

Experiments with a Venue-Centric Model for Personalised and Time-Aware Venue Suggestion

Romain Deveaud, M-Dyaa Albakour, Craig Macdonald and Iadh Ounis
University of Glasgow, UK
firstname.lastname@glasgow.ac.uk

ABSTRACT

Location-based social networks (LBSNs), such as Foursquare, fostered the emergence of new tasks such as recommending venues a user might wish to visit. In the literature, recommending venues has typically been addressed using user-centric recommendation approaches relying on collaborative filtering techniques. Such approaches not only require many users with detailed profiles to be effective, but they also cannot recommend venues to users who are not actually members of the LBSN. In contrast, in this paper, we introduce a venue-centric yet personalised probabilistic approach that suggests personalised and popular venues for users to visit in the near future. In our approach, we probabilistically incorporate two components, a popularity component for predicting the popularity of a venue at a given point in time, as estimated from the attendance of the venue in the LBSN (i.e. number of check-ins), and a personalisation component for identifying its interestingness with respect to the estimated preferences of the user. The popularity of each venue is predicted using time series forecasting models that are trained on the recent attendance trends of the venue, while the users' interests are modelled from the entity pages that they like on Facebook. Using three major cities, we conduct a user study to evaluate the effectiveness of the two components of our approach in suggesting venues for different types of users at different times of the day. Our experimental results show that an approach that combines the popularity and personalisation components is able to consistently outperform the recommendation service of the leading Foursquare LBSN. We also find that combining popularity and personalisation is effective for both new visitors and residents, while former visitors prefer popular venues.

1. INTRODUCTION

With the prevalence of smartphones and ever improving mobile access to the Internet, new behaviours have emerged. People stay connected and rapidly share their opinions, interests, locations and many other aspects of their lives on social networks. The users of these social networks therefore act as Internet-connected sensors of the real world and pro-

vide a wealth of information about what is currently happening around them. At the heart of these user-generated information lie Location-based Social Networks (LBSNs) [40], such as Foursquare, Flickr or to some extent Twitter. All such services allow their users to share their location, whether by tweeting a message, posting a photo on Flickr, or entering a review about a venue on Foursquare.

New applications have emerged [40] with these rising technologies, such as the suggestion of venues that might be of interest to a user. This task of venue suggestion has been mainly tackled in the literature using *user-centric* recommendation approaches that rely on collaborative filtering [3, 12, 38, 39], where suggestions are based on venues that were highly rated by users that have the same preferences as the current user (i.e. similar users). The main drawback of such approaches is that they require many users with detailed profiles to learn individual preferences. Moreover, they are not able to recommend venues to those users that are not members of the same social network. To overcome these limitations, instead of relying on similar users, some recent approaches directly model the likelihood of suggesting venues with *venue-centric* methods [16, 21], which focus on the features that make a venue relevant rather than on venues that similar users like. However, all the applications derived from these user- and venue-centric methods have been restricted to specific sub-problems of venue suggestion. These include the recommendation of only new venues [3] (venues in a city that the user has not visited before), the prediction of the next location of the user [16, 21], or the suggestion of only a certain type of venues [21] (e.g. restaurants). In addition, none of the aforementioned studies explored how the quality of their suggestions vary for different types of users (e.g. residents or visitors), or at different times of the day. More importantly, no study so far has reported results derived from the judgments of users.

In this paper, we address all of these issues by proposing a general venue-centric model, which suggests personalised venues at different times of the day. Along with the modelling of the user's preferences, the proposed model additionally predicts the popularity of a venue at a given time to provide the user with timely and high quality suggestions.

Our model is composed of two components. The first one estimates the popularity of a venue (an indicator of high quality) by measuring its level of attendance: the most popular venues are the most attended ones. Following recent work on estimating venues' attendance at different times of the day [25, 32], we use the data generated by the Foursquare users when they "check-in"¹ at venues to regularly measure

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CIKM'15, October 19–23, 2015, Melbourne, Australia.

© 2015 ACM. ISBN 978-1-4503-3794-6/15/10 ...\$15.00.

DOI: <http://dx.doi.org/10.1145/2806416.2806484>.

¹ A term used by Foursquare to denote users sharing their current location with the LBSN.

the venues’ level of attendance throughout the day. We then predict the future level of attendance of a given venue by applying state-of-the-art time series forecasting models, such as those in [18]. By doing so, we predict the future popularity of venues at a given point in time, thereby adapting the suggestions as a function of the date and the time of the day. Indeed, while harnessing the Foursquare venues’ popularity has already been studied [22], no past work attempted to estimate this popularity at future points in time.

The second component of our model takes advantage of the Facebook profile of the users to suggest personalised venues by inferring their interests. People indeed openly share pieces of their interests on Facebook by “liking” various pages, which can be viewed as records of their personality [20]. Our model uses the pages users like to estimate the likelihood of a venue being of interest to them. With nearly 1.5 billion users in 2014², Facebook is a central and widely used repository of user interests. Leveraging this centrality allows us to avoid the usual cold-start and sparsity problems that plague recommender systems [13].

While we use specific information sources in this study (i.e. Foursquare and Facebook), our model is general enough to be adapted to other sensors of attendance (e.g. CCTV cameras) or sources of user interests (e.g. Google+). Unlike in previous work, we evaluate the effectiveness of our venue suggestions by asking judgments from users. In particular, we conduct a user study, where we examine the quality of venue suggestions over various dimensions such as the time of the day or the profile of the user (i.e. visitor or resident). Hence, this paper presents a three-fold contribution:

- We propose a principled probabilistic model that suggests personalised venues based on their estimated interestingness and popularity over time;
- We separately evaluate the two components of our model by conducting experiments on: (1) the inference of the user’s interests; and (2) the prediction of the venues’ level of attendance by applying time series forecasting models;
- We integrate our observations and analysis from the above experiments to build an effective instantiation of our venue suggestion model, which we validate through a user study that involves 100 participants over three major cities. The experimental results from the user study show that: (1) Taken separately or in combination, both components of our model significantly outperform the Foursquare recommendation service, thus validating our approaches for estimating the popularity of a venue and the user’s interests; (2) New visitors and residents tend to prefer personalised venues, while former visitors prefer popular venues.

The remainder of this paper is organised as follows. Section 2 reviews approaches related to the task of venue suggestion. In Section 3, we describe our proposed model. We experiment in Section 4 with the two components of our model taken separately, before combining them and carrying a user study in Section 5. We conclude and discuss future work in Section 6.

2. RELATED WORK

Suggesting venues or activities to users is an emerging application that is currently receiving attention due to the growing interest in smart cities [6] and location-based social networks (LBSNs) [40]. These applications are often

classified in the literature as Point-of-Interest (POI) [28] (or also sometimes Location [38]) Recommendation. Approaches studied for POI recommendation range from traditional user-centric collaborative filtering [3, 12, 38, 39] to venue-centric topic modelling [16, 21]. User-centric approaches focus on the user and perform recommendation based on the venues that similar users have liked or visited in the past. One weakness of collaborative filtering is that it requires detailed profiles from many users to be able to accurately infer their interests. For example, Bao et al. [3] tackled the sparsity of the user profiles and the cold start problem in collaborative filtering by identifying experts in high-level categories of venues, which are estimated to be of interest to the user from previous interactions on the LBSN (previous visits). Unlike these approaches, we do not rely on previous interactions on LBSNs to model the users’ interests.

In contrast, venue-centric approaches to venue suggestion can abstract from the need of detailed profiles, and also offer the advantage of being able to accommodate a wide range of users by not being restricted to users of a single LBSN. Thus far, the venue-centric approaches have relied on probabilistic topic models [16, 21]. Kurashima et al. [21] showed that a topic model can effectively mine logs of GPS trajectories to infer the next visit of a user. Hu et al. [16] proposed a very similar model that uses text and locations extracted from social media (e.g. Twitter). However, the venue-centric approaches [16, 21] only predict the next location of the user, which may not be a venue, thus leading to a low suggestion precision. Moreover, while all of the aforementioned user- and venue-centric approaches relate to our work in that they can all recommend venues, they only consider two parameters: the users and (sometimes) their current location, but they do not consider time. Our approach is different in the sense that it models the quality of venues with respect to both their popularity and the user’s interests, and also tracks their popularity over time.

A few studies are closely related to ours. Yuan et al. [39] and Gao et al. [12] proposed time-aware recommendation models that incorporate the temporal check-in behaviour of each user in the LBSN. Both use collaborative filtering, and incorporate the similarity between users that exhibit similar temporal patterns (i.e. users that visit the same locations at the same time of the day). Although these two approaches take time into account for POI recommendation, they are both user-centric and thus suffer from the drawbacks mentioned earlier. Our approach distinguishes itself from those studies by being venue-centric, and by modelling the temporal behaviours of the venues instead of those of the users. Brilhante et al. [8] proposed an approach for mining Flickr and Wikipedia to build sequences of visits for users before they start their trip. However, this last approach only generates a static path of venues given a target city. In contrast, our proposed model adapts to the city, the time, the user, and her/his current location.

The TREC Contextual Suggestion track [10] is investigating research questions related to the personalised suggestion of places. The most effective approaches [11, 31, 37] heavily rely on the high quality search results of the recommendation APIs in various LBSNs such as Google Places, Yelp, or Foursquare. Achieving a personalised ranking is then mostly a matter of re-ranking those results to better fit the user’s interests. In the context of the Contextual Suggestion track, these interests are expressed through a list of previously visited places associated with ratings, as for user-

² <http://www.statisticbrain.com/facebook-statistics>

centric approaches. One issue when considering such profiles is that suggestions will always be biased towards categories of venues that the user rated highly, thereby harming the exploration or discovery of new kinds of venues. Again, our approach does not require the user to have a detailed profile of previously visited venues.

The approach we present in this paper differs from all of the above in the sense that it tracks the popularity of venues over time, and proposes a novel method for inferring the interests of the user by using an independent source of preferences (i.e. Facebook). Our proposed model is principled and flexible enough that it could be instantiated using different approaches to represent the popularity and the interestingness of a venue. It is specifically designed to tackle the *zero-query retrieval* task, where suggestions are presented to the user without requiring her/him to issue a query [23].

We present this model and detail the specific instantiations we explore in this paper in the following section.

3. MODEL FORMULATION

For a given user u at a location l and at a certain time t , we aim to model the likelihood of a venue v to be of interest to the user. This model can be formally written as:

$$P(v|u, l, t) \propto P(v|u) \cdot P(v|l, t') \quad (1)$$

and is composed of two components: an *interestingness* component, which estimates the likelihood $P(v|u)$ that a venue v would correspond to the interests of user u ; and a *popularity* component, which estimates the likelihood $P(v|l, t')$ that v would be accessible to the user currently at location l , and that it would be popular at some point in the near future t' , where $t' = t + \epsilon$. In other words, our model can suggest venues that may be of interest in the near future by anticipating the needs of the user. The parameter ϵ controls the length of this “near future”; we will discuss its setup in Section 5. Note that the model in Equation (1) assumes independence between the user u and the location and time (l, t) . This assumption simplifies the model allowing us to study the interestingness and the popularity components, independently or in combination, by making use of judgments from users. It is also consistent with a venue-centric approach that does not require data detailing the locations of the users at different times. Our model is a venue-centric approach in the sense that it models the likelihood of a venue to be of interest given three parameters, namely the user, its location, and the current time. Venues are modelled both in a conceptual space (finding venues that match the interests of a user) and in a physical space (finding nearby venues that are usually popular at certain points in time). The core of our approach lies in the estimation of the two probabilities of Equation (1), which we describe in the following sections: first, we detail in Section 3.1 our approach for predicting the popularity of a venue at a certain point in the future; then, in Section 3.2 we describe how we can leverage elements of the Facebook profile of a user to infer her/his interests.

3.1 Predicting venue popularity

We now address the issue of estimating the popularity of a venue as expressed by $P(v|l, t')$. In our model, this popularity acts as a surrogate for the quality of the venue suggestion for the user, and depends both on the time and on the user’s current location. The quality of a venue would for example decrease if it is far from the current location of the user. We safely assume that a venue will not become popular

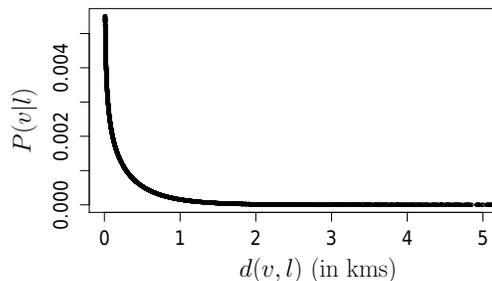


Figure 1: Probability that users check-in at a venue given its distance to their own location. The x-axis represents the distance (in km) between the venue v and the user’s location l .

just because the user is nearby. We thus consider the current location l of the user to be independent of the time t' . The right-hand term of Equation (1) can then be written as:

$$P(v|l, t') \propto P(v|l) \cdot P(v|t') \quad (2)$$

We detail in the two following sections how we model both the venue proximity $P(v|l)$ and the likelihood $P(v|t')$ that the venue will be highly attended in the near future.

3.1.1 Venue proximity

It has been shown in previous studies [30, 38] that users tend to visit venues that are near their current location. Ideally, the probability $P(v|l)$ would thus be high when the distance $d(v, l)$ between the venue v and the user’s location l is short. To do so, we look at pairs of Foursquare venues at which the user has checked in over short periods of time (i.e. less than three hours), and fit a Gamma distribution on this distance distribution. The resulting probabilities, expressed as a function of the distance $d(v, l)$ between the venue v and the user’s location l , are displayed on Figure 1. Finally, the probability that a venue v would be of interest given the current location l of the user is expressed as: $P(v|l) = P(d(v, l))$. Note that evidence other than distance can be used for estimating $P(v|l)$, for example search logs [4], but we leave the use of this additional evidence for future work.

3.1.2 Attendance forecasting

In our model, the popularity of a venue is mainly related to its predicted level of attendance. We follow the ongoing initiatives in the development of smart cities and use social sensors [1, 33] to construct a time series of attendance for each individual venue. Time series are numerical information that are observed sequentially over time. In this work, we use the Foursquare API³ as a social sensor. When asking for information about a venue, the API permits to obtain the number of people currently visiting the venue. We query the API every hour for each venue to build a comprehensive time series of venue attendance. It is important to note that we do not depend on using Foursquare as a particular type of sensors. Indeed, any other kind of sensors (or combination of sensors) that can measure crowd density could be used here, such as CCTV cameras for example [34]. Since we focus on predicting popularity, we build a time series of attendance for each venue and rely on three state-of-the-art time series forecasting algorithms [24]. Time series forecasting encompasses a family of methods that can predict not only the value of the next point of the time series, but the values of the next N points. State-of-the-art forecasting algorithms

³ <http://developer.foursquare.com/docs/venues/herenow>

such as ARIMA (Auto-Regressive Integrated Moving Average), Neural Networks, and Exponential Smoothing use past observations to learn trends, seasonal variations, and recurring patterns in the data [18, 24].

We use these forecasting models to estimate the probability $P(v|t')$ in Equation (2). We apply Bayesian smoothing with a Dirichlet prior to avoid zero probabilities:

$$P(v|t') = \frac{\hat{y}(v, t') + \mu \cdot P(v|\mathcal{T})}{\sum_{v' \in \mathcal{V}} \hat{y}(v', t') + \mu} \quad (3)$$

where $\hat{y}(v, t')$ is a forecast of the number of people attending the venue v at time t' , using either ARIMA, Neural Networks, or Exponential Smoothing as forecasting models. \mathcal{V} is the complete set of venues for a given city. $P(v|\mathcal{T})$ models the popularity of the venue v over an entire timespan \mathcal{T} of past observations, where \mathcal{T} also represents the training period of the forecasting models. This background probability is expressed as $P(v|\mathcal{T}) = \frac{y(v, \mathcal{T})}{\sum_{v' \in \mathcal{V}} y(v', \mathcal{T})}$, where $y(v, \mathcal{T})$ is the total number of people that attended the venue v during \mathcal{T} .

The Dirichlet prior μ is the mean of the counts for all venues in the same city, over the entire timespan \mathcal{T} . The rationale behind applying such a smoothing is analogous to traditional information retrieval: we avoid eliminating the candidate venues with 0 attendances. Without any smoothing, the probability $P(v|t')$ is null when $\hat{y}(v, t')$ is predicted to be 0. However, even if a venue has a low level of attendance, it may still be interesting to some users. For example, we observed in our data that a board game shop is a very unlikely venue to check-in; however, a person who likes such games would be highly interested by this kind of suggestions. Hence, smoothing this probability helps us to tackle the inevitable sparsity of our observations.

3.2 Modelling venue interestingness

The second component of our approach models the interests of the user in order to personalise the suggestions. Users supply a wealth of information about their personality on social networking websites. With its ever growing number of users, Facebook is one of the most popular publicly available personality repository [20]. On Facebook, people can explicitly share their interests by “liking” pages that refer to entities such as public figures, bands, venues, events, or even fictional characters. The personality of users can then be represented as sets of entities covering their various interests [20]. We leverage this entity representation in a probabilistic model, which extracts the interests of a user with the aim of suggesting personalised venues. More specifically, we compute the probability $P(v|u)$ that a venue fits the interests of user u by marginalising over the space of all known entities of the entire Facebook Graph⁴ (with the fair assumption of independence between the user and an entity):

$$P(v|u) = \sum_e P(v|e)P(e|u) \quad (4)$$

We reasonably hypothesise that a “like” is an indicator of relevance, hence a page e will always be relevant for a user u if that user has liked it in the past. Therefore, we define in this work $P(e|u) = \frac{\text{like}(e, u)}{\sum_{e'} \text{like}(e', u)}$, where $\text{like}(e, u) = 1$ if the user u likes the entity e , and 0 otherwise.

Estimating $P(v|e)$ is thus the core part of this component. Since venues and entities do not share the same space, we

map them into a common space of categories and apply machine learning. We detail in the remainder of this section the categorisation and classification approaches we employ.

3.2.1 From interests to categories

Our approach for estimating the probability that a venue is relevant given an entity relies on the categories of the DMOZ⁵ Open Directory Project, a curated directory of web pages that have been manually categorised. This allows us to represent both venues and Facebook entities as sets of categories, thus enabling their comparison to measure how well a venue corresponds to the interests expressed by an entity. Bi et al. [5] recently proposed a method that relies on DMOZ to map Facebook profiles with search profiles in the Bing search engine, using proprietary search logs. However this previous work is limited in that only the number of similar categories are counted, which we overcome by directly computing probabilities. Moreover, Bi et al. only predicted specific personality traits such as the age, the gender, or the political views of the users. Here, we propose a method that can match two types of entities of very different natures, e.g. a Facebook page and a venue, by categorising each in a space of DMOZ categories. Formally, given \mathcal{C} , the set of DMOZ categories, we rewrite the expression of $P(v|e)$ used in Equation (4) as (with the fair assumption of independence between the entity and a category):

$$P(v|e) = \sum_{c \in \mathcal{C}} P(v|c)P(c|e) \quad (5)$$

Using Bayes’ rule, $P(v|c)$ can be expressed as:

$$P(v|c) = \frac{P(c|v)P(v)}{P(c)} \propto \frac{P(c|v)}{P(c)} \quad (6)$$

As we do not consider any prior information about venues in this component of the model, $P(v)$ is constant and does not affect ranking and can be removed from Equation (6).

We can then rewrite Equation (5) as:

$$P(v|e) \propto \sum_{c \in \mathcal{C}} \frac{1}{P(c)} P(c|v)P(c|e) \quad (7)$$

Estimating $P(v|e)$ comes down to estimating the category distribution $P(c|\bullet)$, where $\bullet \in \{v, e\}$. We cover this estimation in the following section.

3.2.2 Learning to categorise

Our approach to entity and venue classification builds upon recent work in query classification. In particular, predicting the categories of queries has been studied in the past, and successful approaches agreed on the fact that context is needed [5, 36] to tackle the sparsity problem. This context is usually obtained by retrieving documents that are topically relevant to the query, and classifying the top results into a predefined set of categories. We take the same approach to predict the category distribution of venues and Facebook entities. We train three state-of-the-art classifiers using web pages previously labelled with the top-level DMOZ⁶ category information. Venues and Facebook Likes can be entities of very different types, we thus intentionally consider the highest level of the DMOZ hierarchy to represent these entities in a general and diverse category space. We show

⁴ <http://facebook.com/about/graphsearch>

⁵ <http://dmoz.org> ⁶ The DMOZ hierarchy has 16 top-level categories. Our training data has a total of 13 categories as it does not contain “Kids and Teens”, “World” and “Regional”.

in Section 4.2 that using the highest-level categories of the DMOZ hierarchy is valid for this task. Future work could however address more refined taxonomies.

Like Bi et al. [5], we transform each Facebook entity e into a query by using its title. Similarly, for each venue v , we concatenate its name and its city to provide some disambiguation, and use this as a query. Let $\mathcal{R}(\bullet)$ be the set of top search results using \bullet as a query. The retrieval model can be of any kind, as long as the retrieved documents are topically relevant to \bullet . Each retrieved document of this set is then classified, yielding a probability distribution over the DMOZ categories. More formally, the probability $P(c|\bullet)$ that \bullet would be of category c is expressed as:

$$P(c|\bullet) = \frac{1}{|\mathcal{R}(\bullet)|} \sum_{D \in \mathcal{R}(\bullet)} P(c|D) \quad (8)$$

where $|\mathcal{R}(\bullet)|$ is the number of retrieved documents. These documents are retrieved from a general repository of web pages that we detail in Section 4.2. The probabilities $P(c|D)$ are directly derived from the predictions of the classifier, which models the category distribution of unseen documents. Different retrieval models and classifiers can be deployed to estimate $P(c|\bullet)$, as we detail in Section 4.2.

We rewrite Equation (7) using Equation (8) as:

$$P(v|e) \propto \sum_{c \in \mathcal{C}} \frac{1}{P(c)} \left(\frac{1}{|\mathcal{R}(v)|} \sum_{D_v \in \mathcal{R}(v)} P(c|D_v) \times \frac{1}{|\mathcal{R}(e)|} \sum_{D_e \in \mathcal{R}(e)} P(c|D_e) \right) \quad (9)$$

It is important to note that our approach could be instantiated using other sources of user preferences in lieu of Facebook (e.g. Google+) to infer the user’s interests. We could also imagine a combination of social networks to improve the coverage of the user personality. We leave these improvements for future work.

4. EXPERIMENTS

As detailed in the previous section, our model is built upon two principal components that model the popularity and the interestingness of a venue. In this section, we evaluate the two components separately before carrying a user study of the entire model in Section 5. More specifically, we seek to answer the following research questions:

RQ1: What is the best time series forecasting algorithm for predicting venue attendance levels? (Section 4.1)

RQ2: Which choice of retrieval model and classifier enables to effectively model the categories of users’ interests? (Section 4.2)

4.1 Forecast accuracy evaluation

We describe in this section the dataset, the measures, the baselines, and the results of our experiments in order to answer our first research question.

4.1.1 Cities dataset

We chose three major cities to evaluate our model and answer our first research question: London, Amsterdam, and San Francisco. London is a major capital with a large population (more than 8 million). Amsterdam has a higher density and counts more than 800,000 inhabitants. We completed our test bed with San Francisco in order to see if the

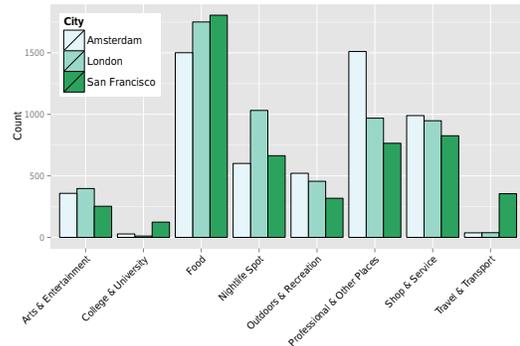


Figure 2: Distribution of venues over their Foursquare categories for the three cities we consider in this study.

observations made on two European capitals generalise to a major American city.

We used 3 different Foursquare API accounts belonging to the researchers involved in this study to measure the levels of attendance of venues. The Foursquare API limits the number of calls per account to 5,000 per hour, thus adding up to 15,000 possible API calls per hour. We reasonably limited the frequency of obtaining the levels of attendance to one measurement per hour for each venue. This rate allowed us to both accurately keep track of each venue’s attendance while considering a large amount of venues. We equally divided the API calls across the three cities, and monitored 4,950 venues for each city⁷ from 5th November 2013 to 17th January 2014. However, tens of thousands of Foursquare venues are available for these cities. Initial experiments showed that the 3,000 venues with the highest total number of check-ins contained the majority of the most popular venues for the three cities. We then performed a selection of venues by choosing the 3,000 ones with the highest number of check-ins, and randomly selected the remaining 1,950 venues from those that had at least 25 check-ins at the time we began our experiments. This threshold of 25 check-ins was used to ensure that we did not include fictitious venues. While performing this selection, we omitted venues falling under the Foursquare categories “College & University” and “Travels & Transport”. While airports and train stations for example are very popular among Foursquare users (who like to check-in at these locations to show that they are travelling), they may not be useful suggestions. Some Foursquare venues can belong to multiple Foursquare categories, hence we kept venues that belong to the two blacklisted categories above if they also belong to other (whitelisted) categories. Overall, this selection policy allowed us to build a diverse dataset that includes the most popular venues, which might be of interest to visitors, while also containing less well attended venues. Statistics on this dataset are available in Figure 2, which shows the number of venues for each category in each city.

4.1.2 Forecast accuracy measures and baselines

We detail here the baselines and evaluation measures we use to answer RQ1. Forecast accuracy is determined by computing the errors of the forecasts with respect to the real observations. We report the results in terms of *root mean squared error* (RMSE), which is a commonly used er-

⁷ The last 50 API calls were used as a safety buffer in case of an intermittent unavailability of the Foursquare service.

ror measure⁸. As already mentioned in Sections 2 and 3.1.2, we explore the effectiveness of three forecasting models. We compare them to 3 widely used baselines that proved to be competitive [24]. The first baseline, denoted by Mean, forecasts all future values for a given venue to be the mean of the past observations. The second one is called Naïve, for which all forecasts are simply set to be the value of the last observation. It was shown to be very effective in the context of economic and financial time series forecasting [24]. The last one is the Seasonal Naïve method, with the forecast being equal to the previous day. All the time series forecasting models were built using the well-known `forecast` package of R [17].

4.1.3 Results

Research question RQ1 aims to explore the accuracy of time series forecasting algorithms for predicting the venue attendance levels. To this end, we evaluate daily forecasts by predicting the attendance of each of the venues for the following 24 hours. In this case, all the days prior to the prediction can be considered as training data. We forecast the attendance of each venue for 10 randomly chosen days (while allowing at least 21 days for training), then average their accuracy. The results are reported in Table 1, where we evaluate the three widely used forecasting models: Neural Networks (Neural), ARIMA, and Exponential Smoothing (Smooth.) with respect to three baselines commonly used in forecasting studies [17]: Mean, Naïve, and Seasonal Naïve (S. Naïve). We vary the number of training days between 7 and 21, and we report the averaged RMSE for all venues in each city. Since RMSE measures a forecast error, the best results are the lowest numbers.

On analysis of Table 1, we see that the Mean method is the most effective method when trained using a small number of observations (7 days). However, with more historical data, ARIMA achieves the best accuracy. The good results of the very basic Mean baseline can be explained by the sparsity of the observations. When building time series of attendance, we consider that a venue is empty if no Foursquare user has checked-in for the current hour. This can result in time series that are mostly composed of zeros for venues that have low levels of attendance, or that attract few Foursquare users. This explains the high results of the Mean baseline, which predicts very low levels of attendance for unattended venues. However, state-of-the-art forecasting models improve their accuracy when we use more historical data, as they are able to model daily and weekly trends even for venues with very few check-ins. ARIMA achieves almost the best results for all cities when using more training days, and Exponential Smoothing models are on par with the three baselines. In contrast, we see that Neural Networks tend to over-fit and hence cannot cope with large daily variations. Nevertheless, we conclude that state-of-the-art time series forecasting models are in general able to outperform these competitive baselines, unless they have a limited amount of training data.

To summarise, in order to address RQ1, we conducted the first systematic series of experiments on the evaluation of forecasting models for predicting the levels of attendance of the venues. Moreover, the forecasting accuracies are very similar across three different cities, which is promising for other potential applications in other cities. It should be

⁸ We noticed in our preliminary experiments that RMSE was highly correlated with other common error measures (such as MAE or MASE).

Table 1: Accuracy of the three used forecasting models compared to the three standard baselines. We report the average RMSE for all the venues of each city, varying the number of training days. Numbers in bold are the best results per city and per number of training days.

	7 days			21 days		
	Lon.	Amst.	S. Fran.	Lon.	Amst.	S. Fran.
Mean	0.633	0.563	0.868	0.533	0.357	0.794
Naïve	0.785	0.686	1.131	0.670	0.440	1.015
S. Naïve	0.784	0.696	1.020	0.651	0.446	0.879
Neural	0.822	0.739	1.107	0.796	0.543	1.093
ARIMA	0.677	0.590	0.908	0.514	0.353	0.737
Smooth.	0.787	0.701	1.108	0.640	0.444	1.031

noted that despite the computational requirements of the state-of-the-art forecasting models, they are still suitable for practical deployments, since the forecasting process is done offline on a regular basis and not in real-time.

In the user study we conduct in Section 5, we use the ARIMA models built from the past 21 days as the instantiation of the *popularity* component of our model.

4.2 Evaluation of models of user interests

We now answer RQ2 by evaluating the component that models the user’s interests. As described in Section 3.2, this component is a two-step approach, where it first retrieves documents related to a Facebook entity, before classifying them into DMOZ categories. We detail in this section the retrieval models and the classifiers that we use, our evaluation methodology, and the results of this experiment.

4.2.1 Classifiers and retrieval models

As already presented in Section 3.2, our method for estimating the interests of a user has two sequential steps: the retrieval of documents related to a Facebook entity, and their classification into 13 DMOZ categories. We detail here the different retrieval models and classifiers we experimented with in order to answer our second research question RQ2.

The retrieved documents come from the category B of ClueWeb09, a corpus of 50M web pages. We use the Terrier IR platform [26] for indexing and retrieval. We experimented with three different well-known retrieval models using their default settings in Terrier: Language Modelling (LM) [15], BM25 [29] and the DPH model [2] (an effective parameter-free model from the Divergence From Randomness framework). It is well-known that ClueWeb09 contains many spam documents, which we overcome by using the Waterloo spam list [9] that provides spam percentiles for all documents. We follow [9] and only retain documents with a percentile score higher than 70.

We use the Weka machine learning software [14] for training and classifying web pages, and test three different classifiers. We take a traditional text classification approach and consider word frequencies as features. Our baseline classifier is the well-known Naïve Bayes [19]. We also experiment with two other state-of-the-art classifiers: Random forests [7] and Decision trees (J48) [27]. Each classifier is trained using 5,000⁹ randomly selected examples from 197,125 documents that were previously labelled in the DMOZ categories.

⁹ Increasing the number of training examples only resulted in reducing efficiency and had no significant effect on classification accuracy.

Table 2: Results for modelling user interests, expressed by the Jensen-Shannon (JS) divergence between the estimated and the reference DMOZ category distributions. We use a two-sided pairwise t-test (with $p < 0.05$) to determine statistically significant differences over: (1) a random baseline that generates random category distributions, and that has a JS divergence of 16.038 (Δ and ∇), and (2) the Naïve Bayes baseline (Δ and ∇).

Classifier	Retrieval Model	Number of retrieved documents					
		5	10	20	30	40	50
Naïve Bayes [19]	DPH [2]	12.846	5.465 Δ	3.113 Δ	2.591 Δ	2.419 Δ	2.002 Δ
	BM25 [29]	9.189 Δ	3.922 Δ	2.693 Δ	2.041 Δ	1.284 Δ	1.262 Δ
	LM [15]	51.530 ∇	32.083 ∇	16.451	10.658	8.776 Δ	6.457 Δ
Random Forests [7]	DPH	13.061 Δ	11.424 $\Delta\nabla$	8.237 $\Delta\nabla$	5.052 Δ	2.804 Δ	2.316 Δ
	BM25	13.405 $\Delta\nabla$	11.386 $\Delta\nabla$	10.351 $\Delta\nabla$	8.698 $\Delta\nabla$	7.437 $\Delta\nabla$	1.803 Δ
	LM	23.659 $\nabla\Delta$	17.588 Δ	15.082	15.073	14.390	13.469 $\Delta\nabla$
Decision trees (J48) [27]	DPH	3.532 $\Delta\Delta$	1.316 $\Delta\Delta$	1.293 Δ	0.577 $\Delta\Delta$	0.542 $\Delta\Delta$	0.540 $\Delta\Delta$
	BM25	1.367 $\Delta\Delta$	0.534 $\Delta\Delta$	0.498 $\Delta\Delta$	0.499 $\Delta\Delta$	0.480 $\Delta\Delta$	0.489 $\Delta\Delta$
	LM	0.441$\Delta\Delta$	0.435$\Delta\Delta$	0.431$\Delta\Delta$	0.408$\Delta\Delta$	0.398$\Delta\Delta$	0.406$\Delta\Delta$

4.2.2 Gold standard and evaluation measure

The nine combinations of classifiers and retrieval models that we described above all output a probability distribution over the DMOZ categories that estimate the interests of the user. In order to quantify the ability of this approach to approximate interests as per RQ2, we first need users to explicitly state their “true” interests in terms of the DMOZ categories. We obtained this gold standard by using the Amazon Mechanical Turk¹⁰ crowdsourcing service. We presented workers with the list of the 13 DMOZ categories and asked them the following question: “*What are your general interests?*”. They had to rate each of these categories in order to answer this question. Ratings ranged from 0 (not interested at all) to 10 (extremely interested). We paid users \$0.05 for this task, which was completed in 32 seconds on average. We asked users to connect with their Facebook account, using Facebook’s software development kit¹¹, in order to obtain the titles of the pages they like while they were performing the task. The users were explicitly asked to authorise access to the titles of pages they like on Facebook and nothing else. This login step also allowed us to entirely prevent our task from being completed by robots or malicious workers. We obtained a total of 25 manually entered user profiles, which constitute our gold standard.

We then converted the ratings of this gold standard into a reference probability distribution over categories, in order to be able to compare this reference to our automatically generated distributions using statistical distances. Categories from which users attributed a high rating received a high probability, and *vice versa*. In order to measure the statistical distance between our estimated category distribution and the reference distribution, we used the Jensen-Shannon (JS) divergence, which is the symmetric version of the well-known Kullback-Leibler divergence. We use this measure as it is a natural and principled choice for dealing with and comparing probability distributions, and quantifying their differences. We also experimented with the Kendall’s τ rank correlation coefficient, from which we made similar observations.

4.2.3 Results

We experiment with the three classifiers and the three retrieval models presented earlier, and vary the number of retrieved documents for each Facebook entity used as a query. For each of these settings, we compute the average JS divergence across the estimated profiles and the crowdsourced

references profiles. We report the results of these experiments in Table 2. As the numbers represent divergences, lower are better. As a simple baseline, we randomly generate the category distributions, which achieves a JS divergence of 16.038 (not shown in Table 2).

On analysis of Table 2, we see that increasing the amount of retrieved documents always improves the quality of the estimation, which stabilises around 40-50 documents for a majority of settings. The DPH and BM25 retrieval models have similar behaviours with all classifiers: using BM25 allows to achieve marginally better results than DPH, with one exception for random forest. Language models are ineffective when used in conjunction with Naïve Bayes and Random Forest, but achieve the best results when the retrieved documents are classified with decision trees. Using decision trees for this task significantly outperforms both the random and Naïve Bayes baselines for all settings (according to a pairwise t-test at $p < 0.05$), except for DPH when using 20 retrieved documents.

To summarise, in this section we conducted a comprehensive experiment, which revealed that our approach can automatically quantify the interests of a Facebook user from the pages she/he likes. In answer to our research question RQ2, we conclude that our approach performs best using language models (LM) to retrieve 40 retrieved documents that are classified by decision trees (J48); we adopt this setting to instantiate our model for the following user study. While the evaluation in this section focused on estimating the interests of the users without considering the venues, in the next section, we evaluate the effectiveness of the personalised venue suggestions through our user study.

5. USER STUDY

We now conduct our user study, where we obtain judgments from users (recruited volunteers), to measure the effectiveness of our model for *time-aware* and *personalised* venue suggestion. Specifically, we seek to answer three further research questions:

RQ3: Is our model effective with respect to a state-of-the-art venue suggestion baseline?

RQ4: Is our model effective for different times of the day?

RQ5: Is our model able to accommodate different types of users (such as visitors or residents)?

To answer these questions, we developed a web interface for assessing venue suggestions and asked 100 volunteers (using word-of-mouth and social networks) from across the

¹⁰ <http://mturk.com/mturk> ¹¹ <http://developers.facebook.com/docs/facebook-login>

world to judge these suggestions. Participants had to judge the relevance of a set of venues with respect to three conditions: their own interests, their location in the city and the time of the day. They first had to connect with their Facebook account, so that our model could infer their preferences. As in the experiment in Section 4.2, the participants were explicitly asked to authorise access to the titles of pages they like on Facebook. Then, they had to choose a city: (London, Amsterdam, or San Francisco). The presumed location of the participant during these assessments was randomly chosen within the city they selected. We used the local time of the participants as the time of the day, and set the ϵ parameter (see Section 3) to 1 hour. We specifically chose this value for two reasons. First, the terms of use of the Foursquare API prevented us from choosing a smaller ϵ without reducing the total number of venues in our dataset. Second, we assume that 1 hour is a reasonable duration for a user to walk in a busy area of a major city, while still having the time to visit a few venues. We further clustered the assessments from different times of the day by “time periods” that represent the main parts of the day. The “morning/lunch” part includes assessments performed between 7am and 1pm, the “afternoon” one focuses on the time period between 2pm and 6pm, and the “evening/night” closes the day by covering assessments between 7pm and 2am. Before displaying the suggestions to be judged, we asked participants for their previous knowledge of the city.¹² We divided this knowledge into three levels, allowing us to explore the effectiveness of our model for different types of users: people who have never been, people who have already visited, or people who live/have lived in the city. We display in Table 3 a summary of the 100 total participants broken down by city, time period, and knowledge. As said earlier, we left the choice of the city to the participants, who showed a high preference towards London. We thus take London as our main test city, and consider Amsterdam and San Francisco as additional cities in order to show generalisation. We also observe that a majority of the participants had already visited the city they chose, while a lower number of residents participated in our study. However, as we will see in the results below, the improvements over the baselines were consistent enough so that we could still identify statistical differences.

After the user had selected a city, the ranking lists of venue suggestions were computed using five different algorithms: Nearby (N), Foursquare (F), Popularity (P), Interestingness (I), and our Full model (Full). Nearby is a simple baseline that selects venues that are located near the user within a 500 meters radius (this value has been set according to the probability density function depicted in Figure 1), and ranks them randomly. Foursquare is a competitive, but a blackbox baseline that uses the API of the Foursquare LBSN, i.e. it uses the same data. It uses its own data to recommend venues near a given location¹³. Since it is currently an industrial application, we do not know the implementation details of the Foursquare recommendation algorithm. However, considering the current leadership role of Foursquare among LBSNs, we reasonably use its recommendation service as a state-of-the-art venue suggestion system. Since less than 5 of the participants had a

¹² Due to the limited space, we have not provided screenshots of the interface used by the recruited volunteers. To preserve anonymity, we will provide a URL to the interface when the paper is published.

¹³ <http://developer.foursquare.com/docs/venues/explore>

Table 3: Number of participants for the three cities, as a function of their level of knowledge of the city and of the time of the day.

		London	Amst.	San Fran.	Total
Participant knowledge	never been	20	10	7	37
	visited	33	10	9	52
	live/lived in	10	1	0	11
Total		63	21	16	100
Time period	morning/lunch	16	3	1	20
	afternoon	28	12	9	49
	evening/night	19	6	6	31

Foursquare account (which would be required to obtain personalised suggestions from Foursquare), these suggestions were not personalised. The (Full) system is the instantiation of the model we propose in this paper as expressed in Equation (1). The (P) and (I) variants aim at exploring the influence of the two components of our model on suggestion effectiveness. For (P), venues are ranked based on their level of forecasted attendance only and according to the following equation: $P(v|u, l, t) \propto P(v|l, t)$. The (I) model is the opposite of (P) and only considers venues that match the user’s interests. In other words, venues are ranked according to $P(v|u, l, t) \propto P(v|u)$ for this model. Details on these probabilities are available in Section 3.

We retained the 10 venues with the highest scores for each of these algorithms. The union of these rankings constitutes the pool of venues to be judged. The average size of the pool for all users is 38.31 venues. We randomised the order of the venues in the pool so that participants could perceive changes of quality in the suggestions. The web interface sequentially presented participants with a single venue, asking them to judge it with respect to the time, the location, and their own interests. Each venue was located on a map, which also shows the participant’s location. Before assessing, the participants were presented with a paragraph that detailed the situation (city, exact location, time of the day), and were provided with instructions for judging the relevance of the venue. We also displayed a description of the venue, which included a photo, a list of categories, and a link to the Foursquare homepage of the venue. Participants were then asked to judge on a 3-point scale how likely they would visit the venue (not likely, likely, highly likely). Once an assessment is submitted, the next venue in the randomised pool is displayed. At the end of the study, participants had the opportunity to enter free comments in a dedicated text box¹⁴. We believe that obtaining user judgments on venues by asking the users to imagine that they are in a certain location of the city is realistic and indeed it has been used for evaluating venue suggestions in the TREC Contextual Suggestion track [10]. We follow the evaluation methodology of previous work in venue suggestion [38, 39] and report performances in terms of precision and nDCG after ranking 10 suggestions.

The first results, reported in Table 4, are averages over the 100 participants. We see that each of the three instantiations of our model outperforms the two baselines (N) and (F) in terms of P@10 and nDCG@10 by a statistically significant margin ($p < 0.05$, according to a pairwise t-test). While (I) gave the best suggestions according to the participants, (P) and (Full) achieve comparable results. The

¹⁴ We received a lot of positive feedback from the participants of our study, who noted the promising applications and thanked us for making them want to travel.

Table 4: Average effectiveness results over all the participants of our study. α and β represent statistically significant improvements (using a two-sided paired wise t-test) over the Nearby and Foursquare baselines (with $p < 0.05$), respectively.

		P@10	nDCG@10
Baselines	Nearby (N)	0.359	0.327
	Foursquare (F)	0.393	0.345
Our approaches	Popularity (P)	0.466 α,β	0.393 α,β
	Interestingness (I)	0.480α,β	0.408α,β
	Full model (Full)	0.459 α,β	0.389 α,β

Table 5: Effectiveness results for the 3 different cities. Notations are identical to those in Table 4.

	London		Amsterdam		San Francisco	
	P@10	nDCG@10	P@10	nDCG@10	P@10	nDCG@10
(N)	0.386	0.355	0.310	0.265	0.313	0.284
(F)	0.398	0.346	0.467 α	0.387 α	0.267	0.273
(P)	0.475 α,β	0.394 α	0.486α	0.427α	0.400 α,β	0.337
(I)	0.514α,β	0.453α,β	0.410 α	0.295	0.433α,β	0.376β
(Full)	0.492 α,β	0.401 α,β	0.405 α	0.375 α	0.393 α,β	0.354 β

Foursquare recommendation service (F) achieves a low performance, even when compared to (N). The full model (Full) achieves a 17% improvement in terms of P@10 over the baseline (F), while (I) achieves a 22% improvement (respectively +13% and +18% in terms of nDCG@10).

We report in Table 5 the suggestion effectiveness results when looking at each city separately. We note that the results are similar to those in Table 4. However, Foursquare achieved similar results to (P) for Amsterdam, and slightly outperformed (I) and (Full). For this particular city, (F) was able to provide participants with diversified suggestions of either “Food”, “Nightlife Spot”, “Arts & Entertainment”, or “Outdoors & Recreation” venues. (F) failed to do the same for London and San Francisco, where 64.8% of its suggestions were “Food” venues. From the observations we make in Tables 4 and 5, and in answer to RQ3, we find that (P), (I), and (Full) are effective and consistently outperform Foursquare in terms of P@10 and nDCG@10.

We explore in Table 6 the quality of the suggestions for different time periods. When looking at the “morning” row, we see that all methods achieve similar results. Often, people’s activities are very personal during this period of the day and highly depend on their activities and the time at which they start their day. While one user may want to have a breakfast, another user may already have visited a museum or might be at her/his office. By suggesting personalised venues, (I) hence achieves slightly better results than the other methods. Moreover, it is interesting to see that (P) achieves the best results on the evening. When looking at the judgments, we see that 32% of the “evening/night” relevant suggestions belong to the “Nightlife Spot” category, 44% belong to “Food”, and 10% belong to “Arts & Entertainment”. The fact that 86% of the relevant suggestions are evening-related activities (even if the food venues are popular for the entire day), along with the high performance of (P) for this time period, demonstrates the validity of our forecasting approach for timely promoting high quality venues. As a summary and in answer to RQ4, we find that venue suggestion is a time-dependent problem and that our model is able to cope with time variations by providing a good balance between mornings and evenings.

Table 6: Effectiveness results for 3 different time periods. Notations are identical to those in Table 4.

	morning/lunch		afternoon		evening/night	
	P@10	nDCG@10	P@10	nDCG@10	P@10	nDCG@10
(N)	0.430	0.319	0.331	0.339	0.355	0.314
(F)	0.480	0.423	0.344	0.303	0.413	0.361
(P)	0.500	0.474α	0.427 α,β	0.359 β	0.503α,β	0.392 α
(I)	0.555α	0.393	0.446 α,β	0.400α,β	0.484 α,β	0.430α,β
(Full)	0.455	0.379	0.450α,β	0.400α,β	0.474 α,β	0.378 α

Finally, Table 7 explores the effectiveness of the venue suggestions for users with various prior knowledge of the city. The results show that while Foursquare is rather effective for people that never visited the city, it is clearly ineffective for residents or former visitors. On the other hand, our model is effective for all types of users. The results achieved by (I) when suggesting venues to residents (i.e. “live/lived in”) are interesting. We observe that this class of users are not interested in popular venues and rather want to go to venues that suit their interests. The venues that residents labelled as relevant have a total number of check-ins of 2707 on average, while visitors are attracted by more popular venues (3419 and 3587 check-ins on average for new and former visitors, respectively). In answer to RQ5, we find that our model is able to accommodate different types of users with suggestions of both personalised and popular venues.

6. CONCLUSIONS

In this paper, we proposed a venue-centric and time-aware model that suggests venues based on their estimated popularity and interestingness to the user. We presented an application of time series forecasting algorithms, and an approach for categorising the interests of a user by using the Facebook pages she/he likes. We evaluated these two different components separately using a dataset of 15000 Foursquare venues spread across three cities. The experimental results showed that state-of-the-art time series forecasting models, especially ARIMA, proved to be effective at predicting the levels of venue attendance, although highly dependent on the amount of training data. By crowdsourcing the categorised Facebook user profiles, we demonstrated that users’ interests can be inferred using the title of the Facebook pages they like as queries and classifying the top retrieved web documents using decision trees. Despite evaluating on Facebook, we argue that our model can generalise to other sources of user interests. For example, “liked entities” can be replaced with URLs or entities from searching/browsing logs.

Through a user study involving 100 users, we evaluated our complete probabilistic model. We also saw through our user study that our model was effective in different cities, at different times of the day, and for different types of users. Moreover, we showed that the level of attendance of a venue is a good indicator of its popularity. Residents were found to prefer personalised suggestions, while participants preferred highly popular venues during the evening. Hence, as future work, we plan to address the problem of balancing the influence of the two components in function of different parameters, such as the knowledge of the user or the category of the venue (e.g. Food, Shop). We also aim to reuse the judgments of our user study to learn this trade-off, for example by applying Learning to Rank techniques.

Overall, the user study showed that our proposed model is versatile by generalising over different types of users, and times of the day. It also demonstrated that we can model venues using social sensors (user check-ins), paving the way

Table 7: Effectiveness results for different levels of knowledge of the participants in our study. Notations are identical to those mentioned in Table 4.

	never been		visited		live/lived in	
	P@10	nDCG@10	P@10	nDCG@10	P@10	nDCG@10
(N)	0.383	0.376	0.369	0.306	0.227	0.272
(F)	0.439	0.316	0.381	0.380 ^α	0.300	0.274
(P)	0.494 ^α	0.364 ^β	0.471^{α,β}	0.434^{α,β}	0.345 ^α	0.289
(I)	0.531^{α,β}	0.411 ^{α,β}	0.452 ^{α,β}	0.397 ^α	0.445^{α,β}	0.449^{α,β}
(Full)	0.508 ^α	0.423^{α,β}	0.433 ^α	0.367 ^α	0.418 ^{α,β}	0.378 ^{α,β}

to a wider use of such data. Note that in our evaluation, we chose large cities with a relatively large number of Foursquare venues and Foursquare adoptions. We aim to consider in future work how our model performs in small cities, rural environments or in developing countries with lesser adoption of LBSNs. Finally, our approach could be extended to combine data coming from other sources, such as the physical sensors of mobile phones [35].

Acknowledgments

This work has been carried out in the scope of the EC co-funded project SMART (FP7-287583).

7. REFERENCES

- [1] M.-D. Albakour, C. Macdonald, and I. Ounis. Identifying Local Events by Using Microblogs as Social Sensors. In *Proc. of OAIR*, 2013.
- [2] G. Amati, E. Ambrosi, M. Bianchi, C. Gaibisso, and G. Gambosi. FUB, IASI-CNR and Univ. of Tor Vergata at TREC 2007 Blog Track. In *Proc. of TREC*, 2007.
- [3] J. Bao, Y. Zheng, and M. F. Mokbel. Location-based and Preference-aware Recommendation Using Sparse Geo-social Networking Data. In *Proc. of SIGSPATIAL*, 2012.
- [4] K. Berberich, A. C. König, D. Lymberopoulos, and P. Zhao. Improving Local Search Ranking through External Logs. In *Proc. of SIGIR*, 2011.
- [5] B. Bi, M. Shokouhi, M. Kosinski, and T. Graepel. Inferring the Demographics of Search Users: Social Data Meets Search Queries. In *Proc. of WWW*, 2013.
- [6] V. Bicer and V. Lopez. Searching in the City of Knowledge: Challenges and Recent Developments. In *Proc. of SIGIR*, 2013.
- [7] L. Breiman. Random forests. *Machine Learning*, 45(1), 2001.
- [8] I. Brillhante, J. A. Macedo, F. M. Nardini, R. Perego, and C. Renso. Where Shall We Go Today?: Planning Touristic Tours with Tripbuilder. In *Proc. of CIKM*, 2013.
- [9] G. V. Cormack, M. D. Smucker, and C. L. Clarke. Efficient and Effective Spam Filtering and Re-ranking for Large Web Datasets. *Information Retrieval*, 14(5), 2011.
- [10] A. Dean-Hall, C. L. Clarke, J. Kamps, P. Thomas, N. Simone, and E. M. Voorhees. Overview of the TREC 2013 Contextual Suggestion track. In *Proc. of TREC*, 2013.
- [11] R. Deveaud, M.-D. Albakour, C. Macdonald, and I. Ounis. On the Importance of Venue-Dependent Features for Learning to Rank Contextual Suggestions. In *Proc. of CIKM*, 2014.
- [12] H. Gao, J. Tang, X. Hu, and H. Liu. Exploring temporal effects for location recommendation on location-based social networks. In *Proceedings of RecSys*, pages 93–100. ACM, 2013.
- [13] G. Guo. Resolving Data Sparsity and Cold Start in Recommender Systems. In *Proc. of UMAP*, 2012.
- [14] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA Data Mining Software: An Update. *SIGKDD Explor. Newsl.*, 11(1), 2009.
- [15] D. Hiemstra. *Using Language Models for Information Retrieval*. PhD thesis, University of Twente, 2001.
- [16] B. Hu and M. Ester. Spatial Topic Modeling in Online Social Media for Location Recommendation. In *Proc. of RecSys*, 2013.
- [17] R. J. Hyndman and Y. Khandakar. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(3), 2008.
- [18] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose. A State Space Framework for Automatic Forecasting using Exponential Smoothing Methods. *Inter. Journal of Forecasting*, 18(3), 2002.
- [19] G. H. John and P. Langley. Estimating Continuous Distributions in Bayesian Classifiers. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, San Mateo, 1995. Morgan Kaufmann.
- [20] M. Kosinski, D. Stillwell, and T. Graepel. Private Traits and Attributes are Predictable from Digital Records of Human Behavior. *Proceedings of the National Academy of Sciences*, 2013.
- [21] T. Kurashima, T. Iwata, T. Hoshida, N. Takaya, and K. Fujimura. Geo Topic Model: Joint Modeling of User’s Activity Area and Interests for Location Recommendation. In *Proc. of WSDM*, 2013.
- [22] Y. Li, M. Steiner, L. Wang, Z.-L. Zhang, and J. Bao. Exploring Venue Popularity in Foursquare. In *Proc. of INFOCOM*, 2013.
- [23] Y. Lv, D. Lymberopoulos, and Q. Wu. An Exploration of Ranking Heuristics in Mobile Local Search. In *Proc. of SIGIR*, 2012.
- [24] S. G. Makridakis, S. Wheelwright, and R. Hyndman. *Forecasting: Methods and Applications*. Wiley, 1998.
- [25] D. Milne, P. Thomas, and C. Paris. Finding, Weighting and Describing Venues: CSIRO at The 2012 TREC Contextual Suggestion Track. In *Proc. of TREC*, 2012.
- [26] I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, and C. Lioma. Terrier: A High Performance and Scalable Information Retrieval Platform. In *Proc. of OSIR at SIGIR*, 2006.
- [27] R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [28] A. Rae, V. Murdock, A. Popescu, and H. Bouchard. Mining the Web for Points of Interest. In *Proc. of SIGIR*, 2012.
- [29] S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford. Okapi at TREC-3. In *Proc. of TREC*, 1994.
- [30] D. Saez-Trumper, D. Quercia, and J. Crowcroft. Ads and the City: Considering Geographic Distance Goes a Long Way. In *Proc. of RecSys*, 2012.
- [31] M. Sappelli, S. Verberne, and W. Kraaij. Recommending Personalized Touristic Sights Using Google Places. In *Proc. of SIGIR*, 2013.
- [32] B. Shaw, J. Shea, S. Sinha, and A. Hogue. Learning to rank for spatiotemporal search. In *Proc. of WSDM*, pages 717–726, 2013.
- [33] J. Soldatos, M. Draief, C. Macdonald, and I. Ounis. Multimedia Search over Integrated Social and Sensor Networks. In *Proc. of WWW*, 2012.
- [34] T. Teixeira and A. Savvides. Lightweight People Counting and Localizing in Indoor Spaces Using Camera Sensor Nodes. In *Proc. of ICDCS*, 2007.
- [35] H. Wang, D. Lymberopoulos, and J. Liu. Local Business Ambience Characterization Through Mobile Audio Sensing. In *Proc. of WWW*, 2014.
- [36] R. W. White, P. Bailey, and L. Chen. Predicting User Interests from Contextual Information. In *Proc. of SIGIR*, 2009.
- [37] P. Yang and H. Fang. An Opinion-aware Approach to Contextual Suggestion. In *Proc. of TREC*, 2013.
- [38] M. Ye, P. Yin, and W.-C. Lee. Location Recommendation for Location-based Social Networks. In *Proc. of SIGSPATIAL GIS*, 2010.
- [39] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. Magnenat-Thalmann. Time-aware Point-of-interest Recommendation. In *Proc. of SIGIR*, 2013.
- [40] Y. Zheng. Location-Based Social Networks: Users. In *Computing with Spatial Trajectories*. Springer, 2011.