

人々の主観とグラフクラスタリングに基づいた画像分類

Image classification based on human intuitions and graph clustering

川本 達郎*¹ 近藤 那央*¹
Tatsuro Kawamoto Nao Kondo

*¹産業技術総合研究所 人工知能研究センター
National Institute of Advanced Industrial Science and Technology

We investigate the classification of human-face images based on similarities that people feel. We conducted an experiment so that the collected data constitutes a sort of random graph, which enables us to classify the human faces using a graph clustering algorithm. The precise criterion of similarity fluctuates depending on individual and each individual answers the similarities among a small fraction of the whole dataset. Nevertheless, as a consensus, we show that a statistically significant global structure of similarity emerges.

1. はじめに

人間の顔認識・識別の問題は、古くから研究されており、機械学習による画像解析の一つの重要な例となっている。しかし本研究では、そのような識別問題とは根本的に異なる問題を考える。機械学習による顔認識は、顔画像の物理的な特徴を抽出することによって識別を行う。一方、本研究では、人間が実際にどのように顔を識別しているかを基準にして分類を行うことを目的としている。すなわち、人間は機械が捉えているほど高い分解能をもって他人の顔の差異を認識していないかも知れないし、物理的特徴としてははっきり現れていない何かを、共通して感じ取っているかも知れない。このような意味で、人間による画像の識別というのは、機械によるタスクとはまったく異なるものである。

もちろん識別の基準は、個人によって厳密には異なる。しかし、多くの人の主観を集めてくることによって、人々が共通して感じる、人間の顔の大域的な類似性構造が浮かび上がって来れば、それは非常に興味深いことである。本講演では、これに対して肯定的な結果を与える。

得られた分類結果について、それが示唆することを研究することにはもちろん意味があると信じている。しかし、本研究では、如何にして人々の主観から大域的な分類を得るかというテクニカルな問題に答えることを目的とし、その分類結果の意味を深く追求することはしない。

2. 結果

最初に結果を示す。121枚の画像データについて、図1のような分類が得られた。結果はウェブサイト [顔顔衰弱 17] でも確認できる。各グループには明示的なラベルは付いておらず、単に人々の回答傾向からグループ構造が見えるだけであるが、その意味は分類結果から容易に解釈できる。すなわち、女性の顔画像は一つのグループとしてまとめられ、詳細な構造は見えないが、男性の顔画像は、中性的な顔とそうでない顔の2グループに分割されていると捉えることができる。後述するように、図1は3分割した結果であるが、4分割までは統計的に有意であると示唆される。しかし、4分割の場合は、女性写真画像・女性イラスト画像・男性写真画像・男性イラスト画像という、ほぼ自明な結果となった。

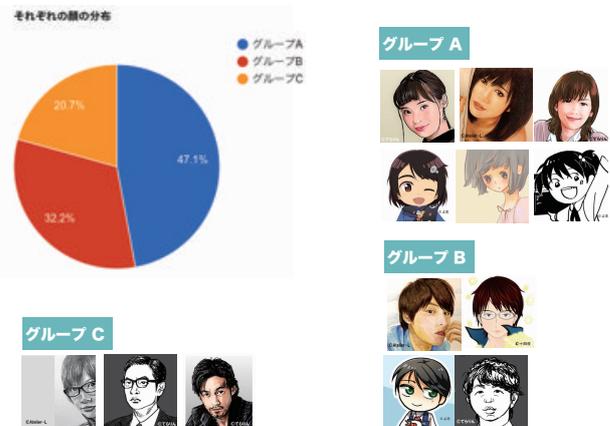


図1: 顔画像の分類結果。左上は、各グループのサイズの分布を示している。表示した各グループの顔画像は、データセットのうちの一部である。

3. 実験

実験のプロセスは非常に単純である。まず、人間の顔画像と顔のイラスト画像をデータセットとして N 枚 ($N = 121$) 用意する。一人の実験協力者は、以下のプロセスを5回繰り返す。

1. データセットからランダムに抽出された9枚の画像から、似ていると思うペアを選択する。
2. 似ているペアが無いと判断した場合、再度1.のプロセスを行う。

これを、多くの実験協力者に行ってもらい、画像データセットのいくつかの組について、類似性データを集める。類似性データの総数を L とする。(これまでの実験では $L = 689$.)

ここで重要なのは、人々の主観情報というのはサンプルするのにとてもコストがかかるということである。画像間の組合せは合計で $N(N-1)/2$ 通りあり得るが、この作業を実験で実行するのは難しい。類似性データの数 L が少ないということは、各々の画像がどのグループに属するかの情報が少ないために、もし背後に何らかの構造があったとしても、それがない場合と見分けがつかない(統計的に有意な差が出ない)かも知れ

ないという難しさを孕んでいる。ここでは、類似性データの数が $L \sim O(N)$ の場合について分類を行うことを考える。

4. 手法: グラフクラスタリング

4.1 Stochastic block model

上記の実験で得られたデータはグラフ（もしくはネットワーク）になっている。各画像を頂点とし、類似しているペアには枝が張られている。枝が張られ得る頂点ペアはランダムに抽出されるので、これは一種のランダムグラフとなっている。同じグループに属する頂点たちが高い確率でつながっているようなランダムグラフモデルとして、stochastic block model がよく研究されている。このモデルを用いて、得られたデータについてのベイズ推論を行う。

Stochastic block model のインスタンスは以下のようにして生成される。グループ数を q とし、頂点 i の属するグループを $\sigma_i \in \sigma$ とする。モデルのパラメーターとして、グループサイズの分布 $\gamma_\sigma \in \gamma$ 、グループ間の結合確率行列 $\omega_{\sigma\sigma'} \in \omega$ (affinity matrix) を持っている。結合確率行列は、頂点 i と頂点 j がそれぞれグループ σ_i と σ_j に属しているとき、確率 $\omega_{\sigma_i\sigma_j}$ で枝が張られることを示している。まず枝が張られていない頂点集合を考え、 γ に従って各頂点のグループラベルをランダムに決定し、その後 ω に従ってランダムに各頂点間に枝を張るか否かを決定する。グラフの生成確率（尤度）は、

$$p(A, \sigma | \gamma, \omega) = \prod_i \gamma_{\sigma_i} \prod_{i < j} \omega_{\sigma_i \sigma_j}^{A_{ij}} (1 - \omega_{\sigma_i \sigma_j})^{1 - A_{ij}} \quad (1)$$

として与えられる。ここで A は隣接行列（頂点 i と j がつながっている場合は行列要素 $A_{ij} = 1$ 、そうでない場合は $A_{ij} = 0$ を持つ $N \times N$ 行列）である。こうしてグラフが一つ生成される。以上は生成プロセスであり、グラフだけが入力として与えられた場合は、各頂点の属するグループ σ （隠れ変数）や γ 、 ω は未知である。また、ハイパーパラメーターとして分割数 q も決定する必要がある。グラフクラスタリングは、これらを推定・学習する、逆問題となっている。

4.2 EM アルゴリズム・信念伝搬法

ベイズ推論としてグラフクラスタリングを行うには、具体的には以下のような自由エネルギーの最小化を考える（対数周辺尤度最大化と等価）：

$$f = -\frac{1}{N} \log \sum_{\sigma} p(A, \sigma | \gamma, \omega). \quad (2)$$

しかし、この最小化問題を直接扱うのは困難であり、様々な近似手法が提案されている。ここでは、EM アルゴリズムによる最適化を行い、その E-step において、信念伝搬法 (BP) を用いる [Decelle 11]。EM アルゴリズムおよび BP の詳細については割愛する。（詳細については [Decelle 11] を参照されたい。）

4.3 検出限界

前述したように、現在考えている状況は $L \sim O(N)$ 程度しか枝がないような、スパースなランダムグラフである。グラフが $L \sim O(N \log N)$ くらいの枝を持てば、漸近的に背後にある構造が完全に推定できることが知られているが、 $L \sim O(N)$ の場合、ある程度のグループ構造の強さがなければ、それは一様ランダムグラフと統計的に有意な差が見られないという検出限界が生じる。このとき、どの程度のグループ構造の強さで検出限界を迎えるのかという理論研究は盛んに行われている。BP による手法は、スパースグラフについて線形時間アルゴリズムとなっており、また検出限界の研究も進んでいる。

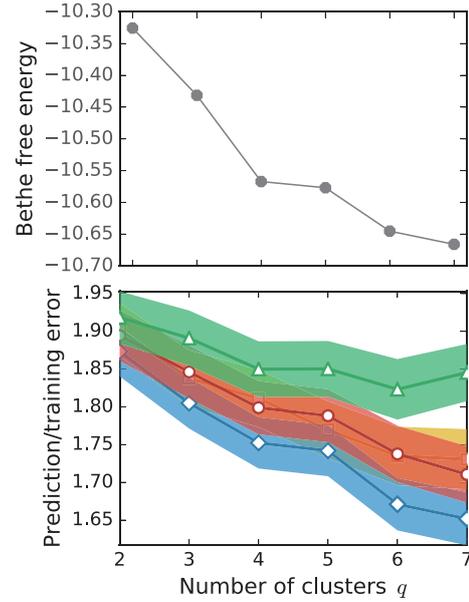


図 2: Bethe free energy と BP を用いた交差検定の結果

4.4 分割数の決定：モデル選択

上記の自由エネルギー最小化問題は、グラフの分割数 q を入力として指定する必要がある。これを決定するもっとも素直な方法は、いくつか異なる q について、自由エネルギー f の近似値であるベーテ自由エネルギー f_{Bethe} を最小化した結果をプロットし、プラトーが見えたところで“正しい”分割数 q^* とすることである。これは、実際にデータが stochastic block model から生成されている場合には確かに上手くいく。しかしながら、実データにおいては、この方法はほとんどの場合上手くいかず、大幅にオーバーフィットする。

“正しい”分割数 q^* を決定する万能の基準は存在しないと思われるが、実データに適用可能な基準として、川本・樺島 [川本・樺島 16] によって提案された交差検定の方法を用いた。通常交差検定には大きな計算コストがかかるが、この手法では、BP の結果を用いることで Leave-one-out cross-validation を解析的に評価することが可能である。また、この交差検定のうち、Bayes 予測誤差 E_{Bayes} と呼ばれる量は、実際に stochastic block model からデータが生成されている場合、(BP が妥当な振る舞いをしている限り) 理論的な検出限界を達成することも示されている。

結果は図2のようになった。ここでは、Bayes 予測誤差 E_{Bayes} (赤色)、Gibbs 予測誤差 E_{Gibbs} (緑色)、Gibbs 予測誤差の MAP 推定 E_{MAP} (黄色)、Gibbs 訓練誤差 E_{training} をプロットしている。各線の幅は、1 standard error を表している。各誤差の詳細は [川本・樺島 16] を参照されたい。多くの実データにおいて、Gibbs 予測誤差 E_{Gibbs} がよい振る舞いを示すことが経験的にわかっている。今回のデータにおいても、Bethe free energy や他の予測誤差がオーバーフィットしているのに対し、Gibbs 予測誤差 E_{Gibbs} はオーバーフィットを防いでいるように見える。予測誤差の測定から、どのようにして最終的な分割数 q^* を選ぶべきなのだろうか。絶対的な方法は存在しないが、単純に予測誤差が最小となるものを選ぶのではなく、one-standard error rule [Tibshirani 13] という基準が経験則としてよく用いられる。これを適用すると、Gibbs 予測誤差 E_{Gibbs} から選ばれる値は $q^* = 4$ となる。しかし、前述したよ

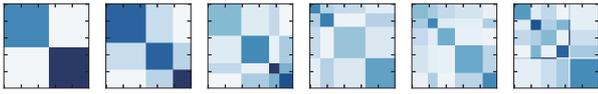


図 3: 学習されたグループ構造. 左から $q = 2, 3, 4, 5, 6, 7$.

うに、4分割の場合は、非常に納得できる結果ではあるものの、あまり興味深い分割にはなっていない。ここで言えることは、4分割までは統計的に有意な結果と言えるということで、2分割や3分割も有意な結果である。従って、今回は $q^* = 3$ に注目することにした。

学習したモデルパラメーター γ, ω はどのようになっているだろうか。各 q における結合確率行列に対し、グループサイズ γ_σ を要素の大きさに反映させたものを図3に示す。色の濃さは、行列要素の値の大きさを表している。これから分かるように、確かに同じグループに属する頂点の場合の結合確率が高い構造（コミュニティ構造）が得られていることが確認できる。

5. おわりに

本研究では、人の主観による顔画像の類似性を実験によりデータ収集し、グラフを構成した。それにグラフクラスタリングを適用することによって、統計的に有意な分割数 q^* の見積もりおよびそのもとでのクラスタリングを行った。現在得られているグラフは非常にスパースであるが、結果は人間が解釈可能な、しかし非自明な、興味深いものとなった。

今回の実験は、ニコニコ闘会議のなかのニコつく 2017 というイベントにて、2日間にわたって行われた。1日目のデータのみから推論を行うと、有意な最大分割数は、 $q^* = 3$ であった。（分割結果は図1とほぼ同じであった。）これは、さらに実験を行いデータを増やせば、さらなる構造が有意なものとして得られる可能性を示唆している。

本研究は、近年研究が進んでいるスパースグラフの研究を積極的に利用した例である。グラフクラスタリング（もしくはコミュニティ検出）のアルゴリズムは、現実のデータに適用してみると上手くいかないということはよくある。本研究は、実データがモデルと近いものになるように実験をデザインすることでこれを克服することを試みた。

実験は、類似性が何であるか明示する必要がないという意味でも面白い。回答している本人すらもはっきりとわからないかも知れない。また、今回は顔画像の分類を試みたが、このフレームワークをベースに、様々な主観ベースのデータを集計できる。グラフクラスタリングの技術的な改善によってさらなる進展が期待できるが、主観データを如何にしてたくさん集めるかという、実験手法の問題を改善することも重要な課題である。

参考文献

[顔顔衰弱 17] <http://img.voteclustering.org/>

[Decelle 11] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, “Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications”, *Phys. Rev. E*, **84**, 066106 (2011).

[川本・樺島 16] T. Kawamoto, Y. Kabashima, “Cross-validation model assessment for modular networks”, arXiv:1605.07915 (2016).

[Tibshirani 13] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: data mining, inference, and prediction*, Springer series in statistics (Springer, New York, 2009).