

人間の能動学習を促すインタラクションデザイン

Interaction Design for Inducing Users Active Learning

水上 淳貴*¹ 岡部 正幸*² 山田 誠二*³
 Junki Mizukami Masayuki Okabe Seiji Yamada

*¹東京工業大学大学院
 Tokyo Institute of Technology

*²豊橋技術科学大学
 Toyohashi University of Technology

*³国立情報学研究所 / 総合研究大学院大学 / 東京工業大学
 National Institute of Informatics, SOKENDAI, Tokyo Institute of Technology

Obtaining user's desired information from large-scale information sources is very important task in data mining. However, it is difficult to realize the correct classification to satisfy the user's request. The constrained clustering is a clustering technique that can deal with user's request. This paper describes GUI for constrained clustering that can induce user's active learning. It is helps users to select effective constraints efficiently during the human active learning process by function of display arrangement. Moreover we introduce how system works, particularly in the function of arrangement by multidimensional scaling and COP-KMEANS.

1. はじめに

情報技術の発展に伴い、Web 上にビッグデータのような大量のデータが蓄積されるようになった近年、個人で扱えるデータ量が急激に増加している。しかし、蓄積されたデータがあまりにも大量であるため、必要な情報を発見する事が困難な状況になっている。それ故、膨大なデータを俯瞰するために、“データのグループ化”への期待が高まりつつある。グループ化する方法としては、一般的にクラスタリングが適用される。クラスタリングは、データマイニングにおいて最も広く使用されている技術の一つである。しかし、コンピュータの行う一度のクラスタリングだけで、ユーザが所望する結果を提示することは難しい。したがって、人間とコンピュータが協調し、人間から効果的なクラスタリングに関する制約を与える制約付きクラスタリングによって分類精度の向上が期待できる。

一方、有効な制約を選択する能動学習は、制約付きクラスタリングにおいても、重要であり、従来計算論的能動学習が研究されてきた。最近では、人間の能動学習に注目が集まりつつあるが、ユーザへの作業負荷の増大、制約の質の不安定性が問題となっている [Okabe 11]。本研究では、この制約付きクラスタリングにおける、人間の能動学習を促し、人間と計算機によるインタラクティブクラスタリングを効率的に行うための GUI の開発を行う。

本研究では、制約とその効果の因果関係を顕在化でき、さらにクラスタリング結果表示の 2 軸を柔軟にユーザが変更できる GUI を開発することで、ユーザの能動学習を促すインタラクションデザインを実現する。

2. ユーザの能動学習を促す GUI

2.1 制約付きクラスタリング

クラスタリングとは、データ解析の手法の一つであり、与えられたデータの類似度を基準に分類する教師なし学習である。クラスタリングアルゴリズムはある特定の観点から分類するために、ユーザの所望する分類になるとは限らない。制約付きク

ラスタリングは、扱うデータに関して、ユーザの持つ背景知識に基づいた制約を与える半教師付き学習である。クラスタリングアルゴリズムは、ユーザから与えられた制約を満たすような分類を行う。制約とは、あるデータペアが同じグループに属するか否かに関する知識であり、同じグループに属するべきデータペアである must-link 制約、同じグループに属すべきでないデータペアである cannot-link 制約の 2 つがある。

本研究では、制約付きクラスタリングとして COP-KMEANS[Wagstaff 01] を用いる。そのアルゴリズムを以下に示す。

COP-KMEANS

1. 任意の k 個、 $C_1 \dots C_k$ を、一様ランダムに初期クラスタ中心とする。
2. データセット D の各ポイント d_i を、VIOLATE-CONSTRAINTS を偽とする、最も近いクラスタ C_j に割り当てる。そのようなクラスタが存在しない場合、割り当てに失敗する（空集合を返す）。
3. 各クラスタ中心 C_i は、構成する各ポイント d_i の距離の平均を計算し、更新される。
4. クラスタに変化がなくなるまで、ステップ 2, 3 を繰り返す。
5. 収束したクラスタ $C_1 \dots C_k$ を返す。

VIOLATE-CONSTRAINTS

- must-link 制約を与えられたデータペアが同じクラスタに属さない場合、真を返す。
- cannot-link 制約を与えられたデータペアが同じクラスタに属する場合、真を返す。
- それ以外の場合、偽を返す。

2.2 多次元尺度構成法

本研究では、3 次元以上を持つデータ集合においても集合内の近接関係をできるだけ保存した形でユーザに提示できるように、多次元尺度構成法 (Multidimensional Scaling, MDS) を

用いて、3次元以上あるデータを2次元上に配置する。MDSは、データ集合の距離行列からデータの配置を決定する方法として用いられる。MDSは、技術的に固有値分解に基づいており、一般的に固有値の大きい順に座標軸を使用していく。

2.3 GUIのメカニズム

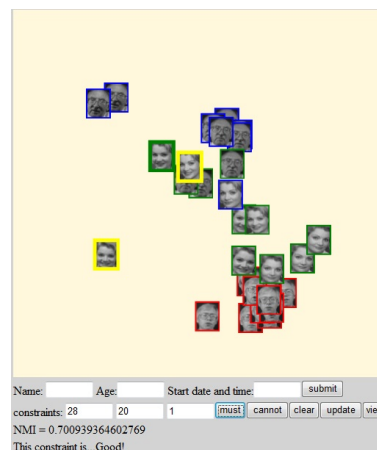
図1に、本研究で作成した試作システムのGUIを示す。構成として、データを描画する領域と制約の追加、制約付きクラスタリングの実行、描画に用いる座標軸の変更を行うボタンなどがある。この試作システムでは、データの内容を視覚的に把握できる画像データを用いており、これを人物別に分類することが目標である。データは、初期状態（ロード時）において制約のないk-meansで一度クラスタリングされる。クラスタリングされた結果に基づき、各画像データが割り当てられたクラスに相当する色（赤、青、緑）で縁どられ、MDSによる座標に基づいて描画されている。これを基に、ユーザはインクリメンタルに制約を追加していく。

制約を追加していくプロセスとして、まず、制約を与えたいデータをクリックすることによって選択する。選択された画像データは、元あったクラスの色から黄色（選択済み）で縁どられることで強調表示される。図1(a)では、女性の画像が選択されている。データペアを選択した後、must ボタン、または cannot ボタンをクリックすることで、そのデータペアに与える制約の種類を決定する(a)の場合、選択されたデータは同じクラスにあるべきなので must ボタンをクリックしている。この時、データの選択、制約の選択を間違えてしまった場合、clear ボタンをクリックすることで、そのステップ中のデータペアと制約を選択し直すことができる。追加する制約が決定すると、update ボタンをクリックすることで、新しく追加した制約と既存の制約を満たす制約付きクラスタリングが行われ、その結果に基づき再度、画像データのクラス（画像データを縁どる色）が決定する。本研究では、ユーザにとって、自身の与えた制約の影響を把握できることが効率的な能動学習を促すきっかけになると考え、1ステップ前とクラスの割り当てに変化が生じたデータは、太枠でデータを縁どるような強調表現を行なっている。図1(b)において、中心部分に位置する画像データの何枚かが、太枠になっており、制約の影響を受けたことがわかる。

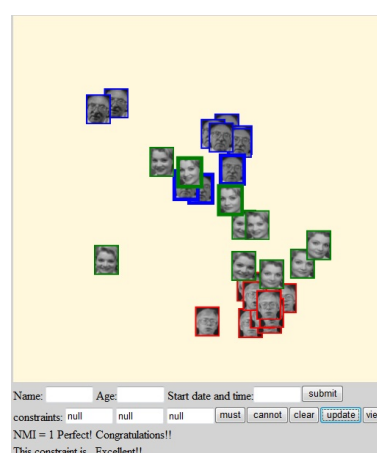
一方、MDSによって固有値の大きい順に選んできた2軸の組み合わせが、必ずしもユーザが最も概観しやすい2次元配置であるとは限らない。そこで、クラスタ分割数に応じて、固有値の大きい順に座標軸を選択しておき、view ボタンをクリックすることで座標軸の組み合わせを変更する。図2(a)と(b)は座標軸の組み合わせ変更の例を示している。図1と図2(a)と(b)は、元のデータ集合は同じものを用いているが、MDSによる軸の組み合わせを変更している。図2(a)では3種類の画像データが不規則に配置されており概観しづらいが、図2(b)では軸の組み合わせが図1と図2(a)の場合と異なり、3種類の画像データがうまく分離されている。

3. まとめ

本論文では、ユーザがインクリメンタルに制約を追加しながら分類学習を行うためのインタラクティブデザインとしてGUIを提案した。今後は、データ数を増やし、大量のデータにおいても、ユーザが効果的な制約を与えられるよう改良し、システムとしての完成度を高め、既存の計算論的能動学習と比較実験を行う予定である。

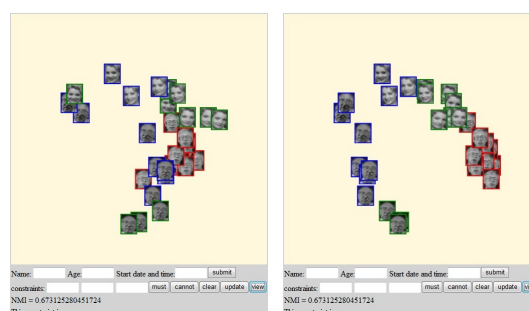


(a) 2ステップ後（制約追加中）



(b) 3ステップ後

図1: 試作システムのGUI



(a) 概観しづらい例

(b) 概観しやすい例

図2: 座標軸の組み合わせ変更

参考文献

- [Okabe 11] M. Okabe and S. Yamada: An Interactive Tool for Human Active Learning in Constrained Clustering, *Journal of Emerging Technologies in Web Intelligence*, Vol. 3, No. 1, pp. 20-27.
- [Wagstaff 01] K. Wagstaff, C. Cardie, S. Schorodl: Constrained K-means Clustering with Background Knowledge, *Proc. 18th International Conf. on Machine Learning*, pp. 577-584, 2001.