

複数メディアの意味・空間距離の検索特性に基づく 携帯端末向け検索手法の提案

A Proposal of a Cross-Media Search Method for Mobile Devices Based on Semantic and Spatial Distance

杉橋 真輝* 栗 達† 熊本 忠彦‡ 河合 由起子*
Masaki Sugihashi Da Li Tadahiko Kumamoto Yukiko Kawai

1. はじめに

画面サイズに制約があるスマートフォン (スマホ) などの携帯端末を対象に, 検索結果に対するクリックやスクロールの操作数を軽減しようという研究 [1][2] が広く行われている. 山田 [1] らは, Web 検索結果を初めから順に閲覧する手間を解決するため, 検索結果として得られたスニペット内の単語と過去に閲覧した Web ページ内の単語をクラスタリングし, 各検索結果とそれに関連する単語を提示する手法を提案している. また, 細川 [2] らは, 検索結果を選別する際の労力軽減を目的に, 検索ログから Word2Vec による単語 (関連語) のリストを作成し, 検索結果と検索ログから作成された関連語リストとの類似度から関連語をタグ付け提示する手法を提案している. しかしながら, これら既存研究は, Web ページの検索結果を効率良く探索することを目的としており, SNS やマップといった他のメディアとの連携には対応していない. また, 画面サイズに制約のあるスマホでは, 複数メディアを同時に閲覧することは難しいという問題もある.

そこで, 本稿では, 複数メディアの検索特性を考慮したスマホ向け検索統合手法を提案する. すなわち, Web と SNS とマップという 3 つのメディアを対象に, Web では Google を用いて文章内容に基づく検索を行い, SNS では Twitter を用いて速報性のある検索を行う. マップでは Google Map を用いて距離を考慮した店舗等の評判情報検索を行う. この異なる検索特性を持つメディアを用いた検索により, 各々のメディアだけでは取得できない検索結果の取得を可能にする. 具体的には, まず, Google 検索より取得したスニペットや Twitter API より取得したツイート本文, Google Maps API より取得した店舗等の評判情報から単語 (名詞) を抽出し, クラスタリングする. 次に, 生成された各クラスタと各々のメディアの各検索結果との間で類似度を求める. そして, メディアの検索特性となる意味・空間距離に対して, ユーザが重みを明示的に (または検索ログに基づく暗黙的なフィードバックから) 付与することで, 類似度に意味・空間距離の重みを積算す

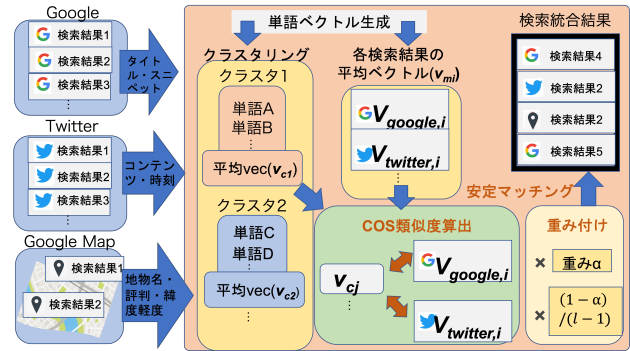


図 1: 複数メディア横断型検索統合提示システム

る. 最後に各クラスタと各検索結果の組み合わせを安定マッチングにより決定し, ランキング提示する. なお, 本稿では, 被験者 30 名による評価実験を行い, 提案手法の妥当性を検証する.

2. 複数メディア検索統合提示システム

本稿では, Web 検索として Google, SNS 検索は Twitter, マップ検索は Google Map を用いる.

2.1. システム概要

図 1 に提案システムの構成を示す. まず, ユーザは検索クエリ入力時にメディア (Google, Twitter, Google Map) に対して重み付けを行う. 初期値は同値とする. 次に, クエリより各メディアの検索結果の上位 n 件から, タイトルとスニペット (Google), ツイート内容と時刻 (Twitter), 現在地から半径 r km 以内にある地物のレビューと緯度経度 (Google Map) を取得する. 取得したテキストを形態素解析し名詞を抽出し, Word2Vec より単語ベクトルを生成する. 単語ベクトルから各検索結果 i ごとに平均ベクトル $v_{media,i}$ を生成する.

その一方で, 抽出した全単語をクラスタリングし, 各クラスタ j の平均ベクトルをクラスタベクトル $v_{c,j}$ として生成する. 最後に各検索結果の平均ベクトル $v_{media,i}$ と各クラスタの平均ベクトル $v_{c,j}$ との \cos 類似度を算出し, ユーザの重みを積算し, 安定マッチングより各クラスタと各検索結果の組み合わせを決定し, 各クラスタから類似度の高い上位 m 件を選出し, 類似度順にランキング提示する.

*京都産業大学

†福岡大学

‡千葉工業大学

2.2. クラスタリングとランキング

提案手法は、複数メディアを統合してランキングする方法として出現単語をクラスタリングし、単語の各クラスタと各検索結果との類似度を算出し、重み付けする。そして、検索結果がいずれのクラスタに属するかを安定マッチングとなる gale-shapley 法 [3] により決定し、各クラスタの上位を選出する。

クラスタリングは、各検索結果のメディアごとの偏りを考慮し、検索結果単位ではなく単語を分類する。まず、Word2Vec を用いて単語 (名詞) から単語ベクトルを生成し、クラスタリングし、各クラスタ j の平均ベクトル $v_{c,j}$ を算出する。次に、各検索結果 i の平均ベクトル $v_{media,i}$ との \cos 類似度 $sim(v_{media,i}, v_{c,j})$ を算出する。その後、ユーザが選択したメディアに対する重みを α ($0 < \alpha < 1$) とするとき、 $sim(v_{media,i}, v_{c,j}) \times \alpha$ を算出する。メディア数が l の場合に選択されなかったメディアの重みは $(1 - \alpha)/(l - 1)$ とする。ただし、ユーザがメディアを選択しない場合 ($\alpha=0$) は全メディアの重みは同値とする。また、マップを選択した場合、現在地からの距離が近い検索結果に対して重みを大きくし、空間的な距離を考慮した検索結果を提示する。ただし、初期値は全て

最後に gale-shapley 法 (安定マッチング) を用いて、検索結果と各クラスタの組み合わせを決定し、ランキングする。なお、安定マッチングは、選べる中では他に希望が高い組み合わせがある不安定対が起こっていないマッチングを実現する手法で、gale-shapley 法が広く用いられている。

3. 検索統合提示手法の実装

前節に基づき、検索統合システムを実装した。クラスタリングは、k-means, x-means, g-means, DB-SCAN (Density-Based Spatial Clustering of Applications with Noise) の 4 種類を実装し、検証では事前実験よりクラスタリングは x-means を採用した。Web 検索には Google Search API を、Twitter 検索には Twitter API を、マップ検索には Google Maps API を用いた。

3.1. フロントとのデータ連動

サーバサイドの検索取得とリランキングは Python を用いて実装し、検索結果を表示するためのフロントエンドは flutter を用いた。図 2 に実装したユーザーインターフェイスの画面を示す。ユーザはクエリ入力と同時に、上記に提示されているメディアのアイコンを選択することでユーザが重視するメディアを決定する。(図 2-①) クエリ入力後にはクラスタのキーワードがサジェストされる。(図 2-②)、そのサジェストを選択すると、検索結果が提示される (図 2-③)。



図 2: 検索入力画面と検索結果提示画面例

3.2. ランキング順位の検証

提案した複数メディアの検索統合手法によるランキング順位の妥当性を検証する。まず、Google, Twitter, Google Map から得られた各検索結果の上位 10 件を提案手法によりリランキングし、その上位 5 件を取得した。次に、この 5 件を被験者 30 人に順位付け評価してもらい、評価結果 (平均) を正解順位とした。正解順位と提案手法による順位結果との nDCG (normalized Discounted Cumulative Gain) を算出し、ランキングの妥当性を検討した。今回用意した 3 つのクエリ (1) 「カレー, ナン」、(2) 「カレー, 料理」、(3) 「カレー, 最高」とした。結果、nDCG @ 5 は (1) が 0.933, (2) が 0.954, (3) が 0.857 となり、平均 0.915 となり良好な結果となり、提案手法の有効性が示唆された。

4. おわりに

本稿では、複数メディアの検索結果を効果的かつ効率的に統合提示できるメディア横断型検索統合手法を提案した。本稿では、提案システムを実装し検索結果順位の妥当性を検証した。今後、各メディアごとの重みを考慮したランキング評価、インタフェースのユーザ評価、LDA や BERT による精度検証、他メディアとの連動を検討する。

謝辞

本研究の一部は、JSPS 科研費 19K12240, 京産大学 HMD センター (M2001) の助成による。ここに記して謝意を表す。

参考文献

- [1] 山田純平, 北山大輔. 網羅的検索のための web 検索結果における拡張情報提示とその評価. DEIM Forum 2021, F13-2.
- [2] 細川涼平, 早川智一, 疋田輝雄. 関連語の分散表現に基づく web 検索結果の自動タギング. DEIM Forum 2018, D4-1.
- [3] D. Gale and L. S Shapley. College admissions and stability of marriage. American Mathematical Monthly Vol. 69, No. 1 1962, pp. 9-15.