

令和元年6月20日現在

機関番号：14301

研究種目：若手研究(A)

研究期間：2014～2018

課題番号：26700009

研究課題名(和文) 潜在的情報を引き出すための情報誘出技術に関する研究

研究課題名(英文) Research on Information Elicitation Technologies for Estimating Latent Information Needs

研究代表者

加藤 誠 (Makoto P., Kato)

京都大学・国際高等教育院・特定講師

研究者番号：00646911

交付決定額(研究期間全体)：(直接経費) 18,600,000円

研究成果の概要(和文)：本研究課題では、ユーザに刺激を与えその反応から検索意図・選好・意見などの潜在的情報を推定する技術の研究を行った。課題「質問回答に基づく検索意図の誘出」では、会話に基づく適合フィードバックの実現、クエリ推薦の配置方法の最適化、再検索における効果的な質問の調査を行った。課題「選好の誘出可能性の検証」では、画像から喚起される印象の推定方法の提案、質問投稿サイトにおけるユーザからの選好誘出方法の実践を行った。課題「複数の質問による信頼性の高い意見誘出」では、ソーシャルメディアからの意見抽出、Webの数値情報およびレビューの集約による意見抽出、ニュースメディアにおける信頼性の高い意見抽出を行った。

研究成果の学術的意義や社会的意義

本研究は、これまで受動的にデータを獲得してきたデータ分析に基づく研究とは異なり、ユーザに対して積極的にデータの提供を要求し、より適した働きかけ方についての検討・実践を行ったところに学術的意義があると考えられる。近年、Google HomeやAmazon Alexaなどの対話型情報アクセス技術が大きな注目を集めており、本研究課題ではこれらの基礎的な研究を行っていたと言える。これに加え、ユーザの反応に基づいてWebサービスを改善していく技術についても、さまざまな企業が近年の研究成果を導入し始めており、特にインターリーピング技術の実践的研究では実利用における有用な知見を得られたと考えられる。

研究成果の概要(英文)：In this research project, I conducted research on information elicitation technologies, which estimate users' information needs, preferences, and opinions by observing responses to a certain stimulus. The first research topic was "elicitation of information needs by question-answering" and included conversational relevance feedback, optimization of search result presentation with query suggestions, and a study on effective questions in refinding search tasks. The second research topic was "elicitation of preferences", including estimation of impressions triggered by pictures and ranking preference estimation in a question-answering Web service. The third topic was "reliable elicitation of opinions based on multiple questions" and focused on opinion elicitation from social media, opinion elicitation by aggregating numerical information and reviews on the Web, and reliable elicitation of opinions from news media.

研究分野：情報検索

キーワード：対話型検索 クエリ推薦 再検索 インターリーピング エンティティ検索 ランキング学習

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

大量のユーザデータから検索ユーザの検索意図(検索目的や情報要求)、商品・ユーザ・Web ページなどの選好、商品やサービス・政治に対する意見を推定する研究は幅広く行われてきており、ビッグデータ時代においては更なる発展が期待される。しかしながら、すべてのユーザ・推定対象について一様な方法をもってユーザ情報を収集し分析するような現在の枠組みでは、個々のユーザの意図・嗜好・意見を正確に推定することはできないと思われる。例えば、数十曲の購買情報、数曲に対する評価値だけで、音楽の嗜好を正確に理解することがそもそも可能であるだろうか。意図や選好、意見などの潜在的な情報を推測するためには、これまでの静的なデータ収集方法から脱却し、個々のユーザから戦略をもって情報獲得を行い、その獲得戦略は個人と推定対象の情報に対し動的に最適化される必要がある。

2. 研究の目的

適応的かつ能動的に刺激を生成しそれに対する反応から人間が潜在的にもっている情報を引き出す技術を情報誘出と名付け、本研究課題ではこの情報誘出に関する研究を行った。図1に情報誘出システムのフレームワークを示す。情報誘出システムは対象となる潜在的情報(検索意図・嗜好・意見)とユーザからの入力(クエリやプロフィール等)を受け取り、そのユーザから特定の潜在情報を引き出すために最も効果的な刺激(質問や音など)を刺激選択器により選択しユーザに与える。システムはユーザからの反応(質問に対する回答や音への生理的反応など)を元にユーザの潜在情報を

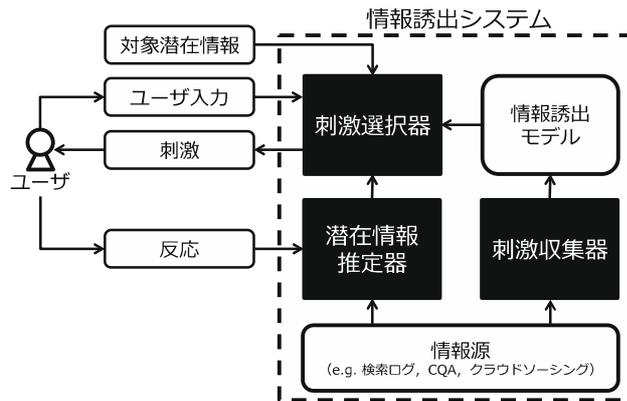


図1. 情報誘出システムのフレームワーク

潜在情報推定器によって推定する。システムは更に推定の精度がより高くなるような別の刺激をユーザへ与え刺激を観測することを繰り返す。情報誘出システムはその動作がユーザ・反応・誘出対象の情報に依存する点において適応的である。また、ユーザの情報をただ観測するのではなく、より正確に潜在情報を推定するために選択的に刺激を与え反応を獲得する点において能動的である。

本研究課題では下記の3種類の情報誘出について研究目標を設定した。

(1) 質問回答に基づく検索意図の誘出

検索ユーザが入力するクエリは非常に短い、その意図は多様であり必要となる情報はユーザによって異なる。そのため、検索ユーザの意図を推定することは、情報検索において最も重要な課題の一つである。検索意図の誘出では、入力されたクエリに対して考え得る検索意図を列挙し、その中から一つに特定できる可能性(特定可能性)が最も高く、かつ、ユーザが回答できる可能性(回答容易性)が最も高い質問を与えユーザに提示する。

(2) 嗜好の誘出可能性の検証

嗜好の誘出においては、どのような刺激に対する反応が最も効果的か、また、どのような順で刺激を与えれば最適なのかを明らかにする。本研究では特に音楽の嗜好の推定に取り組む。比較する刺激としては、質問・音・光・写真・移動物体などを想定しており、ユーザの実際の反応を元に音楽の嗜好推定への有効性を検証する。

(3) 複数の質問による信頼性の高い意見誘出

ある事物についての意見を獲得するためにはアンケートなどが用いられるが、回答が本来のユーザの意見を正確に反映できているかには疑問がある。例えば、政権の支持率は支持する・しないの二択を問うことが多いが、どちらでもないような層の割合や強く支持する層とやや支持する層の違いを区別することができない。一方で、より粒度の細かい選択肢を用意することもできるが、ユーザが自らの意見を客観的な指標に正確に当てはめることも難しい。意見の誘出は検索意図の誘出と同様に、ユーザが回答可能な質問を複数投げかけることによって、ある事物への支持・不支持等の極性値を推定することを目的とする。

3. 研究の方法

(1) 質問回答に基づく検索意図の誘出

会話に基づく適合フィードバック

特定のドメインにおける検索において、特定可能性および回答容易性が高い質問を生成することによって、効率的な検索を実現する方法について提案および実験を行った。

クエリ推薦配置の最適化

一般の検索において、クエリ推薦の利用は暗黙的な質問回答と見なすことが可能であり、どのようなクエリ推薦を配置すべきか、また、どのような検索結果を入力されたクエリと推薦されたクエリに対して提示すべきかを最適化する方法について提案および実験を行った。

再検索における効果的な質問の検討

以前に検索した内容を再度検索することはよく行われており、この再検索において検索の成功確率が高まる質問が何であるのかを明らかにするために大規模なユーザ実験を行った。

(2) 選好の誘出可能性の検証

画像から喚起される印象の推定

刺激としての画像の特性を明らかにするために、どのような画像からどのような印象が喚起されるかについて、その推定方法を提案し実験を行った。

脳構造データからの脳活動データの予測

ある刺激に対するユーザの反応を調べる方法として、fMRIによる脳活動計測が挙げられる。しかしながら、fMRIの利用コストの面から長期間に渡って反応を計測するのは現実的でなく、効率的に刺激を選択するためにはある程度fMRIの結果が予測できることが望ましい。そのため、脳構造データと脳活動データの相関関係を明らかにし、脳構造データからどれほど特定の刺激に対する脳活動を予測できるのかを調査した。

質問投稿サイトの検索ユーザからの選好誘出

質問投稿サイトにおいてどのようなランキングが好まれるのかを、インターリーピングと呼ばれる手法を用いて推定する実験を複数回にわたり行った。

(3) 複数の質問による信頼性の高い意見誘出

ソーシャルメディアからの意見抽出

ソーシャルメディア上において、特定の属性を持ち、ある話題に関して意見を発信しているユーザから意見を抽出する方法について提案および実験を行った。

Webおよびレビューの集約による意見抽出

Webおよびレビューからエンティティに関する比較表現や数値表現を収集し集約することによって、エンティティをある特定の順に並び替える方法について提案および実験を行った。

ニュースメディアにおける信頼性の高い意見抽出

将来の事象に関する予測が複数のニュースメディアで行われている際に、報道時期と報道元を考慮することによって信頼性の高い意見を発見する方法について提案および実験を行った。

4. 研究成果

(1) 質問回答に基づく検索意図の誘出

会話に基づく適合フィードバック

質問応答プロセスを確率モデルによって記述し、ユーザの期待利得を最大化する質問を生成する方法を提案した。この枠組みの中では、ユーザがある質問に対して回答可能であるか、また、その質問に対してどのような回答が行われるのか、そして、その質問に回答することでどれほど検索結果を改善できるのかを推定することによって、ユーザの期待利得を算出している。シミュレーションによる実験では、有名人および飲食店検索タスクにおいて提案モデルが高い精度を達成することを明らかにした。

クエリ推薦配置の最適化

クエリ推薦の使用の仕方によってユーザが持つ検索意図を推定し、可能な限り意図を満たせるような検索結果の提示を行う方法について研究を行った。この問題に対して、我々は128名のユーザのクエリ推薦利用行動をデスクトップおよびモバイル環境において観察し、ユーザモデルを構築した。構築されたユーザモデルによって検索結果の質を予測し、これを最大化するようなクエリ推薦・検索結果を生成する方法を提案した。

再検索における効果的な質問の検討

再検索に関するユーザ実験では、400名程度のユーザに対して事前にニュース記事を読んでもらい、二週間後に同様の記事を検索してもらうというタスクを課した。再検索の際にユーザへ4種類の異なる質問を行い、得られた回答を元にクエリ修正を行い、検索結果の改善率を測定した。実験の結果、4種類中3種類の質問に効果が見られ、有意な差は見られなかったものの、閲覧したニュース中において「行動」に着目した質問を行った際に大きな改善が見られた。

(2) 選好の誘出可能性の検証

画像から喚起される印象の推定

この研究の初期においては、ある画像が特定の印象を喚起するかどうかを深層学習によって推定する方法を提案していた。この研究が発展し画像から詩を生成する方法を

提案するに至った。提案手法では、画像から物体、感情、シーン特徴を抽出し、これによって画像を特徴づけ、画像と詩を同じ空間に埋め込むことによって、画像と詩の対応づけを行なった。また、生成された詩が与えられた画像に適するように、また、生成された詩が詩らしい特徴を有するようにするため複数の敵対的生成ネットワークを利用した。実験では、BLEU や人手による評価に加え、提案手法によって生成された詩が人が作成した詩かを人が判別できるかどうかを評価した。実験の結果、生成元となった画像と詩が提示された場合、詩の専門家でない人は自動生成された詩と人が生成した詩をほとんど区別できないことが明らかとなった。

脳構造データからの脳活動データの予測

脳構造データと脳活動データのペアから脳構造データを用いて脳活動データを予測する方法を開発した。データセットとしては、Web 上に公開されているサンプルタスクや言語タスクに関する fMRI データを利用した。

質問投稿サイトの検索ユーザからの選好誘出

国際ワークショップ NTCIR において、3 年間にわたり OpenLiveQ と呼ばれる共有タスクを運営し、実サービス環境におけるインターリーピングの性能評価を行った。具体的には、ヤフー株式会社が運営するヤフー知恵袋において、合計 6 ヶ月間にわたってインターリーピングを実施し、どれほどのインプレッションが得られればユーザの選好を統計的に有意に推定できるかを調べた。

(3) 複数の質問による信頼性の高い意見誘出

ソーシャルメディアからの意見抽出

ソーシャルメディアにおいて、ある話題に関する意見は発信者の属性と合わせて記述されることは少なく、単純な文字列マッチングによる検索では、特定の属性を持ち、ある話題に関して意見を発信しているユーザを発見することは困難であった。そこで我々はソーシャルメディアが構成するグラフへ適合フィードバックを伝播させることによって、特定の属性を持つユーザを効果的に発見する方法を提案し実験を行った。

Web およびレビューの集約による意見抽出

比較表現の集約に関しては、比較表現の確信度の推定方法およびファジーランキング SVM の提案を行い、カメラ、スマートフォン、イヤフォンの様々なランキングを実現した。数値表現の集約に関しては、文脈誘導型ランキング学習を提案した。この方法では、利用可能な訓練データが少ない場合であっても、Web 上の記述からパラメータの値を推定することによって過学習を回避することが可能である。国、都道府県、カメラの 3 つのドメインに文脈誘導型ランキング学習を適用し、他のランキング学習よりも高い精度を達成できることを確認した。

ニュースメディアにおける信頼性の高い意見抽出

Web 上の記事を用いて将来における知識ベースの変化を予測する問題に取り組んだ。特に、記事の報道元と報道時期が、記事の内容から予測される結果の確信度に大きな影響を与えると仮定し、この仮定を明示的に表現する損失関数および特徴量を提案した。実験では、「サッカー選手の移籍予測」を題材として、記事の報道元と報道時期が予測精度に与える影響、および、提案する損失関数によって結果の確信度が正確に予測できるかを評価した。

5. 主な発表論文等

[雑誌論文](計6件)

Frank Hopfgartner, Allan Hanbury, Henning Müller, Ivan Egel, Krisztian Balog, Torben Brodt, Gordon V. Cormack, Jimmy Lin, Jayashree Kalpathy-Cramer, Noriko Kando, Makoto P. Kato, Anastasia Krithara, Tim Gollub, Martin Potthast, Evelyne Viegas, Simon Mercer, Evaluation-as-a-Service for the Computational Sciences: Overview and Outlook, Journal of Data and Information Quality, 査読有, 10 (4), 2018, 15:1 - 15:32
DOI: 10.1145/3239570

片岡大祐, 加藤誠, 山本岳洋, 大島裕明, 田中克己: SNS グラフデータにおける文脈を考慮した適合フィードバック検索, 日本データベース学会和文論文誌, 査読有, 16-J, 2018
http://dbsj.org/journal/dbsj_journal_j/dbsj_journal_vol_16_21/

Jun-Li Lu, Makoto P. Kato, Takehiro Yamamoto, Katsumi Tanaka, Event Identification for Explicit and Implicit References on Microblogs, Journal of Information Processing, 査読有, Vol. 25, 2017, 505 - 513
DOI: 10.2197/ipsjip.25.505

内田 臣了, 山本 岳洋, 加藤誠, 大島 裕明, 田中 克己, 経験的属性によるオブジェクト検索, 日本データベース学会和文論文誌, 査読有, 14(7), 2016, 1 - 6
http://dbsj.org/journal/dbsj_journal_j/dbsj_journal_vol_14_07/

Bei Liu, Makoto P. Kato, Katsumi Tanaka, Cognition-Aware Summarization of Photos Representing Events, IEICE Transactions, 査読有, E99-D, 2016, 3140 - 3153
DOI: 10.1587/transinf.2016edp7079

Jun-Li Lu, Makoto P. Kato, Takehiro Yamamoto, Katsumi Tanaka , Entity Identification on Microblogs by CRF Model with Adaptive Dependency , IEICE Transactions , 査読有 , E99-D , 2016 , 2295 - 2305
DOI: 10.1587/transinf.2016edp7015

[学会発表](計9件)

Makoto P. Kato, Tomohiro Manabe, Sumio Fujita, Akiomi Nishida, Takehiro Yamamoto , Challenges of Multileaved Comparison in Practice: Lessons from NTCIR-13 OpenLiveQ Task , In Proc. of the 27th ACM International Conference on Information and Knowledge Management (CIKM 2018) , 査読有 , 2018 , 1515 - 1518
DOI: 10.1145/3269206.3269318

Bei Liu, Jianlong Fu, Makoto P. Kato, Masatoshi Yoshikawa , Beyond Narrative Description: Generating Poetry from Images by Multi-Adversarial Training , In Proc. of the 26th ACM international conference on Multimedia (MM 2018) , 査読有 , 2018 , 783 - 791
DOI: 10.1145/3240508.3240587

Daisuke Kataoka, Makoto P. Kato, Takehiro Yamamoto, Hiroaki Ohshima, Katsumi Tanaka , Context-aware Relevance Feedback over SNS Graph Data , In Proc. of IEEE/WIC/ACM International Conference on Web Intelligence (WI 2017) , 査読有 , 2017 , 823 - 830
DOI: 10.1145/3106426.3106527

Shinryo Uchida, Takehiro Yamamoto, Makoto P. Kato, Hiroaki Ohshima, Katsumi Tanaka , Entity Ranking by Learning and Inferring Pairwise Preferences from User Reviews , Proceedings of the 13th Asia Information Retrieval Societies Conference , 査読有 , 2017 , 141 - 153
DOI: 10.1007/978-3-319-70145-5_11

Tomohiro Manabe, Akiomi Nishida, Makoto P. Kato, Takehiro Yamamoto, Sumio Fujita , A Comparative Live Evaluation of Multileaving Methods on a Commercial cQA Search , In Proc. of the 40th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2017) , 査読有 , 2017 , 949 - 952
DOI: 10.1145/3077136.3080687

Rafael López-García, Makoto P. Kato, Katsumi Tanaka , A Propagation-based Method of Estimating Students' Concept Understanding , In Proc. of the 9th International Conference on Social Informatics (SocInfo 2017) , 査読有 , 2017 , 611 - 627
DOI: 10.1007/978-3-319-67217-5_36

Makoto P. Kato, Katsumi Tanaka , To Suggest, or Not to Suggest for Queries with Diverse Intents: Optimizing Search Result Presentation , In Proc. of the 9th ACM International Conference on Web Search and Data Mining (WSDM 2016) , 査読有 , 2016 , 133 - 142
DOI: 10.1145/2835776.2835805

Jun-Li Lu, Makoto P. Kato, Takehiro Yamamoto, Katsumi Tanaka , Entity Identification on Microblogs by CRF Model with Adaptive Dependency , In Proc. of IEEE/WIC/ACM Web Intelligence Conference 2015 , 査読有 , 2015 , 333 - 340
DOI: 10.1109/WI-IAT.2015.94

Shuya Ochiai, Makoto P. Kato, Katsumi Tanaka , Re-call and Re-cognition in Episode Re-retrieval: A User Study on News Re-finding a Fortnight Later , In Proc. of the 23th ACM International Conference on Information and Knowledge Management (CIKM2014) , 査読有 , 2014 , 579 - 588
DOI: 10.1145/2661829.2661920

[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://www.mpkato.net/>

6 . 研究組織

研究代表者のみ

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。