

ネットワーク情報から社会の新しいニーズを発見する 大衆心理分析システム

4 M-7

野村勇治 砂山渡 大澤幸生 谷内田正彦

大阪大学大学院基礎工学研究科

1 概要

社会の新しいニーズを把握することは、マーケティングや政策決定において重要である。本研究では、インターネットの各ページのアクセス回数や、文書データベース中の分野毎の文書数などといった大衆の興味に関する部分的な情報から、未知なる新しい大衆の心理を発見する手法を提案し、その性能を実験評価する。

2 システムの概要

この研究は、我々が作成した、確率ネットワーク上の事象間の依存を計算しそれを説明する未知事象を出力するシステム [1] を用いる。このシステムは、知識ベース中の各事象の生起確率が別々に（あらゆる組合せの結合確率ではなく）与えられた時、世界の未知の構成要素である新しい仮説（未知事象）を生成するものである。具体的には、複数の事象の実際の（知識ベース中で与えられた一部の）結合生起確率と各事象が独立と仮定した時の結合確率値との差を元に、事象間の依存（同時に起こりやすいと判断した事象間に正の依存、同時に起こりにくいと判断した事象間に負の依存）を計算し、それらを説明する未知事象を出力とする。

また、このシステムは、事象の生起確率データが多少抜けていても、それを補間して計算を行うことができる。

我々は、このシステムを用いて大衆の興味の対象間の依存を計算することにより、大衆の興味の傾向を示唆することができると考えている。例えば、100人の母集団がいたとして、20人が野茂に興味を持ち、10人がイチローに興味を持っているとする。野茂とイチローそれぞれに興味を持つ事象が独立事象だとすると、野茂とイチローの両方に興味を持つ人は2人ということになる。もし、実際の人数がそれよりも多いと、野茂とイチロー

は同時に興味を持たれやすいことが予想でき、実際の人数が少ないと、同時に興味を持たれにくいことが予想できる。

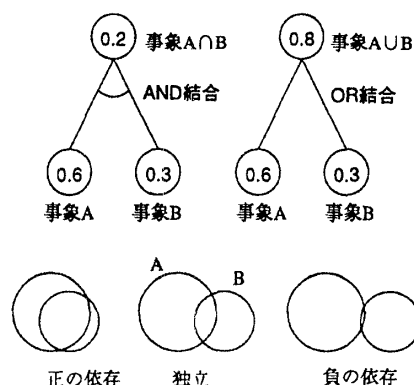


図1: 確率ネットワークと正の依存・負の依存

3 実験例

ここでは、人工知能学会誌第11巻第3号に付録されたCD-ROMの論文データベースを用いて行った実験について述べる。

上記データベースでは、397件の論文が階層的に分類されている。そのデータベースから、各分類枠に属する論文の数と、分類枠間の階層構造を取り出し、分類枠をノードとし、そこに属する論文数の全論文数(397件)に対する割合をノードの確率値とし、分類前のノードを分類後のノードのOR結合とする確率ネットワークを作り、前述のシステムによりノード間の依存を計算した。

このデータベースでは、分類基準として「知識表現技術による分類」「推論技術による分類」「知識獲得技術による分類」「応用による分類」の4つがあり、全論文に対し、これらの基準を次々に適用して、より細かく分類するようになっている。例えば、全論文に対し「知識表現技術による分類」を行った後、それぞれの分類枠の論文の集合に対し「推論技術による分類」が行われている。また、どの分類枠にも当てはまらないものに1つの特別な分類枠を用意している。例えば、「応用による分

Discovering Hidden Social Mind from
Networked Information

Yuji NOMURA, Wataru SUNAYAMA, Yukio OHSAWA
and Masahiko YACHIDA

Graduate School of Engineering Science, OSAKA univ.
Toyonaka, Osaka 560, Japan

e-mail:nomura@yachi-lab. sys. es. osaka-u. ac. jp

類」をした結果「応用では分類できないもの」という分類枠がある。

計算の結果、次の組で強い正の依存が出た(図2, 3)。

- 「知識表現技術」が「制約」のものうち、「仮説推論・アブダクション」と「制約推論・制約充足」
- 「応用では分類できないもの」で「推論技術」が「非単調推論」のものうち、「述語論理・命題論理」と「非単調論理」
- 「応用では分類できないもの」で「推論技術」が「制約推論・制約充足」のものうち、「制約」と「述語論理・命題論理」

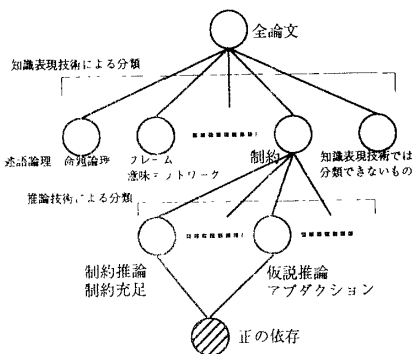


図 2: 実験結果 1

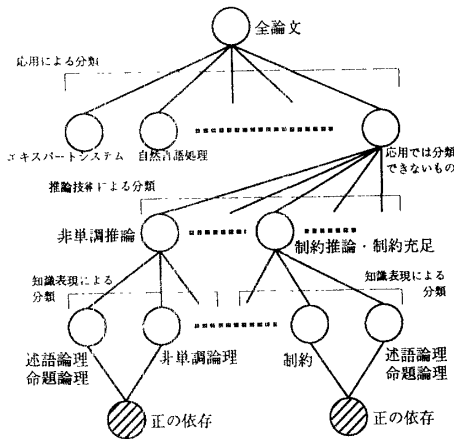


図 3: 実験結果 2

また、次の組で強い負の依存が出た(図4)。

- 「推論技術による分類」のうち、「演繹推論, プロダクションシステム」と「推論技術では分類できないもの」

「仮説推論・アブダクション」と「制約推論・制約充足」との間に正の依存が出たということは、仮説推論が

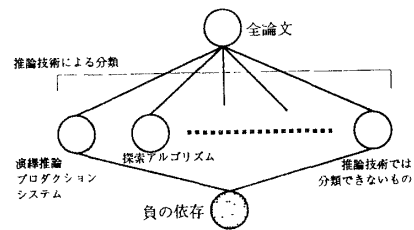


図 4: 実験結果 3

制約充足問題であることが多いことを考えると、納得できる。

「述語論理・命題論理」と「非単調論理」との間の正の依存や、「制約」と「述語論理・命題論理」との間の正の依存からは、非単調推論や制約推論は「述語論理・命題論理」に「非単調論理」や「制約」を加えたものであり、「述語論理・命題論理」が推論には欠かせないことが推測され、興味深い。

負の依存が出た2つの分類枠は、それぞれに属する論文が排反であることが明確であり、負の依存が出るのは当然である。その2つは、それぞれ多数の論文が属していたため、特に強い負の依存が出た。

4 まとめ

本手法の利点は、入力として完全な情報を必要としないことである。大衆の興味を分析を多変量解析で行う場合、データとして被験者の各人が何と何とに興味があるといった、アンケートデータのような完全な情報が必要になる。それに対し本手法は、何に興味がある人が何割、という確率情報を元に解析することができる。さらに、多少確率データが不足していても解析をすることができる。それにより、複数の母集団から採取したデータを混合した解析も可能になる。

今後の課題として、インターネットから効率的に大衆の興味に関する確率データを得る方法を探究することと、確率ネットワークの作成の方法を改良することがある。

参考文献

- [1] 砂山渡, 大澤幸生, 谷内田正彦: 事象毎の生起確率からの未知原因推定による大衆心理の推定, 人工知能全国大会(第11回)論文集, pp. 563-564 (1997).
- [2] 辻 新六, 有馬 昌宏: アンケート調査の方法, 朝倉書店 (1987).
- [3] 浅井晃: 調査の技術, 日科技連出版社 (1987).