重み付きグラフを用いた学習教材の構造化及びクラスタリング*

1G－5

豊田 規人 藤井誠 冨士隆† 三枝 武男‡
(学習情報通信システム研究所) (北海道情報大学)

1. 前書き

教科コース型（教え込み型）のCAIを作成するためには、何らかの意味で、学習教材の構造化が必要となる[4]。更にコンピューター上に実装するためには、単純なアルゴリズム化（自動化）が必要となる。我々は今回、構造化及びクラスタリングを同時に実行する一つの手法を提案する。そこでは、従来のISM法[6][5]などで用いられた２値行列に基ずく方法とは違い、任意の整数値をその隣接行列要素として持つ場合を考える。従って、ブール代数に基付くシステマチックな方法が、一切使えない。更に可到達性をまったく仮定しない。詰まり、関係の推移性仮定しないという点で、従来より一般に広いクラスに適応可能であると期待できる。又、この方法では、ISM法などで問題となった強連結グラフの場合においても良い結果が見られる。ここではそのアルゴリズムを与えるとともに、それに対する若干の評価を与え、更にクラスタリングの大局的構造化に対する展望を試みる[2][3]。

2. 構造化及びクラスタリングアルゴリズム

我々の構造化&クラスタリングアルゴリズムの基本的アイデアを与える。まず、

（１）各N個の教材項目に対しその関連の大きさに応じて$1 \sim n$の整数値を付与して隣接行列$A = [a^i_j]$を作る。

クラスタリングは隣接行列の空間上で行う。そして列ベクトルと行ベクトルを以下で与えられる類似度によって並び替える。

（２）列ベクトルに対しては、類似度

$$R^{ij} = \frac{\sum_k a^i_k a^j_k}{\sum_k \left\{ (a^i_k)^2 + (a^j_k)^2 - a^i_k a^j_k \right\}}$$

の大きい順に並び替える。

（３）行ベクトルに対しては、上で与えられた水平方向の並びに対して、各項目が、どれだけ類似してるかの順に並び変える。そこで、以下のポテンシャル関数[1]の大きい順に並び替える。

$$V(k) \equiv \vec{n} \cdot \vec{a}_k = \sum_{i=1}^{N} n^{N-i} a^i_k.$$

（４）クラスタリング: 行列要素が囲碁的に連結している部分をひとつのクラスターと定義する。

（５）クラスタリングの中の構造化。

（５－１）各項目のレヴェルを

$$L_i = \sum_{k, a^i_k \neq 0} 1 - \sum_{k, a^k_i \neq 0} 1$$

の大きい順に付与する。

（５－２）Aの連結性に従って、リンク付けをする。ここで行と列において、異なった類似度を用いたことにより、一つの項目が一般に、複数のクラスターに属することになる。この意味で通常のクラスタリングと異なる。（このことは、実は我々の目的にとっては、メリットである。）

（６）全体構造：各クラスターに属する共通の項目を、それぞれパッチングすることによって得られる。

3. BPRへの適用とその評価

ここではネットワーク的な構造がより顕著（グラフ理論的にいうと強連結グラフを複雑に含むもの）と思われ、又クラスタリングの効果が伺えると思われる２７項目から成るビジネスプロセスリエンジニアリング(BPR)を例に実行する。このように全体がかなり複雑な構造をしていても程良い規模にクラスタリング可能（10種類のクラスター）であることが分かる。図２は比較的に複雑なクラスター内の構造化の例である。ISM法（図１）と比べて、比較的階層構造が見易くなっている。また、レベルの数が増えていると言う点

---

*Clustering and Structuring of Instructual Materials based on Weighted Graph

†Norihito TOYOTA, Makoto FUJII, Takasi FUJI, Softwaer Research Laboratory, 45 Nisi-Nopporo Ebetu City 069 Japan

‡Takeo SAEGUSA, Hokkaido Information University, 59-2 Nisi-Nopporo Ebetu City 069 Japan

においてもISM法では同一視されるしかなかった強連結グラフに対しても対処でき得ることを示している。

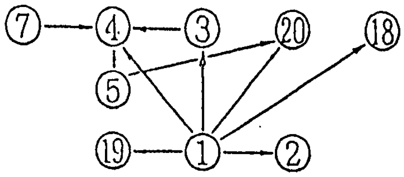

図1. Structuring in a cluster ISM Method.

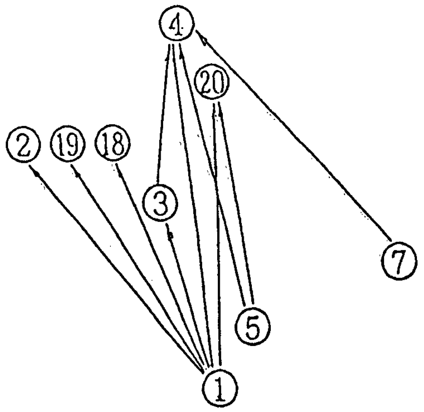

図2. Structuring in a cluster Our Method.

この様子を以下で導入される構造化インデックス（SI）[3] で定量的に見てみる。

$$SI = \frac{\text{有効相対高度の和} + \text{レヴェルの種類の数}}{\text{クラスター内のノード数}}$$

上式で、有効相対高度は、$i \rightarrow j$が存在する時、jとiのレベルの差で定義される。表1で示されているようにすべてのクラスターにおいて、SI値は、ISM法より我々の手法の方が大きく成っている。詰まり、我々の手法の方が構造化が旨く言っていることを暗示している。

4. むすび

本論文では、階層的な構造よりも、ネットワーク的な構造がより顕著と思われる学習教材に対し、構造化及びクラスタリングを同時に実行できる一つのアルゴリズムを提案した。この方法は、従来のISM法などで用いられた2値行列に基ずく方法とは違い、任意の整数値をその隣接行列要素として持つ場合に適用でき、更に、関係の推移性を仮定しない広いクラスに適応可能である。

実際に、BPRの例では有効に作用することがほぼ実証された。特にそのクラスタリングの内部の構造化に関しては、従来よく使用されているISM法より一般にかなり良い（レベルの付け方、及びそれに起因する逆行リンクの減少）構造化が期待できる。特に、ISM法などで問題となった強連結グラフの場合においても良い結果が見られる。これらは、又、定量的にも示されている（SI値）。また、（6）による全体構造制作時においても、（ここでは取り上げれ得なかったが）赤堀の最適解 [4] にコンパラ結果が得られている [3]。

参考文献

[1] 千村 浩靖、佐藤 隆博、“ISM 教材構造化法における教材要素の配列アルゴリズム”、電子情報通信学会 教育技術研究会技報, ET80-9, 5 ~ 10 (1980)

[2] 豊田 規人、藤井 誠、冨士 隆、三枝 武男、“学習教材の構造化及びクラスタリング” 平成7年電気学会全国大会 (1995)

[3] 豊田 規人、藤井 誠、冨士 隆、三枝 武男、“学習教材の構造化及び系列化” 信学技報 Vol.95 No.95 ET95-40(1995)

[4] 赤堀 侃司、“教授設計における学習課題の階層構造表示法”、CAI 学会誌、7, 3, 99 ~ 107 (1900)

[5] 佐藤 隆博、ISM 構造学習法, 明治図書、東京（1987)

[6] J. N. Warfield, “Binary Matrices in System Modeling”, IEEE Trans. Syst., Man & Cybern., SNC-3-5 441 ~ 449 (1973)

cluster	A	B	C	D	E	F	G	H	I	J
Size of cluster	26	19	5	4	13	9	9	6	8	6
Proposed Method	49/9	24/7	4/3	4/3	7/3	11/4	11/4	2	13/5	4/3
ISM-Method	12/9	0	1	1	4/3	1	1/4	4/3	3/5	1

Table 1: SI-values in ISM and our method.