

トピックマップ構造におけるスコープ自動抽出方法の提案

山崎 恭史[†]木村 昌臣[‡]芝浦工業大学院工学研究科[†]芝浦工業大学工学部情報工学科[‡]

1. はじめに

近年、インターネットの普及やメディアの多岐化などにより、人間が生み出す情報量が増加している。そのため、膨大な情報を検索・管理する技術が重要視されており、情報の管理・活用技術としてトピックマップが注目されている。トピックマップは、知識を記号化し、記号化した知識をファイルやWebサイトなどの情報リソースと結び付ける技術である。トピックマップではスコープという仕組みにより、特定の構成要素に対して有効範囲を設定することが出来る。このスコープの設定を手作業で行うとすると非常に手間がかかり、また、トピックマップの作成者がスコープを付与する際に、作成者の主観が影響することが考えられ、トピックマップ内のトピックマップ間で不整合が生じる恐れがある。そのため、スコープを自動的に付与する仕組みが必要であると言える。しかし、調査した限りでは、これに関する研究は行われていない。本研究では、トピックマップの構造からスコープを設定するトピック、及びスコープの自動抽出方法を提案する。

2. トピックマップ

トピックマップは主題と、その主題間の関係を表す関連の集合であり、トピックは主題をコンピュータ上でモデル化したものである。トピックマップでは主題を、「人間が想像出来るもの、言葉に出来るもの、あらゆる手段で表現できるもの」と定義しており、全てのものが主題になるとしている。つまり、全てのものがトピックとなる。関連については、型-インスタンス関係、上位型-下位型関係が規格の中では予め定義されているが、それ以外の関連は自由に定義することが出来る。トピックマップでは、トピックと関連の集合を知識と呼ぶ。また、トピックは、関連に対してどのような立場を取るかという「役割」を持つ。例えば、「著す」という関連において「司馬遼太郎」というトピックは「著者」という役割を持つ。また「司馬遼太郎」の型を「人」とした場合、人のインスタンスとなる。このとき、型である「人」も役割である「著者」もトピックとなる。この型や役割も自由に定義することが出来る。

2.1. スコープ機能

スコープは、構成要素に対して文脈による有効範

Proposal of Method for Automatic Extraction of Scope from Topic Maps Structure

[†]Takashi Yamazaki, [‡]Masaomi Kimura

[†]Graduate School of Shibaura Institute of Technology

[‡]Shibaura Institute of Technology

囲を設定するもので、目的に合わせたフィルタリングに利用される。例えば、スコープを適切に設定することで、トピックマップを言語別や閲覧者別などの目的に合わせたトピックマップを提示することができる。スコープを設定できる構成要素は、トピックの名前を表す「トピック名」、トピックの別名を表す「異形名」、情報リソースとの繋がりを表す「出現」、そして「関連」の4つである。

3. 提案手法

本研究ではトピックマップをネットワークと見なし、ネットワーク分析手法のひとつであるクラスタリングを適用することで、スコープを付与するトピックのクラスタに分類する。

3.1. トピックマップのネットワーク化

トピックをノード、関連をエッジとしてトピックマップをネットワークに変換する。本研究では、型-インスタンス関係におけるインスタンスのトピックをノードとし、型や役割といった要素はトピックとしない。この理由として、本研究では出来る限りシンプルな形式のトピックマップを対象としており、型や役割はネットワークにおいて複数のノードと繋がるハブになる可能性が高く、クラスタリングにおいて大きな影響を及ぼしてしまうためである。

3.2. ファジィクラスタリング

クラスタリングの手法には、ファジィクラスタリングの Fuzzy C-Means 法 (FCM 法) を用いた。ファジィクラスタリングはノードが複数のクラスタに属することを許容する。複数のクラスタに属するノードをファジィノードと呼び、クラスタに属する度合いを、0 から 1 の間の実数値をとるメンバシップ値で表す。各ノードはクラスタごとにメンバシップ値を持ち、その総和は 1 になる。ファジィクラスタリングを用いる理由として、スコープを設定する場合、人間が持つ主観が影響するため、スコープの表現やスコープを付与する集合が多様に存在と考えられる。本研究では、人間がスコープを設定する場合に、視点によるスコープの違いや、同一トピックへのスコープの設定/未設定といった、人間が持つ主観によって発生するスコープの不整合や曖昧性を問題としている。そのため、人間の主観性を考慮するために、曖昧性を持つファジィクラスタリングを用いる。しかし、ファジィクラスタリングの手法である FCM 法ではネットワーク構造ではなく、ノードの位置に対してクラスタリングを行う。そのため、トピックマップのネットワークから、ノードの 2 次

元座標を求める必要がある。そこで本研究では、ネットワーク構造から2次元座標を求める手法として、多次元尺度構成法 (MDS) を用いた。FCM 法には、統計ソフト R のパッケージ e1071 に含まれる cmeans 関数を使用した。

3.1.1. 多次元尺度構成法 (MDS)

ノードの座標を求めるのに計量 MDS を用いた。計量 MDS は、距離データから座標を求める手法である。ノード間の距離データは、全てのエッジの長さを1とするムーアの広さ優先探索のアルゴリズムを用いて求め、計量 MDS は、統計ソフト R のパッケージ stats に含まれる cmdscale 関数を使用した。

3.1.2. ベイズ情報基準量によるクラスタ数決定

FCM 法では、クラスタ数を予め決めておく必要がある。本研究ではクラスタ数を決定するのに、ベイズ情報基準量から決定する手法³⁾を用いた。

4. 実験

実験は、提案手法による分類を評価することを目的とする。実験として、江戸川乱歩に関して知識を持たない被験者 10 人に対して江戸川乱歩に関するトピックマップ¹⁾ (図 1) に含まれる 11 個のトピックを、ウェブサイトを参考にして集合に分けてもらい、集合に名前を付与してもらう作業を実施した。また、集合の記述にはベン図を用いた。被験者にはトピックのみを提示し、トピック間の関連情報は提示していない。この実験により得られた分類と、提案手法による分類を比較し、評価を行った。

5. 結果・考察

提案手法による分類は表 1 の様になった。ノードは、ひとつのクラスタにしか属さないノードと、複数のクラスタに属するファジィノードに大別することが出来る。つまり、「明智小五郎」、「怪人二十面相 (人)」は、クラスタ 1 または 2 に属する可能性があり、「江戸川乱歩」、「名張市」は、全てのクラスタに属する可能性がある。またファジィノードと見なす条件として、ノードの各クラスタへのメンバシップ値の統計を取ったところ、過半数以上が 0.1 未満であることから、本研究では、メンバシップ値が 0.1 以上であれば、クラスタに属しているとした。

江戸川乱歩のトピックマップにある型や役割によらない集合は表 2 の様になった。実験結果の「同シリーズ」、「処女作」の集合は、提案手法のクラスタ 2、クラスタ 3 とそれぞれ一致しており、「小五郎

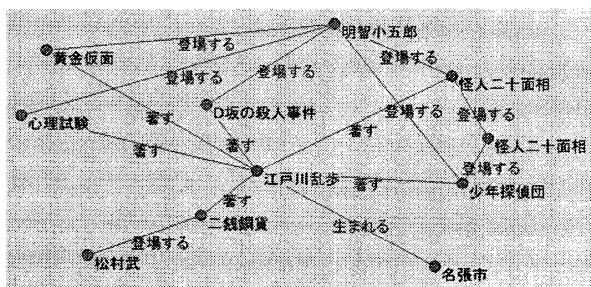


図 1: 江戸川乱歩のトピックマップ

表 1: 提案手法による分類

ノードの種類	クラスタ番号	クラスタに含まれるノード
一つのクラスタに属するノード	クラスタ 1	黄金仮面, 心理試験, D坂の殺人事件
	クラスタ 2	怪人二十面相 (小説), 少年探偵団
	クラスタ 3	松村武, 二銭銅貨
複数のクラスタに属するファジィノード	クラスタ 1, 2	明智小五郎, 怪人二十面相 (人)
	クラスタ 1, 2, 3	江戸川乱歩, 名張市

表 2: 型・役割に基づいた集合以外に見られた集合

集合名	集合に含まれる要素
同シリーズ	少年探偵団, 怪人二十面相 (小説)
処女作	松村武, 二銭銅貨
怪人二十面相	怪人二十面相 (人), 怪人二十面相 (小説)
新青年	二銭銅貨, 心理試験, D坂の殺人事件
小五郎大人時代	明智小五郎, 心理試験, 黄金仮面, D坂の殺人事件
少年倶楽部	明智小五郎, 怪人二十面相 (人), 少年探偵団, 怪人二十面相 (小説)
怪人二十面相の登場人物	明智小五郎, 怪人二十面相 (人)
二銭銅貨の登場人物	明智小五郎, 松村武

大人時代」、「少年倶楽部」では、クラスタ 1、クラスタ 2 にファジィノードである明智小五郎や怪人二十面相 (人) が含まれたクラスタに一致している。このことから、人間がスコープを設定する場合には曖昧性が生じ、提案手法では、その曖昧性を考慮することが出来ているといえる。また、提案手法は、人間の知識による分類が出来ているといえる。提案手法において、「江戸川乱歩」、「名張市」が全てのクラスタに対してファジィになっているが、前者はベン図において他の集合を含む大きな集合の中の要素となっていることから、他の集合の必要条件になっていると考えられ、ファジィであると言える。後者は、ノード間の距離にムーアの広さ優先探索アルゴリズムを用いていることが影響していると考えられ、関連を考慮した距離の設定が必要であると考えられる。

6. まとめ

本研究では、トピックマップの構造から人間が持つ主観によって生じる曖昧性を考慮したスコープの自動抽出・設定を目標としており、本論では、実験により人間の主観による曖昧性がスコープの設定において生じることを明らかにし、提案手法がその分類に有効であることを示した。現段階では、分類は出来たが、設定するスコープは人間が判断しなければならない。今後の課題として、提案手法の結果に基づく適切なスコープの自動的な抽出や、関連を考慮した距離設定、分類精度の向上が挙げられる。

参考文献

- 1) 内藤求: トピックマップ入門, 東京電機大学出版局, (2006).
- 2) Lars Marius Garshol: A Theory of Scope, TMRA2007, (2007).
- 3) 木村昌臣, 大倉典子, 土屋文人: 糖尿病用薬注入デバイスの使用安全性に関する検討, 電子情報通信学会総合大会講演論文集, Vol.2006 年_基礎・境界, p316 (2006).