

令和元年6月6日現在

機関番号：14301  
研究種目：若手研究(B)  
研究期間：2016～2018  
課題番号：16K16156  
研究課題名（和文）検索ユーザの目的に着目したタスク検索技術の研究

研究課題名（英文）A Study on Task-oriented Web Search

研究代表者

山本 岳洋 (Yamamoto, Takehiro)

京都大学・情報学研究科・助教

研究者番号：70717636

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本研究課題の目的は、検索ユーザの背後にある目的を考慮した検索技術を実現することである。この研究を実現することで、現在の検索エンジンがユーザが入力したキーワードに関連した情報のみをユーザに返すのに加え、そのキーワードの背後にある目的を考慮した多様な支援が可能となる。研究期間を通して（1）代替行動マイニング技術、（2）網羅的な情報収集タスクにおけるクエリ推薦技術とユーザ分析、（3）Struggling Searchタスクのためのクエリ推薦技術、（3）（4）特定の目的に合致したオブジェクト検索技術などを実現した。

研究成果の学術的意義や社会的意義

これまで多くの研究がユーザ入力したクエリと文書との適合性を考えていたのに対して、本研究はそのクエリの背後にある目的からユーザに返すべき情報を考える、という点で学術的に新しい試みを行った。また、本研究で得られた成果は情報検索に関するトップカンファレンスのフルペーパーや、複数の論文誌に採録されるなど、国際的・学術的にも高く評価された。

研究成果の概要（英文）：The aim of this project is to develop a search technique that automatically estimates the goal behind a query issued by a searcher, and provides desired information according to the goal. Achieving such a technique enables the search engine not only to return the "relevant" information to the query issued by a searcher but to provide more information that supports the user to achieve the goal behind the query. The outcomes of this project include (1) Mining alternative-actions from community Q&A corpus, (2) Query suggestion interface that visualizes the missing information for intrinsically diverse search task, (3) Query suggestion for struggling search task, and (4) Entity retrieval on the basis of the purpose of using the entity.

研究分野：情報検索

キーワード：情報検索 タスク検索 探索的検索

## 様式 C-19、F-19-1、Z-19、CK-19（共通）

### 1. 研究開始当初の背景

Web上で発信される情報が膨大になるにつれて、Web検索エンジンは単なる調べ物の手段ではなく、「京都旅行の計画を立てたい」や「花粉症を治したい」といった、ユーザの実世界の問題解決の手段として利用されるようになってきている。例えば、「花粉症の対策方法を探す」という目的を持ったユーザは、まず、花粉症対策にどのような方法があるのか様々なキーワード(クエリ)で検索し、「レーザー手術をする」、「減感作療法を受ける」、「ハーブティーを飲む」、「花粉対策マスクをつける」といった方法があることを見つけ、その後、それぞれの詳細を検索し自分に適した方法を取捨選択していただく。

しかし、クエリと文書の内容的な類似性に基づいた、従来の情報検索のパラダイムでは、ユーザの抱える問題や達成したい目的を満足させるには不十分である。実際、「京都観光」というクエリだけでは、京都の観光情報に関するページは検索できるものの、京都の宿泊場所の予約、レストランの調査、観光地の調査、交通手段の調査といった京都観光の計画に関するページは検索することはできない。クエリと文書の内容的な類似性だけでなく、クエリの背後にあるユーザの目的をシステムが理解し、その目的に応じた情報をユーザに返すことが、今後の情報検索に求められる技術である。

### 2. 研究の目的

本研究では、ユーザの実世界での目的を達成するための方法を検索する技術、タスク検索に関する研究を行う。タスク検索では、ユーザの検索ビヘイビアから「花粉症を治したい」といった目的を推定し、「レーザー手術をする」、「減感作療法を受ける」といった方法をWebから発見し、その目的を達成するためのクエリ推薦や文書ランキングを行う。タスク検索における重要な課題は、ユーザの背後にある目的の推定、そしてそれを解決する方法の検索手法である。本研究は、従来の情報検索が対象としてきた、入力されたクエリを満足するための検索技術ではなく、そのクエリの背後にある目的を達成することを目指した検索技術の研究である。

### 3. 研究の方法

タスク検索技術を実現するため、ユーザのタスクの種類として、実世界における目的をマイニングにそれに基づいた検索技術、また、検索という文脈における目的に応じた検索技術について、それぞれアルゴリズムおよびインタフェースの両面から研究を進めた。

### 4. 研究成果

2年間の研究期間を通して得られた主な研究成果は以下の通りである。

#### (1) 代替行動マイニング技術:

ある目的を達成する方法を検索エンジンを通じて探す状況化での、代替行動マイニング技術について取り組んだ。「睡眠薬」というクエリに対して、「ハーブティーを飲む」、「睡眠前に軽く運動する」といった、クエリの背後にある目的を達成することが可能な行動を代替行動(Alternative Action)と定義し、クエリに対する代替行動をYahoo!知恵袋などのコミュニティ質問応答コーパスを利用しマイニングする技術を実現した。これまでの既存研究の多くが、あるタスクを達成するためのサブタスクを発見することに注力していたのに対して、本研究ではタスクからサブタスクを求めるのではなく、与えられたタスクと同等の関係にあるようなタスクを発見する点において大きな新規性がある。

代替行動をマイニングする具体的な手法は、質問応答コーパスの質問-回答構造を利用する。このとき、図1に示すように、類似する質問に対して返された回答中に含まれる行動は代替行動である可能性が高い、また、代替行動が回答として返されるような質問は類似の目的に関する質問である可能性が高い、という仮説のもと質問-行動の2部グラフをコーパスより作成し、作成された2部グラフに対してSimRankアルゴリズムを適用しノード間の類似度を求めることで、代替行動となるような行動ペアを発見する。実験の結果質問応答コンテンツから代替行動をマイニングする提案手法は、既存のクエリ推薦や、ウェブ文書から代替行動をマイニングする手法と比較して精度よく代替行動を発見できることを示した。

本技術には様々な応用例が考えられる。たとえば、図2に示すように、検索ユーザがあるキーワードを入力した際に、システムが背後でそのキーワードに関連する代替行動をマイニングし、クエリ推薦のような形式でユーザに推薦することで、ユーザが検索時には気づかなかったが、ユーザが達成したい目的を実現できる他の方法というものをユーザに伝達することができると考えられる。



図 1: 質問応答コーパスからの代替行動マイニングのアイデア.

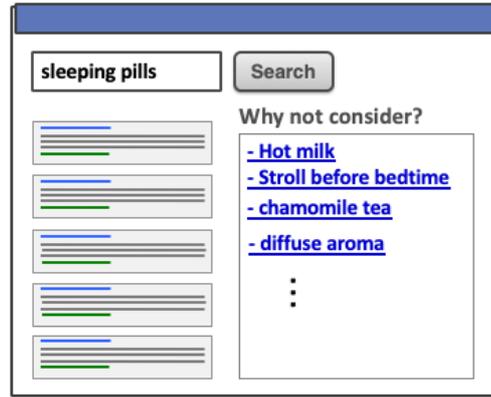


図 2: 代替行動マイニング技術の応用例. 検索ユーザが入力したクエリの背後にある目的を達成するような代替行動に関するクエリを推薦することが可能となる.

## (2) 網羅的な情報収集タスクにおけるクエリ推薦技術とユーザ分析:

網羅的に情報を探索する必要がある検索タスク（網羅性指向タスク）下の検索行動を支援するクエリ推薦インタフェースを提案した. 本研究では未閲覧情報量を検索トピックに関する重要な情報が検索結果の中に残っている度合いと定義した. 提案インタフェースは, 図3に示すように, 現在の検索クエリおよび推薦クエリのそれぞれに対して未閲覧情報量を推定し, その値を棒グラフの形式でユーザに提示する. こうした情報を提示することで, ユーザは適合文書が十分に収集された観点や収集が不十分な観点を視覚的に理解し, 適切なタイミングで検索を終了することが可能になると期待される. 本研究では, ユーザが文書集合から獲得する利得を, 観点の重要性, 観点の新規性, および観点に関する文書の適合性に基づく尺度として定式化し, 各クエリに対する未閲覧情報量をその検索結果中の未閲覧文書集合からユーザが獲得可能な追加利得として算出する.

提案インタフェースがユーザの検索戦略および検索成果に与える影響を明らかにするために, 網羅性指向タスクに関する4種類の検索トピックを用意し, 24名の被験者を対象にユーザ実験を実施した. 被験者の検索ログを用いた分析の結果, 提案インタフェースの利用者は, 未閲覧情報量の推定精度が高いトピックにおいて, 重要な情報を十分に調べた後で個々の検索を終了する, 重要な情報を多く含むクエリを次の検索に利用する, タスクの後半にも重要な情報を継続的に収集する, および, 単位時間あたりに獲得可能な利得が上昇することが確認された. 本研究課題に関する研究成果は情報検索に関するトップカンファレンスに採択されるなど, 国際的にも高い評価を得た.



図 3: 未閲覧情報量を可視化するクエリ推薦インタフェース

### (3) Struggling Search タスクのためのクエリ推薦技術:

情報要求が同一であるがユーザが複数回クエリを修正しなければ適合情報が発見できないような Struggling Search と呼ばれる検索に焦点をあて、このタスク下の検索ユーザを支援するためのクエリ推薦技術について取り組んだ。具体的には、Struggling Search 特有のクエリ修正パターンをクエリログから学習・マイニングし、Struggling Search 特有のフレーズレベルでのクエリ修正パターンを発見する。その後、右図に示すようにフレーズレベルでの修正パターンからグラフを構築し、そのグラフに対して Biased PageRank

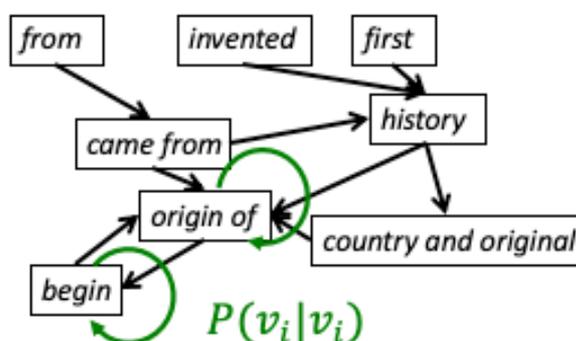


図 4: Struggling Search 特有の修正パターングラフ

アルゴリズムを適用することで、Struggling Search において適切なクエリを探そうとしているユーザが最終的に入力する可能性の高いフレーズの候補を自動的に発見する。既存のクエリ推薦手法と比較した結果、提案手法が既存のクエリ推薦手法よりも Struggling Search タスクにおけるクエリ推薦の精度が高いことを示した。

### (4) 特定の目的に合致したオブジェクト検索:

「夜景がきれいに撮れる」カメラ、「持ち運びやすい」カメラといった、オブジェクトを実際に使用して初めて知ることのできるような属性を経験的属性と定義し、経験的属性に基づくオブジェクトランキング手法を実現した。図 5 は手法の概要を表した図である。具体的には、特定の属性に対するオブジェクト同士の優劣は、そのオブジェクトのレビュー中に比較表現として出現することに注目し、レビュー中から、ある属性に対するオブジェクト間の相対的選好を抽出

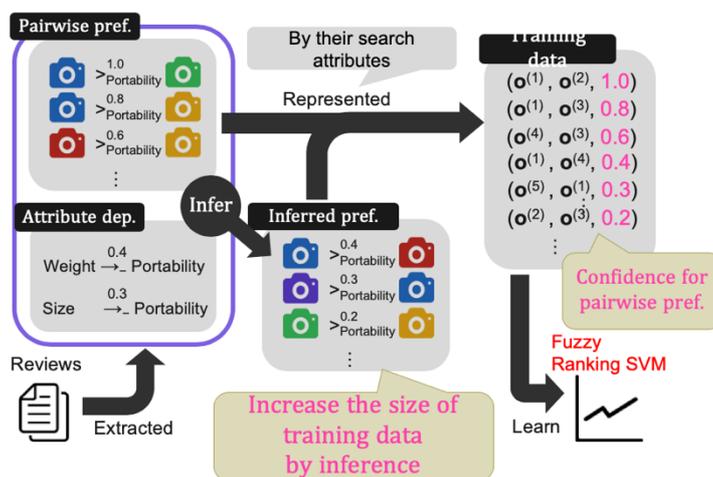


図 5: 相対的選好の抽出とそれに基づくランキング学習

し、学習データとして Learning to Rank に用いることで、オブジェクトのランキングを実現する。さらに、「軽い」カメラは「持ち運びやすい」といったように、属性間には依存関係があることに着目し、属性間の依存関係についてもレビューから自動的にマイニングすることで、ランキング対象となる経験的属性に関する学習データを拡張する手法についても実現した。

## 5. 主な発表論文等

[雑誌論文] (計 3 件)

- ① Suppanut Pothirattanachaikul, Takehiro Yamamoto, Sumio Fujita, Akira Tajima, Katsumi Tanaka and Masatoshi Yoshikawa: Mining Alternative Actions from Community Q&A Corpus, Journal of Information Processing, 査読有, 26, pp. 427-428 (2018). DOI: <https://doi.org/10.2197/ipsjip.26.427>
- ② 梅本和俊, 山本岳洋, 田中克己: 網羅性指向タスクにおける未閲覧情報量の提示, 人工知能学会論文誌, 査読有, 32(1), WII-G, pp.1-12 (2017). DOI: <https://doi.org/10.1527/tjsai.WII-G>
- ③ 福地大助, 山本岳洋, 田中克己: 動詞クエリの語間の関係性に基づくクエリマイニング, 人工知能学会論文誌, 査読有, 32(1), WII-J, pp.1-15, (2017). DOI: <https://doi.org/10.1527/tjsai.WII-J>

[学会発表] (計 10 件)

- ④ Shinryo Uchida, Takehiro Yamamoto, Makoto P. Kato, Hiroaki Ohshima, and Katsumi Tanaka: Entity

Ranking by Learning and Inferring Pairwise Preferences from User Reviews, In Proceedings of the 13th Asia Information Retrieval Societies (AIRS 2017), 査読有, pp.141-153, November 2017.  
DOI: [https://doi.org/10.1007/978-3-319-70145-5\\_11](https://doi.org/10.1007/978-3-319-70145-5_11)

- ⑤ Suppanut Pothirattanachaikul, Takehiro Yamamoto, Sumio Fujita, Akira Tajima and Katsumi Tanaka: Mining Alternative Actions from Community Q&A Corpus for Task-Oriented Web Search, In Proceeding of the 2017 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2017), 査読有, pp.607-614, August 2017.  
DOI: <https://doi.org/10.1145/3106426.3106461>
- ⑥ Zebang Chen, Takehiro Yamamoto, and Katsumi Tanaka: Query Suggestion for Struggling Search by Struggling Flow Graph, In Proceedings of the 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2016), 査読有, pp.224-231, October 2016.  
DOI: <https://doi.org/10.1109/WI.2016.0040>
- ⑦ Kazutoshi Umemoto, Takehiro Yamamoto and Katsumi Tanaka: ScentBar: A Query Suggestion Interface Visualizing the Amount of Missed Relevant Information for Intrinsically Diverse Search, In Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2016), 査読有, pp.405-414, July 2016.  
DOI: <https://doi.org/10.1145/2911451.2911546>
- ⑧ Takehiro Yamamoto, Yiqun Liu, Min Zhang, Zhicheng Dou, Ke Zhou, Ilya Markov, Makoto P. Kato, Hiroaki Ohshima and Sumio Fujita: Overview of the NTCIR-12 IMine-2 Task. In Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-12), pp.8-26, June 2016.
- ⑨ 福地大助, 山本岳洋, 田中克己: Q&A コーパス学習に基づく Web 検索のための意味類似クエリの発見, 第 7 回データ工学と情報マネジメントに関するフォーラム (DEIM 2017), B1-2, 2017 年 3 月.
- ⑩ Takehiro Yamamoto: Supporting Complex Search by Understanding and Designing User Behavior The 11th Korea-Japan (Japan-Korea) Database Workshop, October 2016.

[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

<https://rerank-lab.org/publication/>

## 6. 研究組織

(1) 研究分担者

(2) 研究協力者

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。