

## 学位論文の要旨

論文題目

A Study on Analysis and Utilization of Crowd-sourced Spatio-temporal Contexts from Social Media  
ソーシャルメディアからの群衆ベースの時空間コンテキストの分析と活用に関する研究

印刷公表の方法及び時期 未公表

### Motivation

We have witnessed the great evolution of media environment where a variety of traditional media such as videos, photos, newspapers, TV, etc. are transformed into new types of massive on-line media such as Youtube, Niconico-Doga, Twitter, Facebook, and Foursquare. Characteristically, we can find a critical trend over the evolution that media consumers are no more passive receivers, rather much aggressively attending the whole process of media activity from creation to sharing often through recent popular social network sites. As for social media shared over the social network sites, the most important characteristic of social media is that they are created and shared by numerous crowds' voluntary participations reflecting their real-world lives. Furthermore, compared to the conventional media, social media on the social network services may reflect various aspects of people's social activities. Thereby, the social media are not only just logs of personal lives, but also the valuable source for knowing directly crowd's experiences and indirectly a variety of social phenomena via the media. For instance, user comments to video clips on video sharing websites such as YouTube and NicoNico-Doga are representing users' sentiment, opinions, interests, location information, etc. Therefore, due to the explosive growing social media space, we can take advantages of the extended information of crowd experiences **a) to advance conventional media systems** and furthermore **b) to conduct various social analytics utilizing collective crowd experiences as well as the media data.**

### Our Approaches

In order to exploit the crowd experiences buried in social media, we have to consider the emerging issues:

- Measuring spatio-temporal relations of crowd behavioral logs

In order to exploit the crowd experiences to advance conventional media systems, we need to interpret relations between multiple users' behavior via a media. Different from metadata of the conventional media, the one of social media has been attached from multiple viewpoints by many users and contained lots of noise. Therefore, it would be difficult to deal with it in the same manner as the conventional metadata which is attached by the same person without noise as much as possible. In addition, in order to conduct social analytics by utilizing crowd experiences, we need to extract

useful and valuable crowd experiences and aggregate them in terms of space, temporal aspects. Accordingly, we should explore how to analyze the relations of the crowd viewpoints or activities extracted from crowd behavioral logs over social networking services by focusing on spatial, temporal, and social aspects.

- Utilizing spatio-temporal relations

We develop effective applications which can utilize the extracted relations of the crowd experiences. The relations between spatio-temporal metadata are utilized for media extraction and summarization. Furthermore, the relations between crowd experiences based on areas can measure spatio-temporal characteristics and relations.

In Chapter 2, in order to position our research comparing with others and show the value of our research, we introduce related work.

In Chapter 3 and 4, we describe our approaches for utilizing social media as enhanced media comparing with conventional ones by people's participation and their various logs. Concretely, we measure media relation and popularity with crowd experiences; i) crowd-sourced video scene summarization and ii) crowd-powered TV viewing rates. As for a), we attempted to extract relevant scenes by measuring relations among user comments attached with pointing regions and time intervals. In the case of ii), the interaction from people to media like TV programs shows significant changes. Therefore, the way measuring the value of media is also changing. For this, we challenge to rate TV programs by collecting TV watching logs (experience) from crowd lifelogs over Twitter. Because the site was not designed for this specific goal to collect the TV-related Twitter messages, so-called tweets, we need to identify those that are relevant to TV programs.

In Chapter 5, we show a method to extract crowd-centered urban characteristics by observing crowd activities. Conventionally, in order to grasp characteristics of urban areas, a questionnaire-based survey has been conducted. However, it is time-consuming and hard to know dynamic characteristics. Therefore, we proposed a method to dynamically and lightly crowd-sourced urban characteristics by extracting latent patterns of crowd behavior in urban areas and classify the urban areas.

In Chapter 6, we explain a method to measure crowd-based proximity among urban areas by monitoring crowd movings. Here, we focus on the difference between geographic proximity and cognitive proximity which is deformed by crowd lifestyles depending on transportation infrastructure. Therefore, we measure proximity between urban areas in terms of spatial, temporal, and social aspects and extract vicinity areas.

Finally, we conclude this doctoral thesis in Chapter 7.

### **Scene Extraction System for Video Clips using Attached Comment Interval and Pointing Region**

A method was developed to enable users of video sharing websites to easily retrieve video scenes relevant to their interests. The system analyzes both text and non-text aspects of a user's comment and

then retrieves and displays relevant scenes along with attached comments. The text analysis works in tandem with non-text features, namely, the selected area and temporal duration associated with user comments. In this way, our system supports a better-organized retrieval of scenes that have been commented on with a higher degree of relevancy than conventional methods, such as using matching keywords. We describe our method and the relation between the scenes and discuss a prototype system.

### **Crowd-powered TV Viewing Rates: Measuring Relevancy between Tweets and TV Programs**

Due to the advance of many social networking sites, social analytics by aggregating and analyzing crowds' life logs are attracting a great deal of attention. In the meantime, there is an interesting trend that people watching TVs are also writing Twitter messages pertaining to their opinions. With the utilization of bigger and broader crowds over Twitter, surveying massive audiences' lifestyles will be an important aspect of exploitation of crowd-sourced data. In this paper, for better TV viewing rates in the light of the evolving TV lifestyles beyond home environments, we propose a TV rating method by means of Twitter where we can easily find crowd voices relative to TV watching. In the experiment, we describe our exploratory survey to exploit a large amount of Twitter messages to populate TV programs and on-line video sites.

### **Urban Area Characterization based on Crowd Behavioral Lifelogs over Twitter**

Recent location-based social networking sites are attractively providing us with a novel capability of monitoring massive crowd lifelogs in the real-world space. In particular, they make it easier to collect publicly shared crowd lifelogs in a large scale of geographic area reflecting the crowd's daily lives and even more characterizing urban space through what they have in minds and how they behave in the space. In this paper, we challenge to analyze urban characteristics in terms of crowd behavior by utilizing crowd lifelogs in urban area over the social networking sites. In order to collect crowd behavioral data, we exploit the most famous microblogging site, Twitter, where a great deal of geo-tagged micro lifelogs emitted by massive crowds can be easily acquired. We first present a model to deal with crowds' behavioral logs on the social network sites as a representing feature of urban space's characteristics, which will be used to conduct crowd-based urban characterization. Based on this crowd behavioral feature, we will extract significant crowd behavioral patterns in a period of time. In the experiment, we conducted the urban characterization by extracting the crowd behavioral patterns and examined the relation between the regions of common crowd activity patterns and the major categories of local facilities.

### **Crowd-sourced Cartography: Measuring Socio-cognitive Distance for Urban Areas based on Crowd's Movement**

On behalf of the rapid urbanization, urban areas are gradually becoming a sophisticated space where we

often need to know ever evolving features to take the most of the space. Therefore, keeping up with the dynamic change of urban space would be necessary, while it usually requires lots of efforts to understand newly visiting and daily changing living spaces. In order to explore and exploit the urban complexity from crowd-sourced lifelogs, we focus on location-based social network sites. In fact, due to the proliferation of location-based social networks, we can easily acquire massive crowd-sourced lifelogs interestingly indicating their experiences in the real space. In particular, we can conduct various novel urban analytics by monitoring crowd's experiences in an unprecedented way. In this paper, we particularly attempt to exploit crowd-sourced location-based lifelogs for generating a socio-cognitive map, whose purpose is to deliver much simplified and intuitive perspective of urban space. For the purpose, we measure socio-cognitive distance among urban clusters based on human mobility to represent accessibility of urban areas based on crowd's movement. Finally, we generate a socio-cognitive map reflecting the proposed socio-cognitive distances which have computed with massive geo-tagged tweets from Twitter.

### Conclusions

In this thesis, we described on analysis and utilization of crowd-sourced spatio-temporal contexts from social media. Especially, we defined social media as new media integrating conventional media and crowd experiences. We showed that the extended information of crowd experiences enables us to advance conventional media systems and furthermore to conduct various social analytics utilizing collective crowd experiences as well as the media data.

### 引用文献

1. "Scene Extraction System for Video Clips using Attached Comment Interval and Pointing Region," Shoko Wakamiya, Daisuke Kitayama, and Kazutoshi Sumiya, Multimedia Tools and Applications, Volume 54, Number 1, pp. 7-25, August 2011.
2. "Urban Area Characterization based on Crowd Behavioral Lifelogs over Twitter," Ryong Lee, Shoko Wakamiya, and Kazutoshi Sumiya, Personal and Ubiquitous Computing: Volume 17, Issue 4, pp. 605-620, April 2013.
3. "Crowd-sourced Cartography: Measuring Socio-cognitive Distance for Urban Areas based on Crowd's Movement," Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya, In Proc. of the 4th International Workshop on Location-Based Social Networks (LBSN 2012), pp. 935-942, Pittsburgh, Pennsylvania, USA, September 2012 [**Best Student Paper Award**]

### 参考論文

1. "Towards Better TV Viewing Rates: Exploiting Crowd's Media Life Logs over Twitter for

- TV Rating,” Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya, In Proc. of 5th International Conference on Ubiquitous Information Management and Communication (ICUIMC 2011), pp. 39:1-39:10, Seoul, Korea, February 2011.
2. “Crowd-powered TV Viewing Rates: Measuring Relevancy between Tweets and TV Programs,” Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya, In Proc. of the 2nd International Workshop on Social Networks and Social Media Mining on the Web (SNSMW 2011), Lecture Notes in Computer Science, 2011, Volume 6637, pp. 390-401, Hong Kong, April 2011.
  3. “Tweet 分析による群衆行動を用いた地域特徴抽出”, 李 龍, 若宮 翔子, 角谷 和俊, 情報処理学会論文誌データベース (TOD54), Vol.5, No.2, pp. 36-52, 2012 年 6 月.
  4. “Looking into Socio-cognitive Relations between Urban Areas based on Crowd Movements Monitoring with Twitter,” Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya, 日本データベース学会論文誌 (DBSJ Journal), Vol. 11, No. 2, pp. 19-24, 2012 年 10 月.
  5. “位置ベース SNS を通した群衆の移動経験に基づく都市空間の近接性分析,” 情報処理学会論文誌データベース (TOD58), 若宮 翔子, 李 龍, 角谷 和俊, Vol. 6, No. 3, pp. 159-176, 2013 年 6 月.