Twitter-based Urban Area Characterization by Non-negative Matrix Factorization

Shoko Wakamiya
Kyoto Sangyo University
Kyoto, Japan
shokow@cc.kyoto-su.ac.jp

Ryong Lee
Korea Institute of Science and Technology Information
Daejeon, Korea
lee.ryong@gmail.com

Yukiko Kawai
Kyoto Sangyo University
Kyoto, Japan
kawai@cc.kyoto-su.ac.jp

Kazutoshi Sumiya
Kwansei Gakuin University
Hyogo, Japan
sumiya@kwansei.ac.jp

ABSTRACT

Due to the remarkable growth of various social networks boosted by the pervasive mobile devices, massive crowds can become social sensors which can share microbolgs on a variety of social situations and natural phenomena in urban space in real-time. In order to take advantages of the novel realm of crowd-sourced lifelogs to characterize urban areas, we attempt to explore characteristics of complex and dynamic urban areas by monitoring crowd behavior via location-based social networks. In particular, we define social conditions consisting of crowd's experiential features extracted from the analysis of Twitter-based crowd's lifelogs. Then, we explore latent characteristic faces of urban areas in term of 5-dimensional social conditions by applying Non-negative Matrix Factorization (NMF). In the experiments with massive geo-tagged tweets, we classify urban areas into representative groups based on their latent patterns which enable to comprehensively understand images of the urban areas focusing on crowd's daily lives.

Categories and Subject Descriptors

[**Information systems**]: Location based services; [**Social and professional topics**]: Geographic characteristics

Keywords

Crowd Experience; latent Patterns; LBSN; location-based social networks; NMF; Twitter

1. INTRODUCTION

Urban areas increasingly become a complex and dynamic space where numerous people, structures and a variety of social and natural phenomena are always mixed and interacting with each other. We need to understand social and geographic features of diverse areas in cities well for taking advantages of today's complex urban space for better living and decision makings, e.g. when renting or buying a house, deciding accommodations while travelling, conducting neighborhood marketing, and planning urban development. However, drawing an image of unfamiliar or ever evolving city in mind is a non-trivial task, since overwhelming complexity of big cities makes it hard for us to keep up with the increasing structures and characteristic features of diverse areas.

Conventionally, in order to draw an image of an urban space, we usually refer to various thematic maps which were well investigated and represented regarding urban facilities or characteristics by expert cartographers, or heuristically learning urban towns where we visited [11]. However, due to the recent urbanization, these limited approaches cannot satisfy today's users who would like to follow the most up-to-date urban status and broadly and comprehensively understand complicated urban areas from multiple perspectives intuitively. Meanwhile, the proliferation of mobile devices and the explosive growth of social networks have made a significant convergence bearing location-based social networks (LBSNs), with which people can share their lifelogs possibly with their whereabouts by the location-sensing functions on the mobile devices. In particular, this kind of drastic change has another important implication that massive crowds in an urban space with such capabilities can be regarded as social sensors. Obviously, different from electronic sensors for observing primitive physical status, crowds in LBSNs can be further smarter and talented sensors reporting on a variety of urban status from natural phenomena to social events [9]. Therefore, we are able to explore broader and more comprehensive characteristics of urban areas through massive crowd experiences shared over LBSNs.

In this paper, we attempt to extract urban characteristics in terms of crowd experiences observed through Twitter. Specifically, we explore latent patterns of crowd experiences in urban areas. For this, we define and measure five crowd experiential features as a social condition in each urban area; population, activity, mood, topic, and social relationship as shown in Figure 1. We assume that people

BigDAS '15, October 20-23, 2015, Jeju Island, Republic of Korea
 ISBN 978-1-4503-3846-2/15/10…$15.00
DOI: http://dx.doi.org/10.1145/2837060.2837079

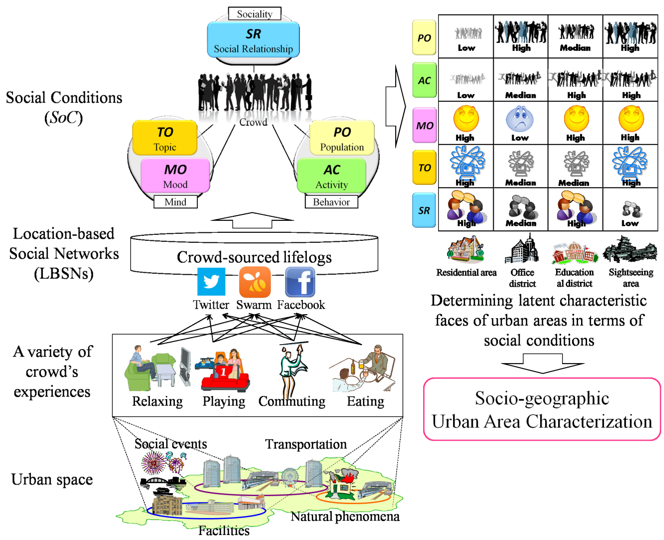


Figure 1: Research model: Crowd-sourced urban characterization based on LBSNs

accumulate partial images which they experienced or perceived when forming urban characteristics. In order to compute urban characteristics based on the process of human's perception, we generate a matrix of social conditions and urban areas by analyzing collected tweets and decompose the matrix by applying Non-negative Matrix Factorization (NMF) [6, 2]. NMF can effectively find out partial latent patterns of social conditions in urban areas because of its non-negativity constraints to the matrix. Then, we classify urban areas by measuring the similarity of their additive combinations of latent patterns. Finally, we reason common urban characteristics in the grouped urban clusters.

Our contributions are summarized as follows:

- to measure **social conditions** indicating the status of local circumstances focusing on crowd experiences such as behavior and mind using massive geo-tagged tweets,
- to determine latent localized patterns of urban areas in terms of social conditions by factorizing a matrix by means of NMF which enables us to consider the process of human's perception when understanding urban characteristics, and
- to classify them into representative groups by determining similar characteristics based on additive combinations of latent patterns.

The remainder of this paper is organized as follows: Section 2 describes related work. Section 3 explains the overall procedure to characterize urban areas and Section 4 illustrates experimental results with massive geo-tagged tweets. Finally, Section 5 concludes this paper and describes future work.

2. RELATED WORK

We summarize recent research work on extraction and utilization of crowd's experiential features as exemplified in this paper such as population, activity, mood, topic, and social relationship, with crowd-sourced dataset.

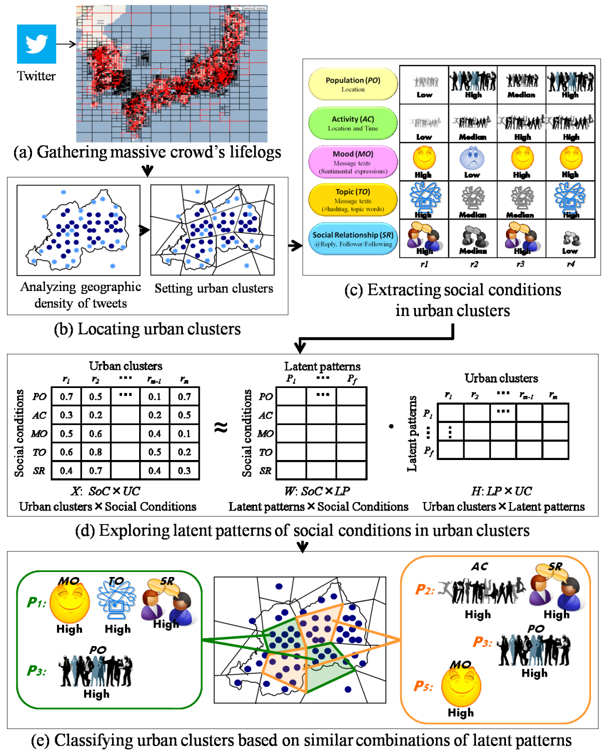


Figure 2: Procedure of socio-geographic urban characterization

There are several researches on urban analyses by monitoring crowd's behavior in an urban space observed using crowd-sourced datasets. In our previous work [7], we proposed a method to detect geo-social events and phenomena such as local festivals and natural disasters by measuring regularity of urban areas in terms of crowd's behavior observed quantitative analysis of geo-tagged tweets without using textual contents. Furthermore, we developed a method to characterize urban areas by measuring usual crowd's behavioral patterns in urban areas [16, 8]. Zheng et al. [18] interestingly analyzed tourist travel patterns by exploiting geo-tagged photos on photo-sharing services. McArdle et al. [12] investigated the usefulness of digital footprints of individual movement for calibrating human mobility models within an urban traffic micro-simulation framework. Hsieh et al. [3] presented a method to recommend a time-sensitive trip route by squeezing a lot of knowledge from check-in data over location-based services. Yuan et al. [17] proposed a framework to discover different functional regions such as educational areas, entertainment areas, and regions of historic interests in a city by analyzing human mobility based on trajectory data of taxies and POIs.

As for analyses of crowd's mind observed from crowd-sourced datasets, O'Connor et al. [14] examined the relations between public opinion from polls and sentiment measured from textual messages on social networks. Mislove et al. [13] generated cartograms for presenting public moods throughout a day in the U.S. by examining textual messages on Twitter. Lampos et al. [5] extracted topics in the real world from Twitter for detecting and predicting events or phenomena. In addition, Lehmann et al. [10] analyzed

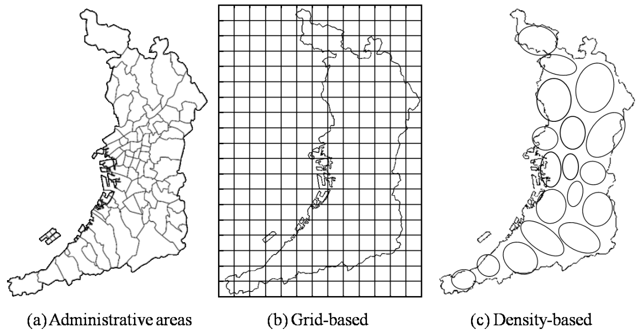


Figure 3: Types of urban cluster generation in Osaka, Japan

temporal hashtag-based topics to find event occurrence and epidemic spreading.

In contrast, in this paper, we extract five-dimensional social conditions which comprehensively include significant crowd experiences in an urban space using geo-tagged tweets for detecting latent characteristic faces of urban areas.

3. SOCIO-GEOGRAPHIC URBAN CHARACTERIZATION

3.1 Gathering Crowd's Urban Lifelogs on Twitter

First of all, we gather geo-tagged tweets from Twitter to observe crowd activities in the real world. Although, with Twitter, we can obtain tweets publicly published through the site's Open API[1], it is not feasible to gather massive geo-tagged tweets for a large scale area in a short time. Specifically, it takes considerable time to acquire a significant number of geo-tagged tweets because of certain practical limitations: In fact, Twitter presents a restricted Open API which solely supports the simplest near-by search based on a specified central location and a radius. Therefore, in order to overcome these restrictions and perform periodic monitoring of any size of user-specified regions, we developed a system to efficiently collect Twitter data in terms of geography based on a quad-tree space partitioning in our precious work [7] as shown in Figure 2(a).

3.2 Setting Urban Clusters

In order to monitor crowd experiences in urban areas for characterizing urban space, we need to set urban clusters by partitioning a given region into appropriate sub-areas as illustrated in Figure 2(b). As for ways of urban space partitioning for examining locally noteworthy areas, there are several different space-partitioning methods; a) administrative areas, b) grid-based space splitting, and c) density-based clustering shown in Figure 3. First, administrative areas can be formed by splitting a target region into prefectures and municipalities based on administrative boundaries; for instance, limits or borders of a geographical area under the jurisdiction of some governmental or managerial entity as shown in Figure 3(a). However, this approach cannot figure out social boundaries which are less relevant to the administrative boundaries, since crowd often easily cross over the

[1]Twitter Search API: http://apiwiki.twitter.com/Twitter-Search-API-Method%3A-search

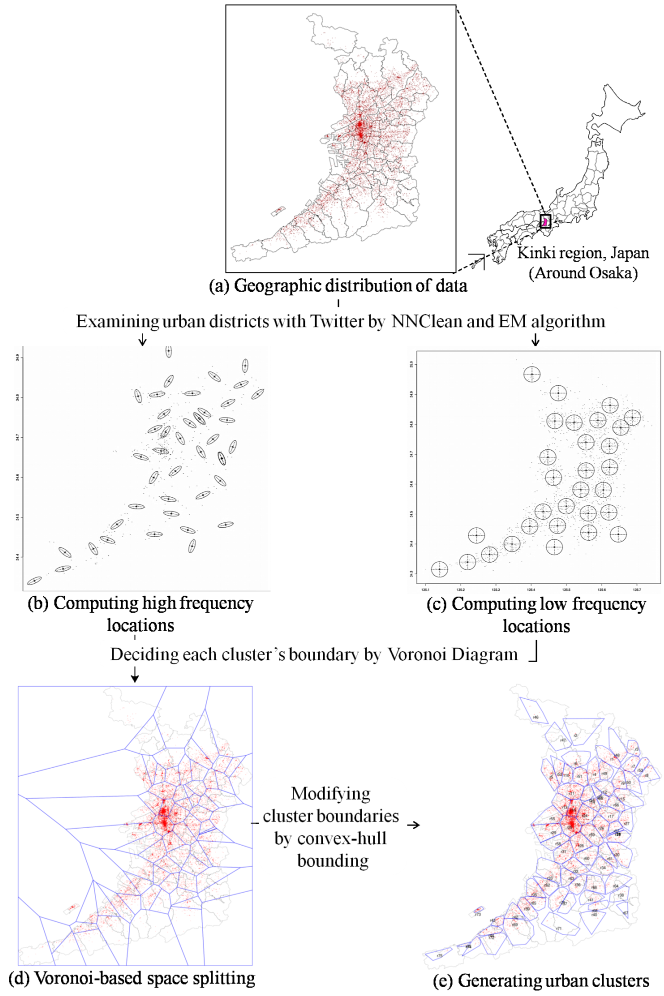


Figure 4: Process of constructing socio-geographic boundaries based on EM algorithm and Voronoi Diagram

administrative boundaries. Accordingly, we consider that this method would be inappropriate for examining crowd experiences. Next, as for the grid approach, it is difficult to decide the adequate cell size because a grid is formed by a lot of equal-sized cells as shown in Figure 3(b). In addition, it would consume considerable costs for observing crowd behavior due to the non-uniform distribution of crowds. On the other hand, in case of the clustering-based space partitioning as illustrated in Figure 3(c), it can reflect the geographical distribution of crowds, in our case, based on location information of geo-tagged tweets in an adaptive way. Thus, we can effectively establish the appropriate socio-geographic boundaries for the target region and partition into urban areas by referring to the actual geographic crowd experience.

In this work, we construct the cluster-based setting of urban clusters which can take into consideration of natural geographic distribution of crowds. Especially, in our experiment in Section 4, since we deal with millions of locational data of geo-tagged tweets obtained from Twitter, it requires enormous computational efforts to find appropriate partitioning. Therefore, we have to reduce the data size in much smaller and computable size without lack of essential quality

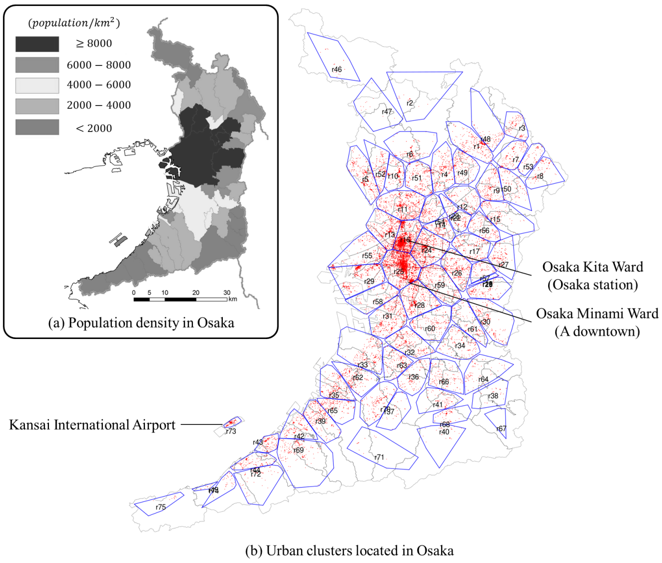


Figure 5: (a) Population density in Osaka and (b) extracted urban clusters

of the original data. For this, we adopt a spatial data summarizing method by Nearest Neighbor Clutter Removal [1] which can split the data into two groups of high-frequency part (dark blue in Figure 2(b)) and low-frequency part (light blue in Figure 2(b)) keeping the original geographic distribution. Although it is used for distinguishing noise from a given data, in the case of tweets, high-frequency points are naturally observed around high-populated areas. Therefore, using only high-frequency points is few clusters in suburban areas. In order to solve this problem, we also utilize the low-frequency parts. We also make clusters for the two datasets by means of EM algorithm as shown in Figure 4(b) and (c). After that, we depict a Voronoi diagram. using the central points of all clusters as shown in Figure 4(d). Finally, we generate convex-hull which creates a minimum convex boundary for each cluster to reduce unnecessary regions as shown in Figure 4(e) and Figure 5(b). Thus, we can effectively establish the appropriate boundaries for the target region and partition into urban clusters by referring to the actual geographic crowd experience.

3.3 Extracting Social Conditions from Crowd Experiences

We extract and utilize crowd experiences through LBSNs for characterizing urban areas as shown in Figure 2(c). Specifically, we compute a social condition for each urban cluster ($r_i \in R$), in terms of the five fundamental aspects. We define a social condition SoC_i for an urban cluster r_i in terms of the five indicators; Population (PO_i), Activity (AC_i), Mood (MO_i), Topic (TO_i), and Social Relationship (SR_i). The indicators are computed as a normalized value between 0 and 1.0 by dividing each value by the sum over all the clusters.

PO **(Population)** : This means the number of distinct users in each urban cluster. In order to get the count of users in an urban cluster r_i, we made a database query with a condition that find outs distinct users inside of the corresponding convex-based polygon.

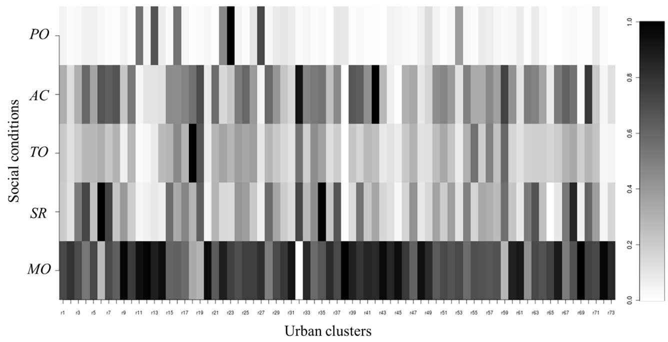


Figure 6: A matrix consisting of social conditions in all urban clusters

AC **(Activity)** : This is designated to measure the degree of active status of crowds for an urban cluster. Thus, we approach the most straightforward way by counting the number of tweets and crowd's movements. We count the inner movements for every moving user, whose the number of tweets at different locations should be at least more than or equal to two.

MO **(Mood)** : This represents the overall atmosphere of local society by measuring their sentiments. We first measure the sentiment of each tweet found in an urban cluster and eventually sum up the degree of crowd's sentiments. We refer to a Japanese dictionary consisted of the list of semantic words and their orientations which indicate whether each word has positive or negative meaning [15] and are represented by a numeric score between -1.0 (negative) and 1.0 (positive).

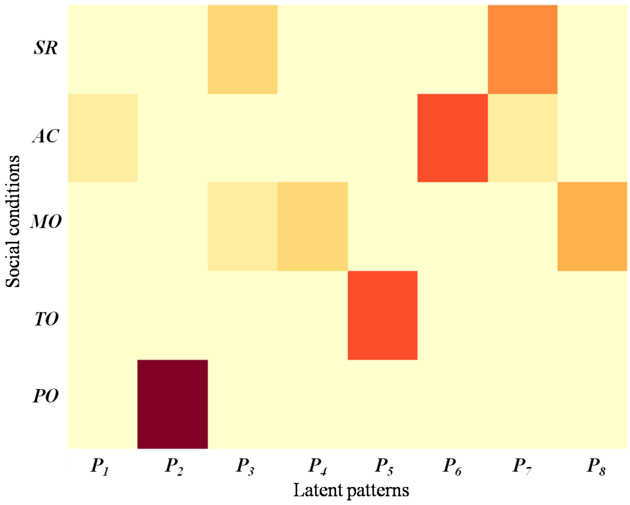
TO **(Topic)** : This is an interesting measure to examine how many different topics crowds are interested in. It may reflect the diversity of crowd's interests or local clusters. For the simplicity, we count the appearances of four types of information regarding hashtags, photos, videos and links, which can be extracted by means of regular expressions.

SR **(Social Relationship)** : The social relationships among people would be an important measurement to indicate the local society, which is eventually affecting local urban cluster's image. In the respect, we utilize two definitive clues from Twitter by 'retweet (RT)' and 'reply (@user_id)'.

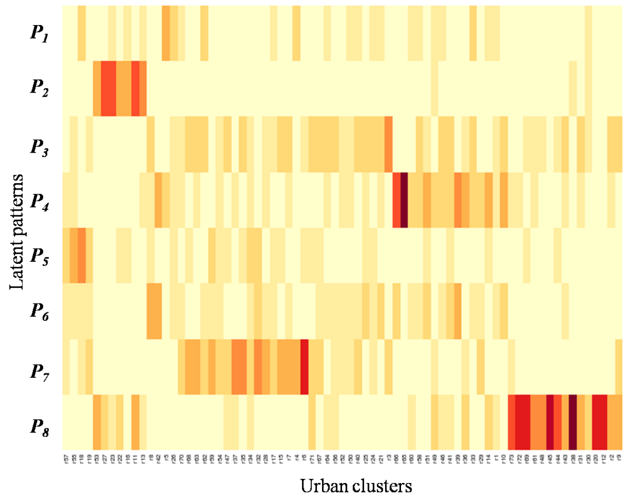
Consequently, we could obtain a matrix of social conditions of these five crowd's experiential features for urban clusters as illustrated in Figure 6, where 73 urban clusters shown in Figure 5 are aggregated above features.

3.4 Exploring Latent Characteristics of Urban Clusters

On the basis of the process of human's urban characterization, we find out latent partial patterns of social conditions in urban clusters by applying Non-negative Matrix Factorization (NMF) [6, 2] as shown in Figure 2(d) and extract urban characteristics as shown in Figure 2(e). Specifically, NMF decomposes a given matrix X $(= n \times m)$ to two matrices, W and H with non-negativity constraints. In fact, the



(a) Basis matrix X $(SoC \times LP)$



(b) Coefficient matrix W $(LP \times UC)$

Figure 7: A result of matrix factorization by NMF

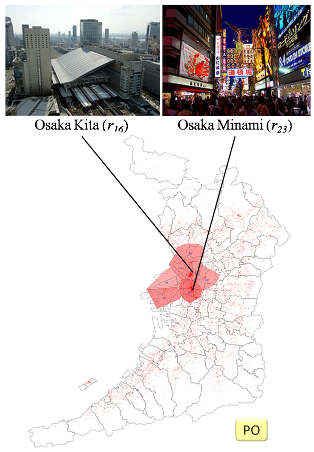
constraints are appropriate to deal with social conditions based on crowd experiences which must be either positive value or 0 quantitatively. In addition, NMF enables to discover a parts-based representation because of allowing only additive combination based on the constraints.

Formally, we define the formula, $X = W \cdot H + \alpha$, where W is a basis matrix, H is a coefficient matrix and α is a residual matrix. The dimensions of the matrix factors W and H are $n \times f$ and $f \times m$, respectively. The product $W \cdot H$ can be regarded as a compressed form of the data in X. In this paper, we first construct a matrix X made of social condition vectors $SoC = (PO, AC, MO, TO, SR)$ and urban clusters ($r_i \in UC$,$1 \leq i \leq m$) as given in Figure 6 and then decompose it into a basis matrix $SoC \times LP$ and a coefficient matrix $LP \times UC$, where LP is a set of latent patterns $P_1, \ldots, P_f$. By this, we represent social conditions in an urban cluster as an additive combination of latent patterns of the social conditions in the urban cluster. Then, in order to extract common urban characteristics, we group urban clusters having similar additive combinations of latent patterns. Specifically, we measure a similarity between urban clusters using Muliti-Dimensional Scaling (MDS) [4] and project them on a 2D space. This method provides a mapping where similar high-dimensional data are positioned close on an output space. Then, we classify urban clusters into k urban groups by means of k-means algorithm (k was empirically set on 4 in the experiments described in Section 4). Finally, we extract and reason representative combinations of latent patterns as characteristics of urban clusters in the same urban group.

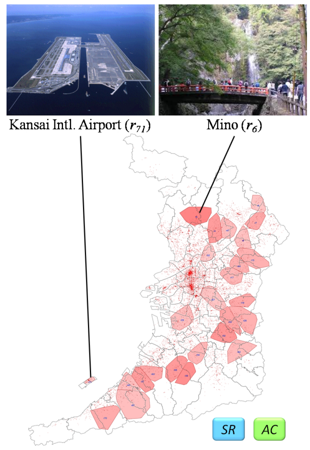
4. EXPERIMENT

4.1 Experimental Setting

In order to extract social conditions of urban areas, we collected geo-tagged tweets from Twitter using our geographical tweets gathering system [7]. Consequently, we could acquire $1,041,449$ geo-tagged tweets from $60,828$ distinct users for one week between August 19-26, 2012 in Osaka,



(a) Urban clusters of latent pattern P_2



(b) Urban clusters of latent pattern P_7

Figure 8: Geographic distribution of urban clusters related to latent patterns

Japan with the latitude range [34.27182: 35.05129] and the longitude range [135.09316: 135.74660]. Next, we located urban clusters in the target space by constructing socio-geographic boundaries based on the density of crowds as shown in Figure 5(a). As a result, we could find 73 polygonal urban clusters (from r_1 to r_{73}) as illustrated in Figure 5(b). In fact, these clusters were successfully created as our expectation to cover most downtown areas in this area. For instance, a populated area in Osaka city was separated into two major clusters; the one is Osaka Kita Ward (r_{16}) which is a busy commercial area around the largest station in Osaka, Osaka station, and the other is Osaka Minami (r_{23}) which is a popular downtown where lots of young people, families and tourists come. In addition, our method could take Kansai International Airport (r_{71}) as one of urban clusters.

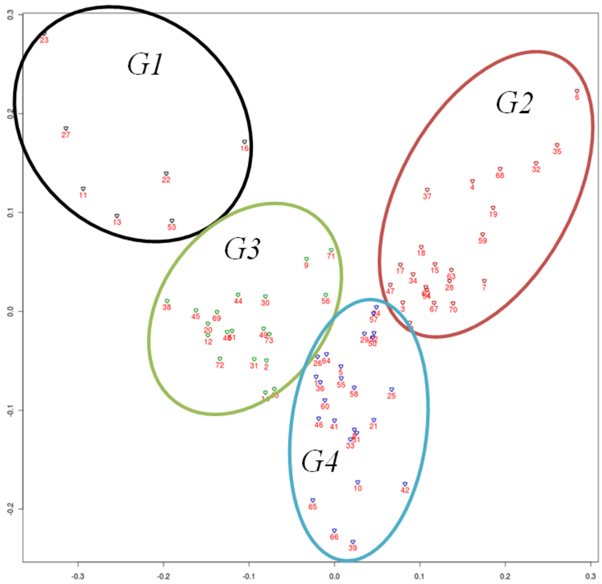


Figure 9: Urban clusters grouped by MDS and k-means algorithm ($k = 4$)

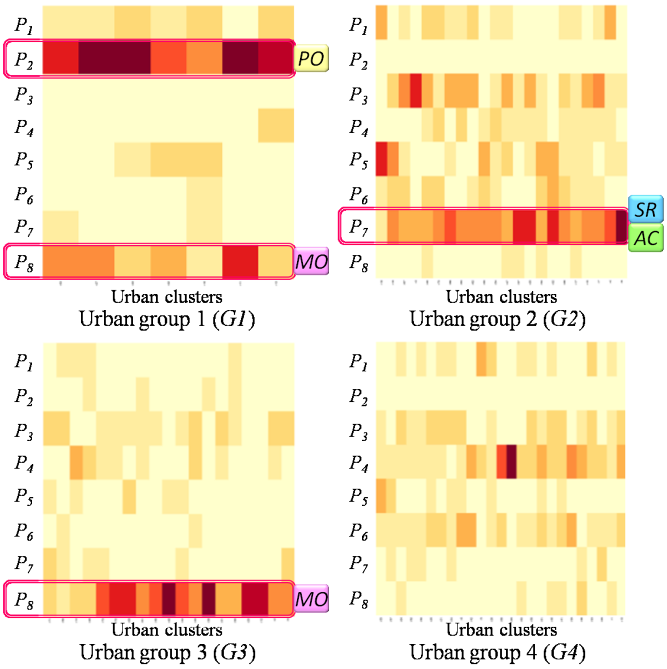


Figure 10: Characteristic latent patterns in urban groups

4.2 Examined Latent Patterns of Urban Clusters

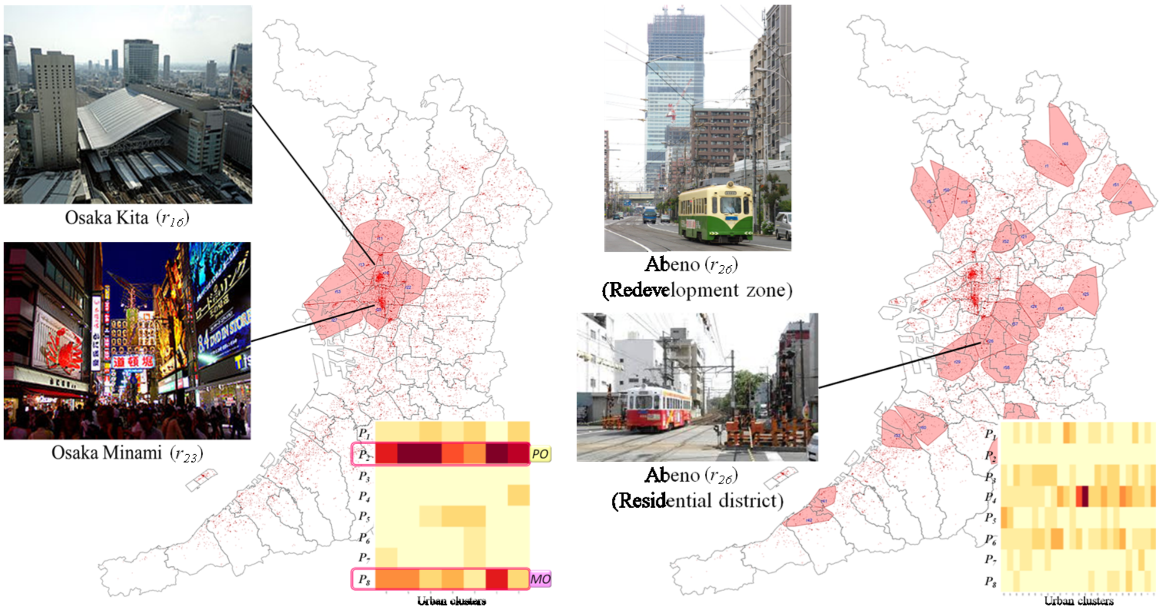
Figure 6 shows a matrix consisting of 5-dimensional social conditions and 73 urban clusters ($SoC \times UC$). This matrix represents a normalized summary for all the urban clusters (in the columns) regarding social conditions (in the rows). With this matrix, we examine the target space in terms of social conditions (SoC) in a simple way. However, we need to conduct further analysis to examine the implicit and latent patterns which can classify the urban clusters UC regarding the possible combination of social conditions. Subsequently, by applying NMF to this matrix, we decomposed it into two matrices $SoC \times LP$ and $LP \times UC$ as shown in Figure 7(a) and (b) respectively, where this result is expressed by means of two heatmaps. The generated latent patterns mean that urban clusters of interests have 8 faces based on social conditions. In detail, the matrix $SoC \times LP$ in Figure 7(a) shows social conditions consisting of each latent pattern. For instance, P_1 consists of AC, while P_3 consists of SR and MO. In the respect of social conditions (in the rows), we can see MO and AC are related to three latent patterns, respectively (the former is P_3, P_4, and P_8 and the latter is P_1, P_6, and P_7). In contrast, SR is relevant to two latent patterns; P_3 and P_7, and PO and TO are strongly related to just one latent pattern; P_2 and P_5, respectively.

Then, we represented the results in the two matrices on a map for examining the relations between social conditions and urban clusters focusing on geographic features. Here, we show two maps detailing the areas of P_2 and P_7 as illustrated in Figure 8(a) and (b), respectively. Urban clusters of P_2 have a characteristic face in terms of population. In fact, we can find the relevant urban clusters in Osaka city abundantly crowded in this whole experimental area. Especially, these urban clusters include the largest station, Osaka station (r_{16}) and some downtowns like Osaka Minami (r_{23}). On the other hand, urban clusters of P_7 are distinguished from others in terms of activity, topic, and mood. Actually, we can see these clusters in sub-urban of this experimental area. Interestingly, we can find r_{71} (Kansai International Airport) where lots of people obviously go to abroad or other cities in Japan, they may feel happy and aggressively write tweets.

4.3 Grouped Urban Areas Based on Combinations of Latent Patterns

Next, we grouped urban clusters which were observed similar combinations of latent patterns. Based on the determined relations in Figure 7, we measured MDS-based similarity of all urban clusters. Then, we generated 4 urban groups (from $G1$ to $G4$) by applying k-means algorithm (k was empirically set on 4) as illustrated in Figure 9.

Subsequently, we observed common combinations of latent patterns in members of each urban group. In fact, as for from $G1$ to $G3$, we could detect significant combinations of latent patterns P_2, P_7 and P_8, respectively, while $G4$ was less clear as shown in Figure 10. The figure also shows definite characteristics of $G1$ contrasting against urban clusters from $G2$ and $G3$. Here, we present two examples of $G1$ and $G4$. In the case of $G1$, urban clusters significantly having the latent pattern P_2 were classified as illustrated in Figure 10. As aforementioned above, these urban clusters were located in the center of Osaka city where lots of people are commuting, working, shopping, sightseeing, etc. as shown in Figure 11(a). In addition, the urban clusters in $G1$ had another characteristic latent pattern P_8. Therefore, the urban clusters were characterized in terms of large population and their sentiments. As for the urban clusters of $G4$ in Figure 10, it would be hard to grasp the image. Different from downtown mainly for visitors around area of Osaka Kita and Osaka Minami, these clusters can be regarded as multifunctional districts for local residents. Indeed, the corresponding clusters covered traditional shopping, entertainment and res-



(a) Urban clusters grouped in $G1$

(b) Urban clusters grouped in $G4$

Figure 11: Urban groups based on similarity of combination of latent patterns

idential districts as shown in Figure 11(b). Finally, we could successfully characterize urban clusters classified into 4 different groups based on common combinations of 8 latent patterns.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a novel method to characterize urban areas by measuring latent characteristic relations between urban areas and social conditions based on crowd experiences. In detail, we defined social conditions in urban clusters consisting of crowd's experiential indicators from geo-tagged tweets in terms of population, activity, mood, topic, and social relationship. Then we analyzed latent patterns by applying NMF to a matrix of social conditions and urban clusters. In the experiment with geo-tagged tweets, we could obtain reasonable results as urban area characterization focusing on crowd experiences in an urban space.

In the future work, we will perform further analyses of latent relations of multi-dimensional components such as space, time, and crowd experiences. In fact, it is a critical challenge to develop such kinds of crowd experience mining functions for the purpose of utilizing the social networks as a fundamental framework to understand the emerging open sharing space and furthermore to utilize the collective experiences of crowds for constructing practical urban life support systems.

Acknowledgments

This research was supported by the JSPS KAKENHI Grants (Nos. 26280042, 15K00162), Japan and the Ministry of Science, ICT and Future Planning (Grant K-15-L03-C01), Korea.

6. REFERENCES

[1] S. Byers and A. Raftery. Nearest-neighbour clutter removal for estimating features in spatial point processes. *Journal of the American Statistical Association*, 93:577–584, 1998.

[2] A. Cichocki, A. H. Phan, and R. Zdunek. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley, 2009.

[3] H.-P. Hsieh, C.-T. Li, and S.-D. Lin. Exploiting large-scale check-in data to recommend time-sensitive routes. In *Proc. of the ACM SIGKDD International Workshop on Urban Computing*, UrbComp '12, pages 55–62, 2012.

[4] J. B. Kruskal and M. Wish. *Multidimensional scaling*. SAGE Publications, 1978.

[5] V. Lampos and N. Cristianini. Nowcasting events from the social web with statistical learning. *ACM Trans. Intell. Syst. Technol.*, 3(4):72:1–72:22, Sept. 2012.

[6] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[7] R. Lee, S. Wakamiya, and K. Sumiya. Discovery of unusual regional social activities using geo-tagged microblogs. *World Wide Web*, 14(4):321–349, July 2011.

[8] R. Lee, S. Wakamiya, and K. Sumiya. Urban area characterization based on crowd behavioral lifelogs over twitter. *Personal and Ubiquitous Computing*, 17(4):605–620, 2013.

[9] R. Lee, S. Wakamiya, and K. Sumiya. Geo-social media analytics: Exploring and exploiting geo-social experience from crowd-sourced lifelogs. *SIGWEB Newsl.*, pages 4:1–4:8, 2014.

[10] J. Lehmann, B. Gonçalves, J. J. Ramasco, and

C. Cattuto. Dynamical classes of collective attention in twitter. In *Proc. of the 21st International Conference on World Wide Web*, WWW '12, pages 251–260, 2012.

[11] K. Lynch. *The Image of the City*. MIT Press, 1960.

[12] G. McArdle, A. Lawlor, E. Furey, and A. Pozdnoukhov. City-scale traffic simulation from digital footprints. In *Proc. of the ACM SIGKDD International Workshop on Urban Computing*, UrbComp '12, pages 47–54, 2012.

[13] A. Mislove, S. Lehmann, Y. Y. Ahn, J. P. Onnela, and J. N. Rosenquist. Understanding the demographics of twitter users. In *Proc. of the 5th Intl. AAAI Conference on Weblogs and Social Media*, pages 133–140, 2011.

[14] B. O'Connor, R. Balasubramanyan, B. Routedge, and N. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *Proc.of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM'10)*, 2010.

[15] H. Takamura, T. Inui, and M. Okumura. Extracting semantic orientations of words using spin model. In *Proc. of the 43rd Annual Meeting of the Association for Computational Linguistics*, ACL '05, pages 133–140, 2005.

[16] S. Wakamiya, R. Lee, and K. Sumiya. Exploring reflection of urban society through cyber-physical crowd behavior on location-based social network. In *Proc. of the 17th International Conference on Database Systems for Advanced Applications*, DASFAA'12, pages 168–179, 2012.

[17] J. Yuan, Y. Zheng, and X. Xie. Discovering regions of different functions in a city using human mobility and POIs. In *Proc. of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 133–140, 2012.

[18] Y.-T. Zheng, Z.-J. Zha, and T.-S. Chua. Mining travel patterns from geotagged photos. *ACM Trans. Intell. Syst. Technol.*, 3(3), May 2012.