Journal of Location Based Services

ISSN: 1748-9725 (Print) 1748-9733 (Online) Journal homepage: https://www.tandfonline.com/loi/tlbs20

Activity event recommendation and attendance prediction

Radu Mariescu-Istodor, Abu S. M. Sayem & Pasi Fränti

To cite this article: Radu Mariescu-Istodor, Abu S. M. Sayem & Pasi Fränti (2019): Activity event recommendation and attendance prediction, Journal of Location Based Services, DOI: 10.1080/17489725.2019.1660423

To link to this article: https://doi.org/10.1080/17489725.2019.1660423

Published online: 16 Sep 2019.

JOURNAL OF LOCATION BASED SERVICES
https://doi.org/10.1080/17489725.2019.1660423

Taylor & Francis
Taylor & Francis Group

Activity event recommendation and attendance prediction

Radu Mariescu-Istodor, Abu S. M. Sayem and Pasi Fränti

School of Computing, University of Eastern Finland, Kuopio, Finland

ABSTRACT

The recommendation problem has been widely studied and researchers are constantly searching for better methods. Recommending events is an even more difficult problem because there is no information such as ratings from past events. In this paper, we introduce a method for recommending activity events: activities hosted by one or more individuals which involve movement: walking, running, cycling, cross-country skiing, and driving to users who have location history such as trajectories, meetings, POI visits, and geo-tagged photos. We tested the method in a real environment in Mopsi platform: http://cs.uef.fi/mopsi/events. Although there are many location-based event recommendation systems in literature, this is to our knowledge the first system that recommends activity events like bicycle and skiing trips. The experiments show that we can predict whether a user is attending the event or not with 80% accuracy, which is significantly better than random chance (50%).

ARTICLE HISTORY
Received 2 April 2019
Accepted 21 August 2019

KEYWORDS
Mobile users; location-based events; recommendation; prediction

1. Introduction

Location-based services (LBS) are becoming very popular. These include travel applications (TripAdvisor, Foursquare), navigation (Google Maps, Waze), fitness (Sports Tracker, Strava), and games (Zombies Run, Pokemon Go, O-Mopsi (Fränti, Mariescu-Istodor, and Sengupta 2017)). Social networking applications like Twitter and Facebook and image-hosting websites (Flickr) also use location to enhance their services. Users of these applications record lots of location data using smart devices such as phones, tablets, and watches, which come equipped with location-sensing capabilities. Most often mobile users record check-ins to various points of interest (POIs), geo-tagged photos and even entire trajectories. The latter are recorded for reasons such as work (taxi and truck drivers), pleasure (updating a travel diary) or for exercise (sports tracking). The amount of location-based data is overwhelming and continues to grow every day.

CONTACT Radu Mariescu-Istodor radum@cs.uef.fi

In this paper, we consider *activity events*. They are social events hosted by an individual and involve user movement in some way. The movement can be walking, running, skiing, cycling, driving, or using public transport. When an activity event is created by a user (host), other users may join this event. An activity event is described by the following attributes:

– Title: short description of the event
– Type: walking/running/cycling/cross-country skiing/driving/bus/train
– Distance: intended distance of the activity
– Duration: intended duration of the activity
– Speed: intended speed/pace during the activity
– Date and Time: of the event
– Location: geo-coordinates and address
– Route: a path drawn on the map by the host (optional)

We use machine-learning techniques to obtain the answer to the following questions:

– What events should be recommended to a given user?
– Which users are likely to attend a given event?

We address both questions. Solving the first problem makes it easier for users to navigate large event databases. Solving the second problem helps the hosts by predicting how many participants are expected to join. This can aid in event planning (see Figure 1).

We use the data from *Mopsi* database to construct user profiles. Mopsi[1] is a location-based social network created by the School of Computing from the University of Eastern Finland. The Mopsi database contains a unique combination of data when compared to other applications. Like sports trackers, Mopsi users collect trajectories by recording points with high frequency (every 2–4 s) (Waga et al. 2013). Transportation mode of the trajectories is automatically inferred [Waga et al. 2012], and a grid-based indexing is applied to allow a fast search of similar trajectories (Mariescu-Istodor and Fränti 2017).

Mopsi has socialising features such as chatting, seeing the current location of friends, and automatic notification of meetings (Mariescu-Istodor and Fränti 2018). As in image-hosting websites, Mopsi users can upload photos, view the photo collections, and rate the content created by others. Mopsi photo collections are fully geo-tagged and can be also displayed as clusters on the map (Rezaei and Fränti 2018). Like local search-and-discovery apps, Mopsi contains

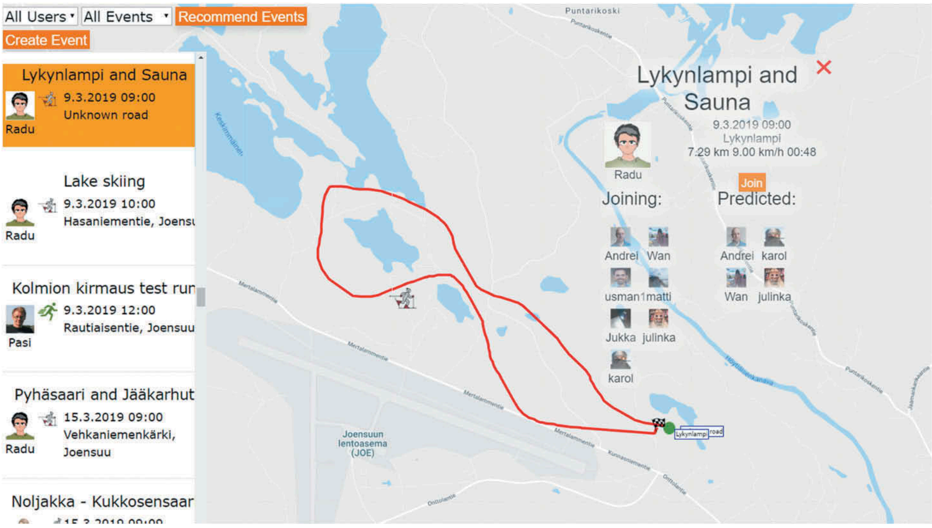


Figure 1. Mopsi Events: a system for planning activity events in Mopsi. The events are shown in a time-ordered list and on the map. Clicking one event displays the route and its properties. The interface offers event recommendation and lists the people who are predicted to attend.

a local database of POIs, which include restaurants, hotels, supermarkets and other enterprises. The database is created and maintained through crowdsourcing. Each entry has a title, keywords, an optional description, and a web link (when available). Users can rate the services. Visits (check-ins) to the services are automatically detected. (Mariescu-Istodor and Fränti 2018).

Just recently (winter of 2018) an event planner[2] was developed into Mopsi (see Figure 1). It serves as a test environment for the activity event recommendation and attendance prediction techniques, which are the focus of this work. We will address how to create a complex profile from Mopsi users' data and how to compute features for matching a given user to a specified event. We will discuss how to deal with the cold-start problem, which is an acute challenge because of two reasons. First, there is no excessive history of earlier events, and many events are one-time ad hoc events, which are not repeated periodically. Second, the system needs to provide recommendation also for new users who do not have any previous Mopsi data.

The motivation of our study is to recommend events from large databases to users. Users can be overwhelmed by the number of events around his/her location, so recommending events which are of interest to the user and filtering out others would help to browse more easily large event databases.

Another motivation is to create a recommendation system for activity events. Literature is currently lacking studies for recommending such events. Activity events differ from other events like concerts, lectures and social gatherings in several ways. First, the location plays a much more important role than events happening in just one static location. Second, activity events have different features like skills and move type preferred by the user.

In addition to this motivation, our proposed method can also be used to predict if users will attend an event. The host can use this information to better prepare for the event. The contributions of this study are the proposed nine-dimensional feature vector and how it can be used for recommendation and prediction of user attendance to the events. Our analysis suggests that a combination of these features gives good prediction results for different users.

2. Location-aware recommendation

Recommendations have been widely used in e-commerce and online services. Once a user clicks to open a product from a catalogue, the system can immediately start to recommend other similar products. The similarity can be measured based on the content, but also by monitoring the behaviour of other users. For example, once a user looked at a product, the system can record what other products the user has looked at afterwards. Recommendation systems originally appeared for books, movies, music, shopping but more recently *location-based recommendation* systems also started to appear.

Nowadays, location is automatically embedded in the pictures taken by smartphone cameras. There are exceptions when users explicitly disable the location setting on their device but the location of the user is widely available and can be used for recommendation. The key problem in such systems is how to define what is *relevant* to a given user at a certain time and place. In (Fränti, Chen, and Tabarcea 2011), relevance was defined by four aspects: *content, time, location,* and *social network*. We next review the existing literature on location-aware recommendation systems from this viewpoint.

Content is by far the most important aspect in recommendation systems. Search engines can be viewed as recommendation systems that rely primarily on the content of a web page and try to match it to a set of user-given keywords and the result depends on the quality of this matching. The location aspect can be utilised here as well. The framework in (Tabarcea, Gali, and Fränti 2017) shows in detail how a search engine can be made to be location-aware. The key component is to detect the address from the text content of the web page, and in this way, tie the web page to a location using a *gazeteer* that includes geo-coding of known addresses. After the location has been extracted, it can be used in ranking the search results, for example, by providing only results from nearby locations. However, recommendation systems lack the keywords and they must conclude the content-relevance in other ways.

Collaborative filtering is a popular method used in recommendation systems, in which personal recommendations are made using the similarity between users' preferences. Takeuchi and Sugimoto (2006) describe CityVoyager, a system for recommending shops based on user's location history. Visited shops are used as input to an item-based collaborative filtering algorithm. In (Bellotti et al. 2008), user activity is predicted from sensing context and from patterns of user behaviour. The recommendation is done automatically using a combination of collaborative filtering, distance, and stated or learned preferences. The method is implemented in a system called Magitti and tested in the urban setting (Tokyo). Other collaborative filtering methods are presented in (Yang and Wang 2009; Levandoski et al. 2012; Zheng et al. 2011).

Content-based filtering (Baudisch 1999) is a less popular way for how recommendation systems work. It utilises the information about an item itself for recommendations. The advantage of this method is that it is not limited to suggesting options that have previously been rated by users. The system showed in (Savage et al. 2012) is an example of content-based filtering that uses a decision tree and a hidden Markov model to mine a person's social network.

The simplest approach to incorporate location in a recommendation system is probably for nearby services such as shops and cafes, and Mopsi has this type of recommendation system as well [Waga, Tabarcea, and Franti 2012]. It is based on content, time, location, and social network as discussed in (Fränti, Chen, and Tabarcea 2011). The system has the capability to recommend three types of items: services stored in local database, photos, and GPS trajectories which lead to locations with more services of interest to the user.

A harder problem is to recommend the locations themselves because the content is not as clearly specified as for services. Tourist locations are recommended in (Clements et al. 2011; Yong 2011) based on visiting history in a geographically remote region. Places that are similar to the user's location history but also novel to the user are recommended. Both location and its frequency were considered in (Leal, Costa, and Galvão 2015). In Rahman [2018] locations to arrange live campaigns are recommended based on Foursquare check-in data. They used counts, consistency, and density of the check-ins, openness of the area, time of the day, and few other features as input to an SVM classifier, which then predicts the suitability of the location.

Some recommendation systems take location a step further and recommend trajectories, which are entire sequences of locations. In (Yoon et al. 2012), attributes are extracted from GPS trajectories to recommend itineraries. These attributes are elapsed time, stay time, interest density, and the social aspect measured as the number of different people recording the same trajectory.

Recommending events differs from the aforementioned types presenting three distinct characteristics:

- Events happen in the future.
- Events might not repeat and therefore have no previous attendance data.
- Attendance might be influenced not only by the content but also who else is participating.

In Minkov et al. (2010), the authors propose a method for event recommendation targeted at upcoming scientific talks. They calculate a persons' likes and dislikes and use collaborative filtering to determine the recommended items. Also in aneducational context, [Neves, Carvalho, and Ralha 2014; Rodríguez et al. 2013] propose a personalised event recommendation based on precomputed user interests. In addition to the content of the event, the models are based on user, location, date, and time. Huijuan, Kejie, and Bai (2007) try to predict the future behaviour of mobile users in order to give push recommendations. They use time, location, environmental factors, and user's current and past behaviour. Bayesian network inference is used to calculate the probability to the future behaviour of the user.

The event recommendation system in Li et al. (2009) uses a multi-stage collaborative filtering to provide event recommendations based on moving patterns. In (Augusto Macedo, Marinho, and Santos 2015), time and geographical preferences are used to give event recommendation. The method uses five types of data: content of the event, information of the invitees, user's memberships to groups, geographical location, time preferences, and the attendance list. The authors address the cold-start problem by exploiting group memberships. The system described in (Schilke et al. 2004) recommends events based on the current location, combining location with multi-dimensional personalisation. Other recommendation methods also attempt to predict user behaviour based on past observations (Kim and Cho 2009; Tuan, Hung, and Kuei 2011).

In (Cena et al. 2016), the authors recommend events by paying attention to its reachability and the attendance list. They consider who the host is and who else is attending. In addition to the typical content-based features, they consider reachability of the location, the reputation of the event in the community, and the participation of the user's friends. If properly weighted, these were reported to improve the prediction accuracy by about 4% when combined with the content-based features. In (Qiao et al. 2014), events are recommended using a Bayesian probabilistic model that uses co-participation of users in past events.

Some patents exist for event recommendation[3, 4].

In the event recommendation methods listed above, the location-based user modelling methods can be categorised into four groups depending on what data they use (Liu 2018): (1) pure check-in data, (2) geographical information, (3) spatiotemporal information, (4) geo-social information. In case of activity events, we must use as detailed information as possible because the location and type of activity are expected to play an important role. We will need to estimate the suitability of an event and its familiarity to the user so that we do not recommend a cycling trip to someone who does not have any past cycling history, implying he may not like or even know how to cycle.

The similarity of the users was considered in (Fränti, Waga, and Khurana 2015) based on their social network and location history. The similarity of the location history was calculated in (Fränti, Mariescu-Istodor, and Waga 2018) by counting the frequency of photo taking and other activities of the users in the vicinity of predefined service locations. The similarity was then calculated using the Bhattacharyya distance between two histograms. The results showed that the location history had only mild correlation to how similar the users consider themselves to their peers (Fränti, Waga, and Khurana 2015). This also confirmed the earlier findings made in (Bao, Zheng, and Mokbel 2012) that the opinions of the local experts are potentially more valuable than just the similarity of the user. In case of organised events, the similarity of the users is expected to lose its significance even more while the location is expected to play a more important role. Users who have similar location history are more likely to join the events organised by each other. The location itself is significant: the user is unlikely to participate in an event located far away, especially if he/she has never been in that place before.

Recommendation problem is intrinsically a cold-start problem when no history data are available. The cold-start is considered a serious issue into the extent that some methods even assume that there must be enough past attended events for the method to work at all. The method in Khrouf and Troncy (2013) uses closeness of the location and utilise external data from DBpedia. Using social information and modelling user diversity was recognised as important future directions to improve recommendation. Quercia et al. (2010) provide cold-start recommendations for users in case their home location was known. They found that instead of recommending only near-by events, it would be better to recommend events that are popular among residents of the area. Motivated by this, we will, therefore, compute an average user profile from the users in the region and investigate its usefulness when recommending events to new users.

3. Mopsi event system

The Mopsi Event system[5] is a platform for creating and joining activity events. The system is free to use, and anybody can create or join events from the database. The events are described using the following attributes:

- Title: short description of the event
- Type: walking/running/cycling/cross-country skiing/driving/bus/train
- Distance: intended distance of the activity
- Duration: intended duration of the activity
- Speed: intended speed/pace during the activity

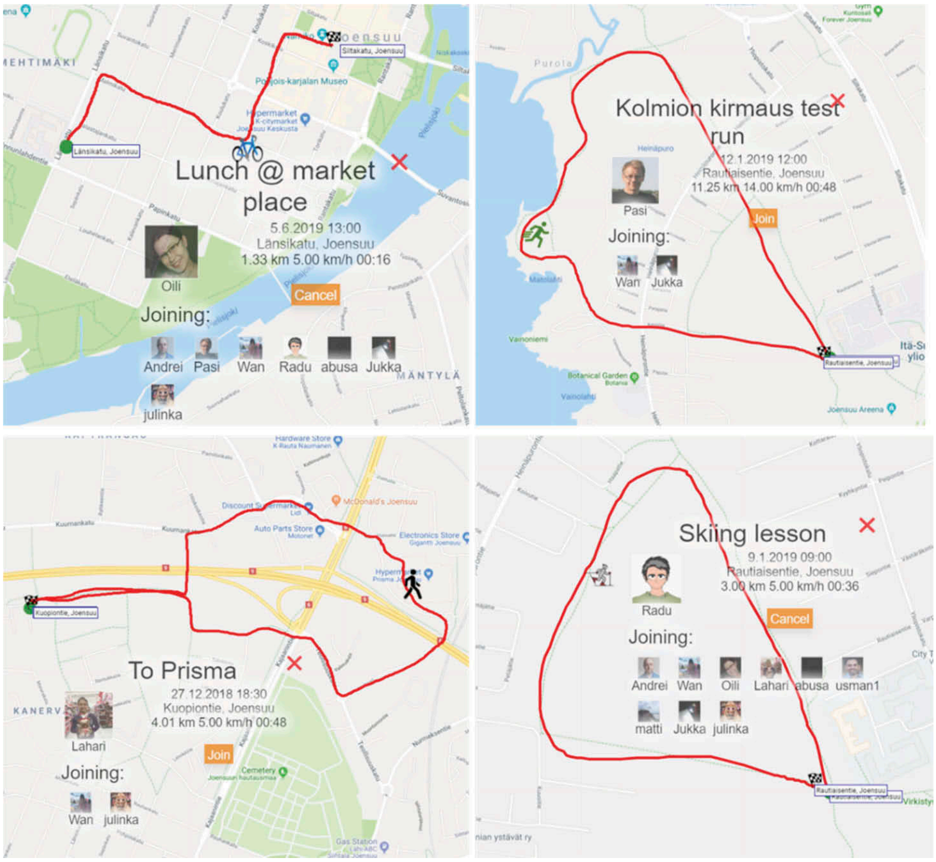


Figure 2. Four examples of activity events in Mopsi. Two of the events (lunch @ market place, To Prisma) are more like social events whereas the other two (Kolmion kirmaus test run and Skiing lesson) are exercise sessions. Two of the events are popular. The shopping trip (To Prisma) is less popular probably because of the start location which is somewhat remote for most users. The test run session is less popular due to its speed parameter which identifies it as an advanced event.

– Date and Time: of the event
– Location: geo-coordinates and address
– Route: a path drawn on the map by the host (optional)

Examples of the Mopsi events are shown in Figure 2. The system relies on the data that the users have recorded in Mopsi: check-ins, photos, and trajectories. It also works for guest users (not logged in), but the recommendation will be limited to using the time, location, and average behaviour of users in the region.

4. Mopsi user profiles

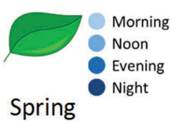
We create user profiles from the Mopsi data shown in Table 1. We can see the three users have different moving characteristics. Pasi records mostly running

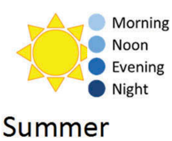
Table 1. Moving characteristics and active times of three sample users. The numbers are the average length, speed and duration of the recordings of the user classified into five move types. The maximum value of each row is emphasised.

		Moving Characteristics		
		Pasi	Andrei	Radu
	Length	**7.8 km**	2.7 km	3.3 km
	Speed	**6.6 km/h**	5.4 km/h	4.9 km/h
	Duration	**1.2 hours**	0.6 hours	0.8 hours
	Proportion	17 %	22 %	**44 %**
	Length	**13 km**	4.8 km	4.7 km
	Speed	**10.8 km/h**	10.1 km/h	10.5 km/h
	Duration	**1.2 hours**	0.5 hours	0.5 hours
	Proportion	**67 %**	19 %	15 %
	Length	12 km	13 km	**25 km**
	Speed	16 km/h	18 km/h	**19 km/h**
	Duration	0.9 hours	0.7 hours	**1.2 hours**
	Proportion	11 %	**34 %**	20 %
	Length	9.8 km	**11 km**	10 km
	Speed	**10.2 km/h**	7.7 km/h	7.4 km/h
	Duration	0.9 hours	**1.4 hours**	**1.4 hours**
	Proportion	2 %	**6 %**	**6 %**
	Length	20 km	**74 km**	54 km
	Speed	51 km/h	53 km/h	**60 km/h**
	Duration	0.6 hours	**1.3 hours**	0.9 hours
	Proportion	3 %	**19 %**	15 %
No. Routes		**2023**	626	1429

Active Times

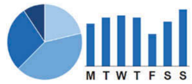
Pasi | Andrei | Radu





(*Continued*)

Table 1. (Continued).

	Active Times		
	Pasi	Andrei	Radu
 Autumn			

tours (67%) and almost no car tours (3%). Andrei and Radu have more bicycle tours (34% and 20%) although Radu's dominant move type is walking (44%). Pasi does less skiing but his speed is significantly faster (10.2 km/h versus 7.7 km/h and 7.4 km/h).

However, the automated move type analysis in [Waga et al. 2012] is not completely reliable. It tends to classify routes as bicycle trips based on single high-speed segments. As a result, many of Pasi's running routes have been miss-classified as bicycle trips due to a single downhill section, tempo run, or even GPS errors. The analysis method also did not support skiing move type, so we made the following modification. Given a trajectory, we calculate its similarity according to (Mariescu-Istodor and Fränti 2017) with the known skiing tracks available in OpenStreetMap. A trajectory is classified as skiing if the similarity exceeds 75%, and was recorded between December and March when skiing conditions are favourable in Joensuu, Finland. The overall results give a good overall estimate suitable for a personal profile. To compute keywords from uploaded photo descriptions, we group together all the text, remove stop words and keep words with frequencies at least two. Table 2 shows statistics about the user profiles for three sample users from Table 1.

For a more detailed analysis of the Mopsi users' data and the event database, we refer to the webpage associated with this manuscript: http://cs.uef.fi/mopsi/events/analysis.html.

5. Feature modelling

We define a binary matching function for every user–event pair, which gives value TRUE if we predict the event is suitable for the user and should be recommended:

$$Match(u,e) = \begin{cases} TRUE & \text{if } Classifier(u \circ e) = Positive\,Class \\ FALSE & \text{if otherwise} \end{cases} \tag{1}$$

where u and e are elements from the users set U and events set E, respectively. The matching function first computes a feature vector by combining a given user profile information with a specified event information $(u \circ e)$ and then

Table 2. Location, meetings and interests inferred from taken photos and visited services for the same three sample users as in Table 1.

	Location					
	Pasi		Andrei		Radu	
Cities	124		19		82	
 Cities	Joensuu	72%	Joensuu	76%	Joensuu	76%
	Kuopio	1.9%	Iasi	8.8%	Kuopio	4%
	Helsinki	1.7%	Suceava	6%	Drobeta	4%
	Singapore	1.7%	Turku	2%	Helsinki	1%
	Tampere	1%	Stockholm	2%	Barcelona	1%
	…		…		…	

	Meetings					
	Pasi		Andrei		Radu	
Meetings	415		544		847	
 Users	Jukka	31%	Radu	28%	Andrei	18%
	Radu	17%	Jukka	8%	Jukka	9.2%
	Oili	12%	Pasi	6.6%	Pasi	8%
	Andrei	8.7%	Julinka	5.5%	Oili	6%
	Minttu	5.3%	Rezaei	5%	Rezaei	3.6%
	…		…		…	

	Photo Keywords					
	Pasi		Andrei		Radu	
Keywords	44,081		3379		3016	
 Photos	Park	1.2%	Road	2.1%	Snow	1.1%
	View	1.1%	Joensuu	1.6%	Beach	1.1%
	Street	1%	SciFest	1.2%	View	1%
	Restaurant	0.9%	Kaislakatu	1.1%	Lake	1%
	Talo	0.9%	Mopsi	1%	Koli	0.8%
	…		…		…	

	Visited Service Types					
	Pasi		Andrei		Radu	
Keywords	651		214		648	
 Services	Kahvila	26%	Kahvila	25%	Kahvila	34%
	Lounas	25%	Lounas	19%	Lounas	33%
	Spa	3.4%	Best	4.7%	Juhlapalvelu	4.9%
	Tanssi	3%	Coffee	4.2%	Library	1.7%
	Kulta	2.8%	Kahvi	3.7%	Book	1.4%
	…		…		…	

uses a generic binary classifier to determine if the match belongs to the positive or negative class. The events that are recommended to a given user u are constructed as follows:

$$Recommended(u) = \{e \in E | Match(u, e) = TRUE\} \quad (2)$$

Similarly, we can construct the set of users that we predict to join a given event *e*:

$$Predicted(e) = \{u \in U | Match(u, e) = TRUE\} \quad (3)$$

To construct the overall match function, we create a nine-dimensional feature vector. Every feature belongs to one of the following five distinct classes: *community, suitability, familiarity, availability* and *interests*. We will elaborate on these in the following subsections.

5.1. Community

Community refers to the people attending the event. There is a lot of uncertainty here since the event will happen in the future and there is no way of knowing who will actually participate. However, the host is expected to participate for sure.

We calculate a feature that measures how likely it is for a given user *u* to meet the event host *h* based on the number of meetings they shared in the past relative to the total number of meetings of *u* (see Equation (4)).

$$Community = \frac{MeetingCount(u, h)}{\sum_{i=1}^{|U|} MeetingCount(u, u_i)} \quad (4)$$

We considered also other features such as the number of times the two have rated each other's photos. However, our experiments showed that these values do not correlate much to participation. Liking the content of a person and joining the event organised by the same person seems to depend on different factors. For instance, Oili often likes Pasi's photos but she does not participate in any of his events because they are mostly exercising like fast running.

Another feature we considered is whether a user participates in an event thinking that a common friend of his and the event host may participate. We calculated the number of meetings they both joined but this feature did not have any effect on the probability that the persons would join each other's events.

5.2. Suitability

Suitability is concerned with the following two aspects:

- How often the user performs the activity specified by the event?
- Does the user have the necessary skill to participate in the event?

For the first, we analyse the user profile and calculate a value for how likely the user has done the activity specified by the event in the past. We have taken into account the season in which the event takes place because user interests change depending on the season: many people do not cycle at all during the winter while the same people would go cycling during the summer is a typical example. The value of this feature is given by the following equation:

$$Suitability_{TYPE} = \frac{1}{|R|}\sum_{i=1}^{|R|}(SameType(e, r_i)\ AND\ SameSeason(e, r_i)), \tag{5}$$

where R is the set of trajectories (routes) recorded by the user, r is a trajectory from the set R, and e is the event.

To answer whether a user is skilled enough to attend an event with specified speed and distance, let us first look at how the activities of a user look like in Figure 3. User Radu has mostly walking, cycling and cross-country skiing activities. We can notice several patterns. On the top right, there is a long-distance cycling cluster with varying speeds. The shorter distance (40 km) corresponds to typical training and the longer (60–70 km) corresponds to long runs, which appear less often. The cross-country skiing cluster contains many observations with around 15–20 km length and few spurious long-distance ones. The bottom clusters are commuting to work by walking and bicycle.

Let us now consider two events: one is 55-km cycling at 23 km/h, and another 35 km skiing at 14 km/h. To determine if Radu is can participate in these events, we look at the top-right corner of the space and investigate

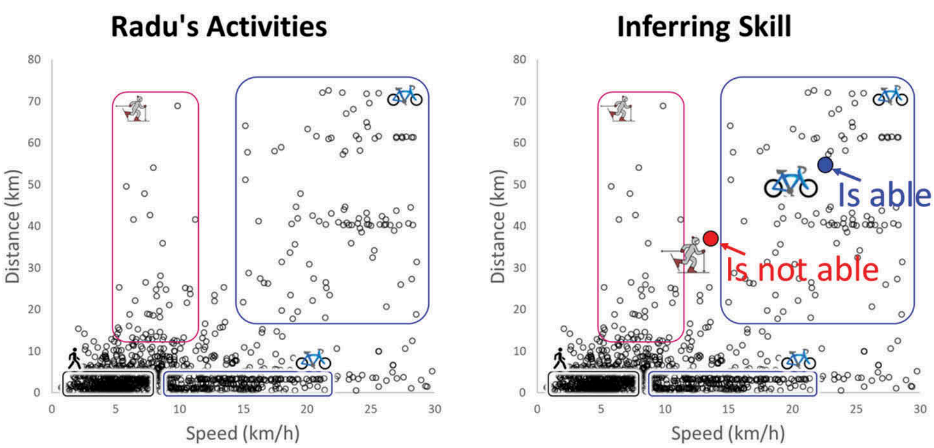


Figure 3. Moving activities of user Radu plotted as a function of the average speed and total distance (left). His most typical activities are skiing, bicycle, and walking, which we emphasised using borders. Two forthcoming events are then shown as red and blue circles in terms of their target speed and distance (right) with respect to Radu's activity data. The cycling event is suitable for Radu, but the skiing event is not.

whether there are observations of the event type in this region. Radu has participated to even more engaging cycling events, so the cycling event is a good fit. The skiing event, however, is too fast paced for his skills. The equation for computing the feature is:

$$Suitability_{SKILL} = \frac{1}{|R|}\sum_{i=1}^{|R|}\left(Speed(e) \lesssim Speed(r_i)\, AND\, Distance(e) \gtrsim Distance(r_i)\right) \tag{6}$$

where R is the set of trajectories (routes) recorded by the user, r is a trajectory from the set R, and e is the event. In other words, if the user has even a single evidence of doing a certain activity with a similar speed (or faster) and distance (or longer) as defined in the event, then the event is considered suitable for the user. We use a 10% tolerance so that slightly slower and shorter events are permitted as well. This is to handle situations when users try to attempt harder events on purpose. This feature is calculated for every movement type separately.

5.3. Familiarity

Familiarity refers to the region where the activity will take place. Some users prefer to join events in familiar places, while others like to explore a new place. We calculate familiarity using the spatial similarity of the event route and all the trajectories the user has recorded in the past. The similarity is calculated using a grid-based similarity measure called *Cell Similarity* (CSIM) (Mariescu-Istodor and Fränti 2017). Familiarity is the percentage of the previous trajectories having high similarity (>75%) with the event trajectory:

$$Familiarity = \frac{1}{|R|}\sum_{i=1}^{|R|}(CSIM(Route(e), r_i) > 75\%) \tag{7}$$

We set the 75% threshold experimentally to account for differences caused by unique trajectory differences caused, for example, by users who start tracking already at their home or workplace when moving towards the meeting point. The path to the event route would be different for users. This threshold allows some tolerance on this matter. The cell length is 25 metres at minimum; however, it is scaled automatically by considering the best-fitting zoom level of the map when the host draws the event route. This is done to account for the differences in drawing precision at different zoom levels of the map.

This feature is also useful for users who prefer to join events in new, unfamiliar places. In that case, we simply trust that during the training stage, the classifier will associate low familiarity with attendance.

5.4. Availability

The availability feature estimates user's availability in the following ways:

– Is the event in the user's current city?
– Is the user usually active on the day of the event?
– Is the user usually active at the time of the event?

For the first, we check if the user is typically active in the city where the event will take place:

$$Availability_{CITY} = \frac{1}{|R|}\sum_{i=1}^{|R|}(City(e) = City(r_i)), \tag{8}$$

where r is a trajectory from the set of trajectories (routes) R recorded by the user, and e is the event. If the event is going to happen in the near future, this should be replaced simply by checking whether the user is in the same city or not. For trajectories that cover several cities, we only consider the city of the start point of the trip in this equation because that is where everybody must gather when the event starts.

For the other two, we check if the user is typically active on the day and at the time of the event, respectively:

$$Availability_{DAY} = \frac{1}{|R|}\sum_{i=1}^{|R|}(Day(e) = Day(r_i)), \tag{9}$$

$$Availability_{TIME} = \frac{1}{|R|}\sum_{i=1}^{|R|}(Time(e) = Time(r_i)), \tag{10}$$

where the Day and Time functions return the day of the week and the time of the day: morning, noon, evening, or night, respectively. We also considered limiting the data with respect to the season when the event takes place with both positive and negative effects. It seems to be slightly more effective when users have a significant amount of data in all seasons; however, for less active users, this dividing would make the subsets too small having a negative effect on the prediction accuracy.

5.5. Interests

We measure the interests of a user by using the keywords describing the services the user has visited in the past, and the text descriptions of the photos the user has uploaded into the system. However, instead of requiring exact matches, we allow small variations in the keywords. We measure the similarity of the words using the Levenshtein edit distance and consider the words matching if their similarity is above 90%. The equations are as follows:

$$Interests_{SERVICES} = \frac{1}{|S_K|}\sum_{i=1}^{|Sk|}(Levenshtein(Title(e), k_i) > 90\%), \tag{11}$$

$$Interests_{PHOTOS} = \frac{1}{|P_K|}\sum_{i=1}^{|P_K|}(Levenshtein(Title(e), k_i) > 90\%), \tag{12}$$

where S_K and P_K are the sets of keywords obtained from the services visited, and from the uploaded photos, respectively. Here k_i is one keyword and *Levenshtein* is the edit distance between two strings converted into similarity value by normalising it with respect to the longer of the two strings. We chose the Levenshtein distance here as it is suitable in the case of annotated names and different languages as demonstrated in (Gali et al. 2019).

6. Experiments

We performed the experiments using real events created by eight volunteers in and around the city of Joensuu, Finland in the next upcoming year (1.12.2018–1.12.2019). The volunteers created a total of 150 events that have a variety of types (see Table 3). The time frame of 1 year was decided to allow season-specific events (cycling in the summer and skiing in the winter) to be created.

We then collected ground truth by asking the 12 most active Mopsi users to mark down for each event if they would participate or not. We tailored a special tool for collecting the ground truth (see Figure 4), which is essentially the same as the real Event system in Mopsi with the difference that two buttons, **Yes** and **No**, were added for each event. When the user

Table 3. The number of events of each type created by the volunteers. The dominant type of event for each user is highlighted in **blue.**

								Total
Pasi	1	**14**			4		1	**20**
Radu	3	4	7	**10**	1			**25**
Oili	**13**		4		2	1	1	**21**
Abu	4	1	**5**		3	4	1	**18**
Lahari	6		**9**		3	1	1	**20**
Sami	**7**	1	4	1	4			**17**
Nancy	**6**	4	4		1	4		**19**
Usman	**10**							**10**
Total	**50**	**24**	**33**	**11**	**18**	**10**	**4**	**150**

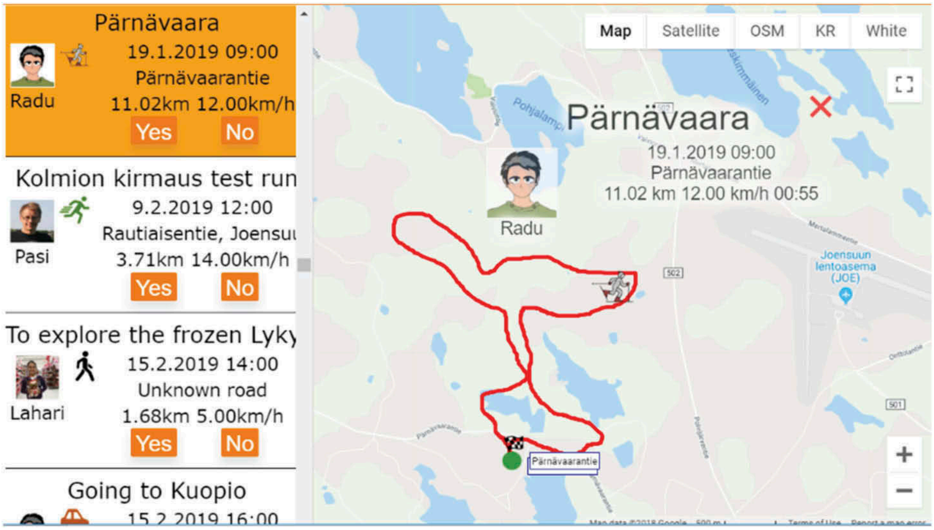


Figure 4. Tool for collecting the ground truth. Events are displayed on the list and on the map. Selecting each event shows the host, the route and its properties on the map.

presses any of the buttons, the answer is registered, and the event is removed from the list. In this way, we made sure that all events were processed by every volunteer.

The 12 volunteers and their choices to attend or not are summarised in Table 4. For the evaluation, we divide the collected data in half: 50% for training and 50% for testing. We do this by sorting the events by date and taking every odd event of every user in the training set and every even event into the test set. In this way, both the training and testing sets cover all four seasons. The volunteers had to choose from a total of 150 events. However, four volunteers (Pasi, Radu, Oili, Sami and Lahari) also host part of those events. For the sake of fair evaluation, we exclude those user-event pairs from the data where the user is also hosting the event.

6.1. Usefulness of the features

We next study how well the features correlate with the decision to participate or not (see Table 5). The importance of these features varies according to the user's personality. For users like Jukka, Karol, Oili, and Pasi, the host of the event seems to correlate most. They participate less in the events whose host they do not often meet. For example, Jukka's participation correlates highly (0.60) to the organiser of the event. Namely, he actively joined Pasi's events. The correlation is also high for Karol (0.35), Oili (0.30), and Pasi (0.28). Simply put, the more the users have joint activities in the past, the more likely they are going to do so in the future. However, this correlation does not apply to all users.

Table 4. The 12 volunteers and their choices to attend the generated events. We show the number of attendances of each user also per training and testing set. Zhentian (emphasised in **blue**) is the most active and is attending most events. Oili and Make (emphasised in **red**) attend the least events. Some names have been altered due to the user's wish for anonymity.

	TRAINING		TESTING		TOTAL	
Name	Attending	Percent	Attending	Percent	Attending	Percent
Andrei	33 / 75	44%	37 / 75	49%	70 / 150	47%
Pasi	9 / 65	14%	10 / 65	15%	19 / 130	15%
Jukka	21 / 75	28%	25 / 75	33%	46 / 150	31%
Karol	29 / 75	39%	35 / 75	47%	64 / 150	43%
Make	**1 / 75**	**1 %**	**0 / 75**	**0 %**	**1 / 150**	**1 %**
Zhentian	**71 / 75**	**95%**	**72 / 75**	**96%**	**143 / 150**	**95%**
Radu	20 / 62	32%	23 / 63	37%	43 / 125	34%
Oili	**3 / 64**	**5 %**	**1 / 65**	**2 %**	**4 / 129**	**3 %**
Matti	24 / 75	32%	30 / 75	40%	54 / 150	36%
Julinka	48 / 75	64%	49 / 75	65%	97 / 150	65%
Sami	3 / 65	5 %	6 / 65	9%	9 / 130	7 %
Lahari	13 / 65	20%	10 / 65	15%	23 / 130	18%
TOTAL	275 / 846	33%	298 / 848	35%	573 / 1694	34%

Table 5. The Pearson correlation of the features with the decision to participate or not in each event. The correlations above 0.3 are highlighted in blue and the exceptionally high ones are also **bold.**

		Suitability			Availability			Interests	
Name	Community	Type	Skill	Familiarity	Season	Time	Day	Photos	Services
Andrei	0.06	0.17	0.18	0.04	0.00	0.33	0.05	0.10	0.10
Pasi	0.28	0.12	0.13	0.14	0.12	0.19	0.08	0.22	0.07
Jukka	**0.60**	0.15	0.21	0.22	0.29	0.21	0.02	0.05	0.18
Karol	0.35	0.06	0.38	0.12	0.21	0.14	0.03	0.02	0.00
Make	0.17	0.09	0.15	0.06	0.02	0.08	0.13	0.05	0.00
Zhentian	0.19	0.04	0.19	0.06	0.33	0.16	0.23	0.06	0.00
Radu	0.11	**0.42**	0.17	0.24	0.31	0.18	0.07	0.20	0.13
Oili	0.30	0.12	0.03	0.16	0.06	0.11	0.16	0.08	0.03
Matti	0.12	0.03	0.19	0.06	0.11	0.35	0.09	0.04	0.08
Julinka	0.08	0.08	0.03	0.03	0.09	0.04	0.09	0.13	0.22
Sami	0.00	0.03	0.04	0.04	0.03	0.08	0.02	0.00	0.00
Lahari	0.00	0.22	0.11	0.04	0.09	0.07	0.14	0.06	0.00
TOTAL	0.19	0.13	0.15	0.10	0.14	0.16	0.09	0.08	0.07
Median	0.15	0.11	0.16	0.06	0.10	0.15	0.09	0.06	0.05

Suitability of the event matters most to Radu, who rarely participates an event that is not his preferred type of activity. He mostly joined events involving skiing and walking and rarely running events or car trips. Karol's participation is affected more on his estimated suitability. For instance, he joined many events but rarely bicycle or running events that did not match to his own speed preferences.

The familiarity of the planned route has only minor effect on the participation. There is no single user whose participation would strongly depend on this feature. The users mostly influenced are Radu (0.24), Jukka (0.22), and Pasi (0.18). A possible reason is that they have the widest coverage of trajectories in their data, and it might be easier to predict the places they are likely to attend.

Other users have collected less data (see Table 6), and it might be harder to predict whether the area of the event is of interest to them.

Availability correlates to the participation of many users. Some users are more active in some seasons than others while other users are active in some parts of the day. For example, Andrei seems to prefer evening activities and Matti daytime activities. The day of the week correlated least even if one might assume some people would have a strong preference on weekend over weekdays.

Interest feature correlates least to the participation. The only few notable exceptions are Pasi (0.22) and Radu (0.20) when using the words from their photo descriptions, and Julinka when using the keywords of the services she has visited. In particular, she has many visits to lunch places and cafeterias, and she also joins most of the events involving such keywords in their description. However, for this feature to become more efficient we would need more data from the users; or create a better feature for the prediction.

For some people like Sami, Julinka, Make, and Lahari, all correlations are relatively low (below 30%). This does not mean that the features are not meaningful to them, but rather that a combination of the features is required to perform a successful prediction for them. It is possible, however, that for some people like Jukka, only one feature can be enough to make a reliable prediction. In his case, he mostly attends Pasi's events.

We note that the correlations are not very high. Nevertheless, we can see that the community, suitability and availability features are the most important, in general.

6.2. Classification task

We use the nine-dimensional feature vectors to perform a binary classification: the user either attends to the event or not. We experiment using four different classifiers: *k*-nearest neighbour (KNN), support vector machine (SVM), naive Bayes and decision tree. We optimise the following parameters: the number of

Table 6. The 12 volunteers and the amount of data they collected in Mopsi.

User	Trajectories	Photos	Meetings	Services
Andrei	687	2 722	1 261	97
Pasi	2 174	22 425	746	330
Jukka	336	1 101	365	183
Karol	765	578	168	34
Make	176	204	58	36
Zhentian	149	270	107	30
Radu	1 676	2 376	1 294	367
Oili	640	20 256	507	208
Matti	442	104	162	13
Julinka	250	4 443	250	327
Sami	40	0	0	15
Lahari	79	23	4	7

neighbours (k) for KNN, for C and γ in SVM we use grid search in the range k= 2 to 10, and a logarithmic grid with basis 10 for C and γ. After optimising on the training set, we found that k= 4 works best for KNN, and C= 10^5 and $\gamma = 10^{-3}$ with RBF kernel best for SVM. Additional classifiers were tested including the multi-layer perceptron and random forest. These, however, did not outperform the four mentioned above and we leave them out from further analysis.

We compute a **personalised** model for every user from his/ her own training data. In addition, we also compute a **general** model using training data of all users. To evaluate the models, we use *accuracy* and *precision*:

$$Precision = \frac{number\ of\ correctly\ predicted\ events}{number\ of\ events\ user\ is\ attending} \tag{13}$$

$$Accuracy = \frac{number\ of\ correctly\ classified\ events}{total\ number\ of\ events} \tag{14}$$

Prediction by random chance would result in an accuracy of 50%, and if we knew the average attendance percentage, random chance would result in an accuracy of 65%. This is because we have 298 participations out of 848 possible. This corresponds to 298/848 = 35% probability to participate. So, if we knew this number, a random guess that a person does **not** participate would lead to 65% accuracy. A good prediction system should, therefore, perform better than these.

The prediction accuracy is higher than random chance no matter which model is used. Table 7 shows the accuracy and precision of every model for each user. On average, the best results are obtained by the decision tree (73–76%), but KNN (68–75%) and SVM (66–75%) are almost as good. Only naive Bayes performs clearly worse (55–63%). A closer inspection reveals that KNN achieves the best prediction performance for half (6/12) of the users. The decision tree was best classified in case of four users, and the other two with one user each.

The precision values provide further insights about the performance. For instance, the personalised models for user Pasi had high accuracy (83% with KNN) in the sense that it predicted that he is not going to attend many events. However, the precision was often poor. For example, Pasi did not participate in any events the personalised KNN model was predicting (0% precision). The general model, on the other hand, worked better providing 2/3 correct prediction, thus, 67% precision. The reverse can also be true. For user Zhentian, the personal model is more accurate because his behaviour is unusual. He is participating in events that do not match at all to his profile. However, he likes to attend many types of events and explore new things. In this case, the personal model manages to capture this behaviour better. Zhentian is also one of the easier users to predict. This matches the

Table 7. Accuracy and precision, the later in parentheses, shown for each user when using general and personal models computed with four different classifiers. The best models are emphasised in blue.

	KNN				Decision Tree			
	General		Personal		General		Personal	
Andrei	56%	(68%)	55%	(64%)	60%	(63%)	59%	(59%)
Pasi	86%	(67%)	83%	(0 %)	72%	(10%)	75%	(20%)
Jukka	65%	(46%)	65%	(45%)	79%	(67%)	77%	(62%)
Karol	60%	(69%)	55%	(55%)	63%	(68%)	51%	(47%)
Make	91%	(0 %)	100%	(0 %)	95%	(0 %)	99%	(0 %)
Zhentian	45%	(100%)	95%	(97%)	80%	(97%)	92%	(97%)
Radu	59%	(20%)	60%	(33%)	59%	(41%)	70%	(61%)
Oili	95%	(0 %)	98%	(0%)	88%	(0%)	94%	(0%)
Matti	59%	(46%)	64%	(59%)	57%	(45%)	61%	(52%)
Julinka	39%	(60%)	56%	(68%)	55%	(68%)	63%	(73%)
Sami	88%	(0%)	91%	(0%)	89%	(33%)	92%	(100%)
Lahari	78%	(30%)	83%	(0%)	77%	(27%)	83%	(33%)
Avg.	68%	(42%)	75%	(35%)	73%	(43%)	76%	(50%)
	SVM				Naive Bayes			
	General		Personal		General		Personal	
Andrei	51%	(50%)	60%	(62%)	51%	(50%)	49%	(49%)
Pasi	78%	(0%)	75%	(25%)	58%	(23%)	51%	(19%)
Jukka	71%	(100%)	79%	(70%)	44%	(35%)	80%	(71%)
Karol	55%	(56%)	53%	(50%)	47%	(47%)	61%	(63%)
Make	92%	(0%)	93%	(0%)	64%	(0%)	100%	(0%)
Zhentian	19%	(100%)	89%	(97%)	80%	(97%)	67%	(98%)
Radu	60%	(25%)	59%	(40%)	37%	(33%)	43%	(38%)
Oili	95%	(0%)	97%	(0%)	25%	(2 %)	82%	(0%)
Matti	59%	(0%)	71%	(75%)	52%	(43%)	53%	(46%)
Julinka	35%	(50%)	56%	(66%)	57%	(63%)	63%	(64%)
Sami	88%	(0%)	88%	(0%)	68%	(14%)	89%	(0%)
Lahari	83%	(0%)	80%	(0%)	75%	(20%)	17%	(15%)
Avg.	66%	(32%)	75%	(40%)	55%	(36%)	63%	(39%)

observations from Minkov et al. (2010), where, in the case of scientific talks, better accuracy and precision is obtained for users interested in a wider spectrum of events than for those who have narrower interests. They also mention that predictions for users who have an interest in different types of events result in more relevant items to be predicted for these users in contrast to users who have narrower interests.

It is also not so surprising to see the decision tree performing well because our data are quite structured; the users are making binary decisions whether to attend or not. It is also interesting to note that the decision tree works best for users with little data (Sami and Lahari), and therefore, it is recommended to be used for cold-start users.

For two users (Make and Oili), the precision is poor no matter which model is used. This is mainly because they do not attend to almost any events in the test data; Make is not attending any events and Oili only once. This shows how challenging the prediction can be with users who are very selective in the participation of the events.

6.3. Classifier ensemble

Since our goal is to build a system that provides the best possible prediction, we briefly investigate whether a classifier ensemble can be constructed from these results. For each user, we chose the classifier and model that works best with his/ her training data, and then calculated the prediction. The results are summarised in Table 8. In this way, we reach 80% accuracy.

We note that two users (Radu and Pasi) are also authors of this paper. We therefore briefly checked whether researcher bias would be a factor by tentatively removing these two from the data. However, the effect on the accuracy or precision was insignificantly small, 1% at most. This is partly because the data collection was made before the recommendation and prediction systems were invented. The behaviour of these users is also not the easiest to predict. Collecting a larger amount of data would be desirable but not practical. Bottlenecks are the large time frame needed, and to make the system popular that more users would start to use it. (There are bottlenecks such as larger time is needed to make the system popular that most users would start to use it.)

To put the numbers in the context of other event recommendation systems, we refer to the results with two existing prediction systems. Mean average precision of 55% was reported in (Minkov et al. 2010) when recommending scientific talks for interested audience. Accuracy of 76% was reported in (Huijuan, Kejie, and Bai 2007) when recommending services for moving users. Thus, our method produces results that are no worse than those existing event recommender systems in different scopes: scientific talks and service visits, respectively.

6.4. Cold-start problem

To study the cold-start problem, we asked three new volunteers who just began to use Mopsi and have no previous data stored in their profile as of now (see Table 9). For these users, we use an average profile obtained, as the

Table 8. Accuracy and precision, the later in parenthesis when using the best classifier and model combination for each particular user.

	Best of all	Classifier	Model
Andrei	56% (68%)	KNN	general
Pasi	86% (67%)	KNN	general
Jukka	80% (71%)	Naive Bayes	personal
Karol	60% (69%)	KNN	general
Make	100% (0%)	KNN	personal
Zhentian	95% (97%)	KNN	personal
Radu	70% (61%)	Decision Tree	personal
Oili	98% (0%)	KNN	personal
Matti	71% (75%)	SVM	personal
Julinka	63% (73%)	Decision Tree	personal
Sami	92% (100%)	Decision Tree	personal
Lahari	83% (33%)	Decision Tree	personal
Avg.	80% (60%)		

Table 9. Attendance information of three new Mopsi users.

Name	Attending	Percent
Abu	32 / 132	24%
Nancy	17 / 127	13%
Usman	46 / 140	33%

name implies, by averaging the properties of the user profiles of the 12 users introduced in Table 4. With the average profile and the event properties, we compute a general model. The prediction results are summarised in Table 10. Even though KNN has higher overall accuracy, decision tree produces significantly higher precision. We, therefore, recommend using the decision tree model with an average profile for cold-start users. The models can then be personalised by updating the average profile with personalised information as soon as the user starts collecting data.

6.5. Qualitative evaluation

Some examples of events are shown in Figure 5. Oili's event Lunch @ market place: three of her friends were correctly predicted but more users have joined. It is likely that the attractiveness of this event comes from the social nature of this event which was not captured by any of the features. The place itself (market place) is popular during the summer time, and people do go to lunch together at times. Pasi's event Kolmion kirmaus test run: Radu and Julinka were wrongly predicted. Both users have run two or more marathons in the past and in this sense, would be fit to join but decided not to – probably because the session was more about speed than endurance which made it less attractive to Radu and Julinka and the models were not good enough to recognise this small difference. For Lahari's event To prisma: prediction is 100% correct. Both Zhentian (Wan) and Julinka have prisma supermarket in their interests and this information was captured by modelling their user profiles. In Radu's event Skiing lesson: Karol is wrongly predicted. He has plenty of skiing history and would have been interested in skiing but not so much in the skiing *lesson*. However, some other unpredicted users are joining because they are new in town and interested to learn new things like skiing. Again, these models could not differentiate between the learning and doing,

Table 10. Prediction accuracy and precision, the later in parentheses for new Mopsi users.

	KNN	Decision Tree
Abu	70% (0%)	70% (50%)
Nancy	88% (9%)	73% (9%)
Usman	70% (71%)	76% (82%)
Avg.	76% (27%)	73% (47%)

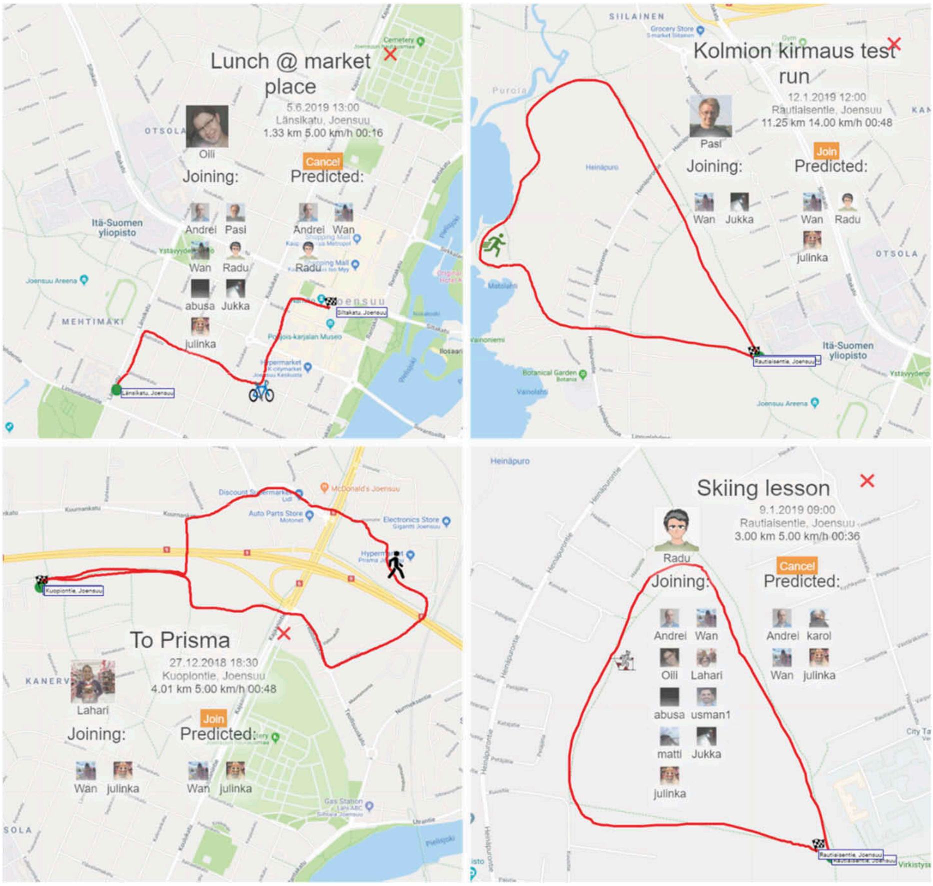


Figure 5. Qualitative examples of the prediction system.

and there was not enough information to conclude the friendship between the participant and host, which is another key factor.

7. Conclusions

We have introduced a method to predict whether a user is going to participate in an activity event. As far as we know, this is the first time such a system has been considered. The method uses nine different features based on *community, suitability, familiarity, availability,* and *interests*. We achieved a prediction accuracy of 80%, which is significantly better than expected results of 50% if the decision was made by random chance.

A general model trained from the data of all users works best for users whose behaviour is more systematic: users who participate only to events that match their skills and are organised in times that match their typical schedule. Personalised prediction worked better for other users.

Among the tested classifiers, decision tree produced the best overall accuracy and worked well for users who had less data. KNN also worked reasonably well and produced the best results for users who have plenty of data collected in Mopsi.

Notes

1. http://cs.uef.fi/mopsi.
2. http://cs.uef.fi/mopsi/events.
3. https://patents.google.com/patent/US20140129505A1/en.
4. https://patents.google.com/patent/US20100325205A1/en.
5. http://cs.uef.fi/mopsi/events.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Pasi Fränti http://orcid.org/0000-0002-9554-2827

References

Augusto Macedo, Q., L. B. Marinho, and R. L. T. Santos. 2015. "Context-aware Event Recommendation in Event-based Social Networks." Proceedings of the 9th ACM Conference on Recommender Systems (RecSys'15), Vienna, Austria, 123–130.

Bao, J., Y. Zheng, and M. F. Mokbel. 2012. "Location-based and Preference-aware Recommendation Using Sparse Geo-social Networking Data." In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12)*, 199–208. New York, NY: ACM. doi:10.1145/2424321.2424348.

Baudisch, P. 1999. "Joining Collaborative and Content-based Filtering." Proceedings of the ACM CHI Workshop on Interacting with Recommender Systems (CHI'99), Pittsburgh, PA, 1–3. doi:(99)00837-7",1,0,0>10.1016/s0021-9673(99)00837-7.

Bellotti, V., B. Begole, E. H. Chi, N. Ducheneaut, J. Fang, E. Isaacs, T. King et al. 2008. "Activity-based Serendipitous Recommendations with the Magitti Mobile Leisure Guide." Proceedings of the ACM sigchi conference on human factors in computing systems (CHI'08), Florence, Italy, 1157–1166.

Cena, F., S. Likavec, I. Lombardi, and C. Picardi. 2016. "Should I Stay or Should I Go? Improving Event Recommendation in the Social Web." *Interacting with Computers* 28 (1): 55–72. doi:10.1093/iwc/iwu029.

Clements, M., P. Serdyukov, A. P. de Vries, and M. J. T. Reinders. 2011. "Personalised Travel Recommendation Based on Location Co-occurrence." *arXiv Preprint arXiv:1106* 5213.

Fränti, P., J. Chen, and A. Tabarcea. 2011. "Four Aspects of Relevance in Sharing Location-based Media: Content, Time, Location and Network." Int. Conf. on Web Information Systems & Technologies (WEBIST'11). Noordwijkerhout, Netherlands, May 2011, 413–417.

Fränti, P., K. Waga, and C. Khurana. 2015. "Can Social Network Be Used for Location-aware Recommendation?" International Conference on Web Information Systems & Technologies (WEBIST'15). Lisbon, Portugal, 558–565.

Fränti, P., R. Mariescu-Istodor, and K. Waga. 2018. "Similarity of Mobile Users Based on Sparse Location History." International Conference Artificial Intelligence and Soft Computing (ICAISC). Zakopane, Poland, 593–603 doi:10.1142/S2424835518720372.

Fränti, P., R. Mariescu-Istodor, and L. Sengupta. 2017. "O-Mopsi: Mobile Orienteering Game for Sightseeing, Exercising and Education." *ACM Trans. On Multimedia, Computing, Communications, and Applications* 13 (4), 56.: 1–25. doi:10.1145/3115935.

Gali, N., R. Mariescu-Istodor, D. Hostettler, and P. Fränti. 2019. "Framework for Syntactic String Similarity Measures." *Expert Systems with Applications* 129: 169–185. doi:10.1016/j.eswa.2019.03.048.

Huijuan, L., C. Kejie, and L. Bai. 2007. "Bayesian Network Based Behavior Prediction Model for Intelligent Location Based Services." 2nd IEEE Conference on Industrial Electronics and Applications, Harbin, China, 703–708. doi:10.1556/OH.2007.28017.

Khrouf, H., and R. Troncy. 2013. "Hybrid Event Recommendation Using Linked Data and User Diversity." In *Proceedings of the 7th ACM conference on Recommender systems (RecSys '13)*. New York, NY: ACM, 185–192. doi:10.1145/2507157.2507171.

Kim, Y., and S.-B. Cho. 2009. "A Recommendation Agent for Mobile Phone Users Using Bayesian Behavior Prediction." Third IEEE International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, Sliema, Malta, 283–288.

Leal, R., P. M. Costa, and T. Galvão. 2015. "Discovering MARS: A Mobility Aware Recommender System." 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Las Palmas, Spain, 1–6. doi: 10.1109/ITSC.

Levandoski, J. J., M. Sarwat, A. Eldawy, and M. F. Mokbel. 2012. "Lars: A Location-aware Recommender System." (ICDE), 2012 IEEE 28th International Conference on Data Engineering, Washington, DC, 450–461.

Li, L.-H., F.-M. Lee, Y.-C. Chen, and C.-Y. Cheng. 2009. "A Multi-stage Collaborative Filtering Approach for Mobile Recommendation." Proceedings of the 3rd ACM International Conference on Ubiquitous Information Management and Communication (ICUIMC'09), Suwon, Korea, 88–97.

Liu, S. 2018. "User Modeling for Point-Of-Interest Recommendations in Location-Based Social Networks: the State of the Art." *Mobile Information Systems* 2018: 13. Article ID 7807461. doi:10.1155/2018/7807461.

Mariescu-Istodor, R., and P. Fränti. 2017. "Grid-based Method for GPS Route Analysis for Retrieval." *ACM Transactions on Spatial Algorithms and Systems (TSAS)* 3 (3): 8.

Mariescu-Istodor, R., and P. Fränti. 2018. "Detecting Location-Based User Actions." Adjunct Proceedings of the 14th International Conference on Location Based Services, ETH Zurich, Zurich, Switzerland, 1–6.

Minkov, E., B. Charrow, J. Ledlie, S. Teller, and T. Jaakkola. 2010. "Collaborative Future Event Recommendation." In Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM'10), Toronto, ON, 819–828. doi:10.4315/0362-028x-73.5.819.

Neves, A. R. D. M., Á. M. G. Carvalho, and C. G. Ralha. 2014. "Agent-based Architecture for Context-aware and Personalized Event Recommendation." *Expert Systems with Applications* 41 (2): 563–573. doi:10.1016/j.eswa.2013.07.081.

Qiao, Z., P. Zhang, C. Zhou, Y. Cao, L. Guo, and Y. Zhang. 2014. "Event Recommendation in Event-based Social Networks." Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Quebec City, Quebec.

Quercia, D., N. Lathia, F. Calabrese, G. D. Lorenzo, and J. Crowcroft. 2010. "Recommending Social Events from Mobile Phone Location Data." IEEE International Conference on Data Mining, Sydney, NSW, 971–976. doi: 10.1109/ICDM.2010.152.

Rezaei, M., and P. Fränti. 2018. "Real-time Clustering of Large Geo-referenced Data for Visualizing on Map"." *Advances in Electrical and Computer Engineering* 18 (4): 63–74. doi:10.4316/AECE.2018.04008.

Rodríguez, A. C., M. Victor, J. M. Alonso Rorís, L. E. Santos Gago, A. Rifón, and M. J. Fernández Iglesias. 2013. "Providing Event Recommendations." In *Management Intelligent Systems. Advances in Intelligent Systems and Computing*, edited by E. Scenarios, J. Casillas, F. Martínez-López, R. Vicari, and F. De la Prieta, 91–98, Vol. 220. Heidelberg: Springer.

Savage, N. S., M. Baranski, N. E. Chavez, and T. Höllerer. 2012. "I'm Feeling Loco: A Location Based Context Aware Recommendation System." In *Advances in Location-Based Services,* 37–54. Berlin, Heidelberg: Springer.

Schilke, S. W., U. Bleimann, S. M. Furnell, and A. D. Phippen. 2004. "Multi-dimensional-personalisation for Location and Interest-based Recommendation." *Internet Research* 14 (5): 379–385. doi:10.1108/10662240410566980.

Tabarcea, A., N. Gali, and P. Fränti. 2017. "Framework for Location-aware Search Engine." *Journal of Location Based Services* 11 (1): 50–74. doi:10.1080/17489725.2017.1407001.

Takeuchi, Y., and M. Sugimoto. 2006. "CityVoyager: an Outdoor Recommendation System Based on User Location History." International Conference on Ubiquitous Intelligence and Computing, Springer, Berlin, Heidelberg, 625–636.

Tuan, C.-C., C.-F. Hung, and T.-C. Kuei. 2011. "Location Dependent Collaborative Filtering Recommendation System." International Conference on Future Network Technologies, Qingdao, China.

Waga, K., A. Tabarcea, M. Chen, and P. Fränti. 2012. "Detecting Movement Type by Route Segmentation and Classification." IEEE Int. Conf. on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom'12), Pittsburgh, PA. doi: PDN",1,0,0>10.1094/PDIS-11-11-0999-PDN.

Waga, K., A.Tabarcea, and P. Franti. 2012. "Recommendation of Points of Interest from User Generated Data Collection." In 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2012), Pittsburgh, PA, 550–555.

Waga, K., A. Tabarcea, R. Mariescu-Istodor, and P. Fränti. 2013. "Real Time Access to Multiple GPS Tracks." Proceedings of the 9th International Conference on Web Information Systems and Technologies (WEBIST '13), Aachen, Germany, 293–299.

Yang, F., and Z.-M. Wang. 2009. "A Mobile Location-Based Information Recommendation System Based on GPS and WEB2. 0 Services." *WSEAS Transactions on Computers* 8 (4): 725–734.

Yong, L. T. 2011. "A Collaborative Awareness Framework for Mobile Tourist Recommender System." 2011 3rd International Conference on Computer Research and Development, Shanghai, China, 329–333. doi:10.1109/ICCRD.2011.5764030.

Yoon, H., Y. Zheng, X. Xie, and W. Woo. 2012. "Social Itinerary Recommendation from User-generated Digital Trails." *Personal and Ubiquitous Computing* 16 (5): 469–484. doi:10.1007/s00779-011-0419-8.

Zheng, Y., L. Zhang, M. Zhengxin, X. Xie, and W.-Y. Ma. 2011. "Recommending Friends and Locations Based on Individual Location History." *ACM Transactions on the Web (TWEB)* 5 (1): 5.