

協調ウェブ検索における ユーザ位置と閲覧コンテンツを用いたグループ推定手法

道念 幹^{1,a)} 高田 秀志^{2,b)}

概要: 近年, インターネット通信が可能な端末が普及することで, 人々がウェブ検索を行う機会が増加している. 学校の授業やワークショップなどでは学習者がグループを形成し, 協調してウェブ検索を行う協調ウェブ検索が取り入れられている. 協調ウェブ検索では情報を共有するためにファイルの交換が行われるが, 既存の情報共有方法ではグループの管理や送信先の選択の際にユーザに手間が生じ, 合意形成や意思決定の効率が低下する. ユーザが用いる端末のグループ形成を容易にすることで, グループ管理やファイルの送信先の選択をより簡便にすることができる. そこで, 協調ウェブ検索時には同一グループに所属しているユーザは同一のテーマで検索を行うため, ユーザ同士は近距離に存在し, 類似したウェブコンテンツを閲覧することに着目し, ユーザ位置と閲覧ウェブコンテンツを用いたグループ推定手法を提案する. 本手法では, ユーザが閲覧したウェブコンテンツを用いてユーザの特徴ベクトルを構成し, ベクトルに基づいてユーザのクラスタリングを行う. さらに, ユーザの位置を考慮し, クラスタリング結果を修正することでユーザがどのグループに属しているかを推定する.

A Grouping Method Using User Position and Browsed Content in Collaborative Web Search

DONEN TSUYOSHI^{1,a)} TAKADA HIDEYUKI^{2,b)}

Abstract: As terminals connected to the Internet become popular, more people are getting involved with searching Web sites. A collaborative Web search performed by people forming groups is implemented in school classes and workshops. When performing the collaborative Web search, people share digital contents. However, since conventional approaches require a lot of operations, efficiency of consensus building and decision making is decreased. By making grouping of user devices easier, operations for sharing digital contents become more efficient. We propose a grouping method using user position and browsed content, focusing on the closeness of people in the same group and accesses to similar web contents. In the method, the clustering of users is performed using feature vectors calculated by browsed contents. Moreover, considering user positions, the clustering result is refined.

1. はじめに

近年, モバイル端末やタブレット端末が広く普及しており, 国内の携帯電話, PHS, スマートフォンを含む端末の

世帯保有率は2016年の時点で94.7%である[1]. また, モバイル端末に限らずインターネット通信が可能な端末が普及することで, 人々がウェブ検索を行う機会が増加している. 人々がウェブ検索を日常的に行うようになるにつれ, 複数人で協調してある共通の目的を達成するためにウェブ検索をする協調ウェブ検索もよく行われている[2]. 教育現場では, 学習者がグループを作り, 協調ウェブ検索を行うことで得られた情報をまとめる授業が展開されている. それぞれのグループでは情報を共有するためにファイルの交換が行われており, 既存の情報共有方法としてDropbox[3]

¹ 立命館大学大学院情報理工学研究科
Graduate School of Information Science and Engineering,
Ritsumeikan University

² 立命館大学情報理工学部
College of Information Science and Engineering,
Ritsumeikan University

a) t_donen@cm.is.ritsumeai.ac.jp

b) htakada@cs.ritsumeai.ac.jp

や Google Dirve[4] などのウェブサービスを用いたものや、Bluetooth[5] や Wi-Fi Direct[6], AirDrop[7] などの近距離無線通信を用いたものが挙げられる。これらの情報共有方法では、グループの管理やファイルの送信先の選択などに手間が生じることが多く、協調ウェブ検索における合意形成や意思決定の効率が低下する。複数グループで行われる協調ウェブ検索において、自動でグループを推定することができれば、グループ管理やファイルの送信先の選択をより簡便にすることができる。

そこで、協調ウェブ検索時には同一グループに所属しているユーザは同一のテーマで検索を行うため、ユーザ同士は近距離に存在し、類似したウェブコンテンツを閲覧することに着目し、グループの構成を推定することを考える。ユーザが閲覧したウェブコンテンツを用いてユーザの特徴ベクトルを構成し、ベクトルに基づいてユーザのクラスタリングを行う。さらに、ユーザ位置を考慮し、クラスタリング結果を修正することでユーザがどのグループに属しているかの判定が可能であると推測できる。本稿では、協調ウェブ検索に参加する各ユーザの位置情報と閲覧したウェブコンテンツを用いてユーザのグループ構成を推定する手法を提案する。

2. グループ形成手法

本節では、協調作業においてグループ形成を行うことを目的とした研究を紹介した上で、本研究におけるアプローチについて述べる。

2.1 関連研究

情報端末を用いた協調作業において情報共有を支援するために AirDrop や Wi-Fi Direct が提供されている。しかし、これらのツールでは、送信時にネットワーク上の端末名をリストから選択する必要があるため、どのユーザがどの名称の端末を使用しているかを把握しておかなくてはならない。また、AndroidBeam[8] のような NFC を利用した情報共有ツールがあるが、端末同士を接触させる必要があるため、参加者間で手が届かない程離れている場合や一斉送信の場合は利用できない。そこで、情報端末を用いた協調作業において、グループ形成やファイルの送信先の選択を行う手間を削減する研究がなされている。

2.1.1 単一端末に対するグループ形成

UDU-L[9] は可視光レーザを用い、レーザポインタのように通信したい端末を指し示すことでペア形成を行う。このとき、レーザ光の明滅によって、対象の端末へ接続に必要な情報を送信し通信を確立する。識別する対象の端末を簡単に選択することができ、一つの端末を識別するための操作時間に優れているが、同時に複数の端末を識別するこ

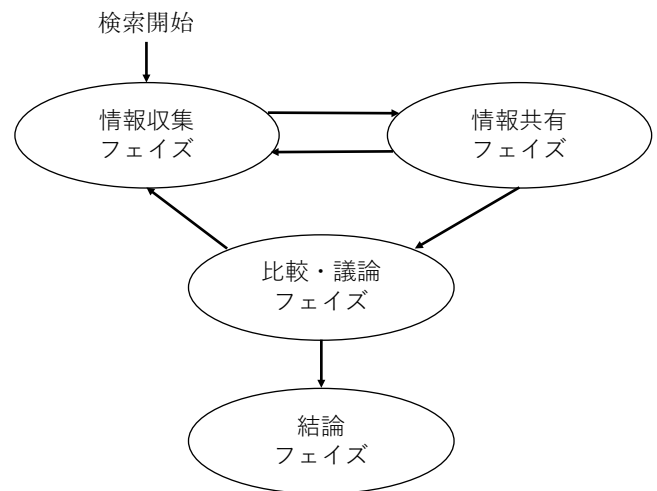


図 1 協調ウェブ検索のサイクル

とは考慮されていない。

2.1.2 複数端末を考慮したグループ形成

タブレット端末を画面をまたぐようにスワイプすることで両端末を識別し、グループ形成を行う研究 [10] や、端末内蔵のカメラとセンサを活用することによって、端末同士を互いのカメラに映るように向き合わせるだけで方向認識を伴ったグループ形成を行なう研究 [11] がある。これらは端末間の方向を認識できるという利点があるが、グループ形成を行う全ての端末にスワイプやカメラで写すなど操作が必要である。

Bluetooth ビーコンを用いたグループ形成手法 [12] では、識別を行う全端末がビーコンの発信および受信を行い、ビーコン信号の RSSI から端末間の距離を推定することで近距離の端末同士でグループ形成を行う。これは同時に複数台でのグループ形成を行うことができるが、手順としてグループに参加する全ての端末を 10cm 程の範囲に集約する必要があるため、グループ形成自体の手間を削減できたとはいえない。端末を集約する範囲については広くすることもできるが、それだけ識別の精度が落ちてしまう。

2.2 タイミングを考慮したグループ形成

前述した関連研究ではグループ形成時に特別な操作を必要とし、ユーザに手間が発生する。そこで協調ウェブ検索中の作業状態に着目し、グループ形成が必要な時点について考える。協調ウェブ検索において作業のプロセスを観察した研究 [13] によると、作業状態は図 1 のように遷移している。観察開始直後では被験者はウェブ検索に集中し、発言はほとんどなかった。検索中に気なるコンテンツを見つけると、近くにいる被験者にコンテンツ内容について意見を求めるなどし、情報の交換が見られている。その後、被験者らは各自保存したコンテンツを持ち寄り比較を行い、これらのプロセスを繰り返すことで最終的な目的情報を決

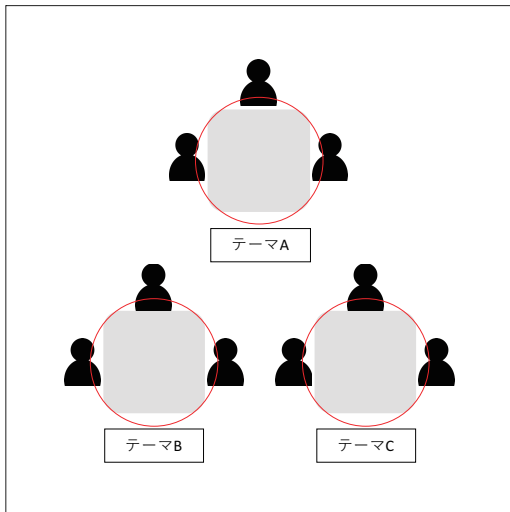


図 2 協調ウェブ検索の作業環境例

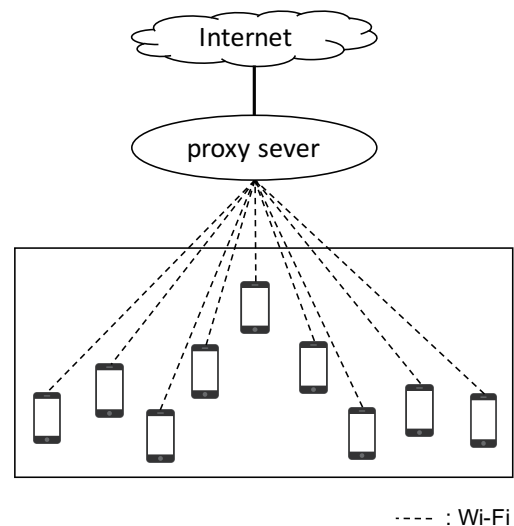


図 3 ネットワーク環境

定していた。

このように、協調ウェブ検索を開始してから情報共有を開始するまでには時間がある。したがって、本研究で提案する手法では、情報収集フェーズにおいて閲覧したウェブコンテンツを用いてグループを推定しておき、情報共有フェーズの時点ではグループ形成が完了していることを目指す。これにより、グループ形成に特別な操作を必要としなくなるため、ユーザの手間を減らすことができると期待される。

3. ユーザ位置と閲覧コンテンツを用いたグループ推定手法

本節ではユーザの位置と閲覧コンテンツを用いたグループ推定手法の内容と実装について述べる。

3.1 想定環境

本手法を用いる協調ウェブ検索の環境の例を図2に示す。ユーザは室内でグループに分かれ、机を囲んでおり、それぞれのグループは別のテーマについて検索している。一例として、学校の授業で行われる調べ学習などが挙げられる。

次に、協調ウェブ検索に参加する各端末のネットワーク環境を図3に示す。それぞれの端末はプロキシサーバを通してインターネットに接続している。プロキシサーバではそれぞれのユーザが閲覧したウェブコンテンツを収集し、本手法を適用しグループの推定を行う。

3.2 グループ推定手法

本手法の全体像を図4に示す。各端末が参照したウェブコンテンツのタイトルを含んだテキストはプロキシサーバに送信され、プロキシサーバは受け取ったテキストを処理する。以下に処理手順を述べる。

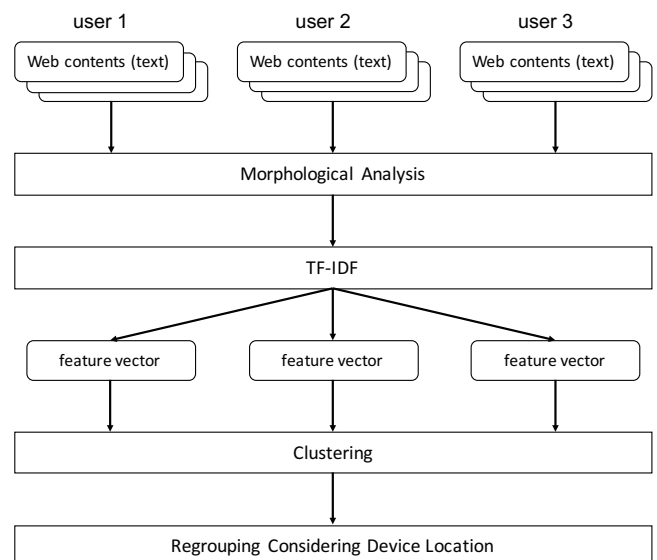


図 4 全体像

- (1) テキストに対して形態素解析を行い、名詞または未知語のみを抽出する。
- (2) 抽出した形態素に TF-IDF[14] を適用し、各ユーザの特徴ベクトルを構成する。
- (3) 特徴ベクトルをクラスタリングすることでグループを推定する。
- (4) 端末の位置を考慮した上で、再度グループ推定を行う。

3.3 閲覧コンテンツ

ウェブコンテンツにはナビゲーション要素やメニュー、広告など本文に無関係のコンテンツも含まれているため、本文のみを抽出する必要がある。本文の抽出には Boilerplate Detection[15] を用いる。また、ウェブコンテンツのタイトルはコンテンツを簡潔かつ具体的に表している場合が多いため、本文とは別のテキストとして処理を行うことで、タイトルに重点を置いたグループ推定ができ、精度の向上が

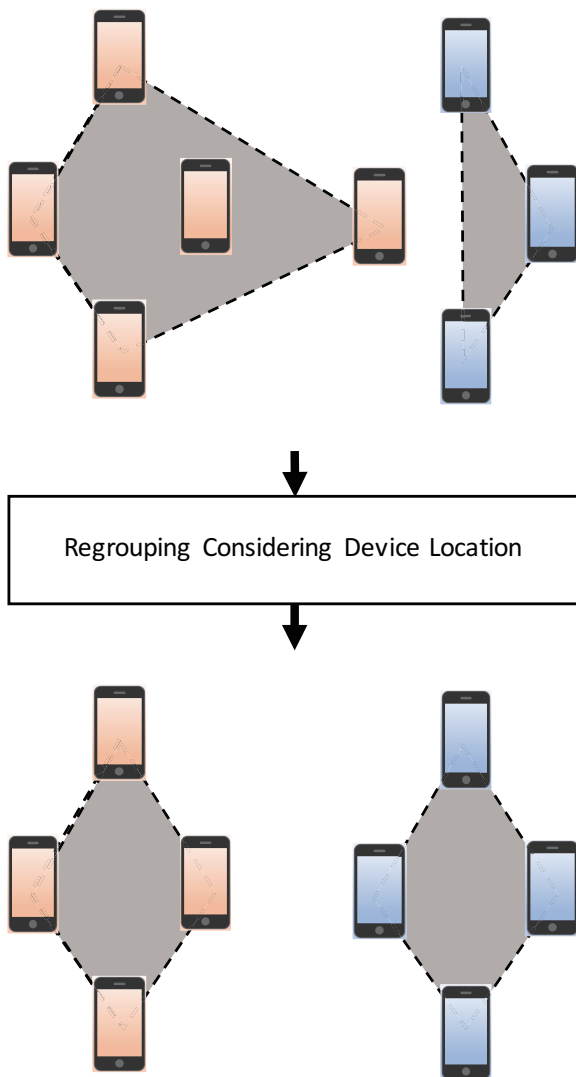


図 5 ユーザ位置を用いた再グループ推定

期待できる。

3.4 ユーザ位置

協調ウェブ検索に参加する全端末は BLE ビーコンを用いて、常に室内での自身の位置を推定しておく。端末の位置推定は、3 点測位や fingerprinting またはそれらのハイブリッド手法 [16] によって推定を行う。

閲覧コンテンツによって推定したグループに対して、ユーザ位置を用いてグループ推定を洗練する方法を図 5 に示す。本研究が対象とする協調ウェブ検索では図中のように各ユーザの間に一定の物理的空間を維持し、他のグループと空間が重なることはない。そこで、グループを構成する端末同士を直線で結び、線の内側に端末が存在しないようにグループ形成を行うことで、グループ推定の精度向上が期待できる。クラスタリングによるグループ推定が完了した時点で、端末を結ぶ線の内側に端末が存在する場合、推定されたグループは誤りであると判定する。その時、他グループに近接している端末についてグループの再編成を行

う。最終的に、端末を結ぶ線の内側に端末が存在しなくなるまでこの処理を繰り返す。

4. 評価

本節では、本手法の精度を検証するための実験と評価について述べる。

4.1 実験内容

実験では、学生 4~5 名を 1 グループとして 4 グループ編成し、それぞれ 30 分間の協調ウェブ検索を行う。協調ウェブ検索のテーマはそれぞれ以下のようなものである。

- (1) 就職活動の選考解禁時期の後ろ倒しについて、賛否をウェブで調べまとめる
- (2) 小学校の段階で英語を学習することについて、賛否をウェブで調べまとめる
- (3) インターネット掲示板に匿名で書き込むことについて、賛否をウェブで調べまとめる
- (4) 定年引き上げについて、賛否をウェブで調べまとめる

今回の実験では、閲覧コンテンツによってグループ推定が可能かどうかを推定することが目的であるため、プロキシサーバを用いず、独自に構築した簡易なブラウザアプリケーションで閲覧ウェブコンテンツと閲覧時刻のみを記録する。使用端末は iPad で OS は iOS 11 とする。

各グループは別の時間に協調ウェブ検索を行うため、実験を行う際のユーザ位置は仮定の値を付与する。そこで、8m 四方の教室内での 4 グループに分かれた協調ウェブ検索を行ったものと仮定し、全ユーザに仮定の位置情報を付与し、本手法の適用をシミュレーションした上で評価を行う。

4.2 評価方法

協調ウェブ検索中にブラウザアプリケーション上で記録した各ユーザの閲覧したウェブコンテンツおよびコンテンツの閲覧時刻、位置情報を用いて本手法のシミュレーションを行う。ツールとして、ウェブコンテンツから本文を抽出する boilerpipe[17]、形態素解析に茶筌 [18] を用いる。

シミュレーションの結果、実際のグループ構成と推定されたグループ構成との差異を比較し、精度の検証を行う。

4.3 実現可能性

テーマ (1) とテーマ (2) の協調ウェブ検索を行った結果に対して、その際の閲覧コンテンツのタイトルを元に本手法の実現可能性について考察する。各グループ 2 名ずつ、それぞれ 3 つのウェブコンテンツを抜き出したものを表 1 に示す。テーマ (1) では「就活」や「就職活動」、テーマ (2) では「英語」や「小学校」といった語の出現頻度が高

表 1 閲覧コンテンツのタイトル

| | | |
|---------|-------|--|
| テーマ (1) | ユーザ A | 後ろ倒しは失敗だった！？ 2015 年、就活解禁時期の変更がもたらしたもの - paiza 開発日誌 |
| | | 後ろ倒しの就活、影響は？ 大手 3 社採用担当者が座談会：朝日新聞デジタル |
| | | 就活スタート時期をめぐる 100 年間の変遷 大正時代から前倒しと後ろ倒しの繰り返し — キャリコネニュース |
| | | ... |
| ユーザ B | | 就職活動 選考解禁 後ろ倒し メリット - Google 検索 |
| | | 就職活動日程 いっそ採用したら：教育ジャーナリスト渡辺敦司の一人社説 |
| | | 3月就活解禁、不安感じる学生たち 気になる「本選考の開始時期」 |
| | | ... |
| テーマ (2) | ユーザ C | 早期英語教育は必要？そのメリット・デメリット Cafeducation |
| | | 小学校英語教育「時間の確保」課題、教員 6 割が英語力自信なし — リセマム |
| | | 小学校 英語教育 教師 - Google 検索 |
| | | ... |
| ユーザ D | | 早期英語教育は必要？そのメリット・デメリット Cafeducation |
| | | 小学校の英語教育義務化はどんな内容？メリットとデメリットは？あなたは賛成？反対？ [ママリ] |
| | | 小学生 英語 - Google 検索 |
| | | ... |

い。「Google」や「検索」といった語も出現頻度が高いが、テーマ (1)(2) をまたいで出現している。TF-IDF を適用することで、検索テーマに沿ったユーザの特徴ベクトルを構成可能であるため、タイトルに重点を置いたグループ推定は十分可能であると推測できる。

5. おわりに

本稿では、協調ウェブ検索に参加する各ユーザの位置情報と閲覧したウェブコンテンツを用いてユーザのグループ構成を推定する手法を提案した。本手法では、ユーザが閲覧したウェブコンテンツを用いてユーザの特徴ベクトルを構成し、ベクトルに基づいてユーザのクラスタリングを行うことで、ユーザがどのグループに属しているかを推定する。

今後は本手法の実装を進め、有用性を評価するために実験を行う予定である。

参考文献

- [1] 総務省：平成 28 年通信利用動向調査の結果，総務省（オンライン），入手先（http://www.soumu.go.jp/johotsusintokei/statistics/data/170608/_1.pdf）（参照 2017-12-20）
- [2] Morris, M. R. and Horvitz, E.: SearchTogether: an interface for collaborative web search, *Proceedings of the 20th annual ACM symposium on User interface software and technology*, pp. 3–12 (2007).
- [3] Dropbox: Dropbox, Dropbox (online), available from (<https://www.dropbox.com/>) (accessed 2017-12-20).
- [4] Google: Google Drive, Google (online), available from (<https://www.google.com/drive/>) (accessed 2017-12-20).
- [5] Miller, B. A. and Bisdikian, C.: *Bluetooth revealed: the insider's guide to an open specification for global wireless communication*, Prentice Hall PTR (2001).
- [6] Camps-Mur, D., Garcia-Saavedra, A. and Serrano, P.:

- Device-to-device communications with Wi-Fi Direct: overview and experimentation, *IEEE wireless communications*, Vol. 20, No. 3, pp. 96–104 (2013).
- [7] Apple: Use AirDrop on your iPhone, iPad, or iPod touch, Apple (online), available from (<https://support.apple.com/HT204144/>) (accessed 2017-12-20).
- [8] Google: NFC Basics — Android Developers, Google (online), available from (<https://developer.android.com/guide/topics/connectivity/nfc/nfc.html>) (accessed 2017-12-20).
- [9] 小木真人, 大木裕太, 吉永努, 入江英嗣: UDU-L: レーザポインティングによる柔軟なデバイス接続手法, 電子情報通信学会論文誌, Vol. J97-D, No. 1, pp. 155–164 (2014).
- [10] 鈴木亮詞, 村瀬隆拓, 白松俊, 大園忠親, 新谷虎松: タブレット端末のためのスマートサイネージシステムの実装について, コンピュータソフトウェア, Vol. 30, No. 2, pp. 176–190 (2013).
- [11] 塩見和則, 高田秀志: 情報共有をとまなうアドホックな対面協調作業のための近接端末間方向認識, 情報処理学会論文誌, Vol. 58, No. 1, pp. 143–152 (2017).
- [12] Masato, W., Sakaguchi, Y., Ozono, T. and Shintani, T.: A Proximity Information Propagation Mechanism Using Bluetooth Beacons for Grouping Devices, *Journal of Intelligent Computing*, Vol. 8, No. 1, pp. 1–9 (2017).
- [13] Komaki, D., Oku, A., Arase, Y., Hara, T., Uemukai, T., Hattori, G. and Nishio, S.: Content Comparison Functions for Mobile Co-located Collaborative Web Search, *Journal of Ambient Intelligence and Humanized Computing*, Vol. 3, No. 3, pp. 239–248 (2011).
- [14] Ramos, J.: Using tf-idf to determine word relevance in document queries, *Proceedings of the first instructional conference on machine learning*, pp. 133–142 (2003).
- [15] Kohlschütter, C., Fankhauser, P. and Nejdl, W.: Boilerplate detection using shallow text features, *Proceedings of the third ACM international conference on Web search and data mining*, ACM, pp. 441–450 (2010).
- [16] Lloret, J., Tomas, J., Garcia, M. and Canovas, A.: A Hybrid Stochastic Approach for Self-Location of Wireless Sensors in Indoor Environments, *Sensors*, Vol. 9, No. 5, pp. 3695–3712 (2009).
- [17] Kohlschütter, C.: Boilerpipe - Google Code Archive, Kohlschütter Search Intelligence (online), available from (<https://code.google.com/archive/p/boilerpipe/>) (accessed 2017-12-20).

- [18] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 浅原正幸: 日本語形態素解析システム「茶釜」 Version 2.0 使用説明書, 奈良先端科学技術大学院大学 (1999).