Multi-Channel Convolutional Neural Networks for Handling Multi-Dimensional Semantic Trajectories and Predicting Future Semantic Locations

Antonios Karatzoglou^[0000-0002-7939-1408]

Robert Bosch GmbH, Chassis Systems Control, Advance Engineering Sector, Abstatt, Germany antonios.karatzoglou@de.bosch.com

Abstract. Current location-aware systems rely increasingly on location prediction techniques in order to provide their services in a timely fashion. At the same time, it has been shown that the use of additional context information, that is, elevating the degree of semantic enrichment of movement data, can lead to significant better results both for analyzing as well as for modeling human trajectories and predicting upon them. In this work, we propose a Multi-Channel Convolutional Neural Network (CNN) based approach for capturing all the existing context dimensions in our semantic trajectory dataset aiming at achieving a higher prediction accuracy compared to a vanilla Single-Channel CNN. Moreover, we investigate whether and to what degree time, activity, companionship and the user's emotional state have an impact on the predictive performance of our multi-dimensional CNN. We evaluated our model on a real-world dataset and compared it to a probabilistic Markov Chain model and a vanilla CNN at two semantic representation levels. It can be shown that especially for the high level case, the present approach is able to outperform the baseline models achieving an overall higher accuracy and f-Score.

Keywords: Semantic Locations and Trajectories · Purpose of Visit · Human Activities · Emotional States · Affective Computing · Multi-Channel Convolutional Neural Networks.

1 Introduction

The market of context-aware, and especially, location-aware computing and services (see *Location-Based Services* (LBS)) has gained enormously in importance over the past few decades. In their attempt to provide timely solutions to their users, LBS providers rely more and more on location prediction methods, a fact that additionally strengthened the demand for accurate location prediction algorithms.

2 A. Karatzoglou

Typical location prediction models are usually merely data driven and depend therefore heavily on the size and the quality of the available training dataset. However, recent research has shown that the use of additional semantic information can help overcome, at least to some extent, the aforementioned issues and can therefore lead to an overall better predictive performance (see Section 2). That is, in the case of modeling and learning human movement patterns, models are fed and trained with the users' (GPS) location trajectories. Additional context information, such as the location type and the user's activity, may be used to enrich them semantically and generate so-called *semantic trajectories* (see Section 3). This type of extensive input helps the model dive even deeper into the users's movement behaviour and can lead to more accurate predictions.

Common approaches used to model and predict human movement include probabilistic methods, such as Markov Chains [2,9], Dynamic Bayes Networks [8], Hidden Markov Models [25,30] and Artificial Neural Networks (ANNs). In the latter case, recurrent neural network architectures (RNNs) have generally proved to perform above the average when it comes to learning sequences and for this reason these are commonly found in the location prediction domain as well. Especially memory-based neural network types, like the Long Short-Term Memory network (LSTM), are capable of achieving high prediction rates and tend to outperform the competition [27, 20].

While recurrent network types are the preferred choice when it comes to modeling 1-dim movement, recent work showed some promising results on the part of Convolutional Neural Networks (CNNs) [19, 24], a model normally used in the 2-dim image classification and object recognition domain. It seems that the locally focused nature of the kernel-based convolution process enables the CNN model to capture existing dependencies between current and future locations. The presented work builds upon the aforementioned work and aims at investigating the use of a *multi-channel CNN* based approach with regard to modeling multi-dimensional semantically enriched location data and predicting the next semantic location of the user. In particular, our semantic trajectories consist of the following feature dimensions: semantic location type, time, human activity, emotional state and companionship. Moreover, this work further explores the impact of the degree of semantic enrichment, that is, whether and to what extent each of the aforementioned dimensions influences the predictive performance of our model. We evaluated our approach using a real-world dataset, which we collected from 21 users by conducting a 2-months long user study. In addition, we selected a 1. Order Markov Chain model and a vanilla CNN as baseline models to compare with.

This paper is structured as follows. Section 2 provides a short overview over some of the most related work in the semantic trajectories and location prediction domain. Next, Section 3 describes the notion of *semantic trajectories* and *semantic locations* with respect to this work. Section 4 goes briefly through the theory behind Convolutional Neural Networks and discusses in detail the proposed approach, while Section 5 provides the respective evaluation outcomes. Finally, Section 6 summarizes the evaluation results and draws some final conclusions.

2 Related Work

There exist many different ways of viewing movement data. Within the scope of mining and analyzing movement patterns, Spaccapietra et al. introduced with [28] one of the first works that make the importance of viewing trajectories of moving objects in a conceptual manner clear. In their work, they highlighted the fact that describing certain aspects of the movement's context by adding semantic information into the available trajectories can significantly support the analysis of the respective movement patterns, as well as the querying process among them. Alvarez et al. came to the same conclusion as they suggested the use of a similar semantic enrichment model to generate semantic trajectories for the same reasons [1]. The added value of working on semantically enriched trajectory data in comparison to working on raw data with regard to mining patterns and supporting decision processes has been underpinned by Elragal et al. as well [7]. Bogorny et al.'s work focuses in mining trajectory patterns as well and introduced in [3] a sophisticated model, which in contrast to former models is capable of handling complex queries over semantic trajectories, while providing different semantic granularities at the same time.

Due to the aforementioned benefits that semantic enrichment brings with it, a number of location prediction papers have recently emerged presenting algorithms that rely on the notion of semantic trajectories. Ying et al. for instance were one of the first to build upon semantic trajectories generated from the users' raw GPS recordings in order to enhance their location prediction framework [32] with promising results. Some years later, they extended their model by taking, apart from geographic and semantic patterns, temporal patterns into account as well [31].

Karatzoglou et al.'s work explores a big variety of models with respect to modeling human semantic trajectories and predicting the user's next semantic location. In [11] and in [18] they evaluate a multi-dimensional Markov Chain model for predicting among activity-enriched semantic trajectories and show that is able to outperform Ying et al.'s framework in terms of accuracy. With regard to recall however, they could identify certain limitations on behalf of the model due to its adverse dependency on the small size and the sparsity of the available training dataset. They attempt to solve this issue by combining the probabilistic Markov Chain model with Matrix Factorization in [12], where they were able to raise the recall scores.

In [20], [17], [13] and [19], Karatzoglou et al. investigate the performance of Artificial Neural Networks using the probabilistic Markov model as baseline. In addition, they explore the role of the semantic granularity of the considered trajectories in the overall performance of the networks. They show that the higher the semantic level is, the better the modeling quality of the networks. While the findings in [20] comply with the results of related work showing that Long Short-

4 A. Karatzoglou

Term Memory networks are generally able to outperform the vanilla Recurrent (RNN) and Feed-Forward ones, [17] indicates no great advantages towards the attention-based application of Sequence to Sequence learning (Seq2Seq) compared to the standard single-input-single-output LSTM model of [20]. Yao et al. propose in [29] a similar to [20] LSTM-based recurrent approach for predicting next semantic locations using an additional embedding input layer, the benefits of which have been also recently shown by Gao et al. in [10]. Other than in [20] and following a similar direction to the approach proposed in the present paper, Yao et al. used beyond location and time the content of the checkins of the users to enrich the users' semantic trajectories, which describe in a way their activity that we're considering in this work as well (among others). However, in contrast to the Reality Mining dataset [6] used in Karatzoglou et al.'s work, they evaluate their approach on rather long-term dependencies using a Foursquare and a Twitter dataset. In [13], Karatzoglou et al. take a look at a gradient-free optimization method for finding the optimal hyperparameter set of a LSTM model based on an evolutionary algorithm. Their work provides some preliminary results indicating among others the temporal efficiency on part of the genetic, population-based optimization method, provided the availability of sufficient computational power.

However, the most striking findings come rather from [19], where a Convolutional Neural Network based approach in combination with an embedding layer as its input is capable of achieving higher prediction scores than the FFNN, the RNN and the LSTM. To our knowledge this represents the only work that explores the use of CNNs with respect to modeling and predicting upon 1-dim semantic trajectories. The closest work to [19] would be the work of Lv et al. in [24], which evaluates the use of a CNN for modeling and predicting large-scale taxi trajectories. Other than in [19] and the present paper, Lv et al. work with raw GPS data without using any semantic information and map past trajectory data onto 2-dim images before feeding them into the CNN model, transforming in this way the trajectory modelling task into an image classification task. In the present paper, following the example of [19], we skip this kind of 1-dim to 2-dim intermediate transformation step and apply our CNN model on the 1-dim semantic trajectory as it is. As in [19], we build our approach upon similar CNN-based work on 1-dim data, coming mostly from the Natural Language Understanding (NLP) domain, such as the framework described in [5] and the multi-channel CNN model of [21].

3 Semantic Trajectories

The term *trajectory* refers to a sequence of consecutive location points traversed by a moving object within a certain time interval. Eq. 1 describes a typical GPS trajectory with each location point being represented by a tuple containing its coordinates $(long_i, lat_i)$ and the corresponding point of time t_i at which this was visited.

$$Traj_{GPS} = (long_1, lat_1, t_1), (long_2, lat_2, t_2), ..., (long_i, lat_i, t_I),$$
(1)

As already mentioned in Section 2, in order to better understand the moving behaviour of moving objects and create more accurate models, Spaccapietra et al. [28] and Alvares et al. [1] went beyond this kind of numerical sequences by adding a semantic view upon them and introduced the so-called *semantic trajectories*. Starting initially with the simple notion of "stops" and "moves", a (human) semantic trajectory can nowadays be defined generally as a sequence of semantically significant locations (*semantic locations*, e.g., "home", "burger joint", etc.) as follows:

$$Traj_{Sem} = (SemLoc_1, t_1), (SemLoc_2, t_2), \dots, (SemLoc_i, t_i)$$

$$(2)$$

A significant location in this case is usually defined by a location within a certain radius (e.g., 200m) where a user stays longer than a pre-defined temporal threshold, e.g. 20min (see [2]). Some researchers add further thresholds, like the loss of the GPS signal due to entering into a building, the GPS recording stop [4] or the popularity, in order to extract the most significant common or public locations [32]. In this work, we evaluate our method using a dataset in which the users annotated their longest visits (> 15min) themselves (see Section 5).

Depending on the number of the considered semantic features, a semantic trajectory can have multiple dimensions. Thus, we could say that the number of dimensions expresses the *degree of semantic enrichment* of the respective semantic trajectory. In this work, we follow the concept of the Location-Specific Cognitive Frames introduced by Karatzoglou et al. in [14, 15] and we consider each stop at a semantic location to be a tuple encapsulating the current location type, the current time, the current activity, as well as the user's current emotional state and whether he is alone or not (companionship). Beyond that, locations can be described differently depending on the semantic representation level, e.g., "restaurant" \rightarrow "fast food restaurant" \rightarrow "burger joint". In this work, we evaluate the modeling performance of CNNs at two different semantic levels.

4 Multi-Channel Convolutional Neural Networks on Semantic Trajectories

This section consists of two parts. The first part gives a brief insight into the theory behind Convolutional Neural Networks and goes briefly through some of the most common CNN steps and layers using the example of image classification. Then, the second and last part describes in detail the architecture of the multi-channel CNN model proposed in this paper for modeling and predicting upon multi-dimensional semantic trajectories.

Convolutional Neural Networks (CNN) represent the standard choice in the image classification and object recognition domain [23]. However, this doesn't mean that it is the only domain in which we can apply them expecting reasonable



Fig. 1. Typical CNN architecture for the image classification task (source: [26]).

results as we saw in Section 2 and can also be seen in [22]. Fig. 1 illustrates a typical CNN pipeline used for classifying images.

A typical CNN consists of many different layers starting usually with the (first) convolutional layer. This layer is responsible for convolving the input image and generating the so-called *feature maps*. This is done by sliding a group of small-sized *filters* (also called *kernels*) with each containing a certain number of learnable weights over the input image and performing element-wise multiplication at each possible position. The generated feature map from each kernel is a new layer and contains the findings of the particular kernel in the input image, ideally with respect to a specific single distinguished feature. The number of kernels defines the number of the generated feature maps (so-called *depth* of the convolutional layer) and represents a CNN hyperparameter which needs to be selected appropriately based on the available data and task. In the next step, this resulting group of layers undergoes a so-called *pooling* process. Pooling refers to a downsampling operation, in which sets of elements in the feature maps are combined and reduced to a single value based on some criterion (e.g., take the maximum value: max pooling) or calculation (e.g., take the average over all values: average pooling). The two aforementioned layers can be repeated multiple times using different kernels of different size and depth. This supports the successive extraction of higher level features and represents one of the strengths of CNNs. Finally, the last pooled layer can be *flattened* into a single vector containing all its weights and connected to a *fully connected layer*, which is further connected to the output layer that contains a field for every possible class and provides us with the classification estimation for the given input.

The multi-channel approach introduced in this paper builds upon the aforementioned typical CNN architecture and extends it by adding a further *embedding layer* into the model and *raising the number of channels* matching the degree of semantic enrichment of our data (see Section 3). Fig. 2 illustrates the architecture of our approach. Our framework takes as input a part of a semantic trajectory, which is in our case a sequence of tuples in the form of (*locationtype*, *purposeofvisit*, *time*, *emotionalstate*, *companionship*) according to a predefined temporal horizon t_n that determines how far backwards in



Fig. 2. Multi-Channel CNN for modeling multi-dimensional semantic trajectories.

the movement history of the user should the model consider for providing an estimation about her next semantic location. In a first step, every single feature type is encoded as a one-hot vector. The additional embedding layer between the one-hot encoded input and the convolutional layer maps the sparse asymmetrical one-hot encoded binary vectors into dense vector representations in a continuous vector space. This fact contributes to a more efficient training and helps improving the prediction accuracy while keeping the model consistent at the same time. The number of dimensions of the vector space is selected based on the properties of the available data, e.g., the number of unique classes of the corresponding feature. In our case, each semantic feature is encoded separately, and therefore the generated vectors may have different number of dimensions.

Raising the number of channels according to the semantic enrichment degree of our trajectories represents an intuitive way of viewing upon them. Each channel handles solely the corresponding semantic dimension. For example the first channel is responsible for the location type, the second channel for the purpose of visiting that location (activity), the third one for covering temporal information and so on. At the end, all channels are merged into a single representation, flattened and forwarded to the output layer in order to provide a final prediction about the next semantic location of the user. It should be noted here that the kernels' depth should match the channel dimension (5 in our case).

Our CNN has one convolutional, one pooling, a flattening, a fully connected and a Softmax output layer. A deeper architecture, that is, adding more layers led in most of the cases to overfitting and reduced the overall performance of our model. Other than the CNN model in Fig. 1, our model executes a 1-dim convolution operation instead of the typical 2-dim operation conducted in the

7

image classification task. Each kernel convolves each semantic dimension in one direction only, namely according the chronological order found in the sequence fed into the model. Thus, the width of each kernel covers the whole row of the CNN input matrices while its height can vary, constituting a further hyperparameter of our model. A higher height indicates a kernel, able to observe a higher number of consecutive locations at the same time, a fact that can be useful when aiming at capturing long-term dependencies in our data, and vice versa.

5 Evaluation

In order to evaluate our approach, we conducted a 8-week long user study tracking 21 users via a tracking and annotation app. The participants of the study were asked to semantically label each significant stop (location type) during their movement, as well as to note the purpose of visiting the certain location (e.g., activity), their companion (if any) and their emotional state by selecting among the following states: happy, hungry, neutral, sleepy, energetic, frustrated, stressed, bored, adventurous, ill, sad, angry and shocked. At the end of the study we end up with approximately 1400 annotated locations covering around 70 unique location types, 53 unique activities, and all 13 emotional states. A thorough description of the user study can be found in [16].

In order to take time into account, we defined $24 \times 7 = 168$ hourly slots, which similar to the other input signals were one-hot encoded first and transformed into an embedding vector in a next step. However, our evaluation results showed that taking time into account had a severe negative impact on the prediction outcome of our model. We saw a similar behaviour in the work of Karatzoglou et al. in [20] and in [18]. This can be mainly attributed to the small size of our dataset which makes it extremely hard for the model to find temporal patterns in this 168-slot temporal granularity. The use of wider time slots, e.g., the use of just daily slots, couldn't yield significantly higher scores either, due to the fact that our 8-week long evaluation dataset contains solely 8 recordings from each day, that is, there exist solely 8 unique Mondays, 8 unique Tuesdays, etc. For this reason, and due to space reasons, this evaluation section neglects to refer thoroughly to the individual results with respect to time. In addition, our users provided very little information regarding their type of companionship (e.g., relative, friend, etc.). Solely the fact whether a user was alone or not can be reliably extracted from our dataset. Therefore, instead of handling the companionship in a separate channel, we integrated the particular information into the emotional state one-hot vector by extending it to a further dimension ('0', when the user is alone and '1' when he is not). Finally and as already mentioned previously in this work, we evaluated our approach at two semantic representation levels, which will be referred to as low and high level, with the latter being more abstract and subsuming the first one. In order to generate these two layers, we built a corresponding location taxonomy based on the Foursquare venue categorization¹. Lastly, a grid search

¹ https://developer.foursquare.com/docs/resources/categories

Kernel	Number	Embedding	Dropout	Batch	Learning	Number of	Sequence	Pooling
size	of Kernels	dimension	Probability	size	rate	epochs	length	size
6	64	100	0.6	16	0.001	100	10	2

Table 1. Optimal hyperparameter set determined via grid search.

helped us to determine the following optimal hyperparameter configuration for our model listed in Table 1.

Fig. 3 compares the result from 5 different models at the higher representation level, a standard 1-channel CNN (*Locations*) that takes just the current semantic location as input, a 2-channel CNN that considers the location type and the purpose of visit (*Location&Purpose*), a 2-channel CNN that considers the location type and the emotional state as well as the companionship status of the user (*Location&Companion&Mood*), a 3-channel CNN that takes location type, purpose of visit, emotional state and companionship (*Location&Purpose&Companion& Mood*), and a probabilistic Markov Chain model of 1.order that serves as our reference. It can be seen that all CNN-based are able to outperform the Markov



Fig. 3. Accuracy and f-Scores at the higher semantic representation level.

model both in terms of accuracy and f-Score. What stands out in the figure is that the 2-channel CNN approach that considers the activity of the user (purpose of visit) can clearly outperform the competition. However, this doesn't hold for the other 2-channel CNN model. On the contrary, it seems that taking the user's mood and emotional state into account affects negatively our prediction performance. Apparently, our model can't establish an association between mood and the users' movement behaviour, a fact that could partly be attributed to the small size of our dataset. The more "sophisticated" 3-channel CNN achieves a similar accuracy to the standard CNN, but a lower f-Score and can't really compete with the *Locations&Purposes* model. Its results are likely to be related

10 A. Karatzoglou

to the aforementioned negative impact of the emotional state when this is taken into account.

Fig. 4 presents the results for the lower semantic representation level. It is



Fig. 4. Accuracy and f-Scores at the lower semantic representation level.

apparent that all models perform worse than in the higher level. This can be mainly attributed to the fact that the lower semantic representation brings more unique classes with it to predict, which makes the learning process of the models much harder. At the same time, another possible explanation for this might be the fact that human movement shows stronger regularities at higher levels, e.g., a user may often visit a food location after going to gym, regardless whether this location is an italian or a greek restaurant, a pizza house or a burger joint. Similar to the high level case, the CNN models outperform in most of the cases the probabilistic baseline. However, this time, other than in the higher level, it seems that the additional channels result in a deterioration of our prediction models. The more channels, the worse the predictive behaviour seems to become. In general, due to the small size of our data, all of our models had to deal with massive overfitting issues (see below). Adding a dropout layer while making our model simpler by reducing the size of our layers could improve significantly our models, but only to a certain extent.

Fig. 5, 6, 7 and 8 illustrate the training behaviour of our 4 CNN models. We can see that the greater the number of channels and thus, the greater the semantic enrichment degree of the trajectory, the faster and smoother the training of the CNN model becomes. Taking additional context dimensions into account seems to contribute to shorter convergence times and result in a more efficient training. The 3-channel CNN is characterized by the shortest convergence, while the vanilla 1-channel CNN straggles with the loss reduction along the whole training process to the 100th epoch. The benefits of the multiple channel approach can be clearly seen during the harder learning task, namely at the lower semantic representation level (see Fig. 6). However, on the other hand, the re-

spective figure also indicates a certain overfitting effect, as mentioned previously, that grows with the number of the CNN input channels.



Fig. 5. Training accuracy and loss curves at the higher representation level. (a): Location, (b): Location&Companion&Mood, (c): Location&Purposes, (d): Location&Purposes&Companion&Mood.

The models presented in this work use 1-dim fixed-sized kernels of size 6. This number was determined through applying a grid search. The problem when using fixed-sized kernels is that these are able to capture only data dependencies of a certain length. Varied-sized kernels as in the work of Kim et al. in [21] could help overcome this issue and capture the individual properties of each semantic dimension in our data.

6 Conclusion

In this work, we explore the performance of a Multi-Channel Convolutional Neural Network (CNN) based approach with respect to its capability in modeling semantic trajectories at different semantic representation levels and predicting the next semantic location of a user. Moreover, we investigate whether and to what extend the degree of semantic enrichment, that is, the number of the context



Fig. 6. Training accuracy and loss curves at the lower representation level. (a): Location, (b): Location&Companion&Mood, (c): Location&Purposes, (d): Location&Purposes&Companion&Mood.

feature dimensions considered in the semantic trajectories, affects the predictive performance of our model. We considered 5 different semantic enrichment dimensions for our trajectories, the location type, the purpose of visit (e.g., activity), the time, the user's mental and emotional state, and his companionship.

We evaluated our model using a 8-week long real-world dataset from 21 users and compared it to a vanilla Single-Channel CNN and a probabilistic Markov Chain model that served as out baselines. We could show that raising the semantic enrichment degree of our trajectory data while increasing the corresponding number of channels at the same time can indeed lead to an improvement in terms of prediction accuracy and f-Score. This could be particularly seen when we attempted to model and predict upon semantic trajectories at a higher representation level. Furthermore, the results of this work indicate a strong correlation between the degree of semantic enrichment, the number of CNN channels and the training behaviour, with our multi-channel based approach being characterized by generally much smoother and faster converging learning curves.

However, our evaluation also identified some limitations leaning mostly on certain overfitting effects, which could be mainly attributed to the small size of our dataset. In our future work, we plan to further explore the use of CNNs in



Fig.7. Training f-Score and loss curves at the higher representation level. (a): Location, (b): Location&Companion&Mood, (c): Location&Purposes, (d): Location&Purposes&Companion&Mood.

the semantic location prediction scenario. In particular, we plan to investigate the use of varied-sized kernels and depthwise separable convolution layers aiming at improving both the predictive performance as well as the computational efficiency of our model.

References

- Alvares, L.O., Bogorny, V., Kuijpers, B., de Macedo, J.A.F., Moelans, B., Vaisman, A.: A model for enriching trajectories with semantic geographical information. In: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems. p. 22. ACM (2007)
- Ashbrook, D., Starner, T.: Learning significant locations and predicting user movement with gps. In: Proceedings. Sixth International Symposium on Wearable Computers,. pp. 101–108 (Oct 2002). https://doi.org/10.1109/ISWC.2002.1167224
- Bogorny, V., Renso, C., Aquino, A.R., Lucca Siqueira, F., Alvares, L.O.: Constant– a conceptual data model for semantic trajectories of moving objects. Transactions in GIS 18(1), 66–88 (2014)
- 4. Cao, Х., Cong, G., C.S.:Mining significant Jensen, seman-Proc. VLDB Endow. 3(1-2), tic locations from gps data.



Fig. 8. Training f-Score and loss curves at the higher representation level. (a): Location, (b): Location&Companion&Mood, (c): Location&Purposes, (d): Location&Purposes&Companion&Mood.

- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. Journal of Machine Learning Research 12(Aug), 2493–2537 (2011)
- Eagle, N., Pentland, A.S.: Reality mining: sensing complex social systems. Personal and ubiquitous computing 10(4), 255–268 (2006)
- 7. Elragal, A., El-Gendy, N.: Trajectory data mining: integrating semantics. Journal of Enterprise Information Management 26(5), 516–535 (2013). https://doi.org/10.1108/JEIM-07-2013-0038, https://doi.org/10.1108/JEIM-07-2013-0038
- 8. Etter, V., Kafsi, M., Kazemi, E.: Been there, done that: What your mobility traces reveal about your behavior (2012), http://infoscience.epfl.ch/record/178426
- Gambs, S., Killijian, M.O., del Prado Cortez, M.N.n.: Next place prediction using mobility markov chains. In: Proceedings of the First Workshop on Measurement, Privacy, and Mobility. pp. 3:1–3:6. MPM '12, ACM, New York, NY, USA (2012). https://doi.org/10.1145/2181196.2181199, http://doi.acm.org/10.1145/2181196.2181199
- Gao, Q., Zhou, F., Zhang, K., Trajcevski, G., Luo, X., Zhang, F.: Identifying human mobility via trajectory embeddings. In: Proceedings of the 26th International Joint

Conference on Artificial Intelligence. pp. 1689–1695. AAAI Press (2017)

- Karatzoglou, A., Koehler, D., Beigl, M.: Purpose-of-visit-driven semantic similarity analysis on semantic trajectories for enhancing the future location prediction. In: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). pp. 100–106 (March 2018). https://doi.org/10.1109/PERCOMW.2018.8480230
- Karatzoglou, A., Lamp, S.C., Beigl, M.: Matrix factorization on semantic trajectories for predicting future semantic locations. In: 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). pp. 1–7 (Oct 2017). https://doi.org/10.1109/WiMOB.2017.8115810
- Karatzoglou, A.: Evolutionary optimization on artificial neural networks for predicting the user's future semantic location. In: Macintyre, J., Iliadis, L., Maglogiannis, I., Jayne, C. (eds.) Engineering Applications of Neural Networks. pp. 379–390. Springer International Publishing, Cham (2019)
- Karatzoglou, A., Beigl, M.: Applying situation-person-driven semantic similarity on location-specific cognitive frames for improving the location prediction (poster). In: 8th International Conference on Knowledge Engineering and Semantic Web (KESW) (2017)
- Karatzoglou, A., Beigl, M.: Enhancing the affective sensitivity of location based services using situation-person-dependent semantic similarity. In: Proceedings of the Eleventh International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM) 2017. pp. 95–100 (2017)
- Karatzoglou, A., Ebbing, J., Ostheimer, P., Hua, W., Beigl, M.: Sentient user modeling using the example of location prediction (under review). User Modeling and User-Adapted Interaction (UMUAI) (2019)
- Karatzoglou, A., Jablonski, A., Beigl, M.: A seq2seq learning approach for modeling semantic trajectories and predicting the next location. In: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 528–531. SIGSPATIAL '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3274895.3274983, http://doi.acm.org/10.1145/3274895.3274983
- Karatzoglou, A., Köhler, D., Beigl, M.: Semantic-enhanced multi-dimensional markov chains on semantic trajectories for predicting future locations. Sensors 18(10), 3582 (Oct 2018). https://doi.org/10.3390/s18103582, http://dx.doi.org/10.3390/s18103582
- Karatzoglou, A., Schnell, N., Beigl, M.: A convolutional neural network approach for modeling semantic trajectories and predicting future locations. In: Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I. (eds.) Artificial Neural Networks and Machine Learning – ICANN 2018. pp. 61–72. Springer International Publishing, Cham (2018)
- Karatzoglou, A., Sentürk, H., Jablonski, A., Beigl, M.: Applying artificial neural networks on two-layer semantic trajectories for predicting the next semantic location. In: International Conference on Artificial Neural Networks. pp. 233–241. Springer (2017)
- Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
- 22. LeCun, Y., Bengio, Y., et al.: Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks **3361**(10), 1995 (1995)
- LeCun, Y., Kavukcuoglu, K., Farabet, C.: Convolutional networks and applications in vision. In: Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on. pp. 253–256. IEEE (2010)

- 16 A. Karatzoglou
- 24. Lv, J., Li, Q., Wang, X.: T-conv: A convolutional neural network for multi-scale taxi trajectory prediction. arXiv preprint arXiv:1611.07635 (2016)
- 25. Mathew, W., Raposo, R., Martins, B.: Predicting future locations with hidden markov models. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. pp. 911–918. UbiComp '12, ACM, New York, NY, USA (2012). https://doi.org/10.1145/2370216.2370421, http://doi.acm.org/10.1145/2370216.2370421
- 26. Mathworks: Convolutional neural network (2018), https://www.mathworks.com/discovery/convolutional-neural-network.html, [Online; accessed 19-February-2018]
- Song, X., Kanasugi, H., Shibasaki, R.: Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level. In: IJCAI. vol. 16, pp. 2618–2624 (2016)
- Spaccapietra, S., Parent, C., Damiani, M.L., de Macêdo, J.A.F., Porto, F., Vangenot, C.: A conceptual view on trajectories. Data Knowl. Eng. 65(1), 126–146 (2008). https://doi.org/10.1016/j.datak.2007.10.008, https://doi.org/10.1016/j.datak.2007.10.008
- Yao, D., Zhang, C., Huang, J., Bi, J.: Serm: A recurrent model for next location prediction in semantic trajectories. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 2411–2414. CIKM '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3132847.3133056, http://doi.acm.org/10.1145/3132847.3133056
- 30. Ye, J., Zhu, Ζ., Cheng, H.: What's Your Next Move: User Prediction Activity inLocation-based Social Networks, DD. 171 - 179(2013).https://doi.org/10.1137/1.9781611972832.19, https://epubs.siam.org/doi/abs/10.1137/1.9781611972832.19
- Ying, J.J.C., Lee, W.C., Tseng, V.S.: Mining geographic-temporal-semantic patterns in trajectories for location prediction. ACM Trans. Intell. Syst. Technol. 5(1), 2:1–2:33 (Jan 2014). https://doi.org/10.1145/2542182.2542184, http://doi.acm.org/10.1145/2542182.2542184
- Ying, J.J.C., Lee, W.C., Weng, T.C., Tseng, V.S.: Semantic trajectory mining for location prediction. In: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 34–43. ACM (2011)