

ジオソーシャルデータに最適化された分析アルゴリズムと実行プラットフォームの研究

| | |
|-------|--|
| メタデータ | 言語: ja 出版者: 静岡大学 公開日: 2019-05-09 キーワード (Ja): キーワード (En): 作成者: 横山, 昌平 メールアドレス: 所属: |
| URL | http://hdl.handle.net/10297/00026467 |

平成 30 年 6 月 25 日現在

機関番号：13801

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00421

研究課題名(和文) ジオソーシャルデータに最適化された分析アルゴリズムと実行プラットフォームの研究

研究課題名(英文) Research on Geo-social Data analysis and platform

研究代表者

横山 昌平 (Yokoyama, Shohei)

静岡大学・情報学部・准教授

研究者番号：20443236

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：本研究は、ビッグデータ分析プラットフォームと、ビッグデータ分析アルゴリズムの二つの研究の柱を同時に進める事で、相補的な効果を得る計画である。この方針に基づき研究期間内において5件の雑誌論文および40件の学会発表を行った。
研究成果の具体的な一例としては、時空間データベースから、エンティティが高密度に集約しているところをクラスタとして抽出するアルゴリズムを提案した。このような用途では、従来は古典的なアルゴリズムであるDBSCANを用いていたが、本研究ではソーシャルビッグデータに最適化する事で、DBSCANの性能を大幅に超える、高速なアルゴリズムの開発に成功した。

研究成果の概要(英文)：This research was planning to obtain mutually benefit by advancing the subjects of the two studies which are related to big data analysis platform and big data analysis algorithm at the same time. Based on this policy, we conducted 4 journal articles and 40 conference presentations during the research period. As an example of the research, we proposed an algorithm to discover highly dense region from a large spatiotemporal database. In this type of application, DBSCAN, which is a classical algorithm, was used for the purpose. but we proposed more efficient algorithm by focusing on trajectory of entities in spatiotemporal database.

研究分野：データ工学

キーワード：ソーシャルデータ

1. 研究開始当初の背景

ジオソーシャルデータ分析に関する関連研究や研究代表者のこれまでの取り組みを俯瞰して明らかになる事は、クラウド時代・ビッグデータ時代と呼ばれる新しい時代に生まれたソーシャルデータの処理が、古典的な時空間データ処理技術に基づいて分析されているケースが多いという事である。

例えば、ジオタグが地理的に密集しているエリアを算出する為に使われる代表的なアルゴリズムとして 1996 年に発表された DBSCAN というアルゴリズムが良く使われている。DBSCAN では全てのジオタグが対等かつ独立した点(Point)として扱われてしまい、データのソーシャル性が失われてしまう。ソーシャル性とは、ソーシャルデータだからこそそのセマンティックの事で、例えばジオタグとはただの緯度経度情報ではなく、SNS ユーザのコンテキストを持った行動のスナップショットであるにとらえるべきである。

しかしながら DBSCAN は、そのソーシャル性を全く考慮しない結果を出力する。そのため、DBSCAN とは別にそれらを考慮する仕組みを取り入れなければならないが、当然ながら効率的とは言えない

本研究では、ジオソーシャルビッグデータと呼ばれる、新しく発生した巨大な情報源と、古典的なデータ分析技術の隙間に着目し、そのギャップを克服する新たなデータ分析プラットフォームとアルゴリズムの開発を行った。

2. 研究の目的

スマートフォンの普及によりジオタグ・チェックイン等地理情報を有するソーシャルデータ(ジオソーシャルデータ)の増加が著しい。本研究計画ではこのジオソーシャルデータの高度な利活用を目指し、収集から分析・可視化までを Web ブラウザ上でワンストップで実現するプラットフォームの構築と、ジオソーシャルデータに最適化された効率的な時空間情報分析アルゴリズムの実現を目指す。既存研究においてジオソーシャルデータの分析は、DBSCAN に代表される、古典的な空間アルゴリズムが使われている。しかしながら、ジオソーシャルデータは単なる空間データではなく、それが膨大なユーザの集合知であるという側面も持っている。そこで、ジオソーシャルデータに内包された人間のアクティビティをも抽出し分析・可視化するデータマイニング基盤を研究開発する。

本研究はこれまでに研究代表者が行ってきた、ジオソーシャルデータのビッグデータ分析技術を抽象化し、分析のための共通基盤となる実行プラットフォームと、ジオソーシャルデータに最適化された分析アルゴリズムの開発に取り組んだ。

そこで、応用研究(ジオソーシャルデータから様々な知見の抽出と可視化)の研究によ

って具現化した問題点を、分析アルゴリズムの問題と実行プラットフォームの問題に切り分け、双方の解決を目指した。すなわち、基礎研究(ジオソーシャルデータに最適化された新たなアルゴリズムの実現)とプラットフォーム開発(ソーシャルネットワークサービスとの親和性の高いデータマイニングプラットフォームの実現)の二つの柱を同時に進めた。

3. 研究の方法

研究代表者および、研究室の学生のべ 13 名により、研究を行った。途中、本研究に関連し JSPS 外国人特別研究員 Ágnes Bogárdi-Mészöly 氏を受け入れた。また本科研費を基課題として国際共同研究加速基金(国際共同研究強化)に採択され、フランス、University of Pau and Pays de l'Adour (UPPA)に 2016 年 10 月から 2017 年 9 月の期間で滞在し、在外国際共同研究を行った。

このように本研究は当初の予定にはない、国際的な環境にて研究を行う事ができた。具体的には、これまでに研究代表者が構築してきた、ジオソーシャルビッグデータのデータベースを Ágnes Bogárdi-Mészöly 氏と共有し、研究代表者が主にデータ処理アルゴリズムの研究、また Ágnes Bogárdi-Mészöly 氏が分析プラットフォームの研究としてアプリケーション事例の提案を行った。

4. 研究成果

本研究計画は、ビッグデータ分析プラットフォームと、ビッグデータ分析アルゴリズムの二つの研究の柱を同時に進める事で、相補的な効果を得る計画であり、研究成果は、大きくはその二つのどちらかの課題に大別される。ここではその中の代表的な成果を例にとって説明する。

まずは分析プラットフォームに関連する成果として[学会発表④]を説明する。本研究では、紅葉あるいは桜の開花という季節性の高いイベントにおいて、クオリティの高い写真の撮影場所と時間が、すなわち、その場所によるそれらのイベントを楽しむ為の最適な時期と、強い相関関係あるという仮説を実験的に検証したものである。

論文中 Fig.1(本資料にも引用)に手法の概要を記している。ここでは、画像の取得容易な特徴量として、色の占有率および、テクスチャの明瞭度を用いている。これらを、写真の SNS としての評価(Flickr の interestingness)と共に用いる事で、画像に対してランキングを行っている。

本研究では、この画像のスコアと、紅葉あるいは桜の開花因果関係がある気温や降水量のデータとの関係性を統計的に求めた。本研究では、画像のスコアと気候情報の二つのファクタを、カイ 2 乗検定によって関係性

を明らかにした(論文中 Table. 3、本資料にも引用)。実験により、この二つの事象が独立しているという、帰無仮説の棄却(論文中 Table 3、本ページにも引用)を確認した。また部分空間同定法により、二つのファクタの高い関係性(気温および降水量を用いる事により 70.84%を達成)を明らかにした。

次に分析アルゴリズムに関する代表的な成果として[学会発表 36]を紹介する。本研究は二次元データのクラスタリングとして良く利用されている古典アルゴリズム DBSCAN を置き換える事を目標とした、新しいクラスタリングアルゴリズム EBSCAN を提案している。DBSCAN に比べ、入力が点群ではなく、軌跡群でなければならないという制約があるものの、ジオソーシャルデータは SNS ユーザでグルーピング可能で、かつコンテンツ生成時間でソート可能な軌跡群であると考えられる事ができ、この制約は障害にはならない。EBSCAN のアルゴリズムのアイデアは論文中 Figure5(本ページにも引用)に示している。古典アルゴリズムである DBSCAN は、任意の点から指定した距離以内に指定した数の点以上あればクラスタを構成するアルゴリズムである。この DBSCAN はジオソーシャル分析の様々な既存研究で用いられている。それらの研究では、点の密度が高い区域を探すために DBSCAN を用いているが、その入力パラメータに密度に依存する値(Eps, MinPts)を指定しなければならない事から、パラメータ設定が非常に困難であるという問題がある。それに対して、提案手法の EBSCAN では、軌跡群に対して、入力パラメータは距離(指定した距離以上の移動は別クラスタに移動したものとみなすという数値)のみで、DBSCAN のようなクラスタリングを行える事を示した。

本論文では、実際に EBSCAN アルゴリズム

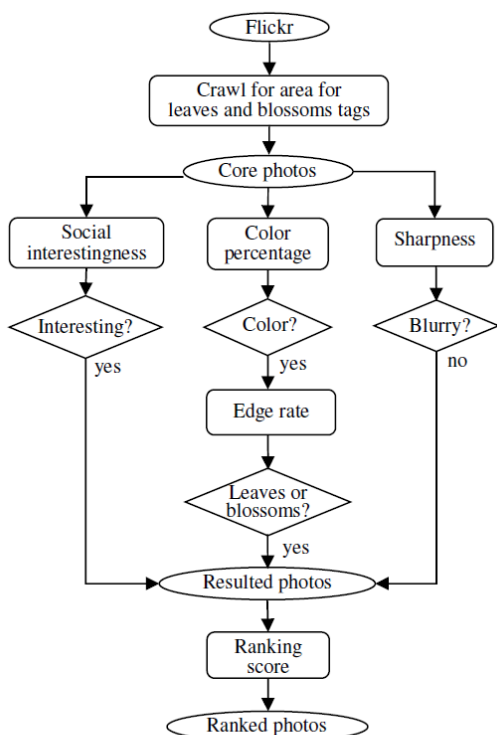


Fig. 1. Flowchart of detecting and ranking scenic leaves and blossoms viewing places.

を実装し、Flickr から取得した写真群に対して性能評価を行った。その結果、クラスタの質、そして処理速度の両面で DBSCAN を上回る事を示した。特に処理速度に関しては、DBSCAN より高速である事を示した(論文中 Figure12、本ページにも引用)。

Table 3. Detailed results of chi square tests.

| Factor | Chi square statistic | Degrees of freedom | Significance level | Critical value | H ₀ |
|----------------------------|----------------------|--------------------|--------------------|----------------|----------------|
| Average temperature in Nov | 438 | 54 | 0.00001 | 110.2 | false |
| Average temperature in Dec | 344.49 | 45 | 0.00001 | 97.37 | false |
| Annual average temperature | 288.45 | 36 | 0.00001 | 84.14 | false |
| Total precipitation in Nov | 243.5 | 45 | 0.00001 | 97.37 | false |
| Total precipitation in Dec | 180.95 | 45 | 0.00001 | 97.37 | false |
| Annual total precipitation | 228.79 | 54 | 0.00001 | 110.2 | false |

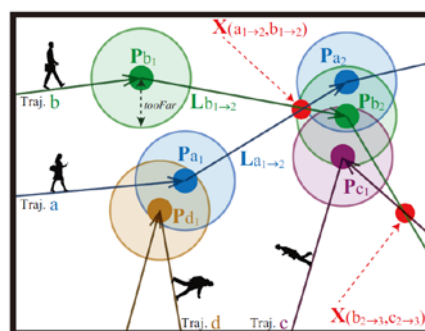


Figure 5: Key idea of entanglement-based approach.



Figure 12: Execution time.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 4 件)

- 菅野 真生, 江原 遥, 廣田 雅春, 横山 昌平, 石川 博, “道路ネットワーク分析を用いた災害時における避難リスクの高い経路の可視化”, DBSJ Journal Vol.15-J, pp.1-8, 査読有, 2017.
- 菊池 栞, 加藤 広大, 山田 竜平, 山本 幸生, 廣田 雅春, 横山 昌平, 石川 博, “深発月震源分類に適した機械学習法の研究”, 宇宙科学情報解析論文誌 第六号, 査読有, pp.51-62, info:doi/10.20637/JAXA-RR-16-007/00

05, 2017

- ③ 加藤 広大, 菊池 栞, 山田 竜平, 山本 幸生, 廣田 雅春, 横山 昌平, 石川 博, “SVM による深発月震分類のための有効な特徴量の分析”, 宇宙科学情報解析論文誌 第六号, 査読有, pp. 39-50, info:doi/10.20637/JAXA-RR-16-007/0004, 2017.
- ④ Ágnes Bogárdi-Mészöly, András Rövid, Shohei Yokoyama, “Performance Modeling of Web-based Software Systems with Subspace Identification”, Acta Polytechnica Hungarica Vol.13 No.7, pp.27-41, 査読有, 2016

[学会発表] (計 40 件)

- ① Chinnapong Angsuchotmetee, Richard Chbeir, Yudith Cardinale, Shohei Yokoyama, “A Pipelining-based Framework for Processing Events in Multimedia Sensor Networks”, The 33rd ACM/SIGAPP Symposium On Applied Computing (SAC 2018), 2018.
- ② Chinnapong Angsuchotmetee, Richard Chbeir, Yudith C. Cardinale, Shohei Yokoyama, “A Dynamic Event Detection Framework for Multimedia Sensor Networks”, the 23rd Asia-Pacific Conference on Communications (APCC2017), 2017.
- ③ Sayaka Tohyama, Yoshiaki Matsuzawa, Shohei Yokoyama, Teppei Koguchi, Yugo Takeuchi, “Constructive Interaction on Collaborative Programming: Case Study for Grade 6 Students Group”, IFIP World Conference on Computers in Education (WCCE 2017), 2017.
- ④ Shiori Kikuchi, Ryuhei Yamada, Yukio Yamamoto, Masaharu Hirota, Shohei Yokoyama, Hiroshi Ishikawa, “Classification of Unlabeled Deep Moonquakes Using Machine Learning”, Special Track on Social and Big Data collocated with the Ninth International Conferences on Advances in Multimedia, 2017.
- ⑤ Kodai Kato, Ryuhei Yamada, Yukio Yamamoto, Masaharu Hirota, Shohei Yokoyama, Hiroshi Ishikawa, “Analysis of Spatial and Temporal Features to Classify the Deep Moonquake Sources Using Balanced Random Forest”, Special Track on Social and Big Data collocated with the Ninth International Conferences on Advances in Multimedia, 2017.
- ⑥ 渡辺 優樹, 廣田 雅春, 石川 博, 横山 昌平, “Twitter を用いた主観的な天気 の可視化”, 第 9 回データ工学と情報マ ネジメントに関するフォーラム (DEIM2017), 2017.
- ⑦ 酒井 健, 廣田 雅春, 石川 博, 横山 昌平, “配列型データベースを用いた宇宙 科学データの可視化”, 第 9 回データ工 学と情報マネジメントに関するフォー ラム (DEIM2017), 2017.
- ⑧ 紋川 雄太郎, 廣田 雅春, 石川 博, 横 山 昌平, “ジオタグ付き写真を用いたお すすめ自撮りルートの提案”, 第 9 回デ ータ工学と情報マネジメントに関する フォーラム (DEIM2017), 2017.
- ⑨ 田邊 哲哉, 大島 純, 廣田 雅春, 石川 博, 横山 昌平, “グラフ型データベース を用いたアクティブラーニングにおけ る会話分析システムの提案”, 第 9 回デ ータ工学と情報マネジメントに関する フォーラム (DEIM2017), 2017.
- ⑩ 植野 友祐, 廣田 雅春, 石川 博, 横山 昌平, “飲料の透過光スペクトルデー タの収集とそれを利用した飲料判定の手 法”, 第 9 回データ工学と情報マネジメ ントに関するフォーラム (DEIM2017), 高山グリーンホテル(岐阜県), 2017.
- ⑪ 遠山紗矢香, 松澤芳昭, 横山昌平, 高 口鉄平, 竹内勇剛, “協調的なプログラ ミングを促す PBL 型学習環境の構築”, HCG シンポジウム 2016, 2016.
- ⑫ Keisuke Mitomi, Masaki Endo, Masaharu Hirota, Shohei Yokoyama, Yoshiyuki Shoji and Hiroshi Ishikawa, “How to find accessible Free Wi-Fi at Tourist Spots in Japan”, 8th International Conference on Social Informatics (SocInfo2016), 2016.
- ⑬ Masaki Kanno, Yo Ehara, Masaharu Hirota, Shohei Yokoyama, and Hiroshi Ishikawa, “Visualizing High-Risk paths using Geo-tagged Social Data for Disaster Mitigation”, 9th ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN16), 2016.
- ⑭ Ágnes Bogárdi-Mészöly, András Rövid, Shohei Yokoyama, “Time Trend Analysis of Scenic Leaves and Blossoms Viewing Places”, The 4th IIAE International Conference on Intelligent Systems and Image Processing 2016 (ICISIP2016), 2016.
- ⑮ Yuya Nozawa, Masaki Endo, Yo Ehara, Masaharu Hirota, Shohei Yokoyama, Hiroshi Ishikawa, “Inferring Tourist Behavior and Purposes of a Twitter User”, Artificial Intelligence for Tourism (AI4Tourism), 2016.
- ⑯ Ágnes Bogárdi-Mészöly, András Rövid, Shohei Yokoyama, “Detect Scenic Leaves and Blossoms Viewing Places from Flickr Based on Social and Image Features”, 1st International

- Conference on Social Informatics and Systems Science (SISS2016), 2016.
- ⑰ 遠藤雅樹, 莊司慶行, 廣田雅春, 横山昌平, 大野成義, 石川博, “位置情報付きツイートを利用した桜の地域別見頃推定結果の比較”, 2016.
 - ⑱ 櫻川直洋, 廣田雅春, 石川博, 横山昌平, “ジオタグ付き写真を用いた撮影スポットの性質を可視化するシステムの提案”, 第7回ソーシャルコンピューティングシンポジウム, 2016.
 - ⑲ 遠山由自, 廣田雅春, 石川博, 横山昌平, “地理的・時間的観点を考慮したジオタグ付きツイートの偏在性及び遍在性の可視化”, ARG 第8回 Web インテリジェンスとインタラクティブ研究会 (Wi2), 2016.
 - ⑳ 伊藤亜依子, 横山昌平, 木谷友哉, “大規模二輪車プローブデータ収集のためのデータベースの設計”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 21 三浦惇貴, 廣田雅春, 野澤浩樹, 横山昌平, “地域に詳しいブロガーの発見”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 22 櫻川直洋, 廣田雅春, 石川博, 横山昌平, “ジオタグ付き写真を用いたイベントとその穴場スポットの発見”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 23 土田崇仁, 遠藤雅樹, 加藤大受, 江原遥, 廣田雅春, 横山昌平, 石川博, “Word2Vec を用いた地域やランドマークの意味演算”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 24 井上貴弘, 廣田雅春, 石川博, 横山昌平, “ジオソーシャルデータを用いたフォトストリームの分割とホットスポットの関連づけ”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 25 三富恵佑, 遠藤雅樹, 江原遥, 廣田雅春, 横山昌平, 石川博, “外国人にアクセシブルな FreeWiFi がない観光スポットの発見”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 26 高木有人, 遠藤雅樹, 江原遥, 廣田雅春, 横山昌平, 石川博, “ジオタグ付き写真の撮影順序を考慮したホットスポットの分析”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 27 加藤広大, 山田竜平, 山本幸生, 横山昌平, 石川博, “既存月震分類の機械学習を用いた妥当性の検証”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 28 菊池栞, 山田竜平, 山本幸生, 横山昌平, 石川博, “月震分類に適した機械学習手法の検討”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 29 菅野真生, 江原遥, 廣田雅春, 横山昌平, 石川博, “道路ネットワーク分析を用いた災害時における避難リスクの高い経路の可視化”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 30 高橋一希, 加藤大受, 遠藤雅樹, 江原遥, 廣田雅春, 横山昌平, 石川博, “マイクロブログユーザの地域による嗜好の違いの分析”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 31 野沢悠哉, 遠藤雅樹, 江原遥, 廣田雅春, 横山昌平, 石川博, “マイクロブログを用いたユーザの訪問目的と動向の推定”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 32 伊藤貴明, 遠藤雅樹, 加藤大受, 江原遥, 廣田雅春, 横山昌平, 石川博, “Twitter を用いた駅イベント検出”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 33 竹内翔太郎, 廣田雅春, 石川博, 横山昌平, “サイクリングのための難易度と寄り道を考慮したルート推薦システムの実装”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 2016.
 - 34 Hiroshi Ishikawa, Masaki Endo, Iori Sugiyama, Masaharu Hirota, Shohei Yokoyama, “Can three-month time-series data of views or downloads predict the highly-cited academic papers in open access journals?”, International Workshop on Informatics (IWIN2015), 2015.
 - 35 青山賢, 廣田雅春, 石川博, 横山昌平, “ジオタグ付き写真を用いた様々な観光地の平均的な滞在時間の推定”, 第8回 Web とデータベースに関するフォーラム (WebDB Forum2015), 2015.
 - 36 Shohei Yokoyama, Ágnes Bogárdi-Mészöly, Hiroshi Ishikawa. “EBSCAN: An Entanglement-based Algorithm for Discovering Dense Regions in Large Geo-social Data Streams with Noise”, 8th ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN 2015), 2015.
 - 37 Ágnes Bogárdi-Mészöly, András Rövid, Shohei Yokoyama, “Performance Prediction of Web-Based Software Systems with Subspace Identification”,

The 4 th International Scientific Conference on Engineering and Applied Sciences, 2015.

- 38 Ágnes Bogárdi-Mészöly, András Rövid, Shohei Yokoyama, "Subspace Identification for Web-Based Software Systems", The 5th International Conference on Engineering and Applied Sciences (ICEAS 2015), 2015.
- 39 井上貴弘, 廣田雅春, 石川博, 横山昌平, "ジオソーシャルデータに基づいたフォトストリーム分割手法の提案", 第6回ソーシャルコンピューティングシンポジウム (SoC2015), 2015.
- 40 遠山由自, 廣田雅春, 石川博, 横山昌平, "ソーシャルメディア上に投影された情報の偏在性及び遍在性の可視化", ARG 第6回 Web インテリジェンスとインタラクション研究会 (Wi2), 2015.

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 0 件)

[その他]

ホームページ等

<http://lab.yokoyama.ac/>

6. 研究組織

(1) 研究代表者

横山 昌平 (YOKOYAMA, Shohei)
静岡大学・情報学部・准教授
研究者番号：20443236

(2) 研究分担者

なし

(3) 連携研究者

なし

(4) 研究協力者

Ágnes Bogárdi-Mészöly
JSPS 海外特別研究員
研究室の大学生・大学院生のべ 13 名