

Article

Identifying Foreign Tourists' Nationality from Mobility Traces via LSTM Neural Network and Location Embeddings

Alessandro Crivellari * and Euro Beinat

Department of Geoinformatics—Z_GIS, University of Salzburg, 5020 Salzburg, Austria

* Correspondence: alessandro.crivellari@sbg.ac.at

Received: 6 June 2019; Accepted: 12 July 2019; Published: 18 July 2019



Abstract: The interest in human mobility analysis has increased with the rapid growth of positioning technology and motion tracking, leading to a variety of studies based on trajectory recordings. Mapping the routes that people commonly perform was revealed to be very useful for location-based service applications, where individual mobility behaviors can potentially disclose meaningful information about each customer and be fruitfully used for personalized recommendation systems. This paper tackles a novel trajectory labeling problem related to the context of user profiling in “smart” tourism, inferring the nationality of individual users on the basis of their motion trajectories. In particular, we use large-scale motion traces of short-term foreign visitors as a way of detecting the nationality of individuals. This task is not trivial, relying on the hypothesis that foreign tourists of different nationalities may not only visit different locations, but also move in a different way between the same locations. The problem is defined as a multinomial classification with a few tens of classes (nationalities) and sparse location-based trajectory data. We hereby propose a machine learning-based methodology, consisting of a long short-term memory (LSTM) neural network trained on vector representations of locations, in order to capture the underlying semantics of user mobility patterns. Experiments conducted on a real-world big dataset demonstrate that our method achieves considerably higher performances than baseline and traditional approaches.

Keywords: neural networks; LSTM; embeddings; trajectories; motion behavior; smart tourism

1. Introduction

The study of human mobility has received significant attention in recent years due to the growing collection and availability of motion data, making it relatively easy to track large numbers of people and create massive trajectory data sets from GPS traces, road sensors, mobile phone traces, social media geo-spatial check-ins, and many more recording tools [1]. This massive amount of mobility data has allowed a better understanding and modeling of travel behaviors and motion patterns [2], leading to significant analysis covering various applications, such as personalized recommendation [3] and preference-based route planning [4].

Human trajectory classification is a well-studied problem in literature, especially for detecting activity patterns and transportation modalities: based on spatial–temporal values and activities, traces are classified into some predefined categories, e.g., walking or driving [5]. However, the use of motion activity for inferring information about individual users is a very recent trend that has room for improvement and expansion. Our work is inserted in this new wave of user profiling, in particular aiming to identify the nationality of individual foreign visitors from their mobility traces. In the big picture of mining user motion behaviors [6], linking anonymous users to their nationality on the basis of only their generated trajectories can be very useful in many scenarios, especially for touristic

purposes [7]. In the context of “smart” tourism, which provides personalized services for improving the experience of travelers and the management and marketing of companies in the sector, the connection between motion traces and user nationalities is indeed helpful in serving tourists more efficiently, making better recommendations, personalized and precise suggestions, and targeted advertisement. Moreover, it may turn out to be a relevant factor for trajectory prediction, particularly if distinctive paths are explored by different nationalities.

The problem of inferring nationalities from motion activity is a challenging task. The hypothesis is that foreign visitors of different nationalities may not only visit different locations over the territory, but also move in a different way between the same locations. The number of classes (nationalities) considered can be on the order of a few tens of units, so typically larger than the number of motion patterns used in the traditional trajectory classification studies. In addition, the motion activity of foreign tourists is naturally characterized by short traces and non-repetitive behaviors, and large-scale mobility can lead to analyzing a very wide territory, hence encountering problems such as a sparsity of trajectory data and high number of locations, entailing the curse of dimensionality.

In this paper, we propose a new method for revealing short-term foreign visitors’ nationality based uniquely on their generated motion traces over the territory. The problem is defined as a multinomial classification using trajectory as an input and the corresponding nationality class as an output. The method consists of three steps: trajectory pre-processing, location embeddings generation, long short-term memory (LSTM)-based model building. More specifically, raw traces are transformed into discrete (in time and space) location sequences and, inspired by word embedding approaches in natural language processing (NLP), fed to a Word2vec-based model for learning the embedding vector of each location according to the motion behavior of people, whereby behaviorally-related locations share similar representations in mathematical terms. Trajectories are therefore defined as sequences of embeddings, which are used as an input to an LSTM neural network for learning the underlying motion patterns of human mobility. The collective motion behavior of people over the territory is used to train the model and associate individual traces to a specific predicted nationality.

To the best of our knowledge, this is the first work to address the above-mentioned problem and propose an effective and efficient machine learning approach leveraging both location embeddings and LSTM networks. Experiments conducted on a real-world large-scale big dataset demonstrate that our method considerably outperforms baseline and traditional approaches.

2. Related Work

The increasing acquisition and availability of mobility data has determined a growing interest in the investigation of human motion activity [8,9]. Trajectory classification (or trajectory labeling) is a central task in understanding mobility patterns—modeling human behaviors to predict the class labels of moving entities is important for many real-world applications in several research fields, such as user recommendations [10], computational health [11], and video surveillance [12].

The goal of trajectory classification is to classify the observed motion behavior into one element of a set of classes. Target classes strongly depend on the application domain and the specific problem addressed. Relying on the extraction of spatial–temporal characteristics, existing works label trajectories as belonging to different motion patterns, e.g., walking/driving/biking in transportation classification [5], or occupied/non-occupied in taxi status inference [13]. Other works use human mobility data to assess the users’ physical and mental health conditions, such as to predict flu-like symptoms [14], daily mood states [15], and stress levels [16]. However, despite the presence of a large number of works on semantic trajectory mining and classification, the problem of inferring nationalities from foreign tourists’ motion traces has never been formally defined and addressed.

Motion activity classification is often based on probabilistic models. In particular, Markov models are the most widely adopted tools, incorporating historical visit locations and sequential patterns: applications comprise movement type classification from GPS routes [17], unusual trajectory detection from surveillance cameras [18], object [19] and human [20,21] activity recognition from trajectory

data. Discriminative methods such as conditional random fields have also been used in activity recognition [22,23]. Other studies have analyzed the features of individuals based on latent Dirichlet allocation and Bayesian models for the purpose of personalized recommendation [3,7]. Finally, the recent trends in machine learning have led to an increasing use of neural network approaches [24,25].

For our task, we utilize specific tools that are particularly known in the NLP domain, namely vector representations of meaning [26] and LSTM neural networks [27].

3. Methodology

In this section, we first formally define the problem and then proceed to present the details of our method.

Given a number of trajectories generated by different anonymous users during a defined time interval, the solution of our model provides a link between each trajectory and the correct user nationality within a set of possible choices. The model is able to learn motion patterns of nationalities from mobility traces, performing a proper trajectory classification without any manual feature extraction or additional information.

The proposed method consists of three steps: trajectory pre-processing, in which the original traces, continuous in time and space, are transformed into discrete location sequences; embeddings generation, in which we define the input variables for the deep learning model; LSTM-based model building, in which we apply and train the model on the processed trajectories to infer the associated user nationality.

3.1. Trajectory Pre-Processing

In mobility data recordings, motion is represented as a mapping function between space and time [28]. Trajectories are modeled as a series of chronologically ordered coordinate pairs enriched with a time stamp: $T = \{p_i | i = 1, 2, 3, \dots, N\}$, where $p_i = (lon_i, lat_i, t_i)$. However, in order to feed the model properly, a pre-processing step is essential. The continuity of space and time needs to be subjected to a discretization process, by which the original traces are transformed into discrete location sequences ($LOC_1, LOC_2, \dots, LOC_N$): continuous longitude and latitude variables are aggregated into discrete locations and time information is encoded in the position along the sequence. Each motion trace is therefore converted into a sequence of locations that unfolds in fixed time steps, and if more than one event refer to the same time step, the one with the most occurrences is chosen to represent the location of the user. The length of the time step depends on both the prediction problem and the data source (different ways of collecting motion data may define different time resolutions), to balance location accuracy with the completeness of the sequences: a long unit affects the accuracy of the actual trajectory representation, a short unit increases fragmentation in cases of discontinuous traces. Moreover, when the traces are very sparse over the territory and there are many locations with a very low number of occurrences, the poor results and the high computational cost could suggest grouping together adjacent locations, where multiple longitude/latitude pairs of individual track points can be mapped to the same discrete location. Raster-based partitioning, clustering, and stop point detection are typical approaches used to convert trajectories into discrete cells, clusters, and stay points [29]. Since human mobility is not usually uniformly distributed over the territory, we recommend methods which avoid cell partitioning when numerous cells contain very few location occurrences, leading to processing a large number of potentially inaccessible and irrelevant places, decreasing computational efficiency and prediction results. We suggest dealing only with locations that are visited by a sufficient number of users, areas with enough tracking of the historical motion behavior of visitors, avoiding bias samples in the dataset. A valuable option may be to choose a number of fixed meaningful reference points over the territory and project the other locations to the nearest reference point. The minimum distance between reference points can vary according to the precision required by different applications (e.g., predicting travel patterns over a country or exploring city-level mobility). The result consists of a number of fixed points, each of them representing a particular area or location.

In conclusion, the pre-processed trajectory is represented by a discrete location sequence $(LOC_1, LOC_2, \dots, LOC_N)$, where, given a time step unit t , locations in the sequence refer to time $(t, 2t, \dots, Nt)$. In the next subsection, we describe how to use these pre-processed trajectories for learning location embedding representations.

3.2. Embeddings Generation

To mitigate the problem of the curse of dimensionality, we represent each location with a low-dimensional dense vector (embedding) instead of using traditional location representations such as one-hot. Similar to word embeddings in NLP [30–32], we generate location embeddings $\theta_i \in \mathbb{R}^d$ (d is the dimensionality of the embedding space) according to the motion behavior of people traveling over the territory, whereby behaviorally related locations share similar representations in mathematical terms. These vectors rely on the concept of “behavioral proximity” between places based on people’s trajectories, not on locations’ geography: two locations are behaviorally similar if they often belong to the same trajectories, they often share the same neighbor locations along the trace [33].

In order to construct location embeddings, we apply Word2vec [26], one of the most efficient techniques to define word embeddings in NLP, on the previously pre-processed trajectories. Based on co-occurrences in large training corpora, each element is represented as a vector with multiple activations, whereby elements occurring in similar contexts have similar vectors.

More specifically, we associate each location with a random initial vector of pre-defined size: the whole list of places refers to a lookup table where each location corresponds to a particular unique row of an embedding matrix of size $num_locations \times vector_size$. To update the matrix, we move a sliding window through every trace, identifying at each step the current focus location and its neighboring context locations along the trajectory. Although we are dealing with an unsupervised model, an internal auxiliary prediction task is defined: each instance is a prediction problem whose goal is to predict the current location with the help of its context (or vice-versa). The task is performed by a neural network model made of a single linear projection layer between the input and the output layers. In our implementation, we adopted the Skip-gram approach, consisting of maximizing the probability of observing the correct context locations cL_1, \dots, cL_j given the focus location L_t , with regard to its current embedding θ_t . The cost function C is the negative log probability of the correct answer, as reported in Equation (1):

$$C = - \sum_{i=1}^j \log(p(cL_i|L_t)). \quad (1)$$

The outcome of the prediction, through backpropagation, determines in what direction the location vectors are updated: the gradient of the loss is derived with respect to the embedding parameters θ , i.e., $\partial C / \partial \theta$, and the embeddings are updated consequently by taking a small step in the direction of the gradient. Prediction here is therefore not an aim in itself, but just a proxy to learn vector representations. The model updates the embedding matrix according to locations’ contexts along the traces using mini-batch stochastic training, until embeddings converge to optimal values.

3.3. Long Short-Term Memory Neural Network Model for Inferring Users’ Nationality

3.3.1. Model Description

In the last few years, remarkable success has been achieved by applying recurrent neural networks (RNNs) to a variety of machine learning problems [34,35]. Their chain-like structure is very suitable for sequences and lists, leading RNN to be particularly used in applications related to text, audio, and video data processing [36–38].

RNNs are composed of a chain of repeating modules of neural networks, processing an input sequence one element at a time. Information flows through the network modules, influencing the output of the subsequent steps of the chain. The repeating RNN module receives two sources of input:

information about the present (current value of the data sequence) and information about the past (output value of the previous RNN module).

LSTM is a complex type of RNN, responsible for many outstanding results in the field of speech recognition, language modeling, and translation [39–41]. Its repeating module is made of four different neural network layers, interacting in a particular way, as shown in Figure 1.

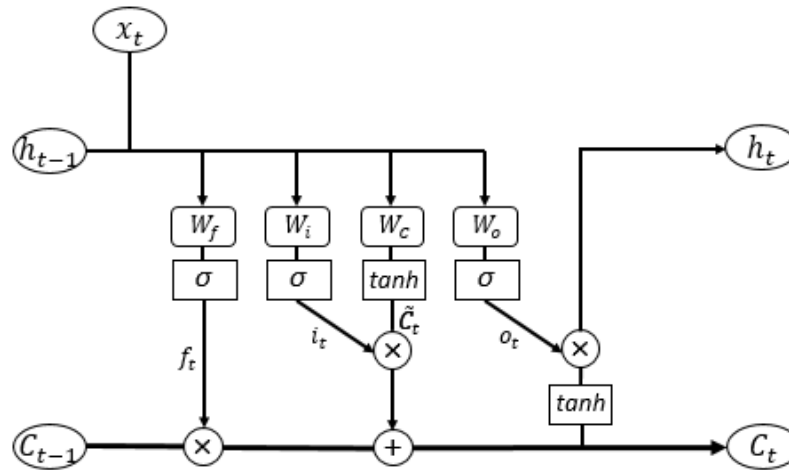


Figure 1. Structure of the long short-term memory (LSTM) repeating module.

Unlike standard RNNs, LSTM is characterized by the presence of the cell state C , the vector containing the information used for executing the machine learning task (e.g., prediction or classification). At each step, the cell state is subjected to some interactions with structures called gates, made of a sigmoid neural network layer and a pointwise multiplication operation. Gates act on the inputs they receive, blocking or passing information on the basis of its strength and relevance, therefore optionally modifying (removing or adding) information in the cell state through their own sets of weights, adjusted via a backpropagation learning process.

Equations (2)–(7) reports the formulas describing the functioning of LSTM. The first gate is called the forget gate layer (2) and defines what information to delete from the cell state. The second gate is named the input gate layer (3) and, before interacting with the cell state, is coupled with a tanh layer (4). They define what new information to store in the cell state: the input gate layer decides which values to update, and the tanh layer determines a vector of new candidate values to be added to the state. The cell state is therefore updated, combining the forgetting action and the updating action: the old cell state C_{t-1} is filtered by the forget gate layer f_t , then the output of the combination between the input gate layer i_t and the tanh layer \tilde{C}_t is added (5). The last gate is the output gate layer (6), which defines what parts of the cell state to output. The output is a filtered version of the cell state, resulting from the multiplication between the output gate layer o_t and the tanh of the new cell state C_t (7).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f); \quad (2)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i); \quad (3)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C); \quad (4)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t; \quad (5)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o); \quad (6)$$

$$h_t = o_t * \tanh(C_t). \quad (7)$$

Since these sequential operations occur at every step in the series, the cell state contains traces not only of the previous state, but also of all those that preceded C_{t-1} .

3.3.2. Model Training

After pre-processing, trajectories are defined as discrete location sequences representing the past time–space transitions of users over the territory. We therefore replace discrete locations with the corresponding embedding vectors, obtaining a new representation of trajectories as sequences of dense vectors (see Figure 2).

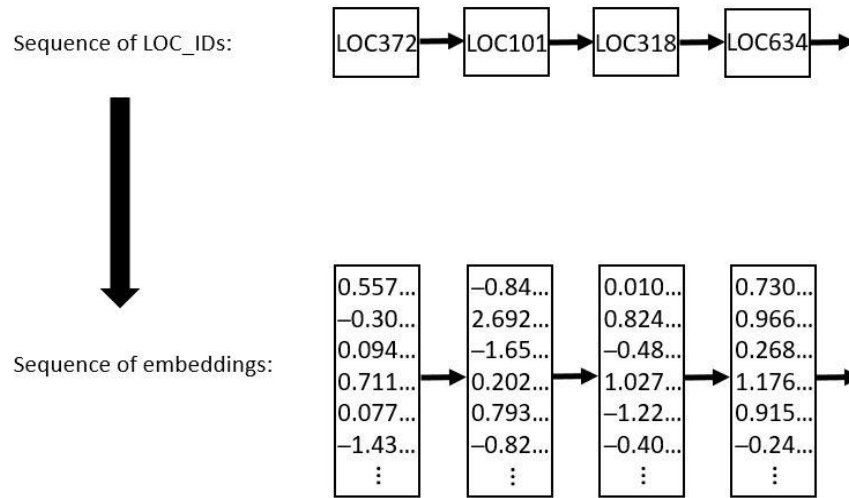


Figure 2. New trajectory representation as a sequence of location embeddings.

Before being fed to the LSTM, sequences are subjected to a segmentation phase, where they are partitioned into multiple segments of fixed length. This is done by a fixed-width sliding window scanning each trajectory. The window moves forward by one location until it reaches the end of the sequence: at each step, the locations in the window are gathered as training features. The segment length depends on both specific purposes and dataset restrictions. Its choice is particularly influenced by the time resolution of the trajectories, whereby a higher time resolution leads to a larger window in terms of locations (e.g., 4 h window contains 16 locations if *time_resolution* = 15 min while just four if *time_resolution* = 1 h).

The LSTM model is finally trained with a collection of these fixed-length trajectory segments, encoded as embedding sequences, where each is labeled with the nationality of the user generating it. For example, if the window length is equal to four locations, the location sequence ($LOC_{t-3}, LOC_{t-2}, LOC_{t-1}, LOC_t$) is identified as input features and the corresponding nationality NAT as a target variable. Therefore, to link trajectories to nationalities, the output of the LSTM is fed into a softmax function, as reported in Equation (8), where h_{last} is the output of the LSTM at the last step and n_{NAT} is the total number of nationalities:

$$P(NAT = j | h_{last}) = \frac{\exp(W_j h'_{last} + b_j)}{\sum_{k=1}^{n_{NAT}} \exp(W_k h'_{last} + b_k)}. \quad (8)$$

Given a trajectory sequence T labeled with a nationality NAT , we train the model to maximize the log-likelihood, with respect to any weight in the network, as reported in Equation (9). The model is trained through backpropagation by mini-batch stochastic training.

$$NAT(T) \rightarrow \sum_{T \in NAT} \log(p(NAT|T)). \quad (9)$$

The prediction is therefore based on both the current sequence of locations and historical trajectories of other users and nationalities. The flowchart of the whole process from raw traces to the final classification is illustrated in Figure 3.

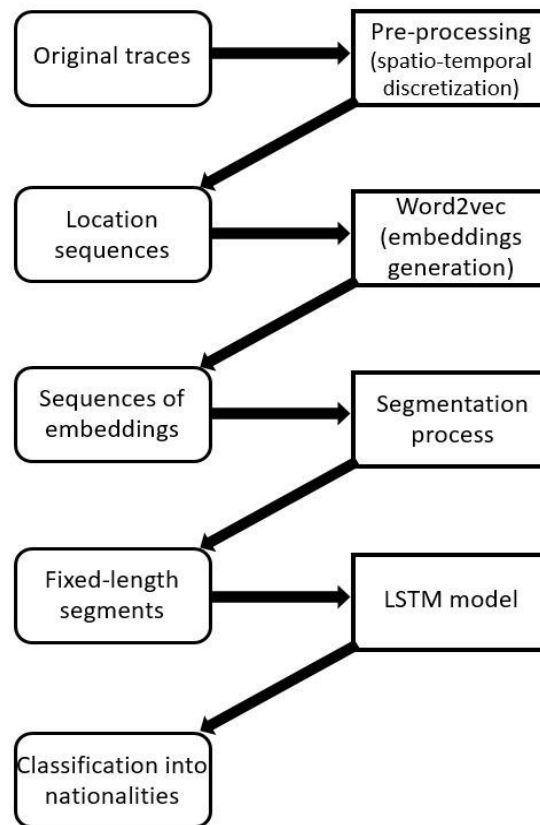


Figure 3. Flowchart of the whole classification process.

4. Experiment

This section first presents the dataset used for the classification task, then describes the experiments conducted and compares the results with baseline approaches. The implementation and training of Word2vec and LSTM was executed on TensorFlow using AWS EC2 p3.2xlarge GPU instance.

4.1. Dataset

To properly depict the general behavior of foreign tourists with a large amount of motion data, we used a real-world dataset of anonymized mobile phone call detailed records (CDRs) of roamers in Italy. Data were provided by a major telecom operator and span the period between the beginning of May to the end of November 2013. Each CDR is related to a mobile phone activity (e.g., phone calls, SMS communication, data connection), enriching the event with a time stamp and the current position of the device, represented as the coverage area of the principal antenna; in addition, each user ID is associated with a mobile country code (MCC). We considered only short-term visitors, located in the country for a maximum of two weeks. CDRs have already been utilized in studies of human mobility to characterize people's behavior and predict human motion [42–45].

The mobile activity pattern of people is usually characterized by an erratic profile of sparse connection events separated by relatively long time gaps. To contrast the resulting trace fragmentation, we pre-processed traces into sequences unfolded in 1 h time steps. If more than one event occurred in the same hour, we selected the location associated with the majority of those events in order to represent the current position of the user. Considering the time step unit chosen, the wide territory under study, and our main interest for large-scale movements, we defined a minimum spatial resolution

of 2 km, aggregating antennas within that distance in a single reference point. We selected the most visited locations as reference points according to the minimum resolution, that is, the antennas with the highest number of connections within 2 km distance, projecting the other antennas to the closest reference point. Furthermore, we removed locations with just a few tens of occurrences. Since they were mostly randomly visited and did not reflect the overall behavior of foreign visitors in Italy, we treated them as a bias in the dataset. In general, the choice of parameters such as time and space resolution can be chosen differently, being highly dependent on the characteristics of the datasets.

We finally obtained 1 h encoded sequences of almost six thousand unique locations over the Italian territory. Since we were interested in categorizing relatively short motion behaviors, which would also allow us to make proper and complete use of the dataset mostly made of short continuous traces, we constructed the fixed-length trajectory segments with a window length equal to 7 h (seven locations). We discarded sequences containing less than seven consecutive locations and also removed those segments that were completely stationary, where the user never moved for the entire 7 h. Our interest is to model large-scale mobility traces representing foreign tourists' motion behavior.

In our classification task, we took into account the motion activity of the top 34 nationalities in terms of amount of data (the nationalities of the great majority of visitors), consisting of 96% of the original dataset. The classification problem was hence defined as associating a 7 h trajectory segment with one of the 34 nationality classes.

The final dataset consists of 12.3 million segments belonging to 1.3 million users. This large number of users and mobility data assured the redundancy of motion patterns related to each nationality. Therefore, the classification task was not performed on the basis of regular schedules of single user behavior, but purely founded on the collective motion of millions of people. Table 1 summarizes the characteristics of the pre-processed dataset.

Table 1. Summary characteristics of the pre-processed dataset.

Num. Users	Num. Trace Segments	Median Displacement per Hour	Median Displacement per Trace Segment	Num. Locations
1.3 millions	12.3 millions	2.6 km	36.5 km	5903

4.2. Experimental Settings

The Word2vec model was implemented with a vector size of 100 dimensions and a window size of three hours (locations) in the past and three in the future. It was trained using a mini batch approach with noise-contrastive estimation loss and Adam optimizer [46,47]. The best parameter combination for the LSTM model was found to be a two-layer stacked LSTM with a hidden size of 4000 neurons per layer, trained using mini-batches, cross-entropy cost function, and Adam optimizer.

In order to evaluate the model, we split the data into a training set and a test set. The test set was used after training the model to determine the performance on previously unseen data and was selected randomly, containing 20% of the users.

To measure the performance, we compared the achieved classification accuracy with three baseline approaches:

- Most Visits. The predicted nationality is the one that visited the locations belonging to the trajectory under observation more times, i.e., summing up for all the nationalities the overall number of visits to each of the seven locations composing the trajectory and selecting the nationality with the highest number of visits. See Equation (10):

$$\hat{NAT}(T) = \underset{NAT_i \in ALL_NAT}{\operatorname{argmax}} \sum_{t=1}^7 \operatorname{count}(NAT_i, LOC_t). \quad (10)$$

- **Most Transitions.** The predicted nationality is the one with the highest number of common transitions with respect to the trajectory under observation, i.e., summing up for each nationality the overall number of transitions between each of the six pairs of consecutive locations in the trajectory and selecting the nationality with the highest number of common transitions. See Equation (11):

$$\hat{NAT}(T) = \underset{NAT_i \in ALL_NAT}{\operatorname{argmax}} \sum_{t=1}^6 \operatorname{count}(NAT_i, (LOC_t \rightarrow LOC_{t+1})). \quad (11)$$

- **Markov model for sequence classification.** The predicted nationality is the result of a Bayesian maximum classifier trained on the output probabilities of each class's first order Markov model. Mobility traces are defined by the concatenation of primitive motion behaviors and transition probabilities are calculated by counting each nationality's transitions between every location. See Equation (12):

$$\hat{NAT}(T) = \underset{NAT_i \in ALL_NAT}{\operatorname{argmax}} p(NAT_i) \prod_{t=2}^7 p((LOC_t | LOC_{t-1}) | NAT_i). \quad (12)$$

4.3. Results

The comparison results are reported in Table 2, showing that the proposed method consistently outperformed the baselines. We evaluated the performances by using accuracy and accuracy in top 3 (if the correct label is in the top three predicted nationalities, the accuracy is 1, otherwise it is 0; the result is the average of those accuracies for each testing trajectory). In terms of exact accuracy, our model yielded a 15% improvement with respect to Markov, the best baseline classifier, and 30% and 33% compared to Most Transitions and Most Visits, respectively. In terms of accuracy in top 3, our model still provided a 12% improvement compared to Markov, 18% to Most Transitions, and 20% to Most Visits.

Table 2. Overall performance comparison.

	Accuracy	Accuracy in Top 3
Most Visits	0.2727	0.5312
Most Transitions	0.2854	0.5457
Markov model	0.3453	0.5864
LSTM model	0.4072	0.6666

Reasonably, Most Visits, which did not consider any location order in the trace, had the lowest scores. However, Most Transition, which took into account the collective common primitive movements, led to only a slight improvement. The Markov model, based on the first order transition probabilities of each nationality, achieved an accuracy of over 7 percentage points greater than Most Visits. On the other hand, LSTM determined a very substantial increment of performance, exceeding the best baseline of over 6 percentage points.

In addition, we studied how the classification performances varied according to different trajectory characteristics. The idea was to evaluate how classification was affected by different values of motion features, such as location changes and traveled distance.

Table 3 reports the accuracy and accuracy in top 3 (in brackets) for different numbers of location changes, in particular if within a time period of 7 h there were one to two changes, three to four changes, or five to six changes. The results show an overall tendency of increasing performance as the number of location changes increases. Comparing baselines, Most Transitions always outperformed Most Visit, and both of them outperformed the Markov model when the number of location changes was very low (one or two changes). On the other hand, the Markov model substantially outperformed

them when the number of changes increased. The LSTM model always outperformed the baselines, but it is worth noting that for very high numbers of location changes, the Markov model lost only 1.2 percentage points of accuracy compared to LSTM.

Table 3. Accuracy (and accuracy in top 3 in brackets) comparison for different numbers of location changes.

Location Changes =	1–2	3–4	5–6
Most Visits	0.2489 (0.4975)	0.2721 (0.5318)	0.3122 (0.5847)
Most Transitions	0.2607 (0.5104)	0.2780 (0.5386)	0.3370 (0.6139)
Markov model	0.2335 (0.4592)	0.3556 (0.6131)	0.5096 (0.7493)
LSTM model	0.3226 (0.5828)	0.4212 (0.6838)	0.5215 (0.7691)

Table 4 reports the accuracies with respect to different values of traveled distance, in particular for bins of ≤ 10 km, 10–25 km, 25–50 km, and ≥ 50 km. In this case, a clear tendency of increasing performance as the traveled distance increases is observable only for the Markov and LSTM models. As in the previous case, Most Transitions always outperformed Most Visits. The Markov model performed very poorly for short distances (< 25 km), but achieved a remarkable performance for very long distances (≥ 50 km). LSTM highly outperformed every baseline for short and long distances, although achieved very similar performances to the Markov model for very long distances.

Table 4. Accuracy (and accuracy in top 3 in brackets) comparison for different values of traveled distance.

Travel Distance =	≤ 10 km	10–25 km	25–50 km	≥ 50 km
Most Visits	0.2443 (0.4849)	0.2757 (0.5368)	0.2953 (0.5604)	0.2753 (0.5386)
Most Transitions	0.2557 (0.4986)	0.2848 (0.5510)	0.3002 (0.5700)	0.2938 (0.5557)
Markov model	0.1813 (0.3779)	0.2398 (0.4999)	0.3078 (0.5890)	0.4922 (0.7287)
LSTM model	0.2868 (0.5385)	0.3533 (0.6267)	0.4023 (0.6768)	0.4939 (0.7439)

Performances can finally be explored with respect to the imbalance of the nationality classes in the dataset. Table 5 reports the macro-average F1-score for nationalities in different ranges of amount of data. The columns from left to right refer to the nationalities, each of them representing, respectively, over 5% of the whole dataset (five nationalities), between 1% and 5% (ten nationalities), between 0.5% and 1% (nine nationalities), and less than 0.5% (ten nationalities). As expected, count-based baselines performed very poorly for rare classes, while LSTM, although dropping some performance with respect to nationalities with a large amount of data, still retained acceptable results even for very rare classes, outperforming the other models.

Table 5. Macro-average F1-score for nationalities in different ranges of amount of data. The percentage value in the first row refers to the amount of data represented by each nationality in that column with respect to the whole dataset.

Amount of Data:	>5% (Five Nationalities)	1–5% (Ten Nationalities)	0.5–1% (Nine Nationalities)	<0.5% (Ten Nationalities)
Most Visits	0.2567	0.0910	0.0247	0.0328
Most Transitions	0.2781	0.1236	0.0306	0.0549
Markov model	0.3776	0.3275	0.2556	0.2344
LSTM model	0.4110	0.3400	0.2602	0.2784

4.4. Discussion

We designed a method for inferring foreign tourists' nationalities from large-scale mobility traces using location embeddings and LSTM neural network. We demonstrated the hypothesis that different nationalities may not only visit different areas over the territory, but also visit the same locations in a different order, hence proving that the way people move is a good indication of their origins.

In particular, results show that baseline approaches relying only on the cumulative number of location visits or transitions, therefore representing the overall presence of nationalities over the territory, perform poorly. The Markov model for sequence classification, taking into account each nationality's probabilities of location changes, achieves better results, but its behavior is highly sensitive to motion characteristics. LSTM, specifically designed to find patterns along sequences, substantially outperforms each of the other models, demonstrating the feasibility of correctly identifying nationalities of individual users based on ordered location sequences representing their mobility traces. This reveals that different nationalities move in different ways over the territory.

Moreover, the influence of motion characteristics in mobility traces suggests a higher predictability for more distinctive trajectories—that is, for a high number of location transitions or a long traveled distance. This means that highly overlapping motion behaviors between different nationalities (e.g., short movements and high stationarity) negatively affects predictability. This trend is particularly visible for the Markov and LSTM models, highly improving performance as the number of location changes increases or the value of traveled distance grows. Distinctive paths and characteristic traces are more predictable than local movements and stationary behaviors: while many nationalities may move in a similar way in the context of urban activities, the frequent routes of each nationality become more specific and recognizable when it comes to larger-scale mobility. However, LSTM outperforms the best baseline results for both low and high stationary traces, and both short and long traveled distances, grasping more information than count-based baselines for short and stationary traces, significantly beating any baseline for longer traces and more location transitions, and slightly outperforming the Markov model for very long and highly non-stationary trajectories.

Another issue that is worth mentioning is related to the class imbalance—although it is preferable to correctly detect the most prevalent nationalities, it is important to verify that the model does not completely drop in performance for very rare classes. The drastic performance imbalance for common and rare classes discloses the capability of a model of correctly detecting only the very few nationalities with a large amount of data. In general, results report a tendency of obtaining a better performance for nationalities with a large amount of data, implying that it is easier to find reliable patterns when the presence of visitors is higher, and harder to properly characterize tourists' motion behavior in cases of rare classes. However, LSTM still performs better than baselines, obtaining acceptable results even for very rare classes. This is especially true when compared to the count-based models, which, relying on cumulative counting, drop their performances significantly for a small number of data points.

In conclusion, LSTM and location embeddings have the advantages of properly identifying individual users' nationalities uniquely on the basis of how tourists move over the territory. This is suitable for applications related to human trajectory analysis, in particular to the study of touristic motion behaviors. Knowing the nationality of a tourist in a foreign country can help in personalized recommendation systems and trajectory prediction models, allowing the management of services and resources on the basis of visitors' profiles. More generally, this work fits in the context of trajectory labeling and user profiling, using mobility traces as a way of inferring information about people, demonstrating how motion behavior can be a useful tool to identify particular user characteristics. Finally, we highlighted the potential of deep learning on mobility traces: the combination of vector representations of meaning for modeling locations and LSTM for analyzing trajectories was revealed to be a powerful methodology for motion pattern recognition.

5. Conclusions

In this paper, we presented a new way to mine human mobility patterns, which aims at identifying short-term tourists' nationalities from location-based trajectories. The proposed model was designed to capture the dependency of track points and to infer the latent patterns of users. We first transformed original trajectories into sequences of locations, unfolding in fixed time steps, then a Skip-gram Word2vec model was used to construct the location embeddings, and finally we applied an LSTM neural network model for correctly labeling each sequence as the nationality of the user generating it.

Defining the problem as a multinomial classification task, the reported methodology was shown to substantially outperform baselines, achieving promising results in terms of correct nationality detection.

Potential extensions of this paper can go in multiple different directions. The first issue that is worth studying is the role of individual travelers and organized groups. Although the dataset used did not contain significant portions of synchronized traces (sequences with same place-time), with the exception of stationary traces, the granularity of the data was insufficient to detect the coordinated motion of groups with certainty. Therefore, the possible role of group motion in some specific situations is a valid motivation for a further investigation, which would require a more granular dataset. In addition, the study of tourists' motion activity at a smaller scale could be an interesting step to evaluate if finer trajectories in space and time (e.g., in an urban environment) still make it possible to identify visitors' information, such as their nationality; in particular, GPS data would allow finer resolutions than telecom data. Another direction could be to integrate explicit time information into the location sequences for assessing a possible performance improvement, or even analyzing detection variation over time (e.g., month by month). A final direction is to explore different types of information detection for user profiling. The same methodology could be utilized to infer user information in different use cases, not limited to tourism analysis.

To conclude, the use of embeddings and LSTM, commonly adopted in the field of NLP, can potentially be successful in a wide range of applications dealing with mobility traces, and therefore extended to various tasks related to trajectory analysis and human motion behavior.

Author Contributions: A.C. conceived and designed the experiments, analyzed the data and wrote the paper. E.B. supervised the work, helped with designing the conceptual framework, and edited the manuscript.

Funding: This research was funded by the Austrian Science Fund (FWF) through the Doctoral College GIScience at the University of Salzburg (DK W 1237-N23).

Acknowledgments: The authors would like to thank Vodafone Italia for providing the dataset for the case study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Feng, Z.; Zhu, Y. A survey on trajectory data mining: Techniques and applications. *IEEE Access* **2016**, *4*, 2056–2067. [[CrossRef](#)]
2. Zheng, Y. Trajectory data mining: An overview. *J. ACM Trans. Intell. Syst. Technol.* **2015**, *6*, 1–41. [[CrossRef](#)]
3. Bhargava, P.; Phan, T.; Zhou, J.; Lee, J. Who, what, when, and where: Multi-dimensional collaborative recommendations using tensor factorization on sparse user-generated data. In Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 18–22 May 2015; pp. 130–140.
4. Zhuang, J.; Mei, T.; Hoi, S.C.; Xu, Y.-Q.; Li, S. When recommendation meets mobile: Contextual and personalized recommendation on the go. In Proceedings of the 13th International Conference on Ubiquitous Computing, Beijing, China, 17–21 September 2011; pp. 153–162.
5. Zheng, Y.; Li, Q.; Chen, Y.; Xie, X.; Ma, W.-Y. Understanding mobility based on GPS data. In Proceedings of the 10th International Conference on Ubiquitous Computing, Seoul, Korea, 20–23 September 2008; pp. 312–321.
6. Dodge, S.; Weibel, R.; Ahearn, S.C.; Buchin, M.; Miller, J.A. Analysis of movement data. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 825–834. [[CrossRef](#)]
7. Chen, D.; Ong, C.S.; Xie, L. Learning points and routes to recommend trajectories. In Proceedings of the 25th ACM International Conference on Information and Knowledge Management, Indianapolis, IN, USA, 24–28 October 2016; pp. 2227–2232.
8. Gonzalez, M.C.; Hidalgo, C.A.; Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779. [[CrossRef](#)] [[PubMed](#)]
9. Schneider, C.M.; Belik, V.; Couronné, T.; Smoreda, Z.; González, M.C. Unravelling daily human mobility motifs. *J. R. Soc. Interface* **2013**, *10*, 20130246. [[CrossRef](#)] [[PubMed](#)]
10. Liu, X.; Liu, Y.; Aberer, K.; Miao, C. Personalized point-of-interest recommendation by mining users' preference transition. In Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, San Francisco, CA, USA, 27 October–1 November 2013; pp. 733–738.

11. Gruenerbl, A.; Osmani, V.; Bahle, G.; Carrasco, J.C.; Oehler, S.; Mayora, O.; Haring, C.; Lukowicz, P. Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients. In Proceedings of the 5th Augmented Human International Conference, Kobe, Japan, 7–9 March 2014; p. 38.
12. Haering, N.; Venetianer, P.L.; Lipton, A. The evolution of video surveillance: An overview. *Mach. Vis. Appl.* **2008**, *19*, 279–290. [[CrossRef](#)]
13. Zhu, Y.; Zheng, Y.; Zhang, L.; Santani, D.; Xie, X.; Yang, Q. Inferring taxi status using GPS trajectories. *arXiv Preprint* **2012**, arXiv:1205.4378.
14. Barlacchi, G.; Perentis, C.; Mehrotra, A.; Musolesi, M.; Lepri, B. Are you getting sick? Predicting influenza-like symptoms using human mobility behaviors. *EPJ Data Sci.* **2017**, *6*, 27. [[CrossRef](#)]
15. Canzian, L.; Musolesi, M. Trajectories of depression: Unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 7–11 September 2015; pp. 1293–1304.
16. Bauer, G.; Lukowicz, P. Can smartphones detect stress-related changes in the behaviour of individuals? In Proceedings of the 2012 IEEE International Conference on Pervasive Computing and Communications Workshops, Lugano, Switzerland, 19–23 March 2012; pp. 423–426.
17. Waga, K.; Tabarcea, A.; Chen, M.; Fränti, P. Detecting movement type by route segmentation and classification. In Proceedings of the 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), Pittsburgh, PA, USA, 14–17 October 2012; pp. 508–513.
18. Mlich, J.; Chmellar, P. Trajectory classification based on hidden markov models. In Proceedings of the 18th International Conference on Computer Graphics and Vision, Moscow, Russia, 23–27 June 2008; pp. 101–105.
19. Bashir, F.I.; Khokhar, A.A.; Schonfeld, D. Object trajectory-based activity classification and recognition using hidden Markov models. *IEEE Trans. Image Process.* **2007**, *16*, 1912–1919. [[CrossRef](#)] [[PubMed](#)]
20. Nascimento, J.C.; Figueiredo, M.A.T.; Marques, J.S. Trajectory classification using switched dynamical hidden Markov models. *IEEE Trans. Image Process.* **2009**, *19*, 1338–1348. [[CrossRef](#)] [[PubMed](#)]
21. Gao, Q.; Sun, S. Trajectory-based human activity recognition with hierarchical Dirichlet process hidden Markov models. In Proceedings of the 2013 IEEE China Summit and International Conference on Signal and Information Processing, Beijing, China, 6–10 July 2013; pp. 456–460.
22. Vail, D.L.; Veloso, M.M.; Lafferty, J.D. Conditional random fields for activity recognition. In Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems, Honolulu, HI, USA, 14–18 May 2007; p. 235.
23. Gao, Q.-B.; Sun, S.-L. Trajectory-based human activity recognition using hidden conditional random fields. In Proceedings of the 2012 International Conference on Machine Learning and Cybernetics, Xian, China, 15–17 July 2012; pp. 1091–1097.
24. Dabiri, S.; Heaslip, K. Inferring transportation modes from GPS trajectories using a convolutional neural network. *Transp. Res. Part C Emerg. Technol.* **2018**, *86*, 360–371. [[CrossRef](#)]
25. Endo, Y.; Toda, H.; Nishida, K.; Kawanobe, A. Deep feature extraction from trajectories for transportation mode estimation. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Auckland, New Zealand, 19–22 April 2016; pp. 54–66.
26. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.
27. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
28. Andrienko, N.; Andrienko, G.; Pelekis, N.; Spaccapietra, S. Basic Concepts of Movement Data. In *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*; Giannotti, F., Pedreschi, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 15–38. [[CrossRef](#)]
29. Urner, J.; Bucher, D.; Yang, J.; Jonietz, D. Assessing the Influence of Spatio-Temporal Context for Next Place Prediction using Different Machine Learning Approaches. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 166. [[CrossRef](#)]
30. Bullinaria, J.A.; Levy, J.P. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behav. Res. Methods* **2007**, *39*, 510–526. [[CrossRef](#)] [[PubMed](#)]
31. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 3111–3119.

32. Pennington, J.; Socher, R.; Manning, C. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
33. Crivellari, A.; Beinat, E. From Motion Activity to Geo-Embeddings: Generating and Exploring Vector Representations of Locations, Traces and Visitors through Large-Scale Mobility Data. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 134. [[CrossRef](#)]
34. Graves, A.; Mohamed, A.; Hinton, G. Speech recognition with deep recurrent neural networks. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 6645–6649.
35. Mikolov, T.; Karafiát, M.; Burget, L.; Černocký, J.; Khudanpur, S. Recurrent neural network based language model. In Proceedings of the 11th Annual Conference of the International Speech Communication Association, Makuhari, Japan, 26–30 September 2010.
36. Sutskever, I.; Martens, J.; Hinton, G. Generating text with recurrent neural networks. In Proceedings of the 28th International Conference on Machine Learning, Bellevue, WA, USA, 28 June–2 July 2011; pp. 1017–1024.
37. Boulanger-Lewandowski, N.; Bengio, Y.; Vincent, P. Audio Chord Recognition with Recurrent Neural Networks. In Proceedings of the 14th International Society for Music Information Retrieval Conference, Curitiba, Brazil, 4–8 November 2013; pp. 335–340.
38. Kahou, S.E.; Michalski, V.; Konda, K.; Memisevic, R.; Pal, C. Recurrent Neural Networks for Emotion Recognition in Video. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, Seattle, WA, USA, 9–13 November 2015; pp. 467–474.
39. Graves, A.; Jaitly, N.; Mohamed, A. Hybrid speech recognition with Deep Bidirectional LSTM. In Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, Olomouc, Czech Republic, 8–12 December 2013; pp. 273–278.
40. Sundermeyer, M.; Schlüter, R.; Ney, H. LSTM neural networks for language modeling. In Proceedings of the 13th Annual Conference of the International Speech Communication Association, Portland, OR, USA, 9–13 September 2012; pp. 194–197.
41. Wu, Y.; Schuster, M.; Chen, Z.; Le, Q.V.; Norouzi, M.; Macherey, W.; Krikun, M.; Cao, Y.; Gao, Q.; Macherey, K.; et al. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv* **2016**, arXiv:1609.08144.
42. de Montjoye, Y.-A.; Quoidbach, J.; Robic, F.; Pentland, A. Predicting Personality Using Novel Mobile Phone-Based Metrics. In Proceedings of the Social Computing, Behavioral-Cultural Modeling and Prediction, Washington, DC, USA, 2–5 April 2013; pp. 48–55.
43. Lu, X.; Bengtsson, L.; Holme, P. Predictability of population displacement after the 2010 Haiti earthquake. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11576. [[CrossRef](#)] [[PubMed](#)]
44. Hawelka, B.; Sitko, I.; Kazakopoulos, P.; Beinat, E. Collective prediction of individual mobility traces for users with short data history. *PLoS ONE* **2017**, *12*, e0170907. [[CrossRef](#)] [[PubMed](#)]
45. Sundsøy, P.; Bjelland, J.; Reme, B.A.; Iqbal, A.M.; Jahani, E. Deep learning applied to mobile phone data for individual income classification. In Proceedings of the 2016 International Conference on Artificial Intelligence: Technologies and Applications, Bangkok, Thailand, 24–25 January 2016. [[CrossRef](#)]
46. Mnih, A.; Kavukcuoglu, K. Learning word embeddings efficiently with noise-contrastive estimation. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 2265–2273.
47. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

