Metadata of the chapter that will be visualized online

Chapter Title	Spatiotemporal Analysis of Social Media Data	
Copyright Year	2016	
Copyright Holder	Springer International Publishing Switzerland	
Author	Family Name	**Yu**
	Particle	
	Given Name	**Rose**
	Suffix	
	Division	Computer Science Department
	Organization	University of Southern California
	Address	CA, USA
Corresponding Author	Family Name	**Liu**
	Particle	
	Given Name	**Yan**
	Suffix	
	Division	Computer Science Department, Viterbi School of Engineering
	Organization	University of Southern California
	Address	Los Angeles, CA, USA
	Email	yanliu.cs@usc.edu

Author's Proof

S

Spatiotemporal Analysis of Social Media Data

Rose Yu[1] and Yan Liu[2]
[1]Computer Science Department, University of Southern California, Los Angeles, CA, USA
[2]Computer Science Department, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA

AU1

Synonyms

AU2

Social media analysis; Spatiotemporal data mining; Spatiotemporal modeling

Definition

Social media provide a convenient platform for users to create and share content or to participate in online social activities. With the development of sensor technologies, it also generates large amount of spatiotemporal data, such as check-in records, user restaurant reviews, and geo-temporal tagged tweets. This entry specifically considers analyzing the spatiotemporal patterns in social media data. The problem involves identifying spatiotemporal correlations, building spatiotemporal models, and making predictions in space and time. Given that spatiotemporal observations have complex correlations, the major challenge of the problem is how to take into account the spatial and temporal correlations within the context of social media.

© Springer International Publishing Switzerland 2016
S. Shekhar et al. (eds.), *Encyclopedia of GIS*,
DOI 10.1007/978-3-319-23519-6_1629-1

Historical Background

Spatiotemporal analysis for social media data is a relatively young area. Many efforts have been focused on geographical topic discovery, e.g., Mei et al. (2006), Yin et al. (2011), and Ahmed et al. (2013). These approaches correlate spatiotemporal context such as geotags and time stamps with topic modeling of social media contents. A related task is spatiotemporal density estimation. For example, the seminal work in Sakaki et al. (2010) estimates event-related Twitter posts for detecting earthquakes; Xu et al. (2012) recovers spatiotemporal event density of animal road killing from tweets. Another line of work extends collaborative filtering and incorporates spatial proximity and temporal components for better recommendation (Lu et al. 2009; Agarwal et al. 2010; Li et al. 2011). More recently, tensor learning have been examined to automatically infer the spatiotemporal dependence structures from social media data (Lin et al. 2009; Bahadori et al. 2014) .

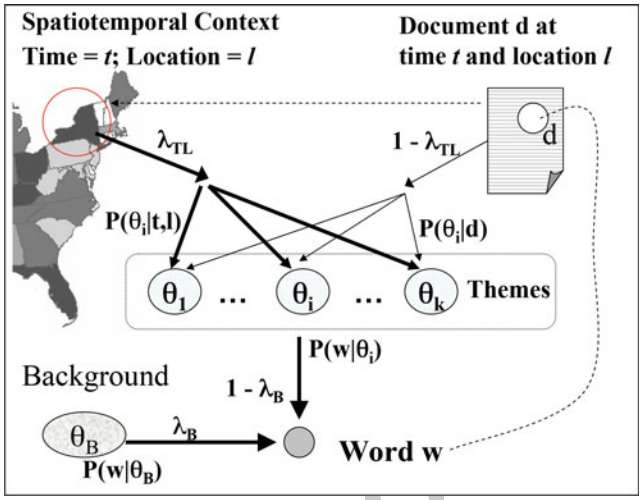
Scientific Fundamentals

Analysis of spatiotemporal data in social media varies drastically according to specific predictive tasks. The following discussion describes a few

UNCORRECTED PROOF

Author's Proof

Spatiotemporal Analysis of Social Media Data, Fig. 1 Graphic model representation for spatiotemporal theme model (Mei et al. 2006)



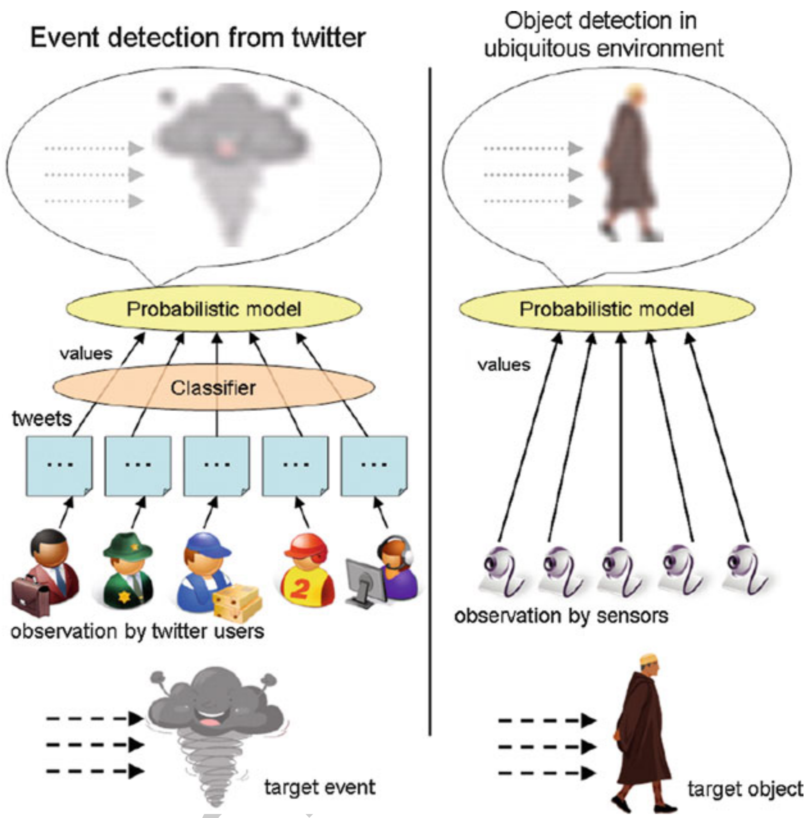
important spatiotemporal analysis techniques and representative algorithms in the field.

Spatiotemporal Topic Modeling Topic modeling is a powerful tool for semantic analysis in texts. Given a collection of social media content as well as the time and location information of the content, spatiotemporal topic modeling builds statistical models to discover the abstract "topics" from the social media content. Popular topic models include latent semantic analysis (LSA), probabilistic latent semantic indexing (PLSI), and latent Dirichlet allocation (LDA) (Blei et al. 2003). Spatiotemporal topic models leverage the spatiotemporal information to summarize the time-varying and location-aware topics in an unsupervised fashion. For example, *spatiotemporal theme model* is proposed in Mei et al. (2006) to discover the geographical topics. Given a finite set of time stamps $\{t_1, t_2, \cdots, t_{|T|}\}$ and locations $\{l_1, l_2, \cdots, l_{|L|}\}$, document collection $C = \{(d_1, \bar{t}_1, \bar{l}_1), (d_2, \bar{t}_2, \bar{l}_2), \cdots, (d_n, \bar{t}_n, \bar{l}_n)\}$, where d_i is the i-th document at time $\bar{t}_i$ at location $\bar{l}_i, \bar{t}_i \in \{t_1, t_2, \cdots, t_{|T|}\}$ and $\bar{l}_i \in \{l_1, l_2, \cdots, l_{|L|}\}$. Figure 1 depicts the proposed spatiotemporal theme model. Specifically, the model defines a **theme** in a text collection C as a probabilistic distribution of words characterizing a semantically coherent topic or subtopic, represented with θ. It treats the time stamp and location as separate variables and assumes that the words at location $l \in L$ and time $t \in T$ are generated as a mixture model of background theme θ_B and global themes $\{\theta_i\}$ with $i \in \{1, \cdots, k\}$. The model defines the probability of a word w as $p(w) = \lambda_B p(w|\theta_B) + (1 - \lambda_B) \sum_{j=1}^{k} p(w, \theta_i|d, t, l)$, where λ_B is the probability of choosing θ_B. The global theme θ_i is a mixture of document factor and spatiotemporal factor, $\theta_i = (1 - \lambda_{TL}) p(\theta_i|d) + \lambda_{TL} p(\theta_i|t, l))$, where λ_{TL} is a parameter to indicate the probability of using the spatiotemporal context to choose a theme. The parameters in the model are estimated by expectation-maximization (EM) algorithm. Yin et al. (2011) further extends the probabilistic model by taking into account the spatial proximity. It assumes that the geography can be divided into N regions. Within each region, the location l_d follows a multivariate Gaussian distribution; thus, words that are geographically close are more likely to be clustered into the same topic. Though many attempts have been made to model the geographical topics, the difficulties in defining locations and capturing the heterogeneity of geographical context still exist. In Mei et al. (2006), locations are prespecified (city) and assumed to be independently identical to each other. In Yin et al. (2011), regions are defined as large grid cells. It is expected that geographical discretization should adapt to regional population as well as the amount of

AU3

UNCORRECTED PROOF

Author's Proof



Spatiotemporal Analysis of Social Media Data, Fig. 2 Correspondence between event detection from Twitter and object detection in an ubiquitous environment (Sakaki et al. 2010)

social media content. Another concern is that LDA-type models are less effective for modeling short documents, such as tweets, since they highly rely on co-occurrences of words (Yan et al. 2013).

UNCORREC

Spatiotemporal Density Estimation Spatiotemporal density estimation treats the social media posts as discrete spatiotemporal events and takes them as inputs. It models the occurrence of those events in observed locations and predicts the event occurrence at unknown spatiotemporal point as outputs. For example, Sakaki et al. (2010) builds a real-time system to detect the center and the event trajectory of an earthquake shake. The system makes the analogy to sensing in ubiquitous environment, i.e., treating users as "social sensors" and tweets as sensory values, as shown in Fig. 2. For each tweet, the algorithm first conducts semantic analysis to classify whether the tweets are related to an event or not. Then it generates binary "sensor readings" of the user for those tweets. It computes the event occurrence probability with respect to the sensor readings to account for the reliability of social sensors. Assuming that the number of sensors at time t is n, the probability of event occurrence is $1 - p_f^n$ where p_f^n is the false-positive ratio of a sensor (user). With a collection of spatiotemporal events, the system applies Kalman filter to predict the event location. In their setting, each location is treated as a hidden state $z_t = (d_{xt}, d_{yt})^t$ with longitude and latitude values. Each event corresponds to a sensory measurement x_t. The Kalman filter operates recursively on sensory measurements to produce an estimate of the event location. Another work by Xu et al. (2013) develops a spatiotemporal signal recovery framework called Socioscope to model the event occurrence distribution. The framework considers binning the events where each bin is a spatiotemporal point, such as "California, day 1." Denoting x_i as the event count in bin i, Socioscope uses

Author's Proof

Poisson distribution for each bin and tries to recover the intensity f_i in the Poisson model. The authors address two challenges, i.e., human population bias by normalizing the occurrence with population intensity z_i and noisy data by imposing a stochastic transition. By adding the graph Laplacian L, Socioscope essentially solves the following penalized optimization problem:

$$\min_{f \in \mathbb{R}^n} - \sum_{i=1}^{m} (x_i \log h_i - h_i) + \lambda \Omega(f; L),$$

where h_i is the Poisson parameter after normalization and transition and Ω denotes the graph regularizer over L.

Research studies have shown that fewer than 0.42 % of the tweets are associated with geographical tags (Cheng et al. 2010), leading to extreme location sparsity. The highly skewed distribution of spatiotemporal events demonstrates that the Gaussian assumption in Kalman filter is inappropriate for spatiotemporal density estimation. Xu et al. (2013) employs Poisson point process estimation. However, it concatenates the spatial dimension with the temporal dimension and regards them as bins. This approach inevitably loses the natural ordering of temporal dimension and fails to capture the evolution of spatiotemporal events.

Spatiotemporal Collaborative Filtering Classic collaborative filtering takes a subset of user ratings as input and predicts the rating values for the missing entries. It works by factorizing a rating matrix **R** of user over item, into two components **P** and **Q**, where the rows in **P** are the user factors and rows in **Q** are the item factors. Spatiotemporal collaborative filtering considers the case when the users' ratings vary across space and time. Lu et al. (2009) investigates the problem of collaboration filtering with additional spatiotemporal information. In their setting, each user has a vector of ratings over time. In addition to the actual rating, extra side information, such as items' feedback and users' demographic profiles, is also available. The spatial component of the proposed model utilizes the side information. The algorithm imposes graph Laplacian prior to user factors **P** and item factors **Q**. The graph Laplacian matrices are constructed from item similarity matrix and user similarity matrix to regularize the spatial proximity. Hence, the MAP estimate of the latent factors is equivalent to minimizing

$$\|\mathbf{B} \circ (\mathbf{R} - \mathbf{P}\mathbf{Q})\|_F^2 \mid + \lambda(\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2)$$
$$+ \alpha(\mathrm{tr}(\mathbf{P}^T \Delta_p \mathbf{P})) + \mathrm{tr}(\mathbf{Q}^T \Delta_q \mathbf{Q})$$

where **B** indicates the location of observed rating **R**, with one as observed and zero otherwise. Δ is the graph Laplacian constructed from similarity matrix **W** as $\Delta = \mathbf{D} - \mathbf{W}$. For temporal modeling, the algorithm assumes the user factors follow a random walk driven by the Gaussian noise, i.e., $\mathbf{p}_t^{(u)} = \mathbf{p}_{t-1}^{(u)} + \mathbf{w}_t^{(u)}\ and\ \mathbf{r}_t^{(u)} = \mathbf{H}_t^{(u)}\mathbf{p}_t^{(u)} + \mathbf{v}_t^{(u)}$, where **p** and **r** correspond to the column vectors of matrix **R**,**P** and **w** and **v** are the Gaussian noise. Parameter estimation of spatiotemporal collaborative filtering follows the standard Kalman filter update. To avoid the expensive matrix inversion at the Kalman filter updating step, mean field approximation is used to estimate the variance matrix given new observations, denoted as $\Sigma_{t|t}$.

Spatiotemporal collaborative filtering extends matrix factorization to include spatiotemporal features, which mitigates the cold start issue. However, the number of latent factors in matrix factorization needs to be decided by hand tuning. Automatic parameter tuning scheme such as Bayesian optimization (Snoek et al. 2012) could be a possible solution. As an alternative for the full singular value decomposition step, fast low-rank matrix approximation techniques such as sampling (Deshpande and Vempala 2006) can also be applied.

Spatiotemporal Analysis via Tensor Learning Spatiotemporal data has complex dependencies between locations and time. Tensor, a multi-dimensional array, provides a convenient way to capture interdependencies along multiple directions. Therefore, it is natural to represent the multivariate spatiotemporal data as tensors. Using tensor as a tool to analyze spatiotemporal

UNCORRECTED PROOF

Author's Proof

data has seen applications in image coding and brain topography modeling; see Kolda and Bader (2009) for a detailed review. Spatiotemporal analysis via tensor learning aims to capture the spatiotemporal correlations by casting the data as tensors and presenting a unified framework for many spatiotemporal analysis tasks. In particular, Bahadori et al. (2014) studies two key tasks in spatiotemporal analysis, namely, cokriging, which is to estimate the measurements in unknown locations, and forecasting, which is to predict the future value of known locations. They proposes to use a 3-mode tensor to represent the spatiotemporal data. $\mathcal{X} \in \mathrm{R}^{P \times T \times M}$ denotes the observations in P locations in T time stamps for M variables. For cokriging task, the algorithm assumes the data tensor $\mathcal{X}$ to be low rank. For forecasting task, the algorithm models the spatiotemporal process using vector autoregressive (VAR) model and imposes low-rank structures on the model tensor. Tensor mode-n rank is used as the low-rank constraint, $\sum_{n=1}^{3} \operatorname{rank}(\mathcal{W}_{(n)}) \leq \rho$, where $\mathcal{W}_{(n)}$ denotes the tensor unfolding at mode n. The low-rank structure captures the spatial clustering effect, temporal periodicity, and commonality among variables. The entry applies the framework to climate data sets, Yelp reviews, and Foursquare check-in data set. It demonstrates that low-rank tensor learning cannot only significantly improve the performance of forecasting and cokriging (5–10 %) but also drastically reduce the computational costs (by at least 300 %).

Generalization from traditional matrix learning to tensor learning is nontrivial due to different mathematical properties of matrix and tensor. Many problems become NP-hard, such as the best rank-K approximation of a tensor. Furthermore, the computational bottleneck of high-dimensional tensors can be restrictive for large-scale data. A natural solution for scaling up machine learning algorithms can be parallelism. However, the complex spatiotemporal correlation and the inherent structure of tensor itself pose significant challenges to parallelism. Parallel factor analysis in low-rank tensor learning has been investigated in signal processing literature (Sidiropoulos et al. 2000); filling the gap between mathematical formulations and practical spatiotemporal applications is needed.

Key Applications

Spatiotemporal analysis techniques can be used in many application domains in social media data. Below, we briefly discuss some of the major applications.

Event Detection Spatiotemporal topic modeling can be applied to discover the trending topics (Munro et al. 2011), highlight events (Weng and Lee 2011), and automatically identify the events from photo tags (Rattenbury et al. 2007) in social media. Viral topics or bursts in social media event occurrence provide hints on potential anomalous events and lead to the development of many real-world systems, for example, alerting earthquake shakes (Sakaki et al. 2010), detecting abrupt crime or preplanned social event (Chae et al. 2012), and flu monitoring (Singh et al. 2010).

Signal Recovery Signal recovery extracts ground truth observations from noisy, incomplete social media data. For example, Mazumder et al. (2013) analyzes the political tweets in Indonesia to recover the degree of radical activities. Xu et al. (2013) recovers the wildlife surveillance statistics from animal road-killing tweets. Sengstock et al. (2013) studies the spatiotemporal context of tweets and recovers the actual geography of certain landmarks.

S

Network Inference The spatiotemporal dependency between social media and human mobility can be used to infer the relation among people. Bahadori et al. (2013) learns the spatiotemporal dependency of tweets and earthquake event thus infers mutual influence graph among the users. Cho et al. (2011) combines periodic short-range movements with travel due to the social network structure. The spatiotemporal co-occurrence of the social media users also has been used to associate the users with each other and cluster them into different groups (Lauw et al. 2005;

UNCORRECTED PROOF

Author's Proof

Crandall et al. 2010). Wang et al. (2013) models the spatiotemporal dependence of the congress roll call data and uses the inferred voting network to predict all votes of new legislation.

Recommendation Spatiotemporal contextual recommendation has contributed to social media content personalization and ads targeting. At the same time, combing social media information improves the accuracy and customer satisfactory of recommendation service. Various applications include recommending movies and news (Lu et al. 2009), point of interests (Lian et al. 2014), photo tags (Sigurbjörnsson and Van Zwol 2008), posted tweets (Liu et al.), and mobile apps (Karatzoglou et al. 2012).

Object Matching and User Localization Object matching refers to matching two entities in the database. Examples include matching user posts to the restaurants (Dalvi et al. 2012). Spatial information available in the social media or inferring the user location via text of their posts can greatly enhance accuracy of the object matching task (Dalvi et al. 2012). It shows that the possibility of friendship increases as users become geographically closer to each other; thus, social network friendship can be used to locate users beyond the resolution of their IP address (Backstrom et al. 2010). The authors in Cheng et al. (2010) show that analysis of tweet contents can also reveal the geographical location of a user.

Future Directions

Spatiotemporal analysis of social media data has attracted significant amount of attention in the community, resulting in a plethora of innovative methodologies and novel applications. However, to a large extent, the problem remains elusive, and there is still a dire need for the future research to take into consideration various open questions in spatiotemporal data analysis. For example, how to handle the extremely sparse observations in social media data? How to extract the information without violating users' privacy? How to naturally integrate social media contents with spatiotemporal side information closely? How to address the fundamental models?

For spatiotemporal analysis, which has a large literature in statistics and earth science, the definition of locations and the specification of spatiotemporal dependence might lead to significant different model performance. In parallel, considering the amount of social media content generated every day, it appears crucial to develop scalable algorithms that can learn spatiotemporal dependence structures efficiently and effectively. Indeed, spatiotemporal analysis of social media should account for the heterogeneity of population across regions and noisy observations. It should also reflect the underlying physical patterns of human mobility. Bayesian approach provides a potential solution to incorporate these domain knowledge into the model. However, the tuning of the prior hyper-parameters can be tedious. Then, in order to systematically compare the performance of different analytical results, benchmark evaluation metric would have to be developed to enable rigorous evaluation of different methodologies.

UNCORRECTED PROOF

Recommended Reading

Agarwal D, Chen B-C, Elango P (2010) Fast online learning through offline initialization for time-sensitive recommendation. In: KDD, Washington, DC. ACM, pp 703–712

Ahmed A, Hong L, Smola AJ (2013) Hierarchical geographical modeling of user locations from social media posts. In: Proceedings of the 22nd international conference on World Wide Web, Rio de Janeiro. International World Wide Web Conferences Steering Committee, pp 25–36

Backstrom L, Sun E, Marlow C (2010) Find me if you can: improving geographical prediction with social and spatial proximity. In: WWW, Raleigh. ACM, pp 61–70

Bahadori MT, Liu Y, Xing EP (2013) Fast structure learning in generalized stochastic processes with latent factors. In: KDD, Chicago. ACM, pp 284–292

Bahadori MT, Yu QR, Liu Y (2014) Fast multivariate spatiotemporal analysis via low rank tensor learning. In: NIPS, Montreal, pp 3491–3499

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022

Chae J, Thom D, Bosch H, Jang Y, Maciejewski R, Ebert DS, Ertl T (2012) Spatiotemporal social media analytics for abnormal event detection and exami-

Author's Proof

nation using seasonal-trend decomposition. In: IEEE conference on visual analytics science and technology, Seattle

Cheng Z, Caverlee, J, Lee K (2010) You are where you tweet: a content-based approach to geo-locating Twitter users. In: CIKM, Toronto. ACM, pp 759–768

Cho E, Myers SA, Leskovec J (2011) Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining, San Diego. ACM, pp 1082–1090

Crandall DJ, Backstrom L, Cosley D, Suri S, Huttenlocher D, Kleinberg J (2010) Inferring social ties from geographic coincidences. PNAS 107(52):22436–22441

Dalvi N, Kumar R, Pang B (2012) Object matching in tweets with spatial models. In: WSDM, Seattle. ACM, pp 43–52

Deshpande A, Vempala S (2006) Adaptive sampling and fast low-rank matrix approximation. In: Approximation, randomization, and combinatorial optimization. Algorithms and techniques. Springer, pp 292–303

Karatzoglou A, Baltrunas L, Church K, Böhmer M (2012) Climbing the app wall: enabling mobile app discovery through context-aware recommendations. In: Proceedings of the 21st ACM international conference on information and knowledge management, Maui. ACM, pp 2527–2530

Kolda TG, Bader BW (2009) Tensor decompositions and applications. SIAM Rev 51(3):455–500

Lauw HW, Lim E-P, Pang H, Tan T-T (2005) Social network discovery by mining spatiotemporal events. Comput Math Organ Theory 11(2):97–118

Li R, Li B, Jin C, Xue X, Zhu X (2011) Tracking user-preference varying speed in collaborative filtering. In: AAAI, San Francisco

Lian D, Zhao C, Xie X, Sun G, Chen E, Rui Y (2014) Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, New York. ACM, pp 831–840

Lin Y-R, Sundaram H, De Choudhury M, Kelliher A (2009) Temporal patterns in social media streams: theme discovery and evolution using joint analysis of content and context. In: IEEE international conference on multimedia and expo (ICME 2009), New York. IEEE, pp 1456–1459

Liu G, Fu Y, Xu T, Xiong H, Chen G, Discovering temporal retweeting patterns for social media marketing campaigns AU4

Lu Z, Agarwal D, Dhillon IS (2009) A spatiotemporal approach to collaborative filtering. In: RecSys, New York. ACM, pp 13–20

Mazumder A, Das A, Kim N, Gokalp S, Sen A, Davulcu H (2013) Spatiotemporal signal recovery from political tweets in Indonesia. In: Social computing, Washington, DC, pp 280–287

Mei Q, Liu C, Su H, Zhai C (2006) A probabilistic approach to spatiotemporal theme pattern mining on weblogs. In: Proceedings of the 15th international conference on World Wide Web, Edinburgh. ACM, pp 533–542

Munro R (2011) Subword and spatiotemporal models for identifying actionable information in Haitian Kreyol. In: CoNLL, Portland, pp 68–77

Rattenbury T, Good N, Naaman M (2007) Towards automatic extraction of event and place semantics from Flickr tags. In: Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval, Amsterdam. ACM, pp 103–110

Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World Wide Web, Raleigh. ACM, pp 851–860

Sengstock C, Gertz M, Flatow F, Abdelhaq H (2013) A probablistic model for spatiotemporal signal extraction from social media. In: SIGSPATIAL, Orlando. ACM, pp 274–283

Sidiropoulos ND, Bro R, Giannakis GB (2000) Parallel factor analysis in sensor array processing. IEEE Trans Signal Process 48(8):2377–2388

Sigurbjörnsson B, Van Zwol R (2008) Flickr tag recommendation based on collective knowledge. In: Proceedings of the 17th international conference on World Wide Web, Beijing. ACM, pp 327–336

Singh VK, Gao M, Jain R (2010) Situation detection and control using spatiotemporal analysis of microblogs. In: WWW, Raleigh. ACM, pp 1181–1182

Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian optimization of machine learning algorithms. In: Advances in neural information processing systems, Lake Tahoe, pp 2951–2959

Wang E, Salazar E, Dunson D, Carin L et al (2013) Spatiotemporal modeling of legislation and votes. Bayesian Anal 8(1):233–268

Weng J, Lee B-S (2011) Event detection in Twitter. ICWSM 11:401–408

Xu J-M, Bhargava A, Nowak R, Zhu X (2012) Socioscope: spatiotemporal signal recovery from social media. In: Machine learning and knowledge discovery in databases, Beijing. Springer, pp 644–659

Xu J-M, Bhargava A, Nowak R, Zhu X (2013) Socioscope: spatiotemporal signal recovery from social media. In: ECML PKDD, Prague

Yan X, Guo J, Lan Y, Cheng X (2013) A biterm topic model for short texts. In: Proceedings of the 22nd international conference on World Wide Web, Rio de Janeiro. International World Wide Web Conferences Steering Committee, pp 1445–1456

Yin Z, Cao L, Han J, Zhai C, Huang T (2011) Geographical topic discovery and comparison. In: Proceedings of the 20th international conference on World Wide Web, Hyderabad. ACM, pp247–256

UNCORRECTED PROOF

Author's Proof

Author Queries

Query Refs.	Details Required	Author's response
AU1	Please check if edit made to the chapter title is okay.	
AU2	Please check if updated author affiliations are okay.	
AU3	Please check if edit to sentence starting "Given a finite set of time..." is okay.	
AU4	Please provide year and publisher details for "Liu et al.".	

Note:
If you are using material from other works please make sure that you have obtained the necessary permission from the copyright holders and that references to the original publications are included.

UNCORRECTED PROOF