

Chapter 42

Data Mining and Knowledge Discovery



Chao Zhang and Jiawei Han

Abstract Our physical world is being projected into online cyberspace at an unprecedented rate. People nowadays visit different places and leave behind them million-scale digital traces such as tweets, check-ins, Yelp reviews, and Uber trajectories. Such digital data are a result of social sensing: namely people act as human sensors that probe different places in the physical world and share their activities online. The availability of massive social-sensing data provides a unique opportunity for understanding urban space in a data-driven manner and improving many urban computing applications, ranging from urban planning and traffic scheduling to disaster control and trip planning. In this chapter, we present recent developments in data-mining techniques for urban activity modeling, a fundamental task for extracting useful urban knowledge from social-sensing data. We first describe traditional approaches to urban activity modeling, including pattern discovery methods and statistical models. Then, we present the latest developments in multimodal embedding techniques for this task, which learns vector representations for different modalities to model people’s spatiotemporal activities. We study the empirical performance of these methods and demonstrate how data-mining techniques can be successfully applied to social-sensing data to extract actionable knowledge and facilitate downstream applications.

42.1 Overview

Our physical world is being projected into cyberspace at an unprecedented rate. People nowadays visit different places and leave behind them million-scale digital traces such as tweets, check-ins, Yelp reviews, and Uber trajectories. The malls they

C. Zhang (✉)
School of Computational Science and Engineering, Georgia Institute of Technology,
Atlanta, USA
e-mail: chaozhang@gatech.edu

J. Han
University of Illinois at Urbana-Champaign, Urbana, USA
e-mail: hanj@illinois.edu

go to, the restaurants they visit, the movies they watch, the concerts they attend—almost everything people do during a day can now result in rich cybertraces. For example, Foursquare has collected more than 8 billion check-ins as of today, Twitter has more than 10 million geo-tagged tweets published every day, and Instagram witnesses more than 20 million geo-tagged photos being shared every day. Such digital data represent a result of social sensing: people act as human sensors to probe different places in the physical world and leave online traces of their spatiotemporal activities.

The availability of massive online social-sensing data provides an unprecedented opportunity for modeling people’s offline spatiotemporal activities. While traditional approaches to urban activity modeling often require costly surveys and field studies, the understanding is often coarse-grained and limited. In contrast, social-sensing data provide a fine-grained coverage of our physical world (Leetaru et al. 2013) and serve as a unique proxy for human activities (Cheng et al. 2011; Jurdak et al. 2015; Noulas et al. 2011). For the first time, it becomes possible to develop data-driven techniques for modeling people’s spatiotemporal activities, which can potentially revolutionize many applications, including urban planning, traffic scheduling, disaster control, and trip planning.

Social-sensing data often comprise modalities (e.g., location, time, and text) that can have totally different representations and distributions. When using massive social-sensing data for spatiotemporal activity modeling, the key is to capture the correlations of these data modalities and make predictions across them. For a subset of the modalities (Fig. 42.1), the model is expected to predict the remaining ones. For example: (1) Given a location and time, what are the typical activities around that location and time? (2) Given an activity and time, where does this activity usually occur? and (3) Given an activity and a location, when does the activity usually occur?

In the remainder of this chapter, we first summarize key data-mining methods for urban analysis tasks (Sect. 42.2). Generally, these methods fall into four broad categories: (1) urban pattern discovery; (2) urban activity models; (3) urban mobility models; and (2) urban event detection. We will describe techniques in each category.

In addition to overviewing how data-mining techniques can address urban-analysis tasks, we introduce the latest development of urban activity modeling techniques based on multimodal embedding (Sect. 42.3). At a high level, multimodal embedding directly captures cross-modal correlations by mapping items from

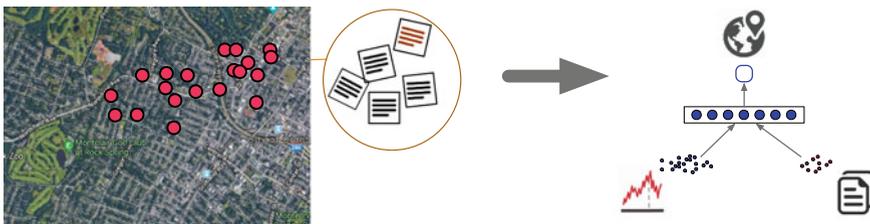


Fig. 42.1 An illustration of spatiotemporal activity modeling using social-sensing data

different modalities into the same latent space. If two elements are correlated (e.g., the JFK airport region and the keyword ‘flight’), their latent representations are encouraged to be close to each other. Compared with existing generative models, multimodal embedding does not impose any distributional assumptions and incurs much lower computational cost in the learning process. We show the performance of the multimodal embedding method and demonstrate its superiority for urban activity modeling.

42.2 Data Mining for Urban Analysis

Generally, data-mining techniques for urban analysis tasks can be categorized into four classes: (1) urban pattern discovery; (2) urban activity modeling; (3) urban mobility modeling; and (4) urban event detection. In the following, we overview these tasks and describe key techniques for each task.

42.2.1 *Urban Pattern Discovery*

Urban pattern discovery aims to discover various forms of spatiotemporal patterns from social-sensing data. Sequential pattern is an important type of spatiotemporal pattern which captures sequential transition regularities of people’s activities. Giannotti et al. (2007) defined a T-pattern as a region-of-interest sequence that appears frequently in the input trajectories. By partitioning the space, they used sequential pattern-mining techniques to extract the T-patterns. Zhang et al. (2014) extracted frequent movement patterns from semantic trajectory data. With a top-down approach, they first discovered coarse-grained sequential patterns, and then partitioned them into fine-grained sequential patterns by clustering pattern-matching snippets. Several studies have investigated how to find objects that frequently move together. Examples in this line include mining flock (Laube and Imfeld 2002), swarm (Li et al. 2010a), and gathering (Zheng et al. 2013) patterns.

Periodic patterns represent user behaviors that regularly occur with one or multiple time periods. To extract periodic patterns, Li et al. (2010b) first extracted reference spots by using density-based clustering, and then detected periodic patterns at those spots. They have also studied how to find periodic patterns from sequences with incomplete observations (Li 2012b). The idea is to partition the time series into small chunks and then overlay them for each candidate period. Cho et al. (2011) found that the mobility of each user usually centers around several regions. Based on this observation, they proposed a periodic mobility model that predicts a user’s location by estimating the regions where a user most likely stays. Following this paper, Tarasov et al. (2013) modeled a region based on radiation models (Simini et al. 2012).

42.2.2 *Urban Activity Modeling*

Urban activity modeling aims to use statistical models to describe people's activity regularities and learn such models from data. There are two subcategories along this line: global activity models and personalized activity models.

Global activity models aim at characterizing people's activities over space and time at the global level without distinguishing personal preferences. Most existing techniques (Hong et al. 2012; Kling et al. 2014; Mei et al. 2006; Sizov 2010; Wang et al. 2007; Yin et al. 2011; Yuan et al. 2013) are latent variable models, which extend the classic topic models (Blei et al. 2003; Hofmann 1999) to handle spatiotemporal contexts. For example, Sizov (2010) extended LDA (Blei et al. 2003) by assuming that each latent topic was characterized by a multinomial distribution over text as well as two Gaussian distributions over latitudes and longitudes. Later, they further extended the model to discover topics that have non-Gaussian distributions (Kling et al. 2014). Yin et al. (2011) extended the PLSA model (Hofmann 1999) by modeling each region with a Gaussian distribution for location generation and a multinomial distribution for text generation.

In contrast, personalized activity models aim at describing spatiotemporal activities at an individual level. Hong et al. (2012) and Yuan et al. (2013) proposed to model the user factor in geographic topic models. In this way, users' individual-level preferences can be inferred. Yuan et al. (2017) later proposed a Bayesian non-parametric model, which can automatically discover the regions a user visits periodically.

42.2.3 *Urban Mobility Modeling*

The task of human mobility modeling is a corner-stone task for various applications, including urban planning, traffic scheduling, location prediction, and personalized recommendation. In the past years, this task has attracted much research attention from the data-mining community.

The first line of human mobility modeling is law-based methods. Such methods study the physical laws that govern human mobility. Brockmann et al. (2006) discovered that human mobility can be approximated by a continuous random-walk model with long-tail distributions. Gonzalez et al. (2008) used mobile phone data for human mobility modeling. They found that people return to a few locations periodically, and such mobility can be modeled by a stochastic process centered on a fixed point. Song et al. (2010) found that more than 93% of human movements are predictable, because of the high regularity of human mobility. They thus proposed a self-consistent microscopic model for individual mobility prediction.

Along another line, many model-based approaches have been explored to learn statistical models from human movement data. For example, Cho et al. (2011) found that a user usually moves around a few center locations (e.g., home, work) in fixed time periods. Based on this observation, they proposed to model user movement

as a mixture of Gaussian distributions. Their model can be further extended by incorporating social influence, as a user is more likely to visit a location that is close to the locations of friends. Wang et al. (2015) proposed a hybrid mobility model, which improved location prediction by using heterogeneous mobility data.

One important area along the line of mode-based approaches is the hidden Markov model (HMM), which is a powerful statistical model for sequential data. In early work, Mathew et al. (2012) first partitioned the space into equally sized triangles using a hierarchical triangular mesh. Based on the assumption that each latent state imposes a multinomial distribution over the triangles, they trained an HMM for the input trajectories. Deb and Basu (2015) proposed a probabilistic latent semantic model. This model uses HMM to extract latent semantic locations from cell-tower and Bluetooth data. Ye et al. (2013) have explored how to use HMM to model user check-in data generated from location-based social networks (LBSNs). Their HMM model can incorporate the category information of places and thereby is capable of predicting the category for the user's next location. Zhang et al. (2016a) have applied HMMs to model people's sequential behaviors. The key idea of their model is that there are a few latent states underlying people's daily activities and that people typically move among these states with strong regularity. Instead of using one model for all the users, they proposed to group users based on their sequential patterns and learn a set of HMMs to characterize group-level activities.

42.2.4 Urban Event Detection

An urban event, such as a protest or a disaster, is an unusual activity occurring in a local area and having a specific time duration, while engaging a considerable number of participants. Detecting urban events in real time was nearly impossible years ago because of the lack of timely and reliable data. However, the recent availability of social-sensing data sheds light on this problem.

Many studies have explored how to detect urban events, which are also termed spatiotemporal events, from social-sensing data (Abdelhaq et al. 2013; Chen and Roy 2009; Feng et al. 2015; Lee et al. 2011; Sakaki et al. 2010; Zhang et al. 2016b). Existing techniques for identifying abnormal events can be categorized into document-based approaches and feature-based approaches. Document-based approaches consider documents as basic units and group similar documents to detect abnormal events. For example, Allan et al. (1998) performed single-pass clustering of the document stream and used a similarity threshold to determine whether a new document is a new topic or should be merged into an existing topic. Aggarwal and Subbian (2012) also proposed to detect events by clustering the tweet stream. However, their similarity measure jointly considers tweet content relevance and user social proximity. Zhang et al. (2016b) first detected geo-topic clusters as candidate events and then employed a z -score to identify abnormal clusters as true events.

The second line of event detection has adopted feature-based approaches (Fung et al. 2005; He et al. 2007; Li et al. 2012a; Mathioudakis and Koudas 2010; Weng

and Lee 2011). The idea is to identify a set of bursty features (e.g., keywords or phrases) from the text stream and then cluster them into events. Specifically, Fung et al. (2005) modeled feature occurrences using a binomial distribution to extract bursty features. He et al. (2007) constructed the stream for each feature and then performed a Fourier transform to identify bursty events. Krumm and Horvitz (2015) monitored the spatiotemporal distributions of tweets and identified spikes in the spatiotemporal signal as abnormal events. There has also been work on detecting specific types of events. Sakaki et al. (2010) investigated real-time earthquake detection. They trained a classifier to judge whether a tweet was earthquake-related or not and then proposed to release an alarm whenever the number of earthquake-related tweets was large. Li et al. (2012a) detected crime and disaster events using a self-adaptive crawler, which can dynamically retrieve crime and disaster-related tweets. Abdelhaq et al. (2013) proposed the EvenTweet model, which could detect local events with the following steps: (1) examine several previous windows to identify bursty words; (2) compute the spatial entropy of each bursty word and discover localized words; (3) group localized words into clusters based on their spatial distributions; and (4) rank the resultant clusters based on event-indicative features such as burstiness and spatial coverage.

42.3 Multimodal Embedding for Urban Activity Modeling

We now describe the latest development of multimodal embedding techniques for urban activity modeling. Different from existing latent variable models that rely on latent states to bridge different modalities indirectly, such embedding-based methods can capture the cross-modal correlations directly. This is achieved by mapping all the modalities into a common vector space. In the following, we first describe the high-level idea (Sect. 42.3.1), then detail the multimodal embedding method for activity modeling (Sect. 42.3.2), and finally present the optimization process (Sect. 42.3.3).

42.3.1 Method Overview

At a high level, our embedding-based method, named CrossMap (Zhang et al. 2017a), maps items from different modalities into the same latent space with their correlations preserved, as shown in Fig. 42.2. Formally, it aims to learn the embeddings L , T , and W where: (1) L is the embeddings for regions; (2) T is the embeddings for hours; and (3) W is the embeddings for keywords. Take L as an example. Each element is a D -dimensional ($D > 0$) vector, which represents the embedding for region l . Once the embeddings are learned, cross-modal predictions can be made by simply searching for items nearest to the given query in the latent space.

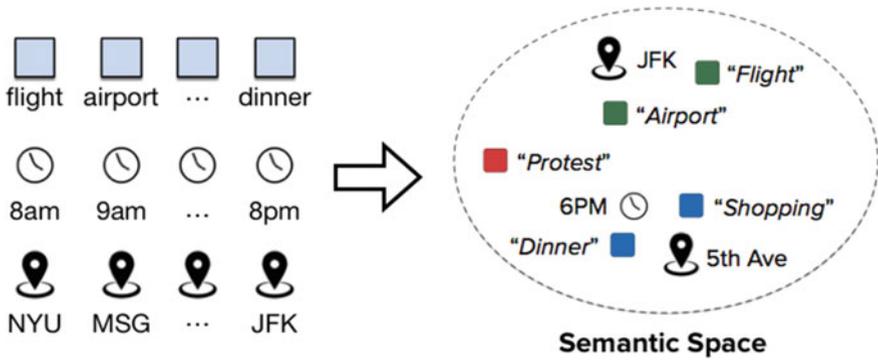


Fig. 42.2 An illustration of multimodal embedding for urban activity modeling. The idea is to map items from different modalities (e.g., location, time, text) into the same latent vector space to preserve their correlations. Their latent representations are then used for cross-modal prediction

42.3.2 Multimodal Embedding via Attribute Reconstruction

The key principle for multimodal embedding is to optimize the embeddings L, T, W such that the observed relationships among location, time, and text can be reconstructed. We thus define an unsupervised attribute reconstruction task. The goal is to learn the embeddings L, T, W such that the attributes of a record r can be reconstructed by assuming that the other attributes are observed.

Let r be a record. Given any attribute $i \in r$ with type X (could be location, time, or keyword), we compute the likelihood of observing attribute i as follows:

$$p(i|r_{-i}) = \exp(s(i, r_{-i}) / \sum_{j \in X} \exp(s(j, r_{-i})))$$

where r_{-i} represents the set of all the attributes in r except for i , and $s(i, r_{-i})$ denotes the similarity between i and r_{-i} .

The key question for the above is how to define $s(i, r_{-i})$. A straightforward idea is to average the embeddings of all the attributes in r_{-i} and then compute $s(i, r_{-i})$ as $s(i, r_{-i}) = \mathbf{v}_i T \sum_{j \in r_{-i}} \mathbf{v}_j / |r_{-i}|$, where \mathbf{v}_i denotes the embedding for attribute i .

However, this simple definition fails to consider spatial and temporal continuities. Consider the spatial continuity as an example. According to the first law of geography, “everything is related to everything else, but near things are more related than distant things.” To achieve spatial smoothness, two spatial items that are close to each other should be considered correlated instead of independent. We thus introduce spatial smoothing and temporal smoothing to capture the spatiotemporal continuities. With the smoothing technique, the method can not only maintain local consistency of neighboring regions and periods, but also alleviate data sparsity. One can refer to Zhang et al. (2017b) for more details about the smoothing techniques.

In addition to the above pseudo-region and period embeddings, we also introduce pseudo-keyword embeddings for notational ease. Given r_{-i} , its pseudo-keyword embedding is defined as:

$$\mathbf{v}_{\hat{W}} = \sum_{w \in N_w} \mathbf{v}_w / |N_w|$$

where N_w is the set of keywords in r_{-i} . With these pseudo-embeddings, we define a smoothed version of $s(i, r_{-i})$ as $s(i, r_{-i}) = \mathbf{v}_i T \mathbf{h}_i$ where if i is a keyword then:

$$\mathbf{h}_i = (\mathbf{v}_l + \mathbf{v}_t + \mathbf{v}_{\hat{W}}) / 3$$

If i is a region then:

$$\mathbf{h}_i = (\mathbf{v}_t + \mathbf{v}_{\hat{W}}) / 2$$

If i is a period, then:

$$\mathbf{h}_i = (\mathbf{v}_l + \mathbf{v}_{\hat{W}}) / 2$$

Let R_U be a collection of records for learning the urban activity model. The final loss function for the attribute reconstruction task is simply the negative log-likelihood of observing all the attributes of the records in R_U :

$$J_{R_U} = - \sum_{r \in R_U} \sum_{i \in r} \log p(i | r_{-i}) \quad (42.1)$$

42.3.3 The Optimization Procedure

To efficiently learn the embeddings, we can use stochastic gradient descent (SGD) and negative sampling (Mikolov et al. 2013) for optimizing the objective function shown in Eq. (42.1). At each step, we can use SGD to sample a record r and an attribute $i \in r$. Based on negative sampling, we then randomly select K negative attributes that have the same type as i but do not appear in r . Then the loss function for the selected samples becomes:

$$J_r = \log \sigma(s(i, r_{-i})) - \sum_{k=1}^K \log \sigma(-s(k, r_{-i}))$$

In the above, $\sigma(\cdot)$ is the sigmoid function. The updating rules for $\mathbf{v}_i, \mathbf{v}_k$, and \mathbf{h}_i can be obtained by taking the derivatives of J_r . We omit the details because of the space limit.

42.4 Experiments

We now demonstrate the empirical performance of different algorithms on three real-life datasets:

- The first dataset, called LA, contains ~ 1.10 million geo-tagged tweets published in Los Angeles. We crawled the LA dataset by monitoring the Twitter Streaming API during 2014.08.01–2014.11.30 and continuously gathering the geo-tagged tweets in the bounding box of LA. We preprocessed the raw data as follows. For the text part, user mentions, URLs, stopwords, and the words that appear less than 100 times were removed. For space and time, we partitioned the LA area into small grids with size 300 m * 300 m and broke the one-day period into 24 one-hour windows.
- The second dataset, called NY, was also collected from Twitter. It consisted of ~ 1.20 million geo-tagged tweets published in New York City during the time period 2014.08.01–2014.11.30.
- The third dataset was called 4SQ. It was collected from Foursquare. It consisted of about 0.7 million Foursquare check-ins posted in New York City, during the time period 2010.08–2011.10. This dataset was mainly used to evaluate the performance of the multi-modal embedding method for the downstream task of activity classification. Similarly, user mentions, URLs, stopwords, and the words that appeared less than 100 times were removed.

We study the following methods for urban activity modeling: (1) the geographic topic model LGTA (Yin et al. 2011); (2) the non-Gaussian geographic topic model MGTM (Kling et al. 2014); (3) the tensor factorization method Tensor (Harshman 1970); (4) the SVD method, which first constructs the co-occurrence matrices between each pair of location, time, text, and category, and then performs singular-value decomposition on the matrices; (5) the TF-IDF method, which constructs the co-occurrence matrices between each pair of location, time, text, and category and then computes the TF-IDF weight for each entry in the matrix; (6) the multimodal embedding method CrossMap (Zhang et al. 2017a) as discussed in the previous section.

We investigated two types of urban activity prediction tasks. The first was to predict locations for a given textual query. Specifically, recall that each record reflects a user’s activity with the following three attributes: a location, a timestamp, and a bag of keywords. In the location-prediction task, the input was the timestamp and the keywords, and the goal was to accurately pinpoint the ground-truth location from a pool of candidates. We predicted the location at two different granularities: (1) coarse-grained region prediction of the ground-truth region that r falls in; and (2)

fine-grained POI prediction of the ground-truth POI that r corresponds to. Note that fine-grained POI prediction was only evaluated on the tweets that had been linked with Foursquare. The second task was to predict activities for a given location query. In this task, the input was the timestamp and the location, and the goal was to pinpoint the ground-truth activities at two different granularities: (1) coarse-grained category prediction of the ground-truth activity category of r (again, such a coarse-grained activity prediction was performed only on the tweets that had been linked with Foursquare); and (2) fine-grained keyword prediction of the ground-truth message from a candidate pool of messages.

To summarize, we studied four urban activity prediction subtasks in total: (1) region prediction; (2) POI prediction; (3) category prediction; and (4) keyword prediction. For each prediction subtask, we first generated a candidate pool by mixing the ground truth with a set of M random negative samples. Take region prediction as a concrete example. For the ground-truth region, we mixed with M randomly chosen regions. Then, we tried to pinpoint the ground truth from the size- $(M + 1)$ candidate pool by ranking all the candidates. Generally, the better a model captures the patterns underlying people’s activities, the more likely it can rank the ground truth for top positions. We thus used mean reciprocal rank (MRR) to quantify the effectiveness of a model.

Tables 42.1 and 42.2 report the quantitative results of different methods for location and activity predictions, respectively. As shown, on all of the four subtasks, CrossMap and its variants achieved much higher MRRs than the baseline methods. Compared with the two geographic topic models (LGTA and MGTM), CrossMap showed as much as 62% performance improvement for location prediction, and 83% for activity prediction. Tensor, SVD, and TF-IDF had better performance than LGTA and MGTM by modeling time and category, yet CrossMap outperformed them by large margins. Interestingly, TF-IDF turned out to be a strong baseline, demonstrating the effectiveness of the tf-idf similarity for the prediction tasks. SVD and Tensor can effectively recover the co-occurrence matrices and tensor, but the raw co-occurrence seems a less effective measure for location and activity prediction.

Table 42.1 MRRs of various methods for location prediction. For each test tweet, we assume its timestamp and keywords are observed, and perform location prediction at two granularities: (1) region prediction retrieves the ground-truth region; and (2) POI prediction retrieves the ground-truth POI (for Foursquare-linked tweets)

Method	Region prediction		POI prediction	
	LA	NY	LA	NY
LGTA	0.3583	0.3544	0.5889	0.5674
MGTM	0.4007	0.391	0.5811	0.553
Tensor	0.3592	0.3641	0.6672	0.7399
SVD	0.3699	0.3604	0.6705	0.7443
TF-IDF	0.4114	0.4605	0.719	0.776
CrossMap	0.5373	0.5597	0.7845	0.8508

Table 42.2 MRRs of different methods for activity prediction. For each test tweet, we assume its location and timestamp are observed, and predict activities at two granularities: (1) category prediction of ground-truth category (for Foursquare-linked tweets); and (2) keyword prediction retrieves the ground-truth message

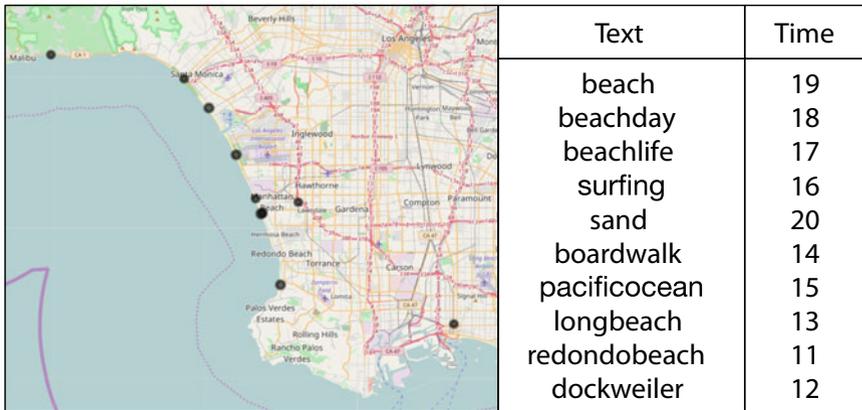
Method	Category prediction		Keyword prediction	
	LA	NY	LA	NY
LGTA	0.4409	0.4527	0.3392	0.3425
MGTM	0.4587	0.4640	0.3501	0.3430
Tensor	0.8635	0.7988	0.4004	0.3744
SVD	0.8556	0.7826	0.4098	0.3728
TF-IDF	0.9137	0.8259	0.5236	0.4864
CrossMap	0.6225	0.5874	0.5693	0.5538

We now performed a set of case studies to examine how well CrossMap predicted across modalities. Specifically, we performed one-pass training of CrossMap for LA and NY, and launched a bunch of queries at different stages. For each query, we retrieved the top-ten most similar items with different types from the entire search space.

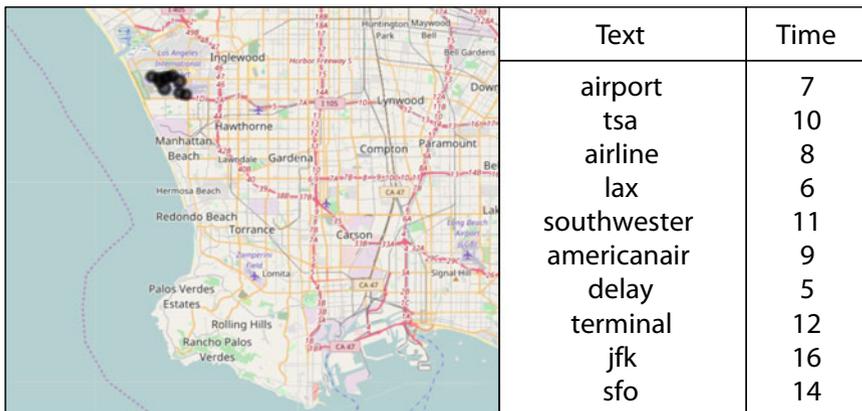
Figure 42.3a shows the results when we queried with the keyword ‘beach’. As shown, the retrieved items in each type are very meaningful: the top locations mostly fall around famous beaches in the Los Angeles area; the top keywords can well reflect people’s activities on the beach, including ‘sand’ and ‘boardwalk.’ Fig. 42.3b shows the results for an example spatial query, at the GPS location of the centroid of LAX airport. One can see that the retrieved top spatial, temporal, and textual elements are closely related to the airport. Given the query at the airport, the top keywords are all concepts that reflect flight-related activities, such as ‘airport,’ ‘tsa,’ and ‘airline.’

Figures 42.4a–c further show temporal-textual queries which can demonstrate the temporal dynamics of people’s urban activities. When we fix the query keyword as ‘restaurant’ and vary the time point in the query, the retrieved top items vary obviously. By examining the top keywords, we can see the query ‘10am’ results in many breakfast-related keywords, such as ‘bfast’ and ‘brunch.’ In contrast, when the query is changed to ‘2 pm,’ many lunch-related keywords are retrieved. When ‘8 pm’ is specified as the query, many dinner-related ones are retrieved. Another interesting observation is that the top locations for the queries ‘10am’ and ‘2 pm’ fall in working areas, while the results for ‘8 pm’ distribute mostly in residential areas. Such results show that the time factor plays an important role in determining people’s activities, and CrossMap captures such fine-grained temporal dynamics.

We proceeded to examine the performance of multimodal embedding models for downstream applications. For this purpose, we chose activity classification as an application. In the 4SQ dataset, every check-in belongs to one of nine categories: Food, College & University, Nightlife Spot, Shop & Service, Travel & Transport, Residence, Arts & Entertainment, Outdoors & Recreation, Professional & Other Places. We used those categories as the labels for people’s urban activities and aimed



(a) Query = 'beach'

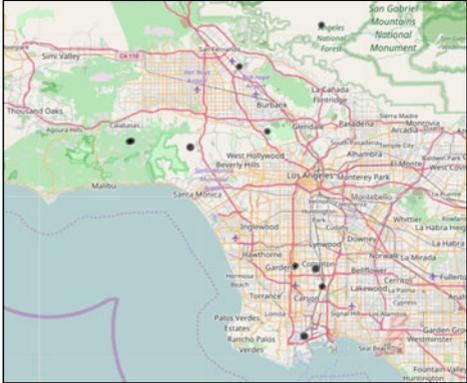


(b) Query = '(33.9424, -118.4137)' (LAX airport centroid)

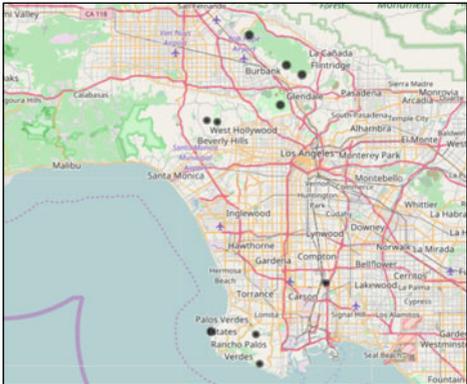
Fig. 42.3 Two example queries and the top-ten results returned by CrossMap

to learn classifiers that can predict those labels for any given check-in. We performed a random shuffling of the dataset, and then randomly chose 80% for training and 20% for testing. For any check-in r , all the studied methods can obtain vector representations for the location, time, and text; we concatenated the vectors as the feature representation of a check-in.

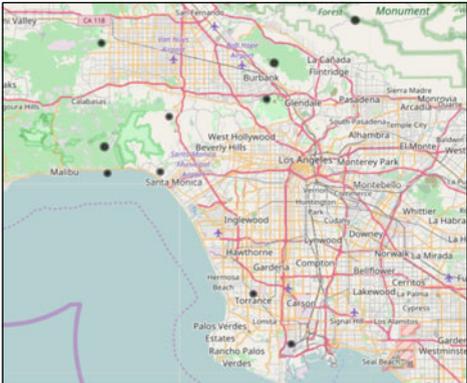
With the above feature transformation, we then trained a multiclass logistic regression for activity classification. Figure 42.5 reports the performance of different methods for the activity classification task. As shown, CrossMap outperformed the other methods significantly. Using the simple linear classification model, the F1 score of the method can reach as high as 0.843. Such results show that the embeddings obtained by multimodal embedding can well distinguish the semantics of different categories. We further verified this fact using data visualization. As shown

	Text	Time
	restaurant	10
	bfast	7
	pastry	6
	brunching	8
	deli	9
	brunch	5
	yummm	11
	bakery	12
	thai	14
foodporn	16	

(a) Query = 'restaurant' + '10am'

	Text	Time
	restaurant	14
	lunch	15
	seafood	13
	deli	16
	foodporn	17
	vietnamese	12
	lunchfood	7
	instafood	6
	dimsum	10
thai	8	

(b) Query = 'restaurant' + '2pm'

	Text	Time
	restaurant	20
	dinner	18
	happyhour	19
	seafood	17
	bartender	16
	thai	7
	server	5
	yummy	14
	dating	20
mexican	15	

(c) Query = 'restaurant' + '8pm'

Fig. 42.4 Three temporal-textual queries and the top ten results returned by CrossMap

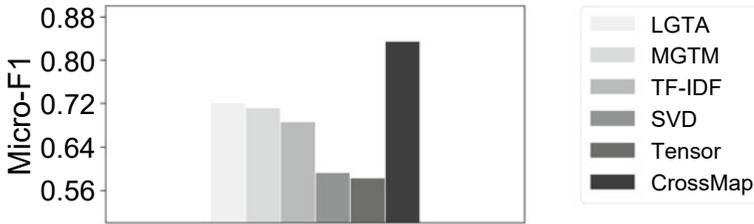


Fig. 42.5 Activity classification performance on 4SQ

in Fig. 42.6, we chose three categories and used the t-SNE method (Maaten and Hinton 2008) to visualize the feature vectors. One can observe that the learnt representations of the multimodal embedding method resulted in much clearer inter-class boundaries compared to the baselines such as geographic topic models.

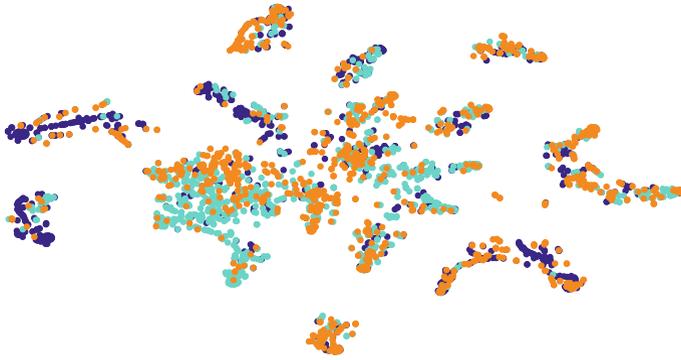
42.5 Summary

We have presented data mining techniques for modeling people’s urban activities from massive social-sensing data. We first overviewed data mining techniques for four important urban analysis tasks: (1) urban pattern discovery; (2) urban activity modeling; (3) urban mobility modeling; (4) urban event detection. Then, we presented the latest development of multimodal embedding techniques for urban activity modeling, which maps items from different data modalities into a common latent space with their correlation preserved. Compared with previous latent variable models, multimodal embedding techniques do not impose distribution assumptions of people’s spatiotemporal activities, and scale well with the data size. We have studied the empirical performance of these methods on real datasets, and demonstrated that these techniques can enable the building of predictive urban activity models and can benefit downstream tasks like activity classification.

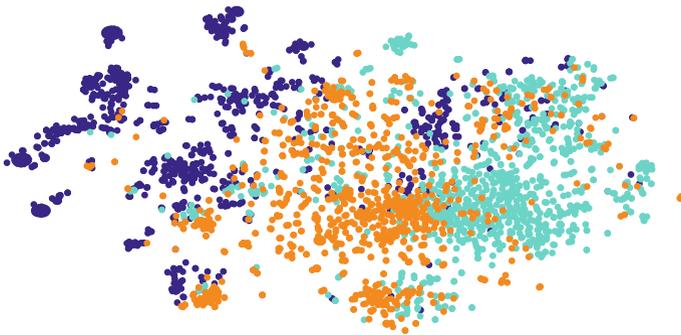
42.6 Future Directions

In the future, social-sensing data will continue to serve as an invaluable source for urban analysis. Data-mining techniques have already shown promising results when acquiring insights from social-sensing data for various tasks. However, there are still challenges that need to be addressed to fully unleash the power of social-sensing data. Below, we list several key challenges in this direction.

Integrating diverse data modalities. Modern social-sensing data often involve multiple modalities, such as text, image, location, and time. Considering the totally different representations of those data modalities and the complicated correlations



(a) LGTA



(b) CrossMap

Fig. 42.6 Visualizing the feature vectors generated by LGTA and CrossMap for three activity categories: ‘Food’ (cyan), ‘Travel & Transport’ (blue), and ‘Residence’ (orange). The feature vector of each 4SQ is mapped to a 2D point with t-SNE (Maaten and Hinton 2008)

among them, how to effectively integrate them for urban activity modeling and prediction remains a challenging problem.

Extracting insights from noisy data. Studies have shown that about 40% social-sensing data are pointless babbles. Even among those informative posts, most are rather short and noisy. It is nontrivial to analyze such noisy and short text messages and distill the information for end tasks.

Real-time data analysis. Many urban-analysis tasks require real-time performance. For instance, when an emergent event happens, it is important to report the event as soon as possible to allow for timely actions. As massive social-sensing data stream in, it is an important yet challenging problem to design on-line learning algorithms that can handle large-scale streaming data efficiently.

References

- Abdelhaq H, Sengstock C, Gertz M (2013) Eventweet: online localized event detection from twitter. *Proc VLDB Endowment* 6(12):1326–1329
- Aggarwal CC, Subbian K (2012) Event detection in social streams. In: *SDM, SIAM international conference on data mining*, pp 624–635
- Allan J, Papka R, Lavrenko V (1998) On-line new event detection and tracking. In: *SIGIR, international conference on information retrieval*, pp 37–45
- Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J Mach Learn Res* 3:993–1022
- Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439(7075):462–465
- Chen L, Roy A (2009) Event detection from flickr data through wavelet-based spatial analysis. In: *CIKM, international conference on knowledge management*, pp 523–532
- Cheng Z, Caverlee J, Lee K, Sui DZ (2011) Exploring millions of footprints in location sharing services. In: *ICWSM, international conference on weblogs and social media*, pp 81–88
- Cho E, Myers SA, Leskovec J (2011) Friendship and mobility: user movement in location-based social networks. In: *KDD, conference on knowledge discovery and data mining*, pp 1082–1090
- Deb B, Basu P (2015) Discovering latent semantic structure in human mobility traces. In: *Wireless sensor networks*. Springer, pp 84–103
- Feng W, Zhang C, Zhang W, Han J, Wang J, Aggarwal C, Huang J (2015) Streamcube: hierarchical spatio-temporal hashtag clustering for event exploration over the twitter stream. *IEEE international conference on data engineering*
- Fung GPC, Yu JX, Yu PS, Lu H (2005) Parameter free bursty events detection in text streams. In: *VLDB, international conference on very large DataBases*, pp 181–192
- Giannotti F, Nanni M, Pinelli F, Pedreschi D (2007) Trajectory pattern mining. In: *KDD, conference on knowledge discovery and data mining*, pp 330–339
- Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding individual human mobility patterns. *Nature* 453(7196):779–782
- Harshman RA (1970) Foundations of the PARAFAC procedure: models and conditions for an “explanatory” multi-modal factor analysis. *UCLA Working Papers Phonetics* 16 (1):84
- He Q, Chang K, Lim EP (2007) Analyzing feature trajectories for event detection. In: *SIGIR, international conference on information retrieval*, pp 207–214
- Hofmann T (1999) Probabilistic latent semantic indexing. In: *SIGIR, international conference on information retrieval*, pp 50–57
- Hong L, Ahmed A, Gurumurthy S, Smola AJ, Tsioutsouluklis K (2012) Discovering geographical topics in the twitter stream. In: *WWW, international World Wide Web conference*, pp 769–778
- Jurdak R, Zhao K, Liu J, AbouJaoude M, Cameron M, Newth D (2015) Understanding human mobility from twitter. *PLoS ONE* 10(7):e0131469
- Kling C, Kunegis J, Sizov S, Staab S (2014) Detecting non-gaussian geographical topics in tagged photo collections. In: *WSDM, conference on web search and data mining*, pp 603–612
- Krumm J, Horvitz E (2015) Eyewitness: Identifying local events via space-time signals in twitter feeds. In: *SIGSPATIAL, conference on advances in geographic information systems*
- Laube P, Imfeld S (2002) Analyzing relative motion within groups of trackable moving point objects. In: *GIScience, international symposium on geographic information science*, pp 132–144
- Lee R, Wakamiya S, Sumiya K (2011) Discovery of unusual regional social activities using geo-tagged microblogs. *World Wide Web* 14(4):321–349
- Leetaru K, Wang S, Cao G, Padmanabhan A, Shook E (2013) Mapping the global twitter heartbeat: the geography of twitter. *First Monday* 18(5)
- Li C, Sun A, Datta A (2012a) Tweet: segment-based event detection from tweets. In: *CIKM, international conference on knowledge management*, pp 155–164
- Li R, Lei KH, Khadiwala R, Chang KC (2012c) Tedas: a twitter-based event detection and analysis system. In: *ICDE, IEEE international conference on data engineering*, pp 1273–1276

- Li Z, Ding B, Han J, Kays R, Nye P (2010b) Mining periodic behaviors for moving objects. In: KDD, conference on knowledge discovery and data mining, pp 1099–1108
- Li Z, Ding B, Han J, Kays R (2010a) Swarm: Mining relaxed temporal moving object clusters. *Proc VLDB Endowment* 3(1):723–734
- Li Z, Wang J, Han J (2012b) Mining event periodicity from incomplete observations. In: KDD, conference on knowledge discovery and data mining, pp 444–452
- Maaten L, Hinton G (2008) Visualizing data using t-SNE. *J Mach Learn Res* 9(85):2579–2605
- Mathew W, Raposo R, Martins B (2012) Predicting future locations with hidden Markov models. In: ACM international joint conference on pervasive and ubiquitous computing, pp 911–918
- Mathioudakis M, Koudas N (2010) Twittermonitor: trend detection over the twitter stream. In: SIGMOD, international conference on management of data, pp 1155–1158
- Mei Q, Liu C, Su H, Zhai C (2006) A probabilistic approach to spatiotemporal theme pattern mining on weblogs. In: WWW, international World Wide Web conference, pp 533–542
- Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: NIPS, conference on neural information processing systems, pp 3111–3119
- Noulas A, Scellato S, Mascolo C, Pontil M (2011) An empirical study of geographic user activity patterns in foursquare. In ICWSM, international conference on weblogs and social media, pp 570–573
- Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes twitter users: real-time event detection by social sensors. In WWW, International World Wide Web Conference, pp. 851–860
- Simini F, González MC, Maritan A, Barabási AL (2012) A universal model for mobility and migration patterns. *Nature* 484(7392):96–100
- Sizov S (2010) Geofolk: latent spatial semantics in web 2.0 social media. In WSDM, conference on web search and data mining, pp. 281–290
- Song C, Koren T, Wang P, Barabási A-L (2010) Modelling the scaling properties of human mobility. *Nat Phys* 6(10):818–823
- Tarasov A, Kling F, Pozdnoukhov A (2013) Prediction of user location using the radiation model and social check-ins. In KDD workshop on urban computing, p 8
- Wang C, Wang J, Xie X, Ma WY (2007) Mining geographic knowledge using location aware topic model. In GIR, AAMC Group on Information Resources, pp 65–70
- Wang Y, Yuan NJ, Lian D, Xu L, Xie X, Chen E, Rui Y (2015) Regularity and conformity: Location prediction using heterogeneous mobility data. In KDD, conference on knowledge discovery and data mining, pp 1275–1284
- Weng J, Lee BS (2011) Event detection in twitter. In ICWSM, international conference on weblogs and social media, pp 401–408
- Ye J, Zhu Z, Cheng H (2013) What's your next move: User activity prediction in location-based social networks. In: SDM SIAM international conference on data mining
- Yin Z, Cao L, Han J, Zhai C, Huang TS (2011) Geographical topic discovery and comparison. In: WWW, international World Wide Web conference, pp 247–256
- Yuan Q, Cong G, Ma Z, Sun A, Thalmann NM (2013) Who, where, when and what: discover spatio-temporal topics for twitter users. In: KDD, conference on knowledge discovery and data mining, pp 605–613
- Yuan Q, Zhang W, Zhang C, Geng X, Cong G, Han J (2017) Pred: Periodic region detection for mobility modeling of social media users. In: WSDM, conference on web search and data mining
- Zhang C, Han J, Shou L, Lu J, La Porta T (2014) Splitter: mining fine-grained sequential patterns in semantic trajectories. *Proc VLDB Endowment* 7(9):769–780
- Zhang C, Zhang K, Yuan Q, Peng H, Zheng Y, Hanratty T, Wang S, Han J (2017a) Regions, periods, activities: uncovering urban dynamics via cross-modal representation learning. In: WWW, international World Wide Web conference
- Zhang C, Zhang K, Yuan Q, Tao K, Zhang L, Hanratty T, Han J (2017b) React: Online multi-modal embedding for recency-aware spatiotemporal activity modeling. In: SIGIR, international conference on information retrieval

- Zhang C, Zhang K, Yuan Q, Zhang L, Hanratty T, Han J (2016a) Gmove: Group-level mobility modeling using geo-tagged social media. In: KDD, conference on knowledge discovery and data mining
- Zhang C, Zhou G, Yuan Q, Zhuang H, Zheng Y, Kaplan L, Wang S, Han J (2016b) Geoburst: Real-time local event detection in geo-tagged tweet streams. In: SIGIR, international conference on information retrieval
- Zheng K, Zheng Y, Yuan NJ, Shang S (2013) On discovery of gathering patterns from trajectories. In: ICDE, IEEE international conference on data engineering, pp 242–253



Chao Zhang is an Assistant Professor at the School of Computational Science and Engineering, Georgia Institute of Technology. His research focuses on data mining and machine learning. He is a recipient of the Google Faculty Research Award (2020), the ACM SIGKDD Dissertation Award Runner-up (2019), the IMWUT Distinguished Paper Award (2018), the ECML/PKDD Best Student Paper Runner-up Award (2015), and the Chiang Chen Overseas Graduate Fellowship (2013).



Jiawei Han is Michael Aiken Chair Professor, University of Illinois at Urbana-Champaign. He is Fellow of ACM, Fellow of IEEE, and has received ACM SIGKDD Innovation Award (2004), IEEE Computer Society Technical Achievement Award (2005), IEEE Computer Society W. Wallace McDowell Award (2009), and Japan's Funai Achievement Award (2018).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

