

インタラクティブクラスタリングのためのマルチタッチインタフェースの提案

Proposal of Multi-Touch Interface for Interactive Clustering

三宅 遼祐^{*1}
Ryosuke Miyake

山田 誠二^{*2}
Seiji Yamada

岡部 正幸^{*3}
Masayuki Okabe

高間 康史^{*1}
Yasufumi Takama

^{*1} 首都大学東京大学院 システムデザイン研究科
Graduate School of System Design, Tokyo Metropolitan University

^{*2} 国立情報研究所/総合研究大学院大学/東京工業大学
National Institute of Informatics, SOKENDAI, Tokyo Institute of Technology

^{*3} 豊橋技術科学大学 情報メディア基盤センター
Information and Media Center, Toyohashi University of Technology

In modern society, it is important to get desired information from large-scale information source. Although recent advance of information processing such as information retrieval and recommendation has provided us with information access support, automatic information processing only is not enough. From this viewpoint, intelligent interactive systems that aim to perform tasks in collaboration with users have been studied, such as active-mining and visual data-mining. This paper proposes an interface for constrained clustering. Constrained clustering uses document pairs (must-links and cannot-links) as constraints. In order to support specifying several constraints, the proposed method employs cluster-level operation rather than object-level operation. A user using the proposed interface operates on clustering structure through multi-touch interface, and results are translated into a set of constraints. This paper describes the proposed interface by introducing the prototype interface.

1. はじめに

現代社会において情報の大規模化が急速に進行しており、情報が持つ価値、重要性が非常に高くなっている。特に、大量の情報の中から自分にとって有益な情報をいかに取り出すかが重要であり、データマイニングなどの計算機による支援技術が研究されている。しかし、大規模情報に対して、計算機による自動処理だけで適切な結果を得ることは困難であるため、人間と計算機の協調によるアクティブマイニングやビジュアルデータマイニングが研究されている。

本稿では制約クラスタリングに基づくビジュアルデータマイニングにおける、人間と計算機によるインタラクティブクラスタリングの支援をするために、直観的な分析作業を可能とするマルチタッチインタフェースシステムを提案する。制約クラスタリングは半教師あり学習の一種であり、アクティブマイニングに有効な手段であるが、オブジェクト単位で制約を指定する必要があり、ユーザの負荷が大きいという問題がある。提案システムではこの問題を解決するため、オブジェクト単位ではなくクラスタ単位で制約を指定可能とする。

本稿ではマルチタッチディスプレイを利用してプロトタイプインタフェースを実装し、提案システムの有効性を検証する。

2. 関連研究

2.1 制約クラスタリング

クラスタリングとは、データ解析方法の一つであり、与えられた類似度を基準にデータを分類する教師なし学習である。対象とするデータは一般に、多様な観点から分類することが可能な場合が多いが、各クラスタリングアルゴリズムはある特定の観点に

従い分類するため、ユーザの求める結果になるとは限らない。制約クラスタリングは、対象データの分類に関して、人間の持つ常識や背景知識などに基づき人手で制約を与える半教師あり学習であり、クラスタリングアルゴリズムは与えられた制約を満たすようにデータを分割する。制約は、*must-link*, *cannot-link* の2種類のオブジェクト対で与える手法が代表的である[Terami 10, wagstaff 01]。 *must-link* を付与されたオブジェクト対は同じクラスタに、 *cannot-link* を付与されたオブジェクト対は異なるクラスタに分類されるべきであることを意味する。

2.2 マルチタッチインタフェース

近年、スマートフォンなどに代表されるマルチタッチインタフェースを備えたデバイスが普及している。一般的なタッチインタフェースとは異なり、2点以上のタッチを認識することができることが特徴であり、単なるオブジェクト選択以上の多様な操作を可能とする。代表的操作にピンチイン、ピンチアウトがあり、これらは操作対象を縮小、拡大するときなどに用いられる。2本の指をスライドしての操作は従来のマウスなどのインタフェースを利用する操作と比較し、入力作業に対する人間とコンピュータとの連動性が高いことが知られている。

3. 制約クラスタリングのためのマルチタッチインタフェース

3.1 クラスタ単位での制約生成

制約クラスタリングは多様なデータ分析に適用可能なアプローチであるが、多様な制約を付与しなくてはならない場合にユーザの負荷が問題となる。すなわち、制約オブジェクト対を指定するために、オブジェクトを一つずつ選択しなくてはならないのでは、多数の制約を付与することは困難である。

この問題を解決するために、本稿ではオブジェクト単位ではなくクラスタ単位でのインタラクションに基づき、複数制約を一括

連絡先: 高間康史, 首都大学東京大学院システムデザイン
研究科, 〒191-0065 東京都日野市旭が丘 6-6,
ytakama@sd.tmu.jp

して指定可能なアプローチを提案する。すなわち、ユーザがあるクラスタを生成した場合、クラスタ内のオブジェクト間に **must-link** 制約を生成する。

ユーザが一からクラスタ構造を生成するのではオブジェクトレベルでの制約指定と負荷は変わらないため、あらかじめ階層的に生成したクラスタ構造を用いて、希望する粒度に近いレベルのクラスタをインタラクティブに選択したり、結合して新しいクラスタを生成したりすることを可能とする。

図1に提案手法で用いる階層的クラスタ構造を示す。各ノードがクラスタに対応し、付与されている数字が **Level** を表す。Level 0 のクラスタを可視状態(**visible**)とし、操作対象としてインタフェース上に表示する。Level 0 のクラスタを基準に、階層が一つ下がるにつれ **Level** の値を+1, 上がるにつれ-1とする。**visible** なクラスタに対して、ユーザは以下の二つの操作を実行することができる。

- **Break**: 一つのクラスタを二つの **visible** なクラスタに分解する
- **Merge**: 二つのクラスタを合成し、一つの **visible** なクラスタを生成する

制約クラスタリングを実行したい場合、その時点で表示されているクラスタの中から、初期には **visible** でなかったクラスタを選択する。その構成オブジェクトの全ての組み合わせを **must-link** として生成する。例えば、クラスタ{A, B, C, D}をユーザが新たに生成し、制約クラスタリングを実行した場合、[A-B], [A-C], [A-D], [B-C], [B-D], [C-D]の6つすべてのオブジェクト対に **must-link** が生成される。これにより、オブジェクト単位で制約付加を行うよりも、ユーザの負荷を軽減できることが期待できる。

生成された制約を用いて距離学習に基づく制約クラスタリング[Okabe 10]を行い、データの分割を再帰的に行うことで、新たな階層的クラスタ構造を得る。

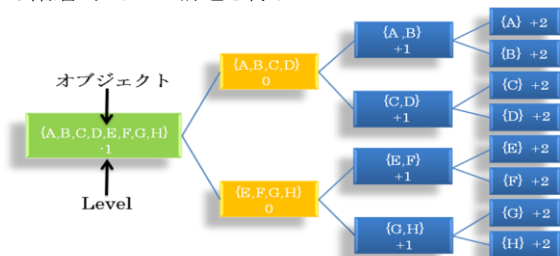


図1 階層的クラスタ構造の例

3.2 マルチタッチインタフェースの概要

開発したプロトタイプインタフェースのスクリーンショットを図2に示す。提案インタフェースは文書クラスタリングを対象としている。提案するプロトタイプインタフェースの開発には Java 言語を用いており、マルチタッチインタフェース用ライブラリ MT4j¹を利用している。各クラスタはサークルで表されており、ユーザがインタフェース上でタッチすることで、作業対象となるクラスタを指定する。サークルの描画に関しては色、大きさ、動きの3つのパラメータがあり、これらを用いてクラスタの特性を表現する。

マルチタッチ操作を利用し、**Break** は対象クラスタを二本指でタッチすることで、**Merge** は二つのクラスタを同時にタッチすることで行う。

また、ユーザによる分析作業が終わり、クラスタ構造の再構築を行いたい場合は、インタフェース上の左下に「Q」と表記されているボタンをタッチすることで、再クラスタリングが実行され、制約クラスタリングの実行結果が取得できる。

サークルの色は赤に近づくほどクラスタ内類似度が低く、青に近づくほどクラスタ内類似度が高いことを表現する。本稿では式(1)をクラスタ内類似度として用いる。ここで、 $DF(x,t)$ は、クラスタ x で単語 t を含む文書数を表す。

$$\text{sim}(x) = \frac{|\{t \mid DF(x,t) > 1\}|}{|\{t \mid DF(x,t) > 0\}|} \quad (1)$$

サークルの大きさは、各クラスタの持つ文書数で決定する。また、サークルの動作は、前回のクラスタリング結果と比較し、構成する文書集合が最も類似しているものとの類似度(Jaccard 係数)を利用し決定する。類似度が高いほどサークルの動きが落ちつきやすく、低いほど活発に動くよう設定する。これにより、さらなるクラスタリングが必要かどうかを、直観的にユーザが判断可能となる。

4. 動作検証

提案手法に基づき実装したプロトタイプインタフェースの動作について、**Break** 操作の実行例を以下に示す。

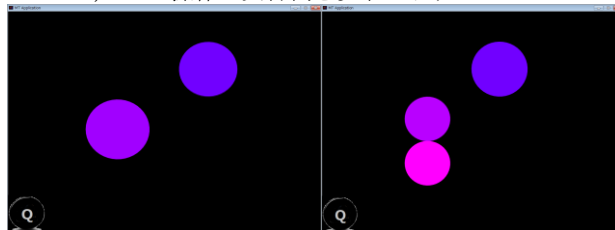


図2 Break 実行例(左:実行前, 右:実行後)

図2において、左図の左に位置するサークルに対する **Break** 実行後が右図である。**Break** 操作により子クラスタの **Level** が 0 に更新され、**visible** な状態となる。新たにインタフェース上に表示されたサークルは親クラスタよりオブジェクト数が少ないため小さく表示されている。また、下方のクラスタの色が赤に近くなっていることから、クラスタ内の類似度が低下していることがわかる。

5. おわりに

本研究では、制約クラスタリングに基づくビジュアルデータマイニングのためのマルチタッチインタフェースを提案し、プロトタイプシステムを実装して動作検証を行った。

今後は、クラスタリングの制約付加方法の検討・改良や、マルチタッチインタフェースの操作性の更なる向上により、インタフェースとしての完成度を高める予定である。

参考文献

- [Okabe 10] M. Okabe and S. Yamada.: Learning Similarity Matrix from Constraints of Relational Neighbors, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.14, No.4, pp. 402-407, 2010.
- [Terami 10] 寺見明久, 宮本定明:階層的クラスタリングにおける対制約の導入のための二つのアプローチ, Fuzzy System Symposium, pp.13-15, September 2010
- [Wagstaff01] K.Wagstaff,C.Cardie,S.Rogers,S.Schroedl: Constrained K-means Clustering with Background Knowledge, Proc. 18th International Conf. on Machine Learning, pp. 577-584, 2001.

¹http://www.mt4j.org/mediawiki/index.php/Main_Page