2007－CH－74 (7)
2007／5／25

近代ストア主義とメスメール主義の思想的類似性に関する グラフ言語学的分析

○赤間啓之，*三宅真紀，鄭在玲
{akama.h.aa, jung.j.aa}@m.titech.ac.jp, *mmiyake@lang.osaka-u.ac.jp
東京工業大学社会理工学研究科 *大阪大学言語文化研究科

赤間(2001)は、語彙の共起頻度データの因子分析を用い、フランス革命期の思想家、カバニスとメスメールの思想的類似性を論証したが、本発表では、このデータを意味ネットワーク化し、グラフクラスタリングにかけることで、新たに解釈学的妥当性を検証する。

Graph-based Linguistic Analysis on the Ideological Similarity between the Mesmerism and the Modern Stoicism

○Hiroyuki Akama, Maki Miyake*, Jaeyoung Jung
akama@dp.hum.titech.ac.jp
Department of Human System Science, Tokyo Institute of Technology
* Graduate School of Language and Culture, Osaka University

By applying factor analysis to word co-occurrence data, Akama(2001) argued in his previous studies the similarity of thinking between two contemporary thinkers apparently too heterogeneous : Cabanis who represented the modern stoicism in a framework of philosophy of medicine and Mesmer who propounded the theory of animal magnetism as a therapy resembling a sort of hypnotism. Here the same data are submitted to a graph clustering to reexamine the appropriateness of this interpretation.

1 研究の背景

赤間(2001)はかつて、因子分析を用い、近代初頭、晩期啓蒙主義の時代のフランスにおける、ストア主義とメスメリスムの、目に見えない思想的類似性を論証した。近代における「ストア主義」とは、コンディヤック流の感覚論哲学から派生したいわゆる「イデオロジー（観念学）」に見られる傾向であり、その頂点にいるのが、ミシェル・フーコーにより再評価された医学哲学者ピエール＝ジャン＝ジョルジュ・カバニス(1758~1808)である。 一方、メスメリスムとは、同じ時期にパリで一世を風靡したウィーン出身の医師、フランツ＝アントン・

メスメール(1734~1815)の名を取ったものであり、その中核をなすのが、「動物磁気」という名の一種の催眠療法である。詳しくは前稿で論じたが、この両者は一方は個人的禁欲の倫理、他方は集団的癒しの医学をそれぞれ指向していたため、一見まったく無縁な存在と受け取られがちである。カバニスは一時期メスメールと師弟関係の契約書を交わしたこともあったが、その後、メスメールの名も著作もまったく引用することがなく、動物磁気について暗黙のうちに言及したくだりさえ存在していない。

しかしながら、カバニスとメスメールは双方とも、下腹部臓器や神経系の受ける刺激印象の違いに基づく、身体的な狂気（ヒステリー、ヒポコンデリー）を、分析と治療の対象にしていた。さらに根本的に、自然界の一元論という発想において両者には共通したものがあり、程度の差はあるものの、ともに賦活論的ないし汎神論的な傾向を有している。そこで赤間(2001)は、両者を比較するため、ストア主義の代表的著作として

Georges Cabanis, Lettre à M.F. sur les causes premières

（ジョルジュ・カバニス、『第一原因についてのF氏への手紙』）を、さらにメスメリスムの代表的著作として、

F.-A. Mesmer, Mémoires de F.-A. Mesmer, docteur en médecine, sur ses découvertes

（F.-A.メスメール、『医学博士 F.-A.メスメールによる、彼の発見についての論文』）

の二つを選択し、そこで用いられた単語の共起データを因子分析にかけることで、潜在的な意味の構成を検出する作業を行った。

表面的に見ると、カバニスには、思想、観念、内省に関わる「哲学」系のタームが多く、メスメールには、物体、流体、身体に関わる「医学」系のタームが多い。しかし当時は、ヒポクラテスらの伝統でもある古典医学の最後の光輝として、「医学哲学」という、現代の私たちの眼から見ると融合的・折衷的な言説領域が存在していた。「医学哲学」とは「身心相関」の理論とでも言い換えられるものであり、両テーマがどのように個別の言表を形成しながら、時代思潮において統一的な言説を形成しているか、興味の尽きないテーマであると言える。

2 方法

2.1 因子分析

本稿では、かつて行った因子分析の結果と今回新たに同種のデータで行うグラフクラスタリングの結果を比較する。因子分析は、ふたつの異なる思想の底部を流れる潜在的な文脈の類似性を論じるため、複雑な構成概念を的確に個別抽出する手段として有効であることがわかっている。われわれは、まず上記の2文書（以降カバニスはC, メスメールはMと略記する）を、長さを規準化しつつ合併連結し、単語数にして長さ10482個のテキスト（ノイズワードは除去）C&Mを生成した。さらに、C、Mのそれぞれ最頻出名詞上位50位まで計77語のキーワードを抽出したが、それらの名詞の内訳は、C、M双方に共通して50位以内のものが26個、Cのみ50位以内のものが26個、Mのみ50位以内のものが25個である。これら最頻出77語のキーワードは、ほぼ1/3ずつに等分されるのが興味深い。それらキーワードの各々のインスタンスを中心に、一定のウィンドウ幅内で共起するすべての

語をオブザベーションとして出現頻度をカウントした。合併連結するにあたり文書の長さを揃えるので、単語は出現回数 1 回につき単純に 1 個とは数えず、各文書領域の長さ（ノイズワードを取った後の総形態素数）に従って適当な重みをかけている。

ウィンドウ幅 5 の共起データの場合、変数（キーワード数）は 77 個、オブザベーション数（共起語の異なり数）は 2134 個であるが、因子分析（主因子法）の結果を見ると、固有値 1 以上という条件では、15 個の因子が抽出され、それらにバリマックス回転を施したところ、因子負荷行列から（全体潜在文脈-）磁気媒体因子、身体内部感覚因子、（全体潜在文脈-）病的催眠因子を始めとする 15 個の因子の解釈がなされた。

この因子分析の詳しい結果については、http://dl.dp.hum.titech.ac.jp/wiki/?plugin=attach&pcmd=open&file=CA-FA.pdf&refer=FrontPage を参照されたい。

2.2 グラフクラスタリング

ドキュメントの意味ネットワークは、作成の前提として、単語の共起情報を利用する。しかし共起行列を多変量解析に直接かけるのではなく、それを単語が点ノードに、共起関係が辺に見立てられるグラフの隣接行列の形にし、まず、大きな単語間のネットワークを構築する。これがドキュメントの意味ネットワークであり、グラフ理論に関する様々な手法をそれに適用することが可能である。特に、本稿で適用するマルコフクラスター・アルゴリズム(MCL)は、グラフクラスタリングの手法であり、Van Dongen (2000)により提案されたもので、ドキュメントの語彙データの場合、MCL は意味ネットワークをいくつかのコヒーレントなサブグラフに分割し、類似語・同一系統語のグループを一個のクラスター（概念）にまとめることができる。MCL では、グラフ全体が重複のない孤立したハード・クラスターに分割されるまで、random walk に基づくクラスタリング計算を繰り返す。

三宅(2006)は福音書の単語共起データの分析に MCL とそれを拡張した RMCL という手法を用いている。赤間(2006)はすでに、現代言語学の祖、ソシュールの第三回講義の概念ネットワーク分析に MCL 系アルゴリズムを利用し、因子分析の結果との比較を行っている。人文科学の文書データ解析に MCL 系アルゴリズムを利用することの有効性が徐々に示されつつある。

本研究では、単語の頻度をもとにキーワードをあらかじめ選別し、それらとの共起データも共起頻度を採用しているので、そうしたデータ構成に合致した隣接行列を生成する必要がある。グラフクラスタリングの元データとしては、因子分析に利用した共起行列から重みつき隣接行列を生成した。ここで注意すべき点だが、変数となるキーワードはすべて共起語サンプルに入っている、つまり、どのキーワードも他のキーワードと必ず少なくとも 1 回は共起していることである。因子分析の場合、必要なデータ構成からして、変数としてのキーワードと、オブザベーションとしてのキーワードは区別せざるをえない。今回重み付き共起行列からグラフの重み付き隣接行列を作る場合は、両者の区別は非現実的であるので(グラフが分裂してしまうので)、変数内での共起もそこに含めることにした。

キーワード間共起	キーワード対共起語
キーワード対共起語	零行列 O 共起語間共起 なし

図1 キーワード中心共起データ隣接行列

もし変数内の共起を含めなければ、グラフの隣接行列は厳密に 2 部グラフになる。しかし、今回は変数*変数の左上部分行列には隣接関係(しかも重み付き)の値が入る。ただし、データの取り方から、共起語同士の（キーワードを交えない）共起は無視するので、右下の部分行列は 2 部グラフの場合と同様、零行列になる。ここに隣接関係を入れた場合は、隣接重みに閾値を設定しないかぎり、結線率の高さから MCL の計算はサブグラフを生まない。今回は、まず因子分析との結果となるべく条件を揃えて比較するため、閾値を用いることは避けた。キーワードは 77、共起語は 2056 あったため、隣接行列の行数は 2133 になる。

2.3 結果比較

この隣接行列を MCL（自己ループの重みは 1 として計算）にかけたところ、興味深い結果が生じた。16 回で結果はほぼ収束し、15 個のクラスターが生成した。この各クラスターサイズは、{1047, 572, 114, 68, 63, 46, 44, 32, 31, 27, 24, 23, 18, 13, 12}であり、サイズの際立って大きいクラスター(これをコアクラスターと呼ぶ)が 2 個生じたことがわかる。どのクラスターにも変数となるキーワードは所属しており、1~77 とナンバリングしたキーワードのみに注目すると、MCL の結果は、{{action,air,animal,corps,degré,éther,feu,fluide,impression,influence,lumière,magnétisme,matière,mécanisme,modification,mouvement,nerf,ordre,organe,organisation,propriété,sens,sensation,sensibilité,substance},{analogie,besoin,critique,esprit,état,être[01],examen,existence,faculté,fait[01],force,habitude,homme,hypothèse,individu,intelligence,loi,magnétisme-animal,moi[01],morale,moyen,nature,observation,phénomene,point[01],principe,question,raison,rapport,résultat,sentiment,sommeil,source,systeme,temps,univers,volonté},{cause,puissance},{crise,maladie},{effet},{effort},{erreur},{expérience},{idee},{objet},{opinion},{partie},{théorie},{vertu},{vie}}([01]は同綴意義区別用タグ；訳語は上記 URL 参考のこと)であり、キーワードに絞ってもコアクラスターが 2 個生じていることがわかる。

これは、因子分析による変数キーワードのクラスタリングとは様相を異にする。確かに、因子分析が固有値 1 以上という基準で 15 の因子を細かく抽出したのに対し、同型のデータを用いた MCL では、同じく 15 個のクラスターが生成した。しかし、変数となるキーワードの帰属様態は著しく異なったものになっている。因子分析では、それぞれの因子を形成する変数の数は、因子番号順に{19,7,10,6,4,6,4,3,5,3,2,4,1,2,2}であった。一方、MCL クラスターのサイズは、変数=キーワードに絞り他の共起語を省くと、{25, 37, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1}であり、概念の重みは 2 個のコアクラスターに集中している。MCL の場合、大きい 2 つ

の意味文脈を切り分けているものの、因子分析ほどは詳細なテーマ系を抽出できないことになる。

これは、グラフクラスタリングにとって弱点となるかは議論の分かれるところであろう。だが、カバニスとメスメールの固有文脈の厳密な判別を逃れる医学哲学の共通コアを同定するうえで、因子分析のもたらす煩雑なまでの細分化は、逆に「木を見て森を見ず」という解釈上の不都合を生み出す恐れがある。グラフクラスタリングの場合と異なり、因子分析の結果は、それ以上再利用が効かない決定的な最終解であり、原データや計算途中の細部の関係性を再導入し粒度を調整し直すことができない。グラフクラスタリング間の手法比較においてさえ、Pujol(2006)らが言うように、分割された群の数が少ないほうが、意味が単純化されてつかみやすくなり、データ構造の全体にわたって包括的な視点で論じることが可能になると考えられる。

2.4　コアクラスター分析から

じっさい、この 2 つのコアクラスターは解釈がしやすい。結線重みを外して単に次数を計算し次数最大の点を代表ノードはすると、コアクラスターの代表ノードはそれぞれ action((活動);次数 269), home((人間);次数 462)であるが、それぞれメンバーを一瞥しただけで、「身体クラスター」、「精神クラスター」と命名することができる。コアクラスター 1 は corps(精神)、コアクラスター 2 は esprit(身体)という 2 大テーマ語を含むことから、この解釈は妥当なものであると言える。

コアクラスターは、因子分析における各因子の特徴的な部分がマージされている面をもつ。特に第 1 クラスターに帰属するキーワードは、5 因子(第 1,2,6,8,11 因子)に見出されるものに限定され、メスメール系因子の示す意味系列が優勢である。しかし、同時にそれがカバニス系因子の示す意味系列と交わる部分を明確にしている。ここでは身体組織の持つ感覚を表す、メスメールにはほとんど見られないカバニス特徴語(sens:（感覚）頻度 43 対 16、sensibilité(感受性):頻度 19 対 1)がメスメール色の強いクラスターの中で異彩を放っている。図 2 は、身体クラスターの内部結線(結線重み 3 以上に限定)を表した部分グラフであり、C,M というラベルはそれぞれ、その単語がカバニス、メスメールのどちらで多く用いられているかを示している。逆に精神クラスターにおいては、カバニス系の示す意味系列が圧倒的に優勢であるが、magnétisme-animal ((動物磁気)頻度 0 対 12）という、カバニスには決して現われるはずのない単語がそのメンバーになっている。

因子分析の場合、カバニス、メスメールに共通であったり、それぞれ独自に結びついたりする因子が生成した。しかし、それらは構成概念として細分化されすぎており、両者の世界の土台と屋台骨を明示するには至っていない。反対に MCL の場合、コアクラスター 2 個はそれぞれ、物質と精神、あるいは身と心の対立に相当する。これは両者に共通なテーマである身心相関を明示していると言ってよい。しかも注目すべきは、身体クラスター側をメスメールが、精神クラスター側をカバニスが押さえているにもかかわらず、前者ではカバニスを特徴付ける senseibilité (感受性)等のストア主義的用

語が、後者ではメスメールを特徴付ける magnétisme animal（動物磁気）等のメスメール主義的用語が異彩を放っているということである。

3 計算結果の評価の問題

3.1 判別のためのグラフクラスタリング?

さて、本節では、対象の特異性を暫時離れ、グラフクラスタリングをテキスト解釈に用いる本質的な意味と、その結果に対する客観的な評価に関して議論する。

本研究で合併テキストの単語共起データから潜在的情報を自動抽出した際、多変量解析(因子分析)にはひとつの成果が期待されていた。それは、テキストの異なる特徴を判別しながら、両者の交錯する地点で潜在する共通概念を、その概念ネットワーク上の位置(トポス-トピック)情報とともに析出することであった。因子分析の結果は、構成概念の細分化によって、判別という観点からは焦点が明瞭でない一般的構造化を示していると言える。それでは、多変量解析に対してグラフクラスタリングは、どのような出力により、本研究におけるような、異質性判別と同質性発見という二重の期待を満たしうるか。そもそも判別にグラフクラスタリングを用いるという発想には違和感が生じる可能性もあるので、この方法論的視点の先行研究における具体例をまず取り上げたい。

グラフクラスタリングに判別力を見出す契機となったのは、Zakary による有名な Karate Club データであると思われる。これは Karate Club という組織の 34 名のメンバーを点に、互いの人間関係の有無を辺によって表したグラフの形で表現されている。さらに教師データとして、現実に起こった内部分裂の結果がラベル付けされている。Karate Club データは、規模や明瞭さの上で、グラフクラスタリングが実際の派閥構成をシミュレーションできるか試すには好適である。そのため、現在グラフクラスタリング関係の多くの論文で引用され、精度を評価するテストデータとして盛んに利用されている。

単語の共起データをもとにふたりの作者を判別する（実際は判別できない部分を明確にする）という点で、グラフクラスタリングを使うのが適当であるという理由もそこにある。実は、Karate Club データは、多変量解析にかけることができるが、グラフクラスタリングとほぼ同じ精度を示すものとしては、Ward 法にもとづくクラスター分析を除くとほとんど存在しないと言ってよい。(他に最近隣法、最遠隣法、グループ内平均連結法、重心法、メディアン法があるがデンドログラムの形が歪になり、クラスター探索には使えない。)

じつは、本研究で利用する MCL は、Karate Club データに対し判別率 100%を誇っている。しかもデンドログラムを使用しないので、Ward 法と違って、どのレベルで木構造を切るのが最適化かという問題も生じない。さらに、次節で紹介する Modularity Q の最適化をクラスタリングの原理とする他のグラフクラスタリングでは、判別率 100%は期待できないこともわかっている。

3.2 Modularity Q

だが、そもそも分類・クラスタリングや構成概念抽出は、教師なし学習の場合が多

く、計算結果の評価がもともと困難である点も否めない。単語共起データの因子分析結果も、その評価の方法としては、発見された文脈を原文に戻って再確認するより他はなかった。同種のデータのグラフクラスタリングの場合は、一般に modularity Q として知られている評価値が存在する。

Modularity Q とは、同じ条件(点の総数、結線総数)のランダムグラフと比較し、結線分布が各クラスター内にどの程度偏っているかを見ることで、グラフクラスタリングの精度を与える指標である。Newman らの定義によれば

$$Q=\sum_i (e_{ii}-a_i^2),$$

であり、ここで iはクラスターc_iの番号, e_{ii}は、グラフ全体に対するクラスター内部リンクの割合、a_iはc_i内の点がもつ辺数のグラフ全体の辺数に対する割合である。Modularity Q が大きいほど、クラスタリングは精度が高いということが言われている。

MCL の結果は、明確に解釈可能な身体系の単語クラスターと精神系の単語クラスターの大きな 2 大クラスターが出現していた。隣接行列の重みを丸めて次数扱いした結果、Q の値は対角重み 1 で 0.305924、対角重み最大(第 4 節参照)で 0.287782 になった。前者は Karate Club のクラスタリング結果に対する Modularity Q の値 0.3714 に近く、クラスタリングの妥当性を示唆している。

一方、因子分析の結果を因子ごとにクラスターに対応すると解釈し、因子分析の結果にも modularity 値の計算を適用した。すなわち、キーワード・共起語のそれぞれが最高因子得点を記録した因子に対応するクラスターに帰属させ、それを意味ネットワークのクラスタリング結果のひとつと同等のものと解釈した。結果は 0.0295 と非常に低い値となった。

4　重み付きグラフのクラスタリング

最後に本研究で扱ったような重み付き隣接行列の扱いに関し、一言触れておかねばならない点がある。MCL の計算においては、隣接行列にあらかじめ自己参照(自己ループ)を加えてから random walk に入ることが前提になる。重みなしのバイナリーな隣接行列の場合、これは対角成分に 1 を入れることに対応する。しかし、本研究では、合併テキストのサイズを考慮した頻度データから、実数の重み付き隣接行列を作成しているので、対角成分となる自己ループにどれだけの重みを与えるかで MCL の結果に微妙な違いが出る。

Van Dongen らが公開している MCL スクリプトでは、対角要素=1 とする我々のスクリプトと異なり、各点の隣接重みの最大値をその点の自己ループの重みとして与えている。その際キーワードからは、esprit(精神)と source(源)の 2 語がコアクラスターを離脱し、新たに独立したクラスターのハブになる。

これらは、別のハブとかなりの重みで結びつく一方で、ぶらさがりノードや小次数ノードをいくつか周囲に抱えており、大きい重みが自己ループのものとなって、まわりの惑星ノードともども小さくまとまった世界を作るようになると言える。普遍的な意味と特殊な意味を双方纏っていると言ってよいかもしれない。

5　まとめ

本研究では、カバニスとメスメールの代表テキストを用い、単語共起に基づくテキストの特徴抽出について、因子分析などの多変量解析と同様、グラフクラスタリングを用いても期待された結果が出力されることを示した。さらに、因子分析の場合以上に、グラフクラスタリングは、包括的な視点から全体構造を明確に際立たせ、その内部での特異な様相を抽出することができることを明らかにした。さらに、それらの結果に対する客観的な評価について、グラフクラスタリングの側から、Modularity Q という指標を利用するという方法を提案した。本研究のもうひとつの特徴として、テキストの意味ネットワークに辺の重みという形で頻度を導入したことが挙げられる。その場合グラフクラスタリングの側で生じる問題点も明らかにした。

【参考文献】

[1] Hiroyuki Akama, Cabanis ou le crepuscle metaphysique, in Actes du XXVIIe Congres ASPLF, J.Vrin, Paris, p.297~305, 2000

[2] 赤間啓之、ベクトル空間モデルに則った、近代ストア主義とメスメリスムの類似性に関する計量文体論的分析、情報処理学会報告, Vol.2001 No.51 1-8, 2001

[3] 赤間啓之、三宅真紀、鄭在玲、テキスト分析における２部グラフクラスタリングの可能性、電子情報通信学会研究会、言語理解とコミュニケーション研究会、情報処理学会研究報告、2006-NL-174, pp.19~24,2006

[4] Jung, J., Miyake, M., Akama, A., "Recurrent Markov Cluster (RMCL) Algorithm for the Refinement of the Semantic Network", LREC2006, pp.1428~1432,2006

[5] Jung, J., Miyake, M., Akama, A., "Markov Cluster Shortest Path Founded upon the Alibi-breaking Algorithm", CICLing-2006, LNCS 3878, Springer Verlag Berlin Heidelberg, pp55-58, (http://dx.doi.org/10.1007/11671299_6) , 2006

[6] 鄭在玲、三宅真紀、赤間啓之、再帰的なグラフクラスタリングを利用した言語連想データの処理について、人工知能学会大会、2006

[7] 三宅真紀、グラフクラスタリングに基づく共観福音書意味ネットワークの実装、じんもんこん2006、人文科学とコンピュータシンポジウム、pp.161~165、2006

[8] 三宅真紀、鄭在玲、赤間啓之、グラフクラスタリングとパターン分類を併用したストーリー・マップ生成の試み、言語処理学会第12回年次大会(NLP2006)、pp.644-647、2006

[9] Josep M. Pujol, Javier Bejar and Jordi Delgado, Clustering Algorithm for Determining Community Structure in Large Networks, http://www.lsi.upc.edu/~jmpujol/public/papers/spectralCluster.pdf,2006

[10] Van Dongen, S. "Graph Clustering by Flow Simulation". PhD thesis, University of Utrecht, 2000.

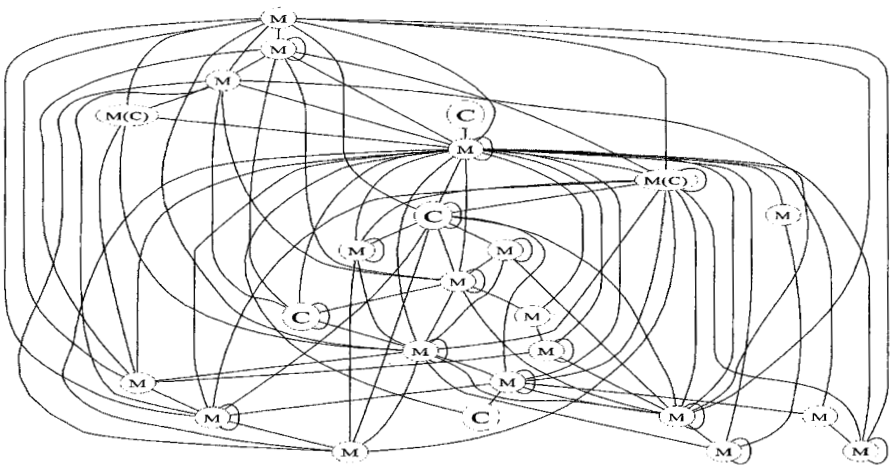

図２　身体クラスターの内部結線