Electronic Commerce Research and Applications 34 (2019) 100832

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/elerap

On-line guest profiling and hotel recommendation

Bruno M. Veloso[b,c,*], Fátima Leal[e], Benedita Malheiro[a,b], Juan Carlos Burguillo[d]

[a] ISEP/IPP – School of Engineering, Polytechnic Institute of Porto, Porto, Portugal
[b] INESC TEC – Porto, Portugal
[c] Univ Portucalense, Research on Economics, Management and Information Technologies – REMIT, Oporto, Portugal
[d] ETSET/UVigo – School of Telecommunications Engineering, University of Vigo, Vigo, Spain
[e] ISQ – Instituto de Soldadura e Qualidade, Grijó, Portugal

ARTICLE INFO

Keywords:
Profiling
Recommendation
Data streams
Post-filtering

ABSTRACT

Information and Communication Technologies (ICT) have revolutionised the tourism domain, providing a wide set of new services for tourists and tourism businesses. Both tourists and tourism businesses use dedicated tourism platforms to search and share information generating, constantly, new tourism crowdsourced data. This crowdsourced information has a huge influence in tourist decisions. In this context, the paper proposes a stream recommendation engine supported by crowdsourced information, adopting Stochastic Gradient Descent (SGD) matrix factorisation algorithm for rating prediction. Additionally, we explore different: (i) profiling approaches (hotel-based and theme-based) using hotel multi-criteria ratings, location, value for money (VfM) and sentiment value (StV); and (ii) post-recommendation filters based on hotel location, VfM and StV. The main contribution focusses on the application of post-recommendation filters to the prediction of hotel guest ratings with both hotel and theme multi-criteria rating profiles, using crowdsourced data streams. The results show considerable accuracy and classification improvement with both hotel-based and theme-based multi-criteria profiling together with location and StV post-recommendation filtering. While the most promising results occur with the hotel-based version, the best theme-based version shows a remarkable memory conciseness when compared with its hotel-based counterpart. This makes this theme-based approach particularly appropriate for data streams.

1. Introduction

Information and Communication Technologies (ICT) are shaping people, businesses and have revolutionised the tourism industry (Aramendia-Muneta and Ollo-Lopez, 2013) as well as the tourist experience (Yu et al., 2017). Namely, the mobile technology has a direct influence on tourist experiences due to its ubiquity in terms of Internet access, services and applications (Wang et al., 2013; Tussyadiah and Wang, 2016). Therefore, the multiple services provided by ICT (*e.g.*, personalised recommendations) condition the tourist behaviour and enhance the tourist experiences. The permanent interaction between tourists and ICT generates large volumes of crowdsourced data. This opinion-based information (ratings, likes, shares and reviews about tourism resources) generates a ripple effect where the decisions of current and future tourists are influenced by the opinions of their predecessors (Book et al., 2016; Zhang et al., 2016).

The data introduced by tourists in crowdsourcing platforms such as TripAdvisor, Expedia or Booking.com, comes in the form of data streams containing the data shared by tourists (*e.g.*, comments, reviews, ratings, videos, photos, posts or likes). A data stream is a sequence of data packets used to exchange information. In terms of personal and strategic decision making, this so-called crowd wisdom is critical for individuals and businesses alike (Howe, 2008). Tourists rely on the crowdsourced opinions to select services, while businesses mine the tourist feedback data to identify both problems and trends (Egger et al., 2016). By applying learning algorithms to crowdsourced data streams, it is possible to predict the tourist behaviour, including his preferences, based on his/her profile, which includes all information introduced by tourist in the platform (ratings, reviews, *etc.*).

This paper couples ICT and crowdsourced information since both are relevant for improving the tourism experience. In this context, we propose a recommendation engine based on crowdsourced ratings and textual reviews as well as on the official hotel information. The recommendation engine models both guests and hotels, and, thus, provides personalised hotel predictions. Moreover, we explore the available information using different mechanisms to refine the profiles and accurate the final recommendations. Specifically, we explore two main profiling approaches: (i) entity-based modelling – hotels and guests are

* Corresponding author at: Campus da Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal.
E-mail addresses: bmveloso@inesctec.pt, bruno.m.veloso@inesctec.pt (B.M. Veloso), mbm@isep.ipp.pt (B. Malheiro), J.C.Burguillo@uvigo.es (J.C. Burguillo).

https://doi.org/10.1016/j.elerap.2019.100832
Received 28 June 2018; Received in revised form 6 December 2018; Accepted 29 January 2019
Available online 21 February 2019

represented by the associated ratings, value for money (VfM) and Sentiment Value (StV); and (ii) feature-based modelling – hotels and guests are represented by the associated hotel themes and corresponding ratings, VfM and StV. The collected hotel themes were inspired in the official Expedia thematic hotel categories (adventure, beach, boutique, business, family, historic, luxury, romantic and spa). In terms of the application of the multi-criteria ratings, we adopted the Personalised Weighted Average Rating (PWRA), which combines multi-criteria ratings, using a personalised weighted average (Leal et al., 2017). Finally, we explore different post-recommendation filters using the location, VfM and StV to improve the accuracy of the final recommendations.

The contributions focus the refinement of guest and hotel profiling as well as the improvement of final recommendations to improve the tourism experience. To refine the guest and hotel profiles, we contribute with a novel profiling approach which combines the available crowdsourced multiple-criteria ratings, textual reviews and hotel information. Finally, to improve the accuracy of the final recommendations, we create post-recommendation filters taking into account crowdsourced information. The proposed hotel and guest profiling and recommendation techniques were evaluated with the HotelExpedia data set,[1] which includes crowdsourced (guest) and official information (hotel).

The results show the relevance of the post-recommendation filters in the recommendation accuracy. Concretely, the hotel-based profiling together with location and StV post-recommendation filter provide the best results. However, while hotel-based approach includes the cold-start problem, the theme-based ensures which all hotels have a theme with associated ratings in order to be recommended.

This paper is organised as follows. The related work section reviews the state of the art on rating-based guest profiling and on-line techniques for guest and hotel modelling and personalised hotel recommendations. The proposed method describes the proposed approach, including the algorithm used. The experiments and results report the data set properties, evaluation metrics and protocol, the tests performed and the results obtained. Finally, the conclusion summarises and discusses the outcomes of this work as well as the future research lines.

2. Related work

While for tourism industries (*e.g.*, airlines, hotels, tour operators, travel agencies, *etc.*) ICT is a key factor regarding competitiveness or marketing strategy, for tourists it simplifies travel planning. Specifically, Internet, mobile technology and wireless computing connect tourists with immediate access to relevant information regarding tourism resources (Farkhondehzadeh et al., 2013). Additionally, tourists rate, comment and share travelling experiences, intrinsically influencing the decision-making process of other tourists (Munar and Jacobsen, 2014; Kim and Fesenmaier, 2017). This behaviour leaves behind significant volumes of crowdsourced feedback data of strategic importance for tourism businesses. In fact, ICT has redefined the five stages of the travel cycle (World Tourism Organization, 2017):

- Dreaming: The tourist starts to consider travelling. In this stage, the Web typically acts as an inspirational source of information where the tourist browses through social networks, blogs, wikis, or travelling sites, going through shared tourism information (texts, photos, or videos) regarding potential destinations.
- Researching: The tourist invests time to research further about the viability of his travelling intentions with the support of the countless tourism services available on the Web, *e.g.*, TripAdvisor, Booking, eDreams, Expedia, airbnb, Wikivoyage, *etc*. Some of these tourism services, which use profiling and recommendation mechanisms, offer registered users automatic support and guidance.
- Booking: The tourist books the selected resources, including transportation, accommodation, events, *etc*. On-line booking, both from desk/laptop and mobile devices, has increased exponentially. In particular, to meet the tourist needs, booking websites should offer desk/laptop and mobile versions and present an intuitive and responsive Web design.
- Experiencing: The tourist searches for location-aware tourism information mainly via smart-phones. In this context, mobile applications provide real time access to routes, recommendations, maps or complementary information.
- Sharing: The tourist shares opinions and information in real time or in deferred time. This valuable feedback information (pictures, reviews, opinions or general travel information) influences the decisions of tourists and businesses alike.

Considering the tourism research literature, this paper aims to facilitate the searching stage using crowdsourced hotel data, *i.e.*, data generated during the sharing stage, to provide personalised hotel recommendations. The gathered data supports, on the one hand, personalised tourist recommendations and, on the other hand, the identification of new tourist trends.

There are a large number of tourism recommendation systems providing routes, maps, tourism resources recommendations, *etc.*, (Borràs et al., 2014; Gavalas et al., 2014). However, scant research integrates the crowd sharing information, *i.e.*, crowdsourced data, to drive the researching stage of the travel cycle. Therefore, this related-work contemplates the on-line approaches for guest modelling and hotel recommendation supported by crowdsourced data, but also an up-to date overview regarding recommendation methodologies in the hotel industry.

In the on-line scenario, stream mining techniques process continuous streams of events also known as data streams to learn models and predict the behaviours in near real-time, *e.g.*, to learn the user behaviour and provide on-line recommendations. Stream mining is the process of discovering knowledge or patterns from continuous data streams (Han et al., 2009). In particular, (Gama, 2010; Amatriain, 2013; Sayed-Mouchaweh, 2016) address the problems of modelling, prediction, classification, data understanding and processing in unpredictable environments exploiting data streams processing techniques. These stream mining and knowledge discovery techniques are used for personalisation in on-line systems. Regarding hotel recommendation, data stream mining enables the continuous updating of the guest and hotel models and, thus, contributes to improve the quality of near real-time personalised recommendations. Regarding this scenario, (Vinagre et al., 2015; Lommatzsch and Albayrak, 2015; Ludmann, 2015; Chang et al., 2017) investigated the problem of recommendation with stream inputs, specifically, the evolution of preferences over time, optimisation, and evaluation of stream-based recommendations.

This related work contemplates, still, a recent overview regarding hotel recommendation systems with profiling approaches. There is a vast set of methodologies, *i.e.*, based on ratings, reviews, context-aware information or using multi-criteria ratings. In terms of recommendation, they rely on data mining or machine learning algorithms, *i.e.*, the related work on hotel recommendation research is based on off-line processing. Specifically, the surveyed approaches make predictions based on:

Ratings are used by guests to express their opinion regarding the different aspects of the service. In terms of hotel industry, crowdsourcing platforms such as TripAdvisor, Expedia or Booking.com allow guest to rate multiple service dimensions (*e.g.*, comfort, cleanliness, value for money, overall, *etc.*).

Song et al. (2016) create hotel clusters according to the multiple criteria ratings available (general-score, cleanliness, comfort,

[1] http://ave.dee.isep.ipp.pt/1080560/ExpediaDataSet.7z.

satisfaction, closeness-to-down-town, and service) and expect the user to provide a set of rating thresholds for receiving recommendations. The system determines the similarity between the user preferences and the hotel cluster average ratings by calculating the Euclidean Distance.

Farokhi et al. (2016) explore single criterion profiling together with collaborative filtering. First, the authors determine the most representative rating based on the correlation between the multiple ratings, then, using the resulting rating (overall), apply data clustering (Fuzzy c-means and k-means) to find the nearest neighbours and, finally, predict the unknown hotel ratings using the Pearson Correlation coefficient. The evaluation is performed with TripAdvisor data.

Textual reviews are yet another type of guest feedback data. These written texts embody a detailed description regarding the guest experience in a hotel, which affects the behaviour of others potential guests. The review-based hotel recommendations rely on opinion mining, natural language processing or topic modelling for learning the guest behaviour and, thus, providing recommendations. Researches have been conducted to explore the processing of textual reviews for hotel recommendation.

Dong and Smyth (2016) apply sentiment analysis to the textual reviews to extract features regarding both hotels and users. The algorithm ranks the recommendations according to the similarity among hotels. To evaluate the system, the authors use TripAdvisor data.

Shrote and Deorankar (2016) propose a review-based recommendation method exploiting Big Data technologies. They apply sentiment analysis to personalise the recommendations. The system is evaluated with an unknown data set composed by hotels in the cities like Dubai, London, Paris, *etc*.

Ratings & **textual reviews** fusion leads to a refined profiling. Regarding this scenario, we found works relying on both ratings and reviews to generate the recommendations.

Ebadi and Krzyzak (2016) develop an intelligent hybrid multi-criteria hotel recommender system which uses both textual reviews and ratings from TripAdvisor. Regarding the ratings, it adopts single criteria profiling to learn the guest preferences and Singular Value Decomposition (SVD) matrix factorisation to predict unknown ratings in off-line mode. Additionally, they perform Topic Modelling to the textual reviews in order to extract the preferences.

Sharma et al. (2015) recommend hotels based on user preferences extracted from textual reviews. The reviews are processed off-line using Natural Processing Language techniques. The evaluation was performed with TripAdvisor data.

Context-aware filtering is another relevant factor in hotel recommendation. The hotel context encompasses information regarding travel purpose (*e.g.*, family, business, *etc.*), hotel amenities (*e.g.*, Wi-Fi, swimming pool, breakfast quality, *etc.*) or themes (*e.g.*, adventure hotels, romantic hotels, luxury hotels, *etc.*).

Hu et al. (2016) propose a hotel recommendation engine considering the travel purpose. The user travel purpose is inferred using textual reviews. The reviews are processed in off-line mode, using Term Frequency-Inverse Document Frequency. The system is evaluated using USA hotels crawled from TripAdvisor.

Hariri et al. (2011) present a context-aware recommender system which extracts the travel purpose, *i.e.*, contextual information from textual reviews. The authors use Labelled Latent Dirichlet Allocation to infer the context and cosine similarity to match the users with hotel contexts. The system was assessed with TripAdvisor data.

This paper explores profiling and prediction using tourism crowdsourced information to support the travel cycle. The main goal is to create the guest and hotel profiles by reusing the multitude of information provided by tourism crowdsourcing platforms. In terms of profiling, we experiment with hotel-based and theme-based profiling as well as with the personalised combination of the multiple crowdsourced information in an attempt to improve the quality of the collaborative predictions. In particular, we argue that theme-based hotel profiling mitigates the new hotel modelling problem. Additionally, we explore recommendation with post-recommendation filters employing the hotel location, value for money and sentiment value in order to improve the recommendation results. To address the problem of recommending hotels in near real-time, we propose a hotel stream recommendation engine.

Table 1
Comparison of hotel recommendation research approaches.

Approach	Evaluation	Profiling	Prediction	Post-Filtering
Song et al. (2016)		Rat	ED	
Farokhi et al. (2016)	TripAdvisor	Rat	*k*-means	
Dong and Smyth (2016)	TripAdvisor	Rev	Similarity	
Shrote and Deorankar (2016)		Rev	SA	
Ebadi and Krzyzak (2016)	TripAdvisor	Rat & Rev	SVD & TP	
Sharma et al. (2015)	TripAdvisor	Rat & Rev	NLP	
Hu et al. (2016)	TripAdvisor	Rev	TF-IDF	
Hariri et al. (2011)	TripAdvisor	Rev	CS + LDA	
Current proposal	Expedia	Rat & Rev	SGD	Loc, StV & VfM

Rat - Ratings
Rev - Textual Reviews
ED - Euclidean Distance
SA - Sentiment Analysis
TP - Topic Modelling
NLP - Natural Language Processing
TF-IDF - Term Frequency–Inverse Document Frequency
LDA - Latent Dirichlet Allocation
CS - Cosine Similarity

Table 1 provides a comparison regarding our proposal and the research found in the literature. While the surveyed systems use on-line processing to generate the predictions, our proposal explores SGD collaborative filtering with incremental updating. In addition, we explore location, the sentiment of textual reviews and the value for money of the hotels as post-filtering in order to create refined profiles and, consequently, obtain improved recommendations.

Ratings and textual reviews have been explored by Shrote and Deorankar (2016) and Sharma et al. (2015) to model guests and hotels, although ignoring the new item problem. Our approach complements these works by investigating both entity-based and theme-based models. In particular, theme-based profiling not only circumvents the new item problem, *i.e.*, the profile of a new hotel is based on the average ratings of its themes, making new hotels recommendable, but also ensures dimensionality reduction, a highly desired property when processing large volumes of data. Moreover, profiles can be further refined by taking advantage of the available information. In this context, Ebadi and Krzyzak (2016) and Sharma et al. (2015) combine ratings with preferences extracted from textual reviews. Our proposal quantifies the sentiment of the textual reviews, as showed by Dong and Smyth (2016) and Shrote and Deorankar (2016), and uses this information as a post-filter to improve recommendations.

The prediction of hotel ratings has been typically performed using on-line processing, *e.g.*, Song et al. (2016) explore on-line content-based filtering and Farokhi et al. (2016) adopt off-line memory-based collaborative filtering. Our implementation is based on SGD, using incremental updating, *i.e.*, on-line processing where the model is incrementally updated with each incoming event, together with post-recommendation filters, using the location, VfM, and StV (extracted from the textual reviews) of the hotels to improve the recommendation accuracy.

To sum up, previous works are focussed on on-line data processing, rely mostly on entity-based ratings to model guests and hotels and tend to use ignore other available information such as crowdsourced hotel sentiment value and value for money. Therefore, the current proposal is a novel approach with scant related work, making it impossible to do benchmarking. In this context, when compared to the research found in the literature, our work contributes with: (i) an on-line

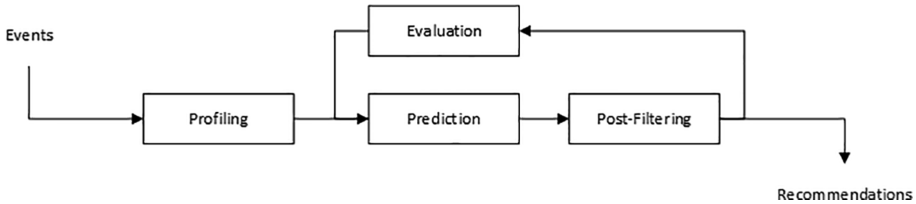


Fig. 1. Stream recommendation engine.

recommendation system for the tourism domain which performs incremental updates based on the crowdsourced data streams; (ii) theme-based profiling to mitigate the new item problem; and (iii) post-recommendation filtering to improve the accuracy of the final recommendations.

3. Proposed method

Tourism crowdsourcing platforms encompass not only hotel-related data (*e.g.*, description, price, location, star rating, amenities, *etc.*), but also tourist-related data (*e.g.*, multi-criteria ratings or textual reviews). Our proposed method addresses the problem of recommending hotels in near real-time using such streams of tourism crowdsourced information. In this context, we design a stream recommendation engine, illustrated in Fig. 1, which incorporates: (i) profiling – entity-based (hotel-based) and feature-based (theme-based); (ii) rating prediction – SGD; (iii) post-recommendation filtering – location, value for money and sentiment value; and (iv) evaluation – Root Mean Square Error (RMSE), Recall@10 and Target Recall@10 (TRecall).

Typically, a collaborative recommendation filter relies on a unique rating to create recommendations, *i.e.*, the user is modelled using a single rating. However, the tourism crowdsourced information typically incorporates multi-criteria ratings. To make the best out of it, we explore single criteria and multi-criteria profiling methods together with hotel-based and the theme-based models. While the hotel-based profiling associates guests with the hotel ratings, the theme-based profiling characterises the hotel using the themes extracted from the official hotel description. Since the themes are platform dependent, there is the need to explore the API of each crowdsourcing platform to identify the set of hotel themes used.

Once the guests and hotels are modelled, the recommendation engine predicts unknown ratings via SGD. SGD represents guests and hotels as vectors of latent factors. We adopt SGD for rating prediction due to its performance and data dimensionality reduction.

To refine the final recommendations, we employ post-recommendation filtering based on the available crowdsourced information, *i.e.*, location, hotel value for money, and hotel sentiment extracted from textual reviews. Finally, the guest is presented with a personalised list holding the top 10 hotels.

The system was evaluated by calculating the incremental RMSE, Recall@10, Target Recall@10 and receiver operating characteristic (ROC) curve. The experiments were repeated 30 times in order to evaluate the variance of the results. Finally, to evaluate the statistical differences between the proposed and the baseline approaches, we applied the McNemar test (McNemar, 1947).

3.1. Profiling

Profiling is the activity of modelling the available information via learning techniques. This paper explores different profiling approaches in order to model guests and hotels. Expedia holds, for each hotel, the hotel official information, including location, price, star rating and themes, as well as the multi-criteria hotel ratings and textual reviews given by each guest. These ratings classify different aspects of the tourist experience at the hotel, *e.g.*, the overall rating, the cleanliness rating, the service rating or the room rating. The exact set of criteria and themes may differ from platform to platform. However, to perform a rating prediction via matrix factorisation, it only possible to have a single rating per guest. Taking advantage of the profiling techniques for crowdsourced multi-criteria ratings explored by Leal et al. (2017), which provide a single hotel rating per guest, we chose the Most Representative Rating (MRR) as single criterion (SC) baseline profiling and the Personalised Weighted Rating Average (PWRA) as multi-criteria (MC) profiling. We argue that the personalised combination of crowdsourced multi-criteria tourism data improves the guest profile and, consequently, the quality of the collaborative predictions. Table 2 presents the notations used in profiling.

Most Representative Rating corresponds to the most influential rating of the available multi-criteria ratings. As first approach, we employ a multiple linear regression to select the most representative rating from the available set of multi-criteria hotel ratings as proposed by Leal et al. (2017).

Personalised Weighted Rating Average explores the combination of the different types of data in order to refine profiles. In this scenario, we employ the personalised weighted rating average approach as proposed by Leal et al. (2017). The PWRA combines the multiple types of ratings into a single rating. This profiling approach introduces the multi-criteria concept in recommendations.

Eq. (1) displays the personalised weighted rating average $r_{g,h}$ where $r_{g,h,c}$ is the non-null criterion c (*e.g.*, cleanliness, hotel condition, service, and staff, *etc.*) rating given by guest g to hotel h. While n_c presents the number of times guest g has rated criterion c, $n_{g,c}$ represents the total number of non-null criterion c ratings given by guest g. Finally, m is the total number of non-null multi-criteria ratings given by guest g.

$$r_{g,h} = \frac{\sum_{c=1}^{m} n_c r_{g,h,c}}{\sum_{c=1}^{m} n_{g,c}} \tag{1}$$

Using the MRR or PWRA, we model guests and hotels using item-based – Hotel-based – or feature-based – Theme-based – profiles. The hotel themes were extracted from the official hotel textual description.

Hotel-based rating profiling associates guests with the given hotel PWRA ratings, resulting in an item vector model (IVM) of the size of the number of hotels in the data set. This IVM guest rating profile $\widehat{P}_g$ holds in each position the average hotel rating $\bar{r}_{g,h}$ given by guest g to hotel h (Eq. (2))

$$\widehat{P}_g[h] = \bar{r}_{g,h} \tag{2}$$

Theme-based rating profiling models guests and hotels using the nine hotel themes – adventure, beach, boutique, business, family, historic, luxury, romantic, and spa. In this case, both the hotel and the guest rating profiles are feature vector models (FVM) of size nine, where each position corresponds to a theme. The theme-based guest rating profile $\widehat{P}_g^T$ holds, for each theme t, the average rating given to all hotels with theme t by guest g, whereas the theme-based hotel rating profile $\widehat{P}_h^T$ maintains, for each theme t, the average rating given by all guests to hotel h with theme t. Eq. (3) represents the theme-based guest rating profile $\widehat{P}_g^T$ where $r_{g,h}^t$ is the (MRR or PWRA) rating given by guest g to hotel h with theme t, T_h is the set of hotel h themes, H_g^t is the number of ratings given by guest g to hotels with theme t and H is the total number of hotels. Theme-based hotel rating profiles correspond to the average hotel rating since, every time a guest rates the hotel, all hotel themes are equally rated. Eq. (4) represents the theme-based hotel rating profile $\widehat{P}_h^T$ where $r_{g,h}$ is the PWRA rating given by guest g to hotel

h with theme t, G is the guests and G_h is the number of ratings given to hotel h.

$$\widehat{P}_g^T[t] = \begin{cases} \frac{1}{H_u^t} \sum_{h=1}^{H} r_{g,h}^t & : t \in T_h \\ 0 & : t \notin T_h \end{cases} \tag{3}$$

$$\widehat{P}_h^T[t] = \frac{1}{G_h} \sum_{g=1}^{G} r_{g,h} \tag{4}$$

3.2. Rating prediction

The rating prediction module implements a collaborative recommendation filtering to predict hotels ratings not yet provided by the tourists. Concretely, this paper explores the SGD technique to create the latent feature vectors, and update incrementally the model using hotel-based and theme-based profiling approaches. We defined the optimal latent feature space k adopting the experimental approach proposed by Sarwar et al. (2002).

Algorithm 1 describes the rating prediction using hotel-based profiling with data streams. The hotel-based stream algorithm: (i) creates an initial rating matrix with the guest ratings; (ii) builds the latent guest and hotel matrices, distributing randomly a component from a small range of -0.02 to 0.02 to ensure different coefficients; and (iii) generates the predictions using the latent matrices. Then, we select randomly 1000 non rated hotels plus the newly rated hotel, *i.e.*, 1001 hotels. The algorithm predicts the ratings of the selected 1001 hotels and sort them in descending order. The hyper-parameter optimisation process determines the optimal learning rate (γ) and over-fitting (λ) using RMSE. These parameters are used to update the guest and hotel latent matrix. Therefore, for each incoming rating, the process: (i) adds the rating to the initial rating matrix; (ii) calculates the prediction error and the accuracy metrics; and (iii) updates the guest and hotel latent matrix using the hyper-parameters. Finally, the method generates new predictions using the updated guest and hotel latent matrix. The method is repeated for each incoming rating. The complexity of Algorithm 1 is $Nlog(N)$.

Algorithm 1. Rating Prediction with Hotel-based Profiling.

Input: $r_{u,i}$, λ, γ, k, P, Q
Output: $\hat{r}_{u,i}$
$p_u \leftarrow getLatentUserVector(P, u)$
$q_i \leftarrow getLatentItemVector(Q, i)$
$\hat{r}_{u,i} = q_i^* p_u$
$e_{u,i} = r_{u,i} - \hat{r}_{u,i}$
$List[] \leftarrow get1000unratedItems() + i$
$SubList[] \leftarrow$ applyPostFilter($List[]$)
for $i \leftarrow SubList$ [] **do**
    $q_i \leftarrow getLatentItemVector(Q, i)$
    $\hat{r}_{u,i} = q_i^* p_u$
Sort(SubList[])
$RMSE \leftarrow computeRMSE(e_{u,i})$
$Recall \leftarrow computeRecall(i_{position})$
for all latent features k **do**
    $p_{u,k} \leftarrow p_{u,k} + \gamma(e_{u,i} q_{i,k} - \lambda p_{u,k})$

Algorithm 2 describes the rating prediction using theme-based profiling with data streams. The profile of a new hotel is based on the average ratings of the hotels classified with the same theme. Therefore, the theme-based stream algorithm: (i) creates an initial rating matrix with the guest ratings associated to the hotel themes; (ii) builds the latent guest and themes matrices distributing randomly a component from a small range of -0.02 to 0.02 to ensure different coefficients; and (iii) generates the predictions using the latent matrices. Then, we select randomly 1000 non rated hotels plus the newly rated hotel, *i.e.*, 1001 hotels. The algorithm predicts the ratings of the themes concerning the selected 1001 hotels and sort them in descending order. The hyper-parameter optimisation process follows the hotel-based algorithm. The complexity of Algorithm 2 is N^2.

Algorithm 2. Rating Prediction with Theme-based Profiling.

Input: $r_{u,i}$, λ, γ, k, P, Q
Output: $\hat{r}_{u,i}$
$p_u \leftarrow getLatentUserVector(P, u)$
for all item themes t **do**
    $q_t \leftarrow getLatentItemVector(Q, t)$
    $\hat{r}_{u,i} += q_t^* p_u$
$\hat{r}_{u,i} /= \#item\_themes$
$e_{u,i} = r_{u,i} - \hat{r}_{u,i}$
$List[] \leftarrow get1000unratedItems() + i$
$SubList[] \leftarrow$ applyPostFilter($List[]$)
for $i \leftarrow SubList$ [] **do**
    for all item themes t **do**
        $q_t \leftarrow getLatentItemVector(Q, t)$
        $\hat{r}_{u,i} += q_t^* p_u$
    $\hat{r}_{u,i} /= \#item\_themes$
Sort(SubList[])
$RMSE \leftarrow computeRMSE(e_{u,i})$
$Recall \leftarrow computeRecall(i_{position})$
for all latent features k **do**
    for all item themes t **do**
        $p_{u,k} \leftarrow p_{u,k} + \gamma(e_{u,i} q_{t,k} - \lambda p_{u,k})$

3.3. Post-recommendation filters

The *a posteriori* recommendation filters aim to improve the final recommendations provided to users taking advantage of the available profile data. Therefore, after the prediction using MRR or PWRA profiling approaches, we employ the following filters:

Hotel Value for Money considers the price, the crowd overall rating, and the hotel official star rating. We confronted the price with the overall and star ratings to find the best crowdsourced value for money. The VfM filters, on behalf of the guest, the hotels with better value for money according to the crowd.

Hotel Sentiment is computed employing sentiment analysis to the hotel crowdsourced textual reviews. Sentiment Analysis classifies attitudes, emotions, or opinions using textual documents. Particularly in tourism domain, the textual reviews contain relevant information at sentiment level. Moreover, tourists check the opinions and experiences published by other tourists on dedicated crowdsourcing platforms. This crowdsourced textual information influences the tourist behaviour in the decision making. Therefore, we apply sentiment analysis to the crowdsourced hotel reviews in order to extract the hotel sentiment value. We use the Alchemy Application Programming Interface (API) from IBM,[2] which provides a set of human language technology tools using Natural Language Processing (NPL). In particular, we use the sentiment analysis tool to quantify the sentiment of the textual reviews. There are five types of sentiment classification: very negative, negative, neutral, positive, and very positive. We rank this classification in a scale between -1 and $+1$. The hotel sentiment value filters, on behalf of the guest, the recommendations with better sentiment value according to the crowd. Table 2 contains two examples of sentiment analysis performed using the Alchemy API.

[2] https://www.ibm.com/watson/alchemy-api.html.

Table 2
Sentiment analysis via the Alchemy API.

Textual Review	Sentiment	Score
The best hotel I have ever been. Friendly and helpful staff, great place, spacious room. There is nothing to complain about.	Positive	0.723
The room was dirty and old. One elevator was out of service for the majority of our stay. Finally, in the third floor the noisy from outside was constant.	Negative	−0.233

experiments involve hotel-based and theme-based profiling using MRR and PWRA as well as the corresponding rating prediction evaluation with the post-recommendation filters (location, VfM, and StV). Finally, we evaluate the system using RMSE, Recall@10, and Target Recall@10, as well as the ROC curve. Moreover, we apply the Wilcoxon and McNemar test to reject the null-hypothesis.

The experiments were made on a cloud OpenStack instance, holding 16GiB RAM, 8 CPU and 160GiB hard-disk. In the following subsections we present the data set used on the experiments, the evaluation protocol as well as the evaluation metrics, the results and a final discussion.

Table 3
Expedia hotel and guest data.

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

Table 4
On-line accuracy and classification results.

	Post-Filter	RMSE	CV	R@10	CV	TR@10	CV
H-b MRR		*0.190*	*0.2*	*0.170*	*1.8*	*0.126*	*1.6*
	Loc	0.190	0.1	0.309	1.4	0.217	1.3
	Loc & VfM	0.190	0.2	0.454	1.7	0.308	1.4
	Loc & StV	0.190	0.2	0.546	1.4	0.360	1.2
H-b PWRA		*0.168*	*0.0*	*0.175*	*2.1*	*0.144*	*2.0*
	Loc	0.168	0.1	0.316	1.4	0.242	1.2
	Loc & VfM	0.168	0.0	0.476	1.5	0.353	1.3
	Loc & StV	**0.168**	0.1	**0.566**	1.5	**0.410**	1.4
T-b MRR		*0.240*	*0.0*	*0.017*	*4.2*	*0.027*	*2.0*
	Loc	0.240	0.0	0.154	1.3	0.131	1.1
	Loc & VfM	0.240	0.0	0.262	1.0	0.186	0.8
	Loc & StV	0.240	0.0	0.462	0.6	0.286	0.5
T-b PWRA		*0.222*	*0.0*	*0.016*	*5.5*	*0.028*	*1.3*
	Loc	0.222	0.0	0.160	1.8	0.141	1.3
	Loc & VfM	0.222	0.0	0.266	0.8	0.209	0.6
	Loc & StV	**0.222**	0.0	**0.469**	0.4	**0.336**	0.5

The baseline algorithm, *i.e.*, without post-filtering, of each profiling approach is identified by the italic type font.

Table 5
Statistical results.

Profiling	Test	Value	*p*-value
Hotel-based	W_{stat}	0.00	1.806×10^{-6}
Theme-based	W_{stat}	0.00	1.629×10^{-6}

4.1. Data set

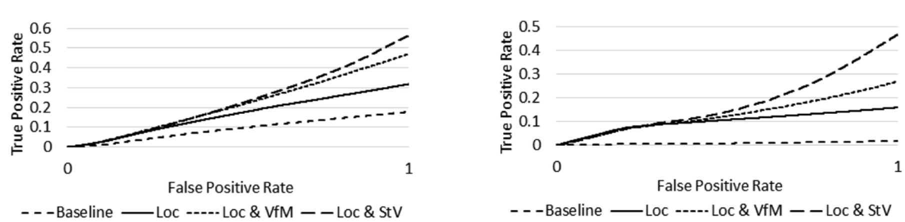
The HotelExpedia data set was proposed by Leal et al. (2017), and it is composed of 6276 hotels, 1090 identified users and 214,342 reviews from 11 different locations. Each user rated at least 20 hotels, and each hotel has a minimum of 10 ratings. However, for our experiments we eliminated the hotels without price. Therefore, the tests were performed with 3809 hotels, 1090 identified users and 187892 reviewers from 11 different locations.

Table 3 describes the content of the data set, highlighting the data used. Our experiments, which rely on the hotel, and user review data, use, specifically, the user nickname, the hotel identification, description, textual reviews and, as multi-criteria ratings, the overall, cleanliness, service & staff, hotel condition and room comfort. This data set does not contain null ratings, *i.e.*, all users have rated the multi-criteria components of the hotels. In the case of the HotelExpedia data set, the MRR is the overall rating.

4.2. Evaluation protocol and metrics

The evaluation protocol defines the data ordering and distribution. The data was ordered temporally in order to apply the proposed stream recommendation engine. The model is updated incrementally, using the entire data set as streams. Therefore, when a guest rates a hotel the method updates not only the predictions and the guest and hotel profiles, but also the evaluation metrics. The adopted evaluation protocol was inspired by the prequential evaluation proposed by Gama et al. (2009).

The evaluation of the proposed stream recommendation engine involves predictive accuracy and classification metrics. In terms of predictive accuracy metrics, we use the RMSE to measure the error between the predicted rating and the real guest rating. The RMSE is calculated incrementally, *i.e.*, in each new incoming guest rating event,

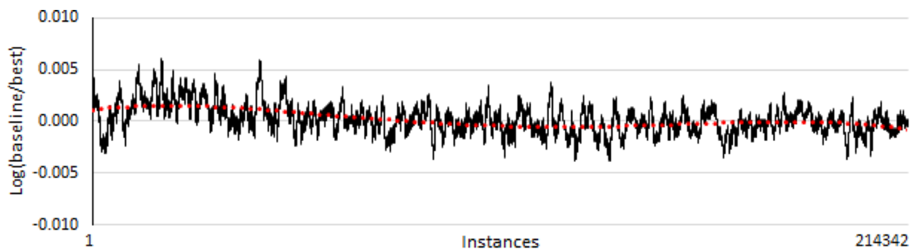


(a) Hotel-based PWRA for HotelExpedia (b) Theme-based PWRA for HotelExpedia

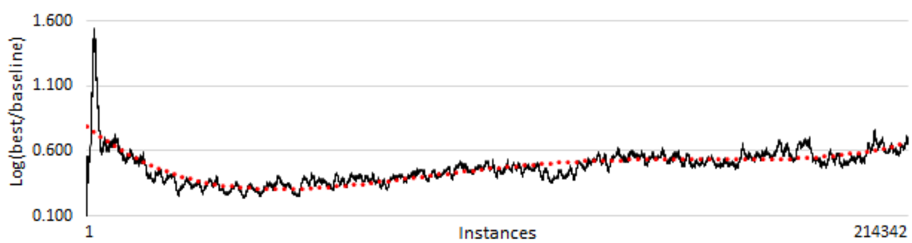
Fig. 2. ROC.

4. Experiments and results

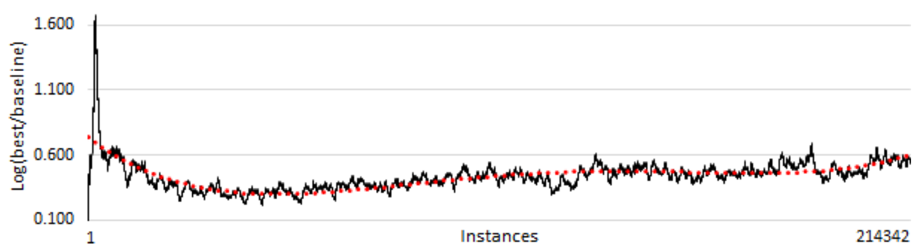
We have conducted several experiments with the Expedia data set to evaluate the effectiveness and usefulness of the proposed method. The



(a) Incremental RMSE



(b) Incremental R@10



(c) Incremental TR@10

Fig. 3. HotelExpedia results with Hotel-based PWRA, location and sentiment value.

as proposed by Takács et al. (2009).

In the case of classification metrics, which evaluate the accuracy of the recommendations, we calculated the Recall@10 (R@10) as proposed by Cremonesi et al. (2010) and the TRecall@10 (TR@10) presented by Veloso et al. (2017). To calculate the Recall@10, first, we predict the ratings of the hotels non rated by the guest, including the newly rated hotel. Then, we select 1000 non rated hotels plus the newly rated hotels and sort them in descending order. If the newly rated hotel belongs to the list of the top 10 guest predicted items, we count a hit. The TRecall@10 evaluates the accuracy of the predictions centred around the target guest rating. It counts a hit if the predicted rating lays within the $\pm \frac{N}{2}$ predictions centred on the real rating. Using these metrics, we plot the ROC curve to show the performance of the different profiling methods. A ROC curve illustrates the capability of a binary classifier by plotting the true positive rate against the false positive rate of the system.

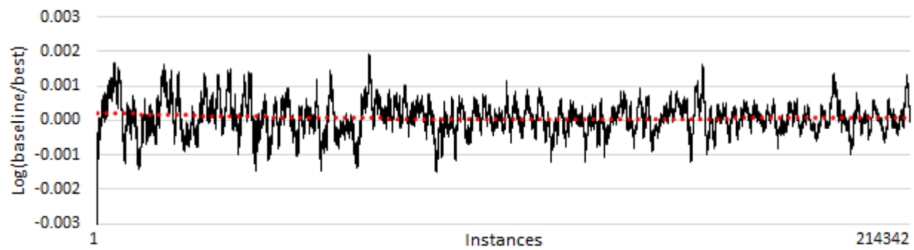
Additionally, we measure the recommendation time – Δt (ms) – and the memory footprint. All tests were repeated 30 times to compute the coefficient of variation (CV) of the evaluation metrics.

4.3. Results

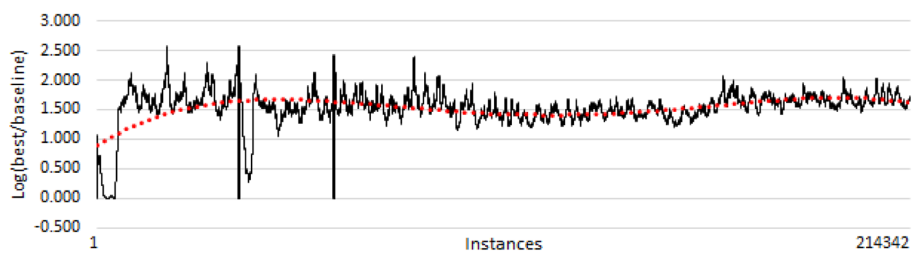
The experiments involve the SGD algorithm together with: (i) hotel-based and theme-based with MRR profiling; and (ii) hotel-based and theme-based with PWRA profiling. Additionally, the previous the different post-recommendation filters – hotel location (Loc), value for money (VfM) and sentiment value (StV) – were applied to the previous profiling approaches. The location is based on the guest destination, the VfM is based on the hotel price, on the overall, and official star ratings, and the StV is extracted from hotel crowdsourced textual reviews. The MRR without post-filtering was used as the baseline approach.

Table 4 depicts the results of the accuracy metrics of the different tests, *i.e.*, RMSE, R@10, and TR@10, together with the corresponding CV. The best accuracy and classification results with hotel-based and theme-based profiling are displayed in bold. Regarding the memory footprint, the hotel-based model with PWRA, Loc & StV occupies 1.696MiB and the corresponding theme-based model 0.353MiB, whereas the baseline versions use 1.606MiB and 0.264MiB, respectively.

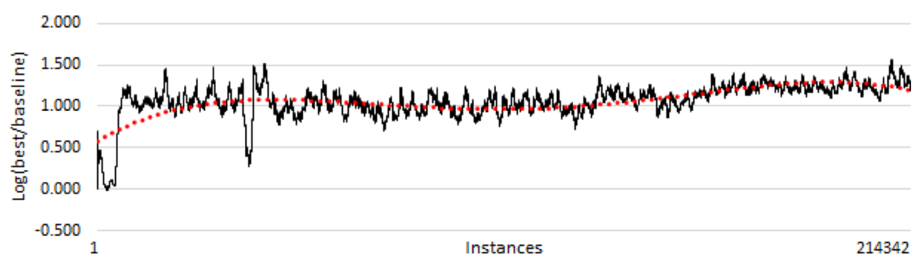
The best results occur using the hotel-based and theme-based together with location & sentiment value post-filtering. As expected, the best accuracy and classification occurs with hotel-based profiling, whereas the lowest memory footprint happens with theme-based profiling. The best hotel-based results, when compared with the corresponding base algorithm (without post-filtering), show an unchanged RMSE and an increase of 2.2 times in R@10, 1.8 times in TR@10, 5.5 times in run time and 0.06 times in memory. With the best theme-based, the RMSE remains unchanged, the R@10 is 28.3 times higher, TR@10 is 11.0 times higher, run time is 0.4 times smaller and the memory footprint is 0.34 times higher than with the corresponding baseline algorithm. These theme-based results, when compared with the best hotel-based counterparts, have higher RMSE (+32%) and run



(a) Incremental RMSE



(b) Incremental R@10



(c) Incremental TR@10

Fig. 4. HotelExpedia results with Theme-based PWRA, location and sentiment value.

time (+7 %) and lower R@10 (−17 %), TR@10 (−18 %) and memory footprint (−79 %). This result reveals that the proposed theme-based profiling is particularly promising for stream recommendation engines,*i.e.*, when memory usage becomes an important factor.

4.4. Significance tests

To detect statistical differences between the baseline and the proposed approaches for hotel and theme-based profiling, we applied the Wilcoxon test. In fact, we compare hotel-based PWRA (baseline) with hotel-based PWRA, Loc & StV profiling and theme-based PWRA (baseline) with theme-based PWRA, Loc & StV profiling. The Wilcoxon test is used to reject the null-hypothesis when the samples do not follow a normal distribution, *i.e.*, reject the hypothesis that the baseline and the proposed profiles have the same performance. We defined a 5 % of significance level *p*. With 5 % significance, the critical value of the Wilcoxon test (W_{crit}) is 137. In the case of the Wilcoxon test, two samples are statistically different if the $W_{stat} < W_{crit}$. Table 5 displays the determined W_{stat}, confirming that the proposed profiling approaches are statistically different from the corresponding baseline versions.

Conclusion and discussion on the results is not provided.

The authors never discuss on the sparsity issue and how the method can solve it. The authors should discuss how the new hotels can be recommended to the new users. This issue must be better explained.

4.5. Discussion

This paper explores different profiling approaches to refine the on-line recommendation of hotels. We conducted a series of experiments regarding entity-based profiling – hotel-based with MRR or PWRA – and feature-based profiling – theme-based with MRR or PWRA. In addition, the guest and hotel profiles were enriched with VfM, StV and location data. As expected, not only hotel-based consistently outperformed theme-based results, but also PWRA was systematically better than MRR rating profiling.

In terms of the prediction of unknown ratings, considering the different profiling approaches tested, the most promising results, regarding RMSE, R@10, and TR@10 evaluation metrics, were achieved with the hotel-based PWRA profiling.

Finally, we applied different post-recommendation filters to improve the final on-line recommendations, namely, hotel location (guest destination), value for money – considering hotel price, crowd overall rating and official star rating – and the hotel sentiment value – extracted from the guest textual reviews. Consistently, the most promising results occurred with location & sentiment value post filtering.

To analyse the multiple experiments concerning the post-recommendation filters, we built the corresponding ROC curves depicted in Fig. 2. Consistently, the most promising results occurred with location & sentiment value post-filtering in both hotel-based and theme-based profiling approaches. The results corroborate our claim that profiling refinement and post-recommendation filters improve on-line personalised hotel recommendations.

On-line evaluation relies on incremental accuracy and classification metrics. These following plots represent the logarithm between the best and the baseline approaches for all data instances regarding the different evaluation metrics used. Fig. 3a–c depict the evolution of the incremental RMSE, Recall@10 and TRecall@10 results concerning hotel-based PWRA profiling with location & StV post-filtering for HotelExpedia.

Fig. 4a–c display the evolution of the adopted evaluation metrics concerning theme-based PWRA profiling with location & StV post-filtering.

These results show that the proposed on-line updating of profiles using the stream of incoming user events, improves the accuracy and classification of the personalised recommendations with both hotel and theme based profiling.

5. Conclusions

The development of tourism information and communication technologies popularised the usage of crowdsourcing platforms (*e.g.*, Expedia, TripAdvisor, Booking.com, *etc.*) among tourists for planning, booking, experiencing and sharing. These tourism crowdsourcing platforms collect large volumes of feedback data regarding tourism resources, including multi-criteria ratings, textual reviews, photos, *etc*.

This paper explores the crowd-sourced information regarding the hotel industry using Expedia data, *i.e.*, multiple criteria ratings and textual reviews as well as official hotel information. The work attempts to create and refine on-line profiles of both guests and hotels, exploring different approaches together with collaborative filtering algorithms.

To address this problem we designed and experimented with two different profiling approaches comparing on-line and off-line execution modes: (i) entity-based modelling using PWRA rating profiles, hotel location, sentiment value and value for money; and (ii) feature-based modelling using hotel themes (extracted from official hotel description) and corresponding ratings, hotel location, sentiment value and value for money. The rating profiles were based on the crowd-sourced guest multi-criteria hotel ratings, the value for money on the hotel crowd-sourced overall rating as well as the hotel official room price and star rating, the sentiment value on the crowd-sourced guest textual reviews and the themes were extracted from the official hotel description. To generate personalised predictions we employed the SGD matrix factorisation algorithms. The experiments were conducted in off-line and on-line scenarios. Additionally, in on-line we employed post-recommendation filters based on the hotel location (guest destination), value for money and aggregated sentiment value to improve the final on-line recommendations accuracy.

The experiments were conducted with the HotelExpedia data streams. As evaluation metrics, we use the RMSE, Recall@10, and TRecall@10. The MRR profiling is the baseline approach for the comparisons. Consistently, the results with post-recommendation filtering improve both in hotel-based and theme-based profiling. The best results, when considering all metrics, are those with hotel-based PWRA profiling and location & StV post-recommendation filtering. However, with theme-based profiling, hotels have one or more themes with associated ratings in order to be recommended, exhibiting lower memory footprint and circumventing the new hotel problem. To verify the differences between the baseline and the proposed hotel-based and theme-based profiling approaches, we use the McNemar test. This test confirms that the proposed approaches present relevant statistic differences.

In terms of contributions, this research work provides a novel profiling approach regarding guests and hotels based on available crowd-sourced information – multi-criteria ratings, textual reviews and hotel information. Furthermore, the paper shows that SGD is a promising technique, which when combined with refined profiles and appropriate post-recommendation filters, is well suited for the on-line recommendation of hotels.

As future work, we intend to explore further with theme-based profiling. We believe that theme-based profiles have potential and require further research, specifically, in terms of theme identification and mitigation of the new hotel cold start problem.

Conflict of interest

There are no conflict of interests.

Acknowledgements

This work has been supported by: (i) the project "TEC4Growth – RL SMILES – Smart, Mobile, Intelligent and Large scale Sensing and analytics NORTE-01-0145-FEDER-000020", financed by the North Portugal regional operational program (NORTE 2020) under the Portugal 2020 partnership agreement and the European Regional Development Fund; (ii) the European Commission Horizon 2020 programme via the ICT COST Action IC1406 High-Performance Modelling and Simulation for Big Data Applications (cHiPSet); and (iii) the European Regional Development Fund (ERDF) together with the Galician Regional Government under agreement for funding the Atlantic Research Center for Information and Communication Technologies (AtlanTTic).

References

Amatriain, X., 2013. Mining large streams of user data for personalized recommendations. ACM SIGKDD Expl. Newslett. 14 (2), 37–48.
Aramendia-Muneta, M.E., Ollo-Lopez, A., 2013. Ict impact on tourism industry. Int. J. Manage. Cases 15 (2), 87–98.
Book, L.A., Tanford, S., Chen, Y.S., 2016. Understanding the impact of negative and positive traveler reviews: Social influence and price anchoring effects. J. Travel Res. 55 (8), 993–1007.
Borràs, J., Moreno, A., Valls, A., 2014. Intelligent tourism recommender systems: a survey. Expert Syst. Appl. 41 (16), 7370–7389.
Chang, S., Zhang, Y., Tang, J., Yin, D., Chang, Y., Hasegawa-Johnson, M.A., Huang, T.S., 2017. Streaming recommender systems. In: Proceedings of the 26th International Conference on World Wide Web, pp. 381–389. International World Wide Web Conferences Steering Committee.
Cremonesi, P., Koren, Y., Turrin, R., 2010. Performance of recommender algorithms on top-n recommendation tasks. In: Proceedings of the fourth ACM conference on Recommender systems ACM, pp. 39–46.
Dong, R., Smyth, B., 2016. From more-like-this to better-than-this: Hotel recommendations from user generated reviews. In: Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization ACM, pp. 309–310.
Ebadi, A., Krzyzak, A., 2016. A hybrid multi-criteria hotel recommender system using explicit and implicit feedbacks. In: Proceedings of 18th International Conference on Applied Science in Information Systems and Technology (ICASIST 2016) Amsterdam.
Egger, R., Gula, I., Walcher, D., 2016. Open Tourism. Springer.
Farkhondehzadeh, A., Robat Karim, M.R., Roshanfekr, M., Azizi, J., Legha Hatami, F., 2013. E-tourism: the role of ict in tourism industry. Eur. Online J. Nat. Soc. Sci. 2 (3s), 566.
Farokhi, N., Vahid, M., Nilashi, M., Ibrahim, O., 2016. A multi-criteria recommender system for tourism using fuzzy approach. J. Soft Comput. Dec. Support Syst. 3 (4), 19–29.
Gama, J., 2010. Knowledge Discovery from Data Streams. CRC Press.
Gama, J., Sebastio, R., Rodrigues, P.P., 2009. Issues in evaluation of stream learning algorithms. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ACM, pp. 329–338.
Gavalas, D., Konstantopoulos, C., Mastakas, K., Pantziou, G., 2014. Mobile recommender systems in tourism. J. Network Comput. Appl. 39, 319–333.
Han, J., Ding, B., 2009. Stream mining. In: Encyclopedia of Database Systems. Springer, pp. 2831–2834.
Hariri, N., Mobasher, B., Burke, R., Zheng, Y., 2011. Context-aware recommendation based on review mining. In: Proceedings of the 9th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems (ITWP 2011). Citeseer, p. 30.
Howe, J., 2008. Crowdsourcing: How the Power of the Crowd is Driving the Future of Business. Random House.
Hu, Y.H., Lee, P.J., Chen, K., Tarn, J.M., Dang, D.V., 2016. Hotel recommendation system based on review and context information: a collaborative filtering appro. In: PACIS, p. 221.
Kim, J., Fesenmaier, D.R., 2017. Sharing tourism experiences: the posttrip experience. J. Travel Res. 56 (1), 28–40.
Leal, F., González-Veléz, H., Malheiro, B., Burguillo, J.C., 2017. In: Profiling and rating prediction from multi-criteria crowd-sourced hotel ratings In: 31st European Conference on Modelling and Simulation. European Council for Modeling and Simulation, pp. 576–582.

Leal, F., Malheiro, B., Burguillo, J.C., 2017. Prediction and analysis of hotel ratings from crowd-sourced data. In: World Conference on Information Systems and Technologies. Springer, pp. 493–502.
Lommatzsch, A., Albayrak, S., 2015. Real-time recommendations for user-item streams. In: Proceedings of the 30th Annual ACM Symposium on Applied Computing. ACM, pp. 1039–1046.
Ludmann, C.A., 2015. Online recommender systems based on data stream management systems. In: Proceedings of the 9th ACM Conference on Recommender Systems ACM, pp. 391–394.
McNemar, Q., 1947. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12 (2), 153–157. https://doi.org/10.1007/BF02295996.
Munar, A.M., Jacobsen, J.K.S., 2014. Motivations for sharing tourism experiences through social media. Tourism Manage. 43, 46–54.
Sarwar, B., Karypis, G., Konstan, J., Riedl, J., 2002. Incremental singular value decomposition algorithms for highly scalable recommender systems. In: Fifth International Conference on Computer and Information Science, pp. 27–28.
Sayed-Mouchaweh, M., 2016. Learning from Data Streams in Dynamic Environments. Springer.
Sharma, Y., Bhatt, J., Magon, R., 2015. A multi-criteria review-based hotel recommendation system. In: Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on. IEEE, pp. 687–691.
Shrote, K.R., Deorankar, A.V., 2016. Hotel recommendation system using hadoop framework. Indian J. Appl. Res. 6 (4).
Song, W.W., Lin, C., Avdic, A., Forsman, A., Åkerblom, L., 2016. Collaborative filtering with data classification: A combined approach to hotel recommendation systems. In: 25th International Conference on Information Systems Development (ISD2016 Poland), Katowice, Poland, August 24–26, 2016.
Takács, G., Pilászy, I., Németh, B., Tikk, D., 2009. Scalable collaborative filtering approaches for large recommender systems. J. Mach. Learn. Res. 10, 623–656.
Tussyadiah, I.P., Wang, D., 2016. Tourists attitudes toward proactive smartphone systems. J. Travel Res. 55 (4), 493–508.
Veloso, B., Malheiro, B., Burguillo, J.C., Foss, J., 2017. Personalised fading for stream data. In: SAC 2017: Symposium on Applied Computing Proceedings. 32nd ACM Symposium on Applied Computing (SAC 2017), Data Streams Track ACM, New York, pp. 870–872.
Vinagre, J., Jorge, A.M., Gama, J., 2015. Evaluation of recommender systems in streaming environments. arXiv preprint arXiv:1504.08175.
Wang, D., Fesenmaier, D.R., 2013. Transforming the travel experience: The use of smartphones for travel. In: Information and Communication Technologies in Tourism 2013, pp. 58–69.
World Tourism Organization: Technology in Tourism, Volume 1. http://cf.cdn.unwto.org/sites/all/files/pdf/unwtodigitalresources_volume1_techtourism_eng.pdf. Accessed on December 2017.
Yu, X., Anaya, G.J., Miao, L., Lehto, X., Wong, I.A., 2017. The impact of smartphones on the family vacation experience. J. Travel Res 0047287517706263.
Zhang, L., Wu, L., Mattila, A.S., 2016. Online reviews: the role of information load and peripheral factors. J. Travel Res. 55 (3), 299–310.

Further reading

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bull. 1 (6), 80–83. http://www.jstor.org/stable/3001968.