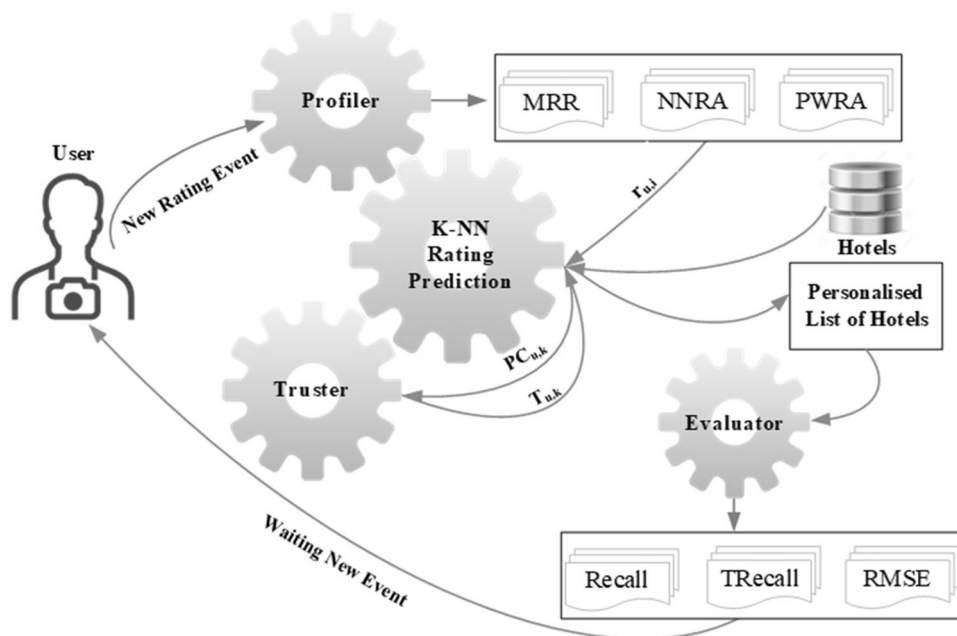
Fig. 1, includes four modules: (1) profiler; (2) k-NN rating predictor; (3) truster; and (4) evaluator.

First, we analyse multi-criteria crowdsourced rating in order to obtain the most refined profiling considering all available ratings. For that, we present three distinct multi-criteria profiling approaches described in Sect. 3: (1) MRR; (2) NNRA; and (3) PWRA. Then, the k-NN rating predictor applies Pearson correlation to find the nearest neighbours. The nearest neighbours are used by the truster module to compute the trustworthiness between users over time. Next, the final rating prediction is calculated using the k-NN algorithm together with trust modelling. In this step, the system orders and recommends, for each user, a list of hotels by descending order. Finally, the system is evaluated. The evaluation protocol includes data ordering, partitions and distribution. To simulate an online scenario, the data were ordered temporally (i.e. data streams) and, then, partitioned. The initial model uses the 20 % of the dataset. The online model uses the 'Stream Data', which correspond to the remaining 80 % of the dataset. When a user rates a hotel, the algorithm uses the new rating to update the predictions for that user as well as for re-evaluating the method. The adopted evaluation method was developed according to Gama et al. [14] proposal. The proposed method is evaluated in terms of RMSE, Recall@10 and TRecall@10 metrics. Once the evaluation process ends, the system is ready to update a new event.

5 Experiments and Results

We conducted several experiments with the HotelExpedia dataset http://ave.dee.isep.ipp.pt/~1080560/ExpediaData Set.7z and the TripAdvisor dataset [36] to evaluate the proposed method.

The experiments involved MRR, NNRA and PWRA profiling with and without trust modelling, as well as the correspondent rating prediction evaluation. The following subsections describe the datasets used and the results obtained.

5.1 HotelExpedia Dataset

Expedia http://www.expedia.com is a powerful platform which contains large volumes of crowdsourced hotel opinions. Moreover, Expedia owns a host of online brands, including TripAdvisor, Hotels.com or trivago. According to Law and Chen [21], Expedia brands cover researching, booking, experiencing and sharing travels. The platform allows choosing flights or hotels, reading personal reviews of hotels, classifying hotels using textual reviews and ratings as well as planning new travels.

Taking into account these characteristics, we have collected different crowdsourced ratings via the Expedia API https://hackathon.expedia.com In the Expedia platform, tourists classify hotels using multi-criteria ratings: *overall*, *cleanliness*, *hotel condition*, *service and staff* and *room comfort*. Based on these multiple criteria classifications, we

Table 2 Expedia hotel and customer reviews data

File	Features
Hotels	**hotelId**, **description**, latitude-longitude, **starRating**, guestReviewCount, **price**, amenity, **overall**, recommendedPercent, cleanliness, serviceAndstaff, roomComfort and hotelCondition
Users and reviews	**nickname**, userLocation, **hotelId**, **overall**, **cleanliness**, **hotelCondition**, **serviceAndStaff**, **roomComfort**, **reviewText** and **timestamp**

Table 3 TripAdvisor dataset

| File	Features
Hotels	name, hotelURL, price, **hotelID**, imgURL
Users and reviews	authorLocation, title, **author**, reviewID, reviewText, date, **overall**, **value**, **rooms**, **location**, **cleanliness**, **service**, **sleepQuality**

Fig. 2 HotelExpedia dataset
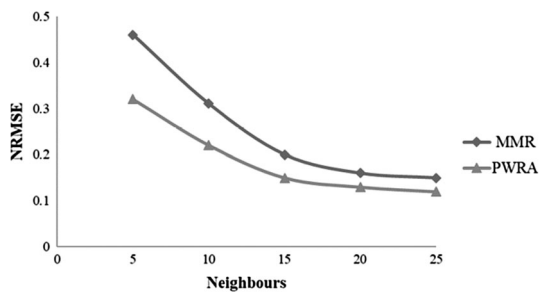
Fig. 3 TripAdvisor dataset

create, using different approaches, unique personalised ratings per tourist and hotel.

Table 2 describes the contents of our dataset. It is composed of 6276 hotels, 1090 identified users and 214 342 reviewers from 11 different locations. Each user classified at least 20 hotels, and each hotel has a minimum of 10 ratings. Our experiments, which rely on the hotel, user and hotel user review data, use, specifically, the user nickname, the hotel identification and, as multi-criteria ratings, the *overall*, *cleanliness*, *service and staff*, *hotel condition* and *room comfort*. This dataset does not contain null ratings, i.e. all users rated the hotels according to the multiple criteria.

5.2 TripAdvisor Dataset

TripAdvisor is a travel website which provides crowd-sourced reviews of travel-related content. Therefore, TripAdvisor data present an ideal scenario to apply our proposed method. Wang et al. [36] provides a TripAdvisor dataset composed by 9114 hotels, 7452 users and 127 517 hotel reviews. Table 3 describes the contents of the dataset. Our experiments reuse the user and hotel identification and, as multi-criteria ratings, the *overall*, *value*, *rooms*, *location*,
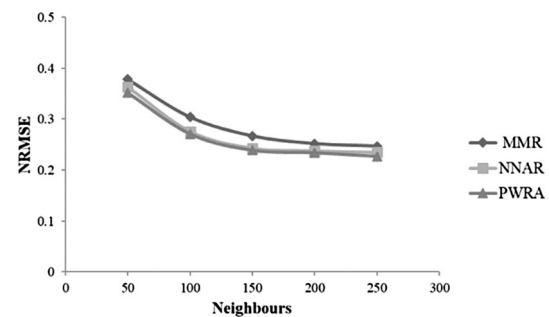
cleanliness, *service* and *sleep quality*. This dataset contains 14% of null ratings.
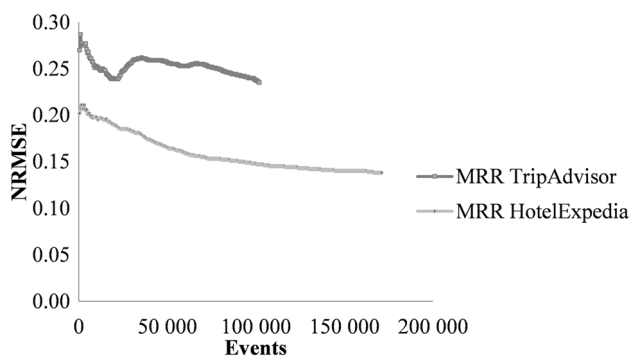
5.3 Neighbours Analysis

The number of neighbours (k) influences the performance of our proposed method. Therefore, we test the method using different k in order to obtain the best results. The experiments involve two distinct datasets. The results present different behaviours due to the distinct composition of each dataset.

On the one hand, Fig. 2 plots the system behaviour in terms of prediction accuracy facing different number of neighbours using HotelExpedia dataset. The NRMSE decreases monotonically and converges over time for 20 neighbours in the different profiling approaches presented. In this scenario, we selected $k = 20$ for the different tests performed and analysed for Expedia data.

On the other hand, Fig. 3 plots the system behaviour in terms of prediction accuracy facing different number of neighbours from the TripAdvisor dataset. The NRMSE decreases monotonically and converges over time for 200

Table 4 MLR results for the overall rating

| Dataset	Independent features	β_i	R^2
Hotel Expedia	Service and Staff	0.32	0.80
	Hotel condition	0.30	
	Room comfort	0.29	
	Cleanliness	0.11	
Trip advisor	Value	0.23	0.78
	Service	0.22	
	Rooms	0.18	
	Cleanliness	0.14	
	Location	0.12	
	Sleep quality	0.10	

Fig. 4 NRMSE of the predictions with MRR profiling

neighbours. In this scenario, we selected $k = 200$ for the different tests performed and analysed for TripAdvisor data.

5.4 Profiling Analysis

First, we analysed the available multi-criteria guest ratings per hotel and, then, applied the proposed method to predict the unknown hotel ratings. The rating analysis comprised two different approaches: (1) the identification of the most representative hotel rating and (2) the combination of the multi-criteria guest ratings per hotel into a unique guest rating per hotel.

MRR estimates and quantifies the relationship between the overall rating (dependent variable) and the remaining ratings (independent variables) using multiple linear regression for both HotelExpedia and TripAdvisor datasets. Table 4 displays the OLS MLR results where β_i are the regression coefficients and R^2 quantifies the response variable variation that is explained by the model. In the case of HotelExpedia, the results show that the independent variables (*cleanliness*, *hotel condition*, *room comfort* and *service and staff*) are capable of explaining approximately
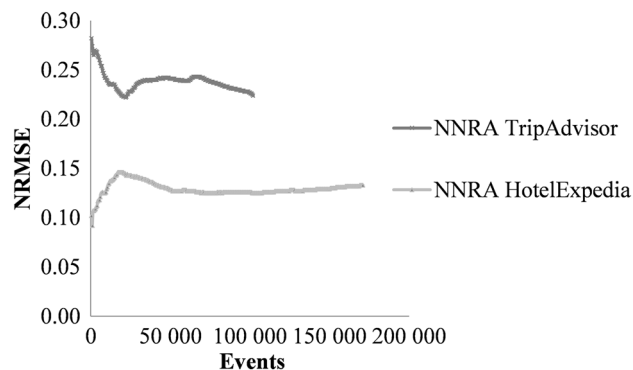
Fig. 5 NNRA profiling

80% of the dependent variable. The regression was performed with 214343 multi-criteria ratings. In the case of TripAdvisor, [22] report that the independent variables (cleanliness, location, rooms, service, sleep quality and value) are capable of explaining approximately 78% of the dependent variable (*overall*).

Based on these results, we chose the *overall* rating as the most representative rating (MRR) of both HotelExpedia and TripAdvisor and, then, performed the *overall* rating prediction using k-NN algorithm. Figure 4 plots the normalised RMSE (NRMSE) of the predictions for both datasets. In both cases, the NRMSE decreases monotonically and converges over time to approximately 0.138 and 0.235 using HotelExpedia and TripAdvisor data, respectively.

NNRA is applied according to Eq. 1. It is the first approach of rating analysis which combines the multi-criteria guest ratings per hotel into a single guest rating per hotel. For example, if a user u rates a hotel i with 5.0 for *cleanliness*, 3.0 for *serviceAndstaff*, 4.0 for the *overall* rating and provides no rate regarding *hotelCondition* and *roomComfort*, then, the combined NNRA hotel guest rating $r_{u,i}$ is 4.0, i.e. $\frac{5.0+3.0+4.0}{3}$. Figure 5 plots the NRMSE of the predictions for HotelExpedia and TripAdvisor datasets using NNRA profiling approach. In both cases, the NRMSE decreases monotonically and converges over time to 0.133 and 0.224 using Expedia and TripAdvisor data, respectively.

PWRA was applied as an alternative combination approach according to Eq. 2. For example, if a user u rates a new hotel i with 4.0 for *cleanliness*, 3.0 for *roomComfort*, 3.5 for *overall* and has a past rating history of 20 *cleanliness* ratings, 20 *roomComfort* ratings and 60 *overall* ratings, the combined PWRA hotel guest rating $r_{u,i}$ is 3.6, i.e. $\frac{21}{101}4.0 + \frac{21}{101}3.0 + \frac{61}{101}3.5$. Figure 6 plots the NRMSE of HotelExpedia and TripAdvisor data predictions based on the PWRA rating. The NRMSE decreases monotonically
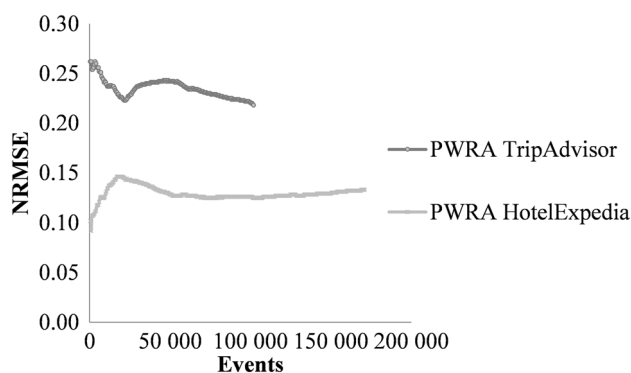
Fig. 6 PWRA profiling

and converges over time to 0.133 and 0.218 for Expedia and TripAdvisor, respectively.

Comparison The previous profiling approaches explore crowdsourced multi-criteria rating profiling together with collaborative filtering to provide hotel recommendations. The predictions were performed with k-NN algorithm executed using data streams. The MMR profiling corresponds to the usage of the standard *overall* rating. The NNRA and PWRA results with the HotelExpedia dataset, which has no null ratings, are naturally equal, whereas, with the TripAdvisor dataset, which includes 14% of null ratings, they are not only distinct, but favourable to PWRA. In terms of the accuracy of the rating predictions, these results show that: (1) NNRA and PWRA are preferable to MRR profiling and (2) PWRA, when faced with null multi-criteria user ratings, outperforms both MMR and NNRA profiling in both datasets.

5.5 Trusted Recommendations

The trustworthiness computation aims to refine the final recommendations. Therefore, the system models the users

using Eq. 3 to identify the nearest neighbours and Eq. 4 to compute the trustworthiness over time among users. For example, if k has been selected as a neighbour of u 9 times and 8 of these times the list of top 10 recommendations included recommendations based on neighbour k, according Eq. 4, the trustworthiness attributed by u to k is 0.89, i.e. $\frac{8}{9}$.

The final prediction with trust is calculated according to Eq. 5, i.e. the standard k-NN prediction of the user ratings together with the trustworthiness values of all selected neighbours. Therefore, we perform a trust-based neighbour selection. In short, with this data stream approach we recommend hotels to potential guests with the support of k-NN predictions and trust modelling. Our profiling approach reuses the complete collection of multi-criteria hotel ratings available.

The effectiveness of this recommendation engine was measured using the, NRMSE, Recall and TRecall considering the top 10 hotel recommendations per user. Table 5 compares the global predictive (NRMSE) and classification (Recall and TRecall) accuracy with the MRR, NNRA and the PWRA profiling approaches. Lower error values and higher classification values indicate higher prediction accuracy. The MMR profiling, which corresponds to the usage of the standard *overall* rating, is the base profiling approach.

In the case of HotelExpedia, the NNRA and PWRA profiling, when compared with the MMR approach, improve the NRMSE 19%, the Recall 60% and the TRecall 6%. When trust modelling is applied, it additionally improves the NRMSE 4%, the Recall 110% and the TRecall 9%.

In the TripAdvisor case, the results of PWRA profiling, when compared with those of the MMR approach, improve the NRMSE 40%, Recall 69% and the TRecall 35%. When trust modelling is applied, it further improves the NRMSE 13%, the Recall 68% and the TRecall 21%.

Table 5 Comparison of prediction metrics results

	Profiling	Neighbours	Recall@10	TRecall@10	RMSE
Hotel Expedia	MRR	20	0.190	0.186	0.159
	NNRA	20	0.304	0.198	0.129
	PWRA	20	0.304	0.198	0.129
	MRR + Trust	20	0.637	0.191	0.187
	NNRA + Trust	20	0.639	0.215	0.124
	PWRA + Trust	20	**0.639**	**0.215**	**0.124**
Trip advisor	MRR	200	0.289	0.379	0.252
	NNRA	200	0.269	0.344	0.239
	PWRA	200	0.489	0.513	0.152
	MRR + Trust	200	0.703	0.414	0.261
	NNRA +Trust	200	0.714	0.382	0.192
	PWRA +Trust	200	**0.823**	**0.621**	**0.132**

Best results are highlighted in bold

In terms of the accuracy of the rating predictions, these results show that: (1) NNRA and PWRA are preferable to MRR profiling; (2) PWRA, when faced with null multi-criteria user ratings, outperforms both MMR and NNRA profiling; (3) trust modelling influences positively the final recommendations and, together with PWRA, is the best approach.

6 Conclusions

Tourism crowdsourcing platforms, e.g. Expedia and TripAdvisor, collect large volumes of feedback data regarding tourism resources, including multi-criteria ratings, textual reviews and photographs. The crowdsourced tourist profile corresponds to this individual digital footprint. This information, introduced voluntarily by tourists, has a direct influence in the final tourist decision making. Therefore, it is crucial exploring not only the available crowdsourced information, but also the trustworthiness established among users over time for creating accurate user-based recommendations. Facing this scenario, the tourism crowdsourced data analysis became a relevant research topic in tourism domain mainly to discover the unknown tourists' needs using the crowd impact. Regarding this research challenge, this paper explores trust-based modelling of crowdsourced multi-criteria ratings to create user-based online recommendations using collaborative filtering. Therefore, we propose a method that considers: (1) multi-criteria profiling; (2) online rating prediction that relies on trust modelling and the k-NN algorithm; and (3) a evaluation, including the metrics, which assess the proposed method.

The profiling module aims to use all available ratings to obtain refined profiles to use in collaborative filtering. In order to apply standard collaborative filtering, it is necessary to provide only a single classification per user and item to the filter. To be able to use multi-criteria ratings for profiling, we designed and experimented with two main approaches: (1) the identification of the most representative rating (MRR) with MLR and (2) the combination of the multi-criteria ratings into a single rating, per user and item, using NNRA and PWRA.

The online rating prediction module processes the stream of ratings to predict unknown hotel ratings and, thus, provide online recommendations. We implemented a user-based collaborative filter employing the k-NN algorithm, using the Pearson correlation to determine neighbours. Additionally, we computed the trustworthiness between the current user and each neighbour analysing both the number of items actually recommended to the user

based on the neighbour and the number of times that user was chosen as a neighbour of the current user. The final predictions were obtained using the k-NN algorithm together with the trustworthiness among users.

The proposed methodology was tested and evaluated with Expedia and TripAdvisor crowdsourced multi-criteria hotel ratings using RMSE, Recall@10 and TRecall@10 as evaluation metrics. First, we analysed the number of neighbours, i.e. the k to be used in the experiments. This method presented different results for the HotelExpedia and TripAdvisor datasets. While the RMSE converged in $k = 20$ with HotelExpedia data, in the case of the TripAdvisor data, the RMSE stabilised at $k = 200$. Then, in terms of profiling, we tested the three proposed approaches, i.e. MRR, NNRA and PWRA. The results showed that the highest k-NN prediction accuracy occurs with the PWRA multi-criteria profiling. Finally, we model trust based on Pearson correlation and k-NN predictions. The PWRA profiling together with trust modelling presented the best results concerning prediction and recommendation accuracy. In the case of the HotelExpedia dataset, PWRA and trust modelling, when compared with PWRA, improve the RMSE 4%, the Recall 110% and the TRecall 9%. In the case of the TripAdvisor dataset, PWRA and trust modelling, when compared with PWRA, improve the RMSE 13%, the Recall 68% and the TRecall 21%. Therefore, we can conclude that trust has a relevant impact in user-based recommendations using multi-criteria crowdsourced data.

To sum up, this research work proposes a new profiling approach based on crowdsourced multi-criteria ratings data streams together with trust modelling that improves the k-NN hotel rating prediction accuracy. As future work, we intend to: (1) explore multi-criteria recommendation using both textual reviews and multi-criteria ratings and (2) explore the users' reputation factor to model incoming ratings.

References

1. Adomavicius G, Kwon Y (2007) New recommendation techniques for multicriteria rating systems. IEEE Intell Syst 22(3):48–55

2. Adomavicius G, Kwon Y (2015) Multi-criteria recommender systems. In: Recommender systems handbook, chapter 24. Springer, Berlin, pp 847–880

3. Amatriain X (2013) Mining large streams of user data for personalized recommendations. ACM SIGKDD Explor Newsl 14(2):37–48

4. Bilge A, Kaleli C (2014) A multi-criteria item-based collaborative filtering framework. In: JCSSE 2014, Pattaya, May. IEEE, pp 18–22

5. Breese JS, Heckerman D, Kadie C (1998) Empirical analysis of predictive algorithms for collaborative filtering. In: UAI'98, Madison, July 1998. Morgan Kaufmann, pp 43–52

6. Cremonesi P, Koren Y, Turrin R (2010) Performance of recommender algorithms on top-n recommendation tasks. In: RecSys'10, Barcelona, September. ACM, pp 39–46

7. Davoudi A, Chatterjee M (2016) Modeling trust for rating prediction in recommender systems. In: MLRec 2016, Miami, May. SIAM, pp 1–8

8. Ebadi A, Krzyzak A (2016) A hybrid multi-criteria hotel recommender system using explicit and implicit feedbacks. Int J Comput Electr Autom Control Inf Eng 10(8):1377–1385

9. Egger R, Gula I, Walcher D (2016) Open Tour Open Innov. Crowdsourcing and co-creation challenging the tourism industry, Springer, Berlin

10. Ekstrand MD, Riedl JT, Konstan JA et al (2011) Collaborative filtering recommender systems. Found Trends Hum Comput Interact 4(2):81–173

11. Farokhi N, Vahid M, Nilashi M, Ibrahim O (2016) A multi-criteria recommender system for tourism using fuzzy approach. J Soft Comput Decis Support Syst 3(4):19–29

12. Fuchs M, Zanker M (2012) Multi-criteria ratings for recommender systems: an empirical analysis in the tourism domain. In: EC-Web 2012, Vienna, September. Springer, pp 100–111

13. Gama J (2010) Knowledge discovery from data streams. Data mining and knowledge discovery series. CRC Press, Boca Raton

14. Gama J, Sebastião R, Rodrigues PP (2009) Issues in evaluation of stream learning algorithms. In: KDD'09, Paris, June. ACM, pp 329–338

15. Herlocker JL, Konstan JA, Terveen LG, Riedl JT (2004) Evaluating collaborative filtering recommender systems. ACM Trans Inf Syst 22(1):5–53

16. Jannach D et al (2012) Recommending hotels based on multidimensional customer ratings. Information and communication technologies in tourism. Springer, Helsingborg, pp 320–331

17. Jhalani T, Kant V, Dwivedi P (2016) A linear regression approach to multi-criteria recommender system. In: DMBD 2016, volume 9714 of LNCS, Bali, June. Springer, pp 235–243

18. Jia D, Zhang F, Liu S (2013) A robust collaborative filtering recommendation algorithm based on multidimensional trust model. J Softw 8(1):11–18

19. Jøsang A, Ismail R, Boyd C (2007) A survey of trust and reputation systems for online service provision. Decis Support Syst 43(2):618–644

20. Lathia N, Hailes S, Capra L (2008) Trust-based collaborative filtering. In: IFIP international conference on trust management, volume 263 of trust management II, Boston. Springer, pp 119–134

21. Law R, Chen F (2000) Internet in travel and tourism-part II: expedia. J Travel Tour Mark 9(4):83–87

22. Leal F, Dias JM, Malheiro B, Burguillo JC (2016) Analysis and visualisation of crowd-sourced tourism data. In: C3S2E'16, Porto, July. ACM, pp 98–101

23. Leal F, Malheiro B, Burguillo JC (2017) Prediction and analysis of hotel ratings from crowd-sourced data, volume 570 of advances in intelligent systems and computing, vol 570. Springer, Madeira, pp 493–502

24. Lee HH, Teng WG (2007) Incorporating multi-criteria ratings in recommendation systems. In: IRI 2007, Las Vegas, August. IEEE, pp 273–278

25. Linden G, Smith B, York J (2003) Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput 7(1):76–80

26. Liu L, Mehandjiev N, Xu DL (2011) Multi-criteria service recommendation based on user criteria preferences. In: RecSys'11, Chicago, October. ACM, pp 77–84

27. Manouselis N, Costopoulou C (2007) Analysis and classification of multi-criteria recommender systems. World Wide Web 10(4):415–441

28. Nilashi M, Bin Ibrahim O, Ithnin N, Sarmin NH (2015) A multi-criteria collaborative filtering recommender system for the tourism domain using expectation maximization (EM) and PCA-ANFIS. Electron Commer Res Appl 14(6):542–562

29. Randall Brandt D (1988) How service marketers can identify value-enhancing service elements. J Serv Mark 2(3):35–41

30. Sayed-Mouchaweh M (2016) Learning from data streams in dynamic environments. SpringerBriefs in applied sciences and technology. Springer, Heidelberg

31. Sedgwick P (2012) Pearson's correlation coefficient. Br Med J 345:e4483

32. Shambour Q, Hourani M, Fraihat S (2016) An item-based multi-criteria collaborative filtering algorithm for personalized recommender systems. Int J Adv Comput Sci Appl 7(8):275–279

33. Sykes AO (2000) An introduction to regression analysis. In: Posner EA (ed) Chicago lectures in law and economics. Foundation Press, New York

34. Takács G, Pilászy I, Németh B, Tikk D (2009) Scalable collaborative filtering approaches for large recommender systems. J Mach Learn Res 10:623–656

35. Veloso B, Malheiro B, Burguillo JC, Foss J (2017) Personalised fading for stream data. In: SAC 2017, Marrakech, April. ACM, pp 870–872

36. Wang H, Lu Y, Zhai C (2010) Latent aspect rating analysis on review text data: a rating regression approach. In: KDD'10, Washington, July. ACM, pp 783–792

37. Zhang L, Jiang S, Zhang J, Ng WK (2012) Robustness of trust models and combinations for handling unfair ratings. In: IFIPTM 2012, volume 374 of IFIP advances in information and communication technology, Surat, May. Springer, pp 36–51