様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

# 科学研究費助成事業　　研究成果報告書

令和　２ 年　 ６ 月 ２９ 日現在

機関番号：６２６１５
研究種目：若手研究(B)
研究期間：2016～2019
課題番号：１６Ｋ１６０５８

研究課題名（和文）Participatory Sensing and Felicitous Recommending of Venues

研究課題名（英文）Participatory Sensing and Felicitous Recommending of Venues

研究代表者
ユ　イ（Yu, Yi）

国立情報学研究所・コンテンツ科学研究系・特任助教

研究者番号：００７５４６８１

交付決定額（研究期間全体）：（直接経費）　　2,900,000円

研究成果の概要（和文）：ビジュアル・コンテキスト・アウェア・アプリケーションは、ユーザーコンテキストに適合したサービスを提供するには、非常に有望である。本研究では、それに関して、1）ユーザーは、インターネット上のソーシャル共有プラットフォームから取得した場所の写真に興味を持っているが、この写真がどこで撮られたのか正確にはわからない。2）ユーザーが初めてある場所を訪れる。彼は自分がどこにいるか正確に知らず、アーバンキャニオン、屋内、または地下にいるため、モバイルデバイスGPSモジュールも位置を算出できない。マルチモーダル内容解析に基づいてきめ細かい場所発見というフレームワークで検討し、場所発見のデモシステムも開発した。

研究成果の学術的意義や社会的意義
Fine-grained venue discovery relies on the correlation analysis between images and text description of venues. Our research focuses on developing various methods to discover knowledge and relation from more complicated and challenging venue-based heterogeneous multimodal data generated by users.

研究成果の概要（英文）：Visual context-aware applications are very promising because they can provide suitable services adapted to user context. We consider two kinds of scenarios. 1) A user is very interested in a venue photograph obtained from social sharing platform on the Internet, but he does not exactly know where this photograph was taken. 2) A user visits a venue for the first time. He does not know exactly where he is, and the GPS module of his mobile device fails to compute a position, because the user is in an urban canyon, in a building or underground.
We study 1) exact venue search (find the venue where the photograph was taken) and 2) group venue search (find relevant venues that have the same category as the photograph) in a joint framework for fine-grained venue discovery based multimodal content association and analysis. Moreover, we also developed a venue discovery demo system based on proposal methods.

研究分野：マルチモーダル内容解析，人工知能・深層学習

キーワード：Venue discovery　Cross-modal retrieval　Multimodal learning　Deep Learning　CCA

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

## １．研究開始当初の背景

Online advertising, which usually shows the pros but not the cons of goods/services, is not so convincing to users without actual experience. On the other hand, many users visit different venues (e.g. restaurant, shopping center, scenic spot) in person, and share on social networks (e.g. on Foursquare, Facebook) their experiences at venues, in the formats of text, images, and videos etc. Such multimedia data not only implies user preferences, but also provides a lot of comments about venues. In this sense, users are involved in the participatory sensing of venues. When users decide places to visit or goods to purchase, they can refer to others' opinions. With an unprecedented growth of venues and venue-related multimedia data online, it is impractical for users to manually navigate venues. Therefore, the function of venue discovery and recommendation becomes necessary, and accurate recommendation calls for exploiting all kinds of user-generated venue-related multimodal data.

## ２．研究の目的

Visual context-aware applications are very promising because they can provide suitable services adapted to user context. We consider two kinds of scenarios. 1) A user is very interested in a venue photograph obtained from social sharing platform on the Internet, but he does not exactly know where this photograph was taken. 2) A user visits a venue for the first time. He does not know exactly where he is, and the GPS module of his mobile device fails to compute a position, because the user is in an urban canyon, in a building or underground. We study 1) exact venue search (find the venue where the photograph was taken) and 2) group venue search (find relevant venues that have the same category as the photograph) in a joint framework for fine-grained venue discovery based multimodal content association and analysis.

## ３．研究の方法

Few efforts focus on fine-grained venue discovery with more complicated real images generated by users such as venue photographs containing objects, geographic categories, and more meaningful semantic descriptions. Fine-grained venue discovery relies on the correlation analysis between images and text description of venues. Our research focuses on developing various methods including deep learning to discover and the knowledge and relation over venue-based heterogenous multimodal data. Given a venue photo as input, the system can predict the exact venue where the photo was taken and predict a group of relevant venues that have the same category as the input photo.

## ４．研究成果

In the past three years, this project focused on multimodal multimedia analysis and multimodal learning applied to venue discovery and recommendation over user-generated multimodal multimedia data. We took Wikipedia as reliable knowledge source about venues, and only take English Wikipedia into account in this work. We collected 1994 pairs of image-article from 5 cities in Wikipedia. With the same venue name and same geo-coordinate, we integrated about 19,792 Foursquare photos with categories in the area of Los Angeles and London. We studied to deal with semantics understanding and automatically generate a multimedia summary for a given event in real-time by leveraging different social media such as Wikipedia and Flickr. A representative work based on the proposed C-DCCA method can be found in [1]. Moreover, we also developed a venue discovery demo system as shown in Figure 1 based on proposal methods.

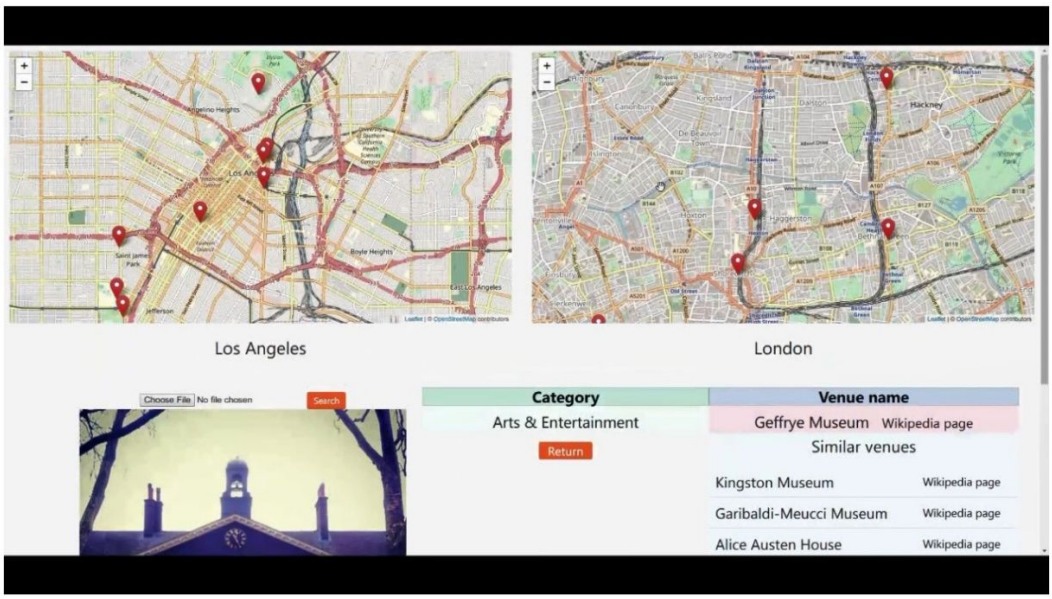

Figure 1: Demonstration screenshot of fine-grained venues discovery

Referred publication

[1] **Yi Yu**, Suhua Tang, Kiyoharu Aizawa, and Akiko Aizawa, “Category-based deep CCA for fine-grained venue discovery from multimodal data,” IEEE Transaction on Neural Network and Learning System (TNNLS), Vol. 30, No.4, pp.1250-1258, 2019. [CORE A*], IF: 11.683

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　2件／うち国際共著　1件／うちオープンアクセス　1件）

| | |
|---|---|
| １．著者名<br>Yi Yu, Suhua Tang, Kiyoharu Aizawa, and Akiko Aizawa | ４．巻<br>30 |
| ２．論文標題<br>Category-Based Deep CCA for Fine-Grained Venue Discovery from Multimodal Data | ５．発行年<br>2019年 |
| ３．雑誌名<br>IEEE Transaction on Neural Network and Learning System | ６．最初と最後の頁<br>1250-1258 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>10.1109/TNNLS.2018.2856253 | 査読の有無<br>有 |
| オープンアクセス<br>オープンアクセスではない、又はオープンアクセスが困難 | 国際共著<br>－ |

| | |
|---|---|
| １．著者名<br>Rajiv Ratn Shah, Yi Yu, Akshay Verma, Suhua Tang, Anwar Dilawar Shaikhe, Roger Zimmermann | ４．巻<br>108 |
| ２．論文標題<br>Leveraging Multimodal Information for Event Summarization and Concept-level Sentiment Analysis | ５．発行年<br>2016年 |
| ３．雑誌名<br>Journal of Knowledge-Based Systems | ６．最初と最後の頁<br>102-109 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>http://dx.doi.org/10.1016/j.knosys.2016.05.022 | 査読の有無<br>有 |
| オープンアクセス<br>オープンアクセスとしている（また、その予定である） | 国際共著<br>該当する |

〔学会発表〕　計5件（うち招待講演　2件／うち国際学会　5件）

| |
|---|
| １．発表者名<br>Yi Yu, Suhua Tang, Kiyoharu Aizawa, Akiko Aizawa |
| ２．発表標題<br>VenueNet: Fine-Grained Venue Discovery by Deep Correlation Learning |
| ３．学会等名<br>The 19th IEEE International Symposium on Multimedia (ISM 2017)（招待講演）（国際学会） |
| ４．発表年<br>2017年 |

| |
|---|
| １．発表者名<br>Chunlin Zhong, Yi Yu, Suhua Tang, Shin'ichi Satoh, Kai Xing |
| ２．発表標題<br>Deep Multi-label Hashing for Large-Scale Visual Search Based on Semantic Graph |
| ３．学会等名<br>APWeb/WAIM　2017（招待講演）（国際学会） |
| ４．発表年<br>2017年 |

１．発表者名
Rajiv Ratn Shah, Anupam Samanta, Deepak Gupta, Yi Yu, Suhua Tang, Roger Zimmermann

２．発表標題
PROMPT: Personalized User Tag Recommendation for Social Media Photos Leveraging Personal and Social Contexts

３．学会等名
IEEE International Symposium on Multimedia（国際学会）

４．発表年
2016年

１．発表者名
Rajiv Ratn Shah, Yi Yu, Suhua Tang, Shin'ichi Satoh, Akshay Verma, Roger Zimmermann

２．発表標題
Concept-Level Multimodal Ranking of Flickr Photo Tags via Recall Based Weighting

３．学会等名
Multimedia COMMONS Workshop at ACM Multimedia 2016（国際学会）

４．発表年
2016年

〔図書〕　計0件

〔産業財産権〕

〔その他〕
http://research.nii.ac.jp/~yiyu/
https://www.nii.ac.jp/faculty/digital_content/yu_yi/

６．研究組織

| | 氏名<br>（ローマ字氏名）<br>（研究者番号） | 所属研究機関・部局・職<br>（機関番号） | 備考 |
|---|---|---|---|