

## 自己評価報告書

平成 23 年 5 月 23 日現在

機関番号：12613

研究種目：若手研究（B）

研究期間：2008 ～ 2011

課題番号：20700129

研究課題名（和文） 時系列文書における主題遷移パターンの抽出と利用

研究課題名（英文） Extraction and reorganization of topic transition patterns in recorded documents

研究代表者

田中 克明（TANAKA KATSUAKI）

一橋大学・情報基盤センター・助教

研究者番号：80376657

研究分野：知能情報学

科研費の分科・細目：情報学・知能情報学

キーワード：情報システム、人工知能、可視化、設計工学

## 1. 研究計画の概要

人々の活動の記録として時間経過とともに蓄積される文書群には、活動に参加した人々がどのような知識を用い、活動を行ったかの痕跡が含まれている。

本研究では、これら文書群の解析を行うことにより、そこに記述された主題の変化パターンを抽出し、1. 変化をもたず定常的に記述される知識（オントロジー的なもの）、2. ある時点でのみ使われる知識（アドホックに使われるヒューリスティクス的なもの）を抽出することを目指している。

とくに後者の抽出を行うことにより、明示的な構造を持ち、なおかつ時間経過と共に変化してしまわないという性質を持つことで知識システムの拠り所となっているオントロジーに対し、オントロジーではカバーしきれなかった知識の利用側面をとらえ、利用者の文脈に寄り添った知識支援を行うことを目的とする。そのために、文書からの記述内容の時間変化パターンの抽出と、それらを抽出システム利用者の視点に基づいて動的に再構成を行うことができるシステムの構築を行う。

## 2. 研究の進捗状況

本研究では、比較的長期間にわたって行われた設計活動などの記録文書を集め、一定時間分ごとの文書にグループし、文書によっては細かな断片へと分けたいうえで、グループごとにクラスタリングを行い、生成されたクラスタの文書グループ間、つまり時間経過に伴う遷移を主題遷移として抽出し、つづいて利用者の視点から再構成するシステムを構築している。システムの構築に伴い、これまでに以下のような検討を行った。

(1) 時系列文書以外の時系列データとして、個人の行動履歴に着目し、個人が持つ携帯電話から、位置の履歴、閲覧・記述したコンテンツの履歴などからなる行動履歴の取得を行った。取得した履歴に基づき、個人が日常的な行動をとっているのか、非日常的な行動をとっているのかを判別する仕組みを構築し、文書以外にも本手法が適用可能であるかを確認した。

(2) 文書内に記述されている内容の忘却を行うモデルの検討、および時間経過に伴い忘却する割合のパラメータ検討を実施した。具体的には、文書内に記述されている内容を分類し主題とするためのクラスタリングにおけるパラメータ、忘却を行う仕組み、忘却に際してのパラメータなどの検討を行った。

(3) 設計活動にまつわる記録以外に、近年 Web 上に数多く蓄積されるようになった、多数の人間が参加するコミュニケーションの記録を対象として、本研究手法の適用を行った。

その結果、なんらかの趣旨に基づき蓄積されたデータへ、本研究の手法を応用できることがわかった。たとえば、同一の目的を持って複数の対象の検討を行った議事録からは、忘却の度合いを高くすることで各対象に関する議論を抽出できた。

一方、特定の対象も共通の目的もたず、多人数が多数の対象について並行して議論を行っているような文書群では、議論にまつわる複数の時間が並行して進んでいる状態となり、一律に忘却を行う本研究の手法では、主題繊維の抽出、および再構成の結果が、解析者に理解しづらいものとなる場合があることが分かった。

### 3. 現在までの達成度

②おおむね順調に進展している。

(理由)

変化をもたず定常的に出現する事項、ある時点でのみ出現する事項の抽出を、複数の時系列文書から行っている。個人の行動履歴という、文書以外の時系列データからの抽出も行えた。また、抽出結果を利用するために、並び替える、組み合わせるなどの検討も行い、そのためユーザインタフェースの整備を進めている。これらより、本研究はおおむね順調に進展していると判断する。

### 4. 今後の研究の推進方策

まず、以下の2つのシステムの研究開発、実装を引き続き行う。

(1) 時間経過に沿って蓄積された文書群のある時点において出現した主題が、一時的な主題か、あるいは定常的な主題であるかの分類を、システム利用者の設定に応じて行うシステム。

(2) 出現する主題をそこに含まれる内容、出現期間などの指定に基づいて抽出し、システムを利用する解析者の視点に基づいて組み合わせおよび再構成を行うシステム。

つぎに、これまでにおこなった検討をふまえて、蓄積された文書群のうち、本研究が適用可能な領域、すなわち文書群への知識の集約が起きる領域がどのようなものであるかについて検討を行う。この際、たとえば一律に忘却を行う部分の変更など、本研究手法の調整も検討する。

同時に、活動に参加した人々の知識が集約される形で文書を残すためにはどのような場の設定が適しているのか、すなわち本研究の手法にあった情報の蓄積を起こすのに適している状況を用意する方法について、調査をすすめる。

### 5. 代表的な研究成果

(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 8 件)

1. Katsuaki Tanaka, Koichi Hori, Masato Yamamoto, Development of a Recommender System based on Extending Contexts of Content and Personal History, Journal of Emerging Technologies in Web Intelligence, Vol2, No.3, 2010, pp.197-203, 査読有
2. 田中克明, 堀浩一, Twitter ハッシュタグに基づく Tweet 群からの変化抽出, 電子情報通信学会第二種研究会資料 WI2-2010, 2010, pp.31-32, 査読無,
3. 田中克明, 堀浩一, 山本真人, 個人行動履歴に基づく情報推薦システムの開発,

人工知能学会論文誌, Vol.23, 2008, pp.412-423, 査読有

4. Katsuaki Tanaka, Mina Akaishi, Koichi Hori, Reorganizing Topic Transitions in Design Process Records, Proc. of the Third International Conference on Knowledge, Information and Creativity Support Systems, Vol.1, 2008, pp.148-155, 査読有