

影響変数を伴う関係データの分析と応用についての研究

博士（情報学）

金城敬太

総合研究大学院大学複合科学研究科

情報学専攻

2009 年度

本論文は総合研究大学院大学複合科学研究科情報学専攻に
博士（情報学）授与の要件として提出した博士論文である。

審査委員会

市瀬龍太郎（主査） 総合研究大学院大学/国立情報学研究所
相澤彰子 国立情報学研究所
井上克己 総合研究大学院大学/国立情報学研究所
大山敬三 総合研究大学院大学/国立情報学研究所
小林哲郎 総合研究大学院大学/国立情報学研究所
孫媛 総合研究大学院大学/国立情報学研究所

Research on analysis of relational data
with effected variables and its application

Keita Kinjo
Graduate University for Advanced Studies,
School of Multidisciplinary Sciences

2009

A dissertation submitted to the Department of Informatics,
School of Multidisciplinary Sciences,
The Graduate University for Advanced Studies(SOKENDAI)
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Advisory Committee

Ryutaro Ichise	National Institute of Informatics/ The Graduate University for Advanced Studies
Akiko Aizawa	National Institute of Informatics
Katsumi Inoue	National Institute of Informatics/ The Graduate University for Advanced Studies
Keizo Oyama	National Institute of Informatics/ The Graduate University for Advanced Studies
Tetsuro Kobayashi	National Institute of Informatics/ The Graduate University for Advanced Studies
Yuan Sun	National Institute of Informatics/ The Graduate University for Advanced Studies

概要

今日、ウェブやソーシャルネットワークなどにおいて、さまざまな「関係」が注目されている。

関係は、実世界の実体をノードとして、ノード同士をつなぐリンクで表現することができる。すなわち、関係は二つのノードとリンクをもって定義される。ただし、関係は単独で存在しているのではなく、他のノードやリンクに付随する変数（これを「内部」変数とよぶ）や、関係によって構成されるネットワーク全体に付随する変数（これを「外部」変数とよぶ）に対して影響をあたえている。ある会社組織でいえば、構成員同士のつながりがツリー状かネットワーク状かといった違いにより個々の人の生産性（内部）が異なり、さらに組織全体の生産活動（外部）も異なる。このような関係の内部もしくは外部にある変数を、本研究では「影響変数」と呼ぶ。この変数は既知であり観察可能なものとする。

本研究では、関係と影響変数との関連について分析を試み、それを人や集団に関する課題に応用して有効性を確認した。人や集団に関連した具体的な課題としては、A.既知の関係と影響変数が既知の関係を通じて相互に作用しているかという、仮説の検証、B.明示的に関係が与えられていない状況のもとで、影響変数を説明する関係を獲得、C.関係と影響変数が動的に変化している状況で影響変数の変化を説明する関係を抽出、という3つが想定される。

本研究ではまず影響変数を伴う関係データに関する問題群を整理して、上記の人や集団に関する課題を位置づけ、各課題に対する具体的な問題と解法を検討した。

Aの課題では、商品に対する信頼と新聞や友人などの情報源に対する信頼を影響変数として、これらが「XはYの情報源である」という関係を通じて依存していることを検証した。統計的な検定を可能にするためにネットワーク自己相関モデルを利用した。その結果、商品に対する信頼に情報源に対する信頼が寄与していることや、情報源同士の信頼も、互いに情報源であるという関係に基づいて相互に依存していることがわかった。以上により、関係と影響変数の相互作用について統計的に検証することができた。この具体的な課題を通じて、本研究で利用した手法が関係の影響の検証に有効であることが示せた。

つぎにBの課題については、熟達者によるチェロ演奏のタスクを影響変数として与え、それを実現する筋肉の協調関係を抽出する問題を扱った。この課題でははじめに、筋肉の動きの特徴を抽出するために、時系列モデルと情報量規準を利用した筋電波形のセグメンテーションを行った。次に抽出された筋肉の動きの協調関係（時区間関係）を扱うために帰納論理プログラミングを利用した。実際に獲得した協調関係は、チェロの演奏に関する既存の知見とも一致した結果となった。これにより、関係が明確でないものから、背景知識を用いて新たに影響変数を説明する関係の抽出する手法を示すことができた。その際、関係に関する制約などが事前にある場合は、帰納論理プログラミングを用いた本手法が有

効であることがわかった。また派生的に時系列データのセグメント方法も開発しその有効性を確認した。

最後に C の課題については、不況の際の景気認識を影響変数、集団同士でのクチコミを関係として、景気認識の変化に対応したクチコミ関係の抽出を行った。方法としては景気認識が大きく変化した時点を特定し、次に帰納論理プログラミングを利用してリンクやノードの時間的な依存関係を扱う背景知識を導入したうえで、大きく変化した時点前後の関係の差分を抽出した。その結果、どの時点でクチコミが大きく変化をしているのか、変化の前後でどのような層の間でコミュニケーションの変化が起きたかといったことがわかった。これらは既存の手法では得られなかった知見である。これにより、関係と影響変数が動的に変化するなかで、影響変数に対応した関係をみつけるという新たな問題を解くことができた。

3つの課題を通じて、それぞれで提示した方法が、A.関係の影響変数への効果の検証や、B.影響変数に応じて未知の関係を獲得すること、そして C.関係と影響変数が動的に変化する、といった人や社会に共通みられる一連の課題に対して有効であることがわかった。

以上を通じた本研究の意義は、影響変数を伴う関係についての統一的な問題の枠組みを設定して人や集団における主要な課題に対して適用可能な方法を提案し、実験や実データの分析をすることによって、これまでは獲得が困難であった知見が得られるような方法を示したことにある。

Abstract

Today, with the popularity of web and social networks, various “relations” are attracting attention.

Relations are made from nodes and links. These relations do not exist independently. They affect the inside variables that are being connected by the relation and variables that are outside this relation. For example, in an organization, the effect on each member(inside) and the productive activity of the organization as a whole(outside) differ with the structure of the organization. We call those variables “effected variables”. In this research, we conduct an applied analysis of the effects these relations have.

The 3 main issues of the effects of the relations in real world are :

A. verification of the effect of the relations, investigating whether the relations do actually have an effect,

B. acquisition of unknown relations that correspond to the target effects from situation where relations are not given,

C. characteristic extraction in cases where relations and the effects are dynamically changing.

These can be considered as a common problem of analyzing the related data with effected variables. Therefore, we analyzed the issues and the domain of problems concerning the related data with effected variables each issue is responsible for. Furthermore, we have proposed concrete solution for each issue.

For issue A, we have specifically focused on the interdependent relationship between the trust of products and that of information source as effected variables. In this case, we discovered that the trust of information source has a large influence on that of the products, and the trust of information source depended on the relation between information sources. To verify the effect, we have used the network autocorrelation model. As a result, we have statistically verified the internal effects that the relations have.

For issue B, we have selected the issue of extracting the cooperative relation between muscles when cello experts are performing a given task as effected variables, and successfully extracted the rules as relations. The extracted result matched the existing knowledge. In this process we have adopted Inductive Logic Programming to deal with temporal-interval relation between muscle movements. Finally, we successfully extracted new relationships from unclear structures using background knowledge from the effected variables.

As for the last issue C, we analyzed the change in the word of mouth concerning the economy during recession as relations. As a result, we were able to identify at what point the opinion changes and in which demographic group change of communication occurred before and after the opinion shift. We have used the background knowledge, and also adopted Inductive Logic Programming to deal with temporal relation. So, we have redefined the issue C to a problem of finding common rules between dynamically changing relations and effected variables, and successfully solved it.

As previously stated, the contribution of this research is that we redefined the issues concerning the main effects of relations, and through analysis of concrete examples we have proposed new solutions to each issue and acquired new knowledge.

目次

1 章 序論	1
1.1 研究の目的	2
1.2 問題の種類と形式化	8
1.3 本論文の課題の位置づけとアプローチ	18
1.4 本論文の成果	21
1.5 本論文の構成	22
2 章 商品および商品の情報源に対する信頼の統計的ネットワークモデル	23
2.1 背景と問題設定	24
2.2 構成概念の定義	25
2.3 信頼のネットワーク仮説	26
2.4 検証	32
2.5 関連研究	39
2.6 本章のまとめ	40
3 章 帰納推論による時系列データからの関係構造の抽出	42
3.1 背景と問題設定	43
3.2 提案手法	44
3.3 実験	51
3.4 関連研究	57
3.5 本章のまとめ	58
4 章 調査データに基づく社会構造変化の抽出	60
4.1 背景と問題設定	61
4.2 提案手法	62
4.3 実験	69
4.4 関連研究	78
4.5 本章のまとめ	80
5 章 考察	82
6 章 結論	86
謝辞	87
参考文献	88
研究業績	95
付録	97

図表索引

図 1-1	関係のもつ主要な作用の概念図	3
図 1-2	関係とその表現	7
表 1-1	関係に関する諸問題の構造	9
図 1-3	関係に関する諸問題の構造	10
表 1-2	諸問題に関する解法とその特徴	17
図 1-4	本研究の課題とそのアプローチ	20
図 2-1	信頼のネットワーク仮説	29
表 2-1	被験者の構成	32
表 2-2	仮説と調査項目	33
図 2-2	各食品の信頼	34
図 2-3	食についての情報源に対する信頼	35
図 2-4	情報源同士の情報源としての依存関係	35
図 2-5	食パンにおけるネットワーク自己相関	36
表 2-6	各食品でのネットワーク効果	36
図 2-7	商品（食パン）への影響	37
図 2-8	各情報源の全体への影響	37
図 3-1	提案手法	45
図 3-2	再起分割推定アルゴリズム	47
表 3-1	テスト結果	49
図 3-3	テスト結果図	49
表 3-2	時区間論理	51
表 3-3	各筋肉のクラスタ数	52
表 3-4	前腕筋の KL 値	52
表 3-5	記号系列	53
図 3-4	筋電波形と動作図	54
図 3-5	該当する楽譜部分	56
図 4-1	提案手法の概要	64
図 4-2	本研究の一貫性制約の事例	69
図 4-3	シミュレーションプロセス	71
図 4-4	認知数の時系列変化	72
表 4-1	特定結果	72
図 4-5	シミュレーションプロセス 2	73
図 4-6	景気悪化認識率の推移	75
表 5	本研究における各問題の新規性と今後の発展	85

1. 序論

1.1 で本研究の目的をのべる。まず本研究の研究課題について記述し、それを解く意義をのべる。こうした課題が着目されるようになった背景の説明や用語の定義についてもここで行う。

1.2 では、本研究に関連する機械学習や統計の問題とその解法の整理を行う。

1.3 で1.1でのべた研究課題を1.2で整理したなかに位置づけたうえで本研究のアプローチを提示する。

1.4 では研究の結果得られた成果をまとめ1.5では論文の構成についてのべる。

1.1 研究の目的

1.1.1 研究課題

今日、さまざまな分野で「関係」が注目されている。この理由には、単純なシミュレーションでこれまで社会的に観測されてきたネットワークのベキ則などの関係の構造が明らかにされるようになってきたこと、ソーシャルネットワークサービス(SNS)やブログの発達により関係についての大規模なデータが手に入り易くなったことが挙げられる。本研究でもこの関係に着目する。

本研究でいう「関係」とは、ある実体と実体同士のつながりのことである。たとえば、人同士のつながりであったり、運動であればある動きと別の動きの連携、商品とその消費者といった異種の実体同士のというつながりもある。数学的には、同じ集合同士もしくは別の集合同士の直積の部分集合である。なお、このような関係について調査などを利用し、世界の事実を集めた集合のことを本研究では関係データと呼ぶ。また、この集合について実世界にあるものをより強調して示す場合ネットワークとも呼ぶ。

このような関係を扱うデータについては、その分析方法も着目されており、すでにくつかの技術が存在している。古くは社会ネットワーク分析という研究があった。近年では関係データマイニングの研究開発が進んでいる。詳細については後述するが、1990年代は関係データの分析をする帰納論理プログラミング(ILP) [Muggleton 94], 2000年ごろからは、グラフマイニング[Washio 03]をはじめとしてリンクマイニングといった技術も数多く開発されている。こうした流れのなか、確率的な関係データのモデリング技術もすすんでおり、近い将来これらネットワークやグラフ、ILPのモデリング技術が統一化される可能性もある[Domingos 02;Mooney 04]。

こうしたなか本研究で扱うのは、関係の内部もしくは外部にある変数に対して関係が影響しているという事象を分析することである。関係というものは、それ単独で存在しているのではなく、その関係によって「外部、もしくは内部」に対して目的や指標といった変数が存在しており、これらに影響している場合が多い。

ここで、この外部や内部に限らず関係によって影響をうける変数のことを「影響変数」と本論文では定義する。

「外部」の影響変数とは、実体とその実体同士のつながりである関係を合わせた要素全体に付随している変数である。

「内部」の影響変数とは、実体とある関係が存在した場合に、実体のもつ属性値や、ある関係以外の関係などの実体と関係の集合のなかにすでに含まれる変数のことをさす。

たとえば、会社組織を一つの人と人同士の関係の集合と考えると、このような組織がもつ関係にとって外部の変数とは、全体としての売上や、その組織の株価などをさす。こうした売上は、組織がもつその関係によって影響を受ける。またこの組織がもつ内部の変数

には、個々の人がもつ生産性などの組織内部の実体ももつ属性などが該当する。このような個々ももつ変数も関係から影響を受ける。

本研究は、このような影響変数をともなう関係データを分析するという視点から以下で取り上げる課題にアプローチする応用研究である。こうした視点からの問題設定やその応用はあまり確立していなかったため本研究で取り上げた。

本研究では関係に関する応用問題として、人や社会に関する課題を取り上げる。関係がもたらす影響は、社会や人間の活動のさまざまな場面でみうけられる。

特に、人や集団における応用を考えたときに重要な課題は

- A. 関係の影響変数に対する影響の検証,
 - B. 影響変数を説明する関係の獲得,
 - C. 影響変数と関係が動的に変化するなかで、影響変数を説明する関係を抽出
- といった3つがある。具体的な例をあげよう。

『ある企業組織がある。この組織では、目的を達成するために役割や上下関係などの関係を形成しており、個々の人の活動に対してこの関係が影響をあたえている。たとえば、あるチーム内で友好的コミュニケーションや連携がとれていれば、それぞれの人の問題発生時のカバーもスムーズになり、個々の生産性は向上する。また、個々の活動の結果を集めた、企業全体としての売上や、生産性などにも関係が影響を与えている。

ただし、このような、関係はあらかじめ存在したのではなく、ある目的を達成するために未知の状態から形成（獲得）されてきたものである。さらに、この関係や企業全体がもつ売上などの指標は時間がたつにつれ変化している。』

こうした状況下で3つの問いを考える。

- A. 関係が個々の人（もしくは全体の売上）に対してどのような影響を与えているか。
- B. ある目的を達成するためには、人同士がどのような関係であればいいのか。

ただし、関係には先輩後輩関係のように、年齢などの属性により関係の形成可能か否かの“制約やルール”がある。

- C. 人同士の関係が動的に変化するなかで、売上の変化を説明する関係をしらべるにはどうすればいいか。この際、人のもつ年齢や、性別といった属性も考慮する必要があるほか、ある社員とある社員の間にもどのような関係（上司一部下つながりなのか、友人つながりなのか）が存在しているかといった関係の属性や種類についても考える必要がある。

この3つの問題は、1.1.2でのべるようにさまざまな分野で見受けられる重要な問題である。この問題を解くことが本研究の課題である。

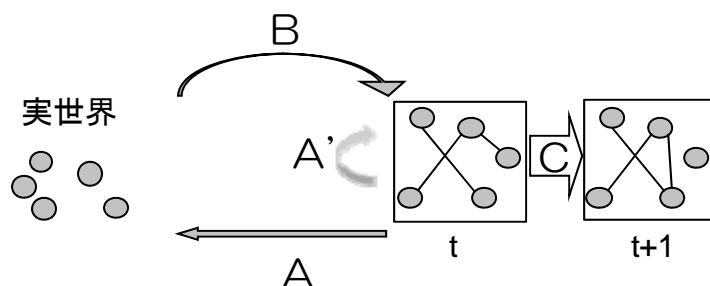


図 1 - 1. 人や集団に関する関係がもつ影響についての概念図

この3つの問題を他の分野でも対応づけ可能にするために再度より一般化した表現をすると

- A. 関係がもつ内部もしくは外部の変数に対する影響について検証すること。
(例：個々の社員に関係が影響しているか)
- B. 明示的な関係が存在しない場合に影響変数（例：生産性など）があるなかでそれを説明する関係を抽出すること。
- C. 時系列で変化する関係と影響変数があるなかで、影響変数を説明する関係を抽出すること。（例：売上と組織）

になる(図 1 - 1). 上記の3つの問題をみてもわかるようにいずれも A は内部もしくは外部の変数を影響変数ととらえることができる (図 1 - 1 において関係の内部に向かっている矢印 A' は内部への影響を示し、実世界へ向かっている矢印 A は外部への影響を示す), B にも影響変数が必要で、さらに C にも影響される変数が必要である. 全体として影響変数を伴うデータを扱う必要がある. つまり本研究は、「具体的に設定した関係についての課題を関係と影響変数のある問題という視点から解く手法を提案すること」を主な目的とする.

1.1.2 意義

この課題を解く意義をのべる. まず、さまざまな分野で同様の問題がみられる. A. 関係の影響の検証, B. 未知の関係の獲得, C. 関係と影響変数が動的に変化しているなかで影響変数を説明する関係抽出は世の中のさまざまなスケールで出現している. たとえば、人間に関しては身体内、認知、社会的相互作用などであられる. そのため、この課題をとくことは多くの分野での問題の解明に結びつく.

具体的には関係やその集合、そしてそれらに対応する影響変数に対してどのような働きをもつかかわるほか、新しい関係の予測にも使用できる. 組織の例をあげるのであればどのような人間関係が生産性を向上させたのかがわかり、ある人間関係をつくることで生産

性をより向上させることも可能となると考えられる。

・運動スキル[中村 03]

運動の学習で、人はある初歩的な動作（歩く、走るなど）を獲得する。次に、これらを組み合わせることにより、この複雑な動作を行う。このとき、個々の動きはたとえば楽器演奏などを実現するため、関係の中でそれぞれ何らかの機能をしている(A)。またこれらはあらかじめ学習されたのではなく、試行錯誤などをしながら動きを組み合わせることで特定の演奏を獲得する(B)。さらにこれらの組み合わせは一定ではなく、熟達をするにつれて変化していく(C)。

・認知／知識獲得[Deacon 97]

知識獲得のプロセスにおいても、同様の現象がある。人は言語や記号の獲得段階では、視覚聴覚などのセンサーデータを入手し、そのデータを汎化して記号を割り当てる。次にこれらの記号同士の組み合わせにより、我々は個々の記号に対して知識と呼ばれるような複雑な意味を与え、またなんらかの推論を行い行為などを行った場合、外の変数に対して関連を持つ(A)。このような記号を同士の組み合わせ（つまり関係）は未知の状態であったり、外部から知識が与えられることを通じて獲得される。たとえば、「花は美しい」などの知識がはいり、「花」と「美しい」が組み合わせられたものができる(B)。最後に、新しい記号の導入や、新しい組み合わせが生じた結果として、時間とともにこの関係は変化していく(C)。

・社会／組織[安田 09]

実社会では、組織は、全体としてある製品を作ったり、事業を行うなど、外に対してひとつの機能を果たす。また、組織の人間関係は、個々の社員に権限や圧力などの影響を与える(A)。このような組織はあらかじめあったものではなく個々の人間が、それぞれある目的のために相互に協力関係をつくることで成立する(B)。さらに、これらの組織は一定ではなく、時間がたつにつれて、組織の構造や、その影響も変化していく(C)。

1.1.3 関係と影響変数

関係についての分析は1.1.1でものべたように近年多くの分野で注目されているがそもそも「関係」というものはなにか(図1-2)。ここでは本研究における定義を述べる。

一般に世の中のデータは、実体-関連図(E-R図)のように実体と、その関連、そしてそれらの属性によって多くは表現できる[Chen 76]。たとえば、ある商品の販売というものを考えると、ある商品には属性があり、また販売の対象となる顧客にも属性がある。これらの中で販売という関係が成立し、そこには数量や日時などの属性が含まれる。これだけではなく、顧客同士でも相互に情報交換をするなどの関連がある場合もある。このように、「関係」とは実体同士のつながりに着目したものといえよう。これら実体と関係は「ネットワーク」という名称で表現されることもある。

以上は、実世界で関係が扱っているものについて説明したが、それでは関係はどのような形式で取り扱われるのだろうか。

一般に数学では、関係は直積の部分集合として表現される[野口 07 他]。ひとつの集合の要素間の関係については、要素もふくめグラフとして表現されることも多い。グラフとは、ノードとエッジ(リンクとも呼ばれる)、すなわち点と線で構成されたもので、主に実体と実体同士の関係性を表現したものである(図1-2①)。また、このグラフにおけるエッジは、つながりの有無を隣接行列として表現が可能である(図1-2②)。関係データベースで扱う関係もある(図1-2③)。これはグラフであつかわれるような1対1の関係だけではなく、複数の関係を扱うことができる。たとえば $R(A, B, C)$ という関係スキーマを考えると3項以上の関連を扱える[Codd 70]。また1対1の関係に属性を加えたものを表現することも可能である。この関係データベースで用いられる関係は、述語論理における述語とも密接に関連しており、その表現はほぼ同等である。複数の関係については数学でハイパーグラフとしても扱われる[Horvath 06]。

これまでの関係の研究をふまえて、本研究で共通に扱う関係および影響変数の意味と内容について、再度定義する。

本研究では、上記と同様にある実体と実体同士のつながりを関係と呼ぶ。また、事実にもとづいた関係の集合を関係データと呼ぶ。なお、実世界にある関係全体を示す際には、ネットワークという名称も用いる。関係の具体例としては、ある動きとある動き、人や組織、集団と集団の間のつながりを扱う。特に一対一の実体間のつながりを中心に扱う。表現としては、グラフおよび述語論理表現を扱う。特にBの問題やCの問題については関係の種類も扱うため、述語表現を用いている。

また本研究は、ある関係に着目し、関係全体が影響を与える外部や内部にある変数を取り扱う。これを影響変数と呼ぶ。

影響変数とは、販売における例をあげると、ある属性をもった顧客とある属性をもった商品との間に成立する販売を関係とすると、その商品の販売数量が影響変数にあたる。

これらは二つの異なる実体間の関係であるが、本研究では主に一つの実体の中の相互の関係、たとえば顧客同士の関係と売上を中心に扱う。

さて、既存の関係を扱った研究と異なっているのは、このように影響変数のある関係を扱っているという点にある。

これまでの研究では関係を表現した隣接行列などの分析を通じて、たとえばエッジの次数分布や、ネットワーク内に存在するクラスター（エッジのかたまり）、ネットワーク分析にあるようなネットワーク中心性（関係からみたノードの重要性をはかる指標）の計算といった「関係」の特徴の分析が中心であった。また統計分析においては、確率的効用関数の推定のように、ある実体をもつクラスとその属性の関係について焦点をあてたもの（たとえば、購買するか否かを、その属性によって判断する）や、変数間の因果関係を検証していくもの（パス解析など）が多かった。

これらとは異なり、関係だけではなく、影響変数も同時に扱うのが本研究の特徴である。

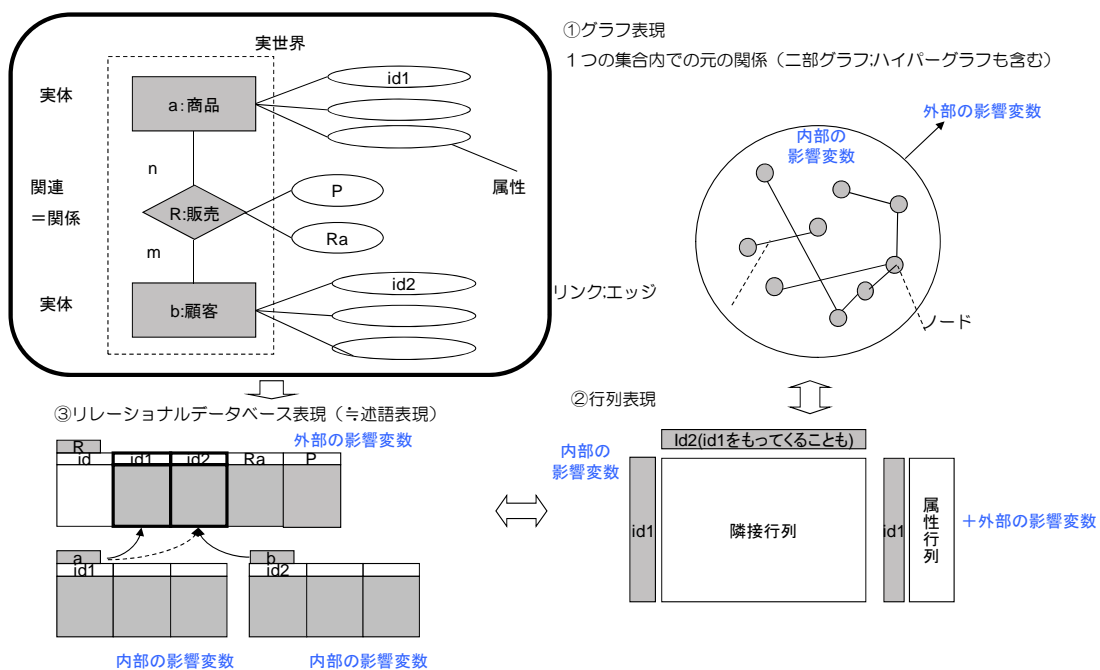


図1-2. 関係と影響変数およびその表現

1.2 問題の種類と形式化

本論文で扱う関係を伴う問題の分類と問題の形式的な定義を行い、その解法についてのべ、1.1.1でのべた具体的な関係についての主要な研究課題を位置づけるための準備をおこなう。これを通じて、どのような既存の問題にあたるか、新しい問題なのかを整理し、既存手法のどれが利用できるか、またその応用が可能かの検討の土台とする。

1.2.1 関係に関する問題の種類

関係を扱う分析課題は、主に以下の条件により複数の種類が想定される。

- a. 影響変数があるかないか
- b. 時間を考慮するか、時間を考慮しないか
- c. 関係データが事前に与えられているか、いないか
- d. 背景知識や制約を利用するか、利用しないか

個別に詳細をのべる。

a は、影響変数を設定し、関係を説明変数とするか否かで問題は異なることを示す。影響変数を伴う場合は、機械学習においては教師あり学習といわれている問題である。今回の課題 A, B, C はいずれも関係とそれに関連する変数との関連をみていく。

また、細かくみると 1.1 でのべたように影響変数がある場合でも「内部」の変数を用いるか、「外部」の変数を用いるかでも問題は異なっている。

他方、関係を扱った影響変数を設定しない記述的な問題とは、たとえば頻出パターンマイニングやクラスタリングのように正の事例のみが与えられているなかでその抽象化をおこなうという課題である。機械学習では教師なし学習と呼ばれているものである。

b は、関係とその属性、そして影響変数などのデータが時間情報を有しているか否かで問題は異なることを示す。一般には、時間を考慮していない場合が多いので、時間を考慮していないタイプの拡張として時間を考慮したタイプの問題が想定される。

c は、関係についての情報が事前には与えられているか否かで異なることを示す。ただし関係についての情報がない場合でも、実体をもつ属性などの情報は事前に与えられているなど、影響変数と関連した関係以外のデータがあることを想定している。このようなデータが与えられたもとの、影響変数をうまく説明する関係を探索をしながらみつけることが課題になる。一方、関係が与えられている場合は、関係の中から予測に必要な

な要素をみつけたり、影響変数と関係との間の関数をみつけることが課題となる。

d は、事前に制約や知識が与えられているかどうかで問題は異なることを示す。これは「ある条件が備わった場合、ある関係が成り立つ」というような知識を与えるか否かをさす。つまり、完全に探索的に関係抽出をするか、もしくは事前に関係に対して制約条件が存在するか否かである。詳細については、1.2.2 の I.i.ii でのべる。

以上の a, b, c, d の観点から問題について、既に問題や解決法が存在しているものや、本研究に関連するものについて分類すると以下の表 1-1 および図 1-3 のようになる。なお、分類上、問題設定は可能であるが関連がない問題については記述していない。

表 1-1. 関係に関する諸問題の構造

I.	影響変数を伴う時間が含まれない関係・属性データ分析の問題
I.i.	影響変数があり、関係は明示的に与えられてない問題
I.i.i	関係は与えられておらず、関係を推量しながら学習する問題
I.i.ii	関係は与えられていないが、属性などに対して関係をつくる制約（ルール）がある問題
I.ii.	影響変数があり、関係はすでにあたえられている問題
I.ii.i	影響変数とグラフの組が複数与えられている問題
I.ii.ii	影響変数がノードの属性で、グラフが 1 組与えられている問題
I.ii.iii	影響変数がグラフで、グラフの集合が 1 組与えられている問題
I.ii.iv	関係が完全に与えられていない 1 組のグラフが与えられており、一部の関係が影響変数である問題
II.	影響変数を伴う時間が含まれる関係・属性データ分析の問題
II.i	影響変数があり、関係が明示的に与えられていない問題
II.ii	影響変数があり、関係が与えられているときに 関数やルールを推定する問題
III.	記述的な問題やその他の問題

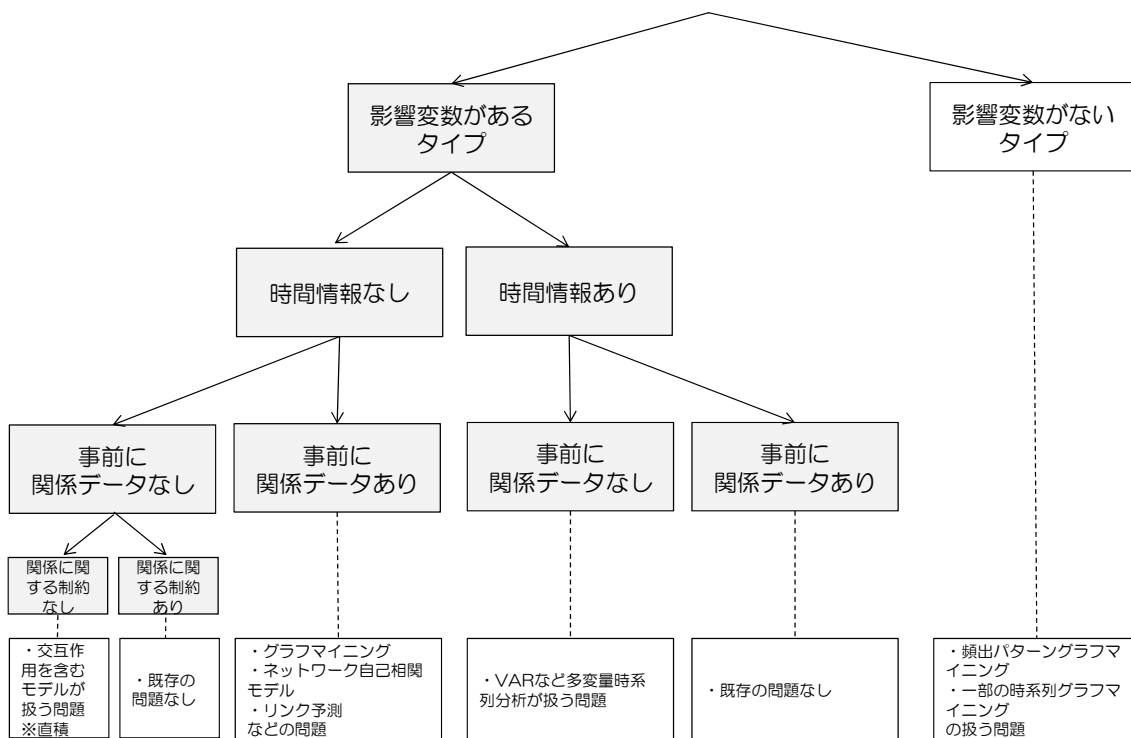


図 1 - 3. 関係に関する諸問題の構造

1.2.2 影響変数を伴う関係を扱う問題一般についての形式的定義

本研究で主に扱う、影響変数を伴う関係を伴った一般的な問題(1.2.1でのべた問題分類がI.iiやII.iiに該当するもの全般. I.iやII.iはこれらから一部情報が除外された問題とみることができる)について形式的な定義を行う. 関係に関しては、グラフ表現を用いた定式化を採用する.

本研究で扱う問題における入力データは $\langle G_s, h_s \rangle$ で定義される.

まず、属性のある関係データ G_i について定義する. G_i は $f_i, V_i, E_i, \phi_{ei}, \phi_{vi}$ で定義される. V は、頂点. または実体と呼ばれるものである. 属性などを有することもある. E は、リンク (エッジ). 実体-関連図においては関連と呼ばれるもの. D_e, D_v は、それぞれ、リンクラベル、頂点ラベルの集合である. これら G_i に加えて、影響変数 h_i を導入する. h_i は、 G_i の内部にある影響をうける変数 (V や E のもつ変数) もしくは、 G_i の外部で影響をうける変数である. 企業を事例としてとりあげるのであれば内部の変数としては個々人の生産性などの属性、外部は営業利益率などが該当する. 後者は、関係やシステムをもったものを外部から観測したときに用いられる量のことである.

$$\begin{aligned} G_i &= (f_i, V_i, E_i, \phi_{ei}, \phi_{vi}) \\ f_i: E_i &\rightarrow V_i \times V_i \\ \phi_{ei}: E &\rightarrow D_e \\ D_e &= \{\text{実数, カテゴリー}\} \\ \phi_{vi}: V_i &\rightarrow D_v \\ D_v &= \{\text{実数, カテゴリー}\} \end{aligned}$$

次に定義されるようにこれら、 G_i と h_i の組 $\langle G_i, h_i \rangle$ が n 個あつまった $\langle G_s, h_s \rangle$ が本研究における一般的な入力データである.

$$\langle G_s, h_s \rangle = \langle G_1, h_1 \rangle, \langle G_2, h_2 \rangle, \dots, \langle G_n, h_n \rangle$$

集合 G_s から h_s への関数を F とする.

$$F: G_s \rightarrow h_s$$

F には G に対応した h への関数である. 場合により、論理式も含まれる. このとき本論文における問題の一般的な形式は、「データ $\langle G_s, h_s \rangle$ が与えられたもとで $F \cong f'$ となる f' を学習すること」にある.

時間を伴う関係データ分析の問題については時間を $1 \leq t \leq T$ とすると入力データは

$$\langle G_T, h_T \rangle = \langle G_1, h_1 \rangle, \langle G_2, h_2 \rangle, \dots, \langle G_T, h_T \rangle$$

で定義される. F は m 期過去のグラフデータ G を予測に用いた関数

$$F: G_t, G_{t-1}, \dots, G_{t-m} \rightarrow h_t$$

で定義され、 A と同様に推定した関数 f' を学習することにある.

1.2.3 各問題の詳細

1.2.2 を参考にして、各問題について詳細な説明を行う。

I. 影響変数のある時間が含まれない関係・属性データ分析の問題

I. i 影響変数があり、関係は明示的に与えられてない問題

G_s から E と ϕ_e を抜いたデータを G_s' とすると、 $\langle G_s', h_s \rangle$ が与えられているもとで h に関連する G の特定や G_s から h_s を説明する関数 f' を学習する。

I. i. i 関係は与えられておらず、関係を推量しながら学習する問題

$\langle G_s', h_s \rangle$ のみが与えられている問題である。この中で G を特定しながら h_s の予測を行う。これには、たとえば交互作用を組み込んだモデルで扱う問題が該当する。交互作用を組み込んだモデルには、分散分析などがある。

分散分析の場合、ある属性 A とその属性 B から、ある影響変数を説明する。このとき、 A と B の組み合わせ、すなわち $A \times B$ による説明力についても検定を行う。このとき、 n 個の属性があるとして、一对の組み合わせを考えると、ひとつの n 個の頂点とその関連を完全グラフとして表現可能である。このことは、 V の要素を変数と考えると、あらかじめ関係が与えられていない中で、有効な関係を発見するタスクとして考えることも可能である。ただし、実世界との対応では交互作用の場合は、実体というよりも実体のもつ属性間の関連をみていることが多い。具体的に説明する。以下の影響変数が Y 、独立（説明）変数が X_a, X_b, X_c のものがあつたとするとこれらの二つの変数間の交互作用を考慮した線形モデルを以下で表す。

$$Y = f(X_a, X_b, X_c, X_a X_b, X_a X_c, X_b X_c)$$

このとき、 X_a, X_b, X_c といった変数の間で作用が存在するか（有意か否か）否かを G ととらえることができる。

I. i. ii 関係は与えられていないが、属性などに対して関係をつくる制約（ルール）がある問題

$\langle G_s', h_s \rangle$ が与えられているもとで、必要な E や ϕ_e を得ながら f' を学習する問題である。ただし、この際、関係を学習するためにある知識（制約） K を利用する。 K とはノードの属性 $\phi_e(V)$ が、ある条件にあつた場合、関係が存在するとする

ようなルールのことである。たとえば、時区間論理のように時区間のある実体同士が、ある条件が満たされた場合、どのような時間的隣接関係にあるかを事前に定義しておく。属性が類似している場合、リンクがあると定義することもある。

形式的に定義する。通常はノード V の変数 $\Phi_v(V)$ 同士の組み合わせから、有用な組み合わせを選ぶため任意のノードを $V1, V2$ とすると $\Phi_v(V1) \times \Phi_v(V2)$ というデータと hs との関連をみていく。しかし、これら直積をすべてみていくのではなく、たとえばあるノード間で

$$R(X, Y) := \text{node}(X), \text{node}(Y), \Phi_v(X) > \Phi_v(Y).$$

といった条件が成立するか否か、これが人間同士でいうならば、年齢の上下関係が成立しているか否かといったある特定の関係(述語) R のみをみていくということにあたる。

I . ii 影響変数はあり、関係はすでにあたえられている問題

$\langle Gs, hs \rangle$ が与えられているもとで h に関連する G を特定、もしくは $\langle Gs, hs \rangle$ が与えられているもとで Gs から hs を説明する関数 f' を学習する。

I . ii . i 影響変数とグラフの組が複数与えられている問題

この問題の場合、 G の集合 Gs が与えられているなかで頻出するパターンを列挙し、 h を説明するのに、重要なデータのみを選ぶ課題である。これは影響変数のあるグラフマイニングの課題と等しい。グラフマイニングは、ある性質をもったグラフの集合が与えられたときに、そのグラフのなかから共通する関係、リンクやノード、およびそれらのラベルを抽出する問題である。

近年ではグラフ回帰というタスクもある[Saigo 08]。これは $\langle Gs, hs \rangle$ というグラフと影響変数の組を複数与えたうえで、グラフマイニングによる特徴抽出を利用して、グラフをその特徴ベクトルへと変換して、影響変数に対する関数を学習していく課題である。

I . ii . ii 影響変数がノードの属性で、グラフが1組与えられている問題

G のなかのノードの属性ベクトル $\Phi_v(V)$ を h としてみる課題である。このとき、線形なモデルを仮定するとネットワーク自己相関というモデルになる[Leenders 02]。このモデルは、ノードがもつ属性 $\Phi_v(V)$ に対して、関係 E を介して属性 $\Phi_v(V)$ 同士がどれだけ影響されているのかを形式化したものである。もともとは、空間計量経済学において、空間的な相関を調べるために導入された空間的自己相関モ

デルを拡張してネットワークに適用したものである。詳細については2章でのべる。

I. ii. iii 影響変数がグラフで、グラフの集合が1組与えられている問題

影響変数 h をグラフ G (ただし、複数の E が存在) に設定して予測を行う問題がある。このとき回帰(線形回帰)を用いる方法があり [Krackhardt 88; Dekker 03], これはネットワーク回帰とも呼ばれる。具体的には入力 G のなかから E を複数抜き出し、これらの集合 E_s をもとに別の E へと回帰を行う。結果、ある関係を説明するのに重要な関係はなにかを知ることができる。

I. ii. iv 関係が完全に与えられていない1組のグラフが与えられており、一部の関係が影響変数である問題

G の部分集合 $G' (\in G)$ のみが与えられている状況下で、その要素となる E' や $\Phi_v(V)$ を利用して全体の E を求める問題である。このとき影響変数は E' 自体に設定する。これはリンク予測問題と呼ばれる [Getoor 03]。属性情報を用いたリンク予測を例にあげると、ある人 A とある人 B が繋がっているとす。このとき、 A と B がどのような属性のとき、繋がるかのパターンを学習する。次に A と C など未知の人同士の関係が、その属性に基づいて実際に繋がっているかどうかを予測する。この問題は、あらかじめ全体の関係が与えられていないという意味では、I. i に近いが一部の関係を用いて関数を学習するという意味で I. ii に分類する。

II. 予測変数のある時間が含まれる関係・属性データ分析の問題

II. i 影響変数が与えられており、関係が明示的に与えられていないときに関係やルールを推定する問題

GTではなく、 $\Phi_v(V)$ の時系列のみが与えられている状況下で、 $\Phi_v(V)$ 自身が hT すなわち影響変数となり、 V 同士の関連である E を推定する問題がある。これは通常の変数の時系列分析で扱う問題も該当する。複数の時系列データが与えられたうえで、ある変数を説明するために別の変数を用いて分析を行う。この場合相互の依存関係を学習しながら推定を行う。関連する手法としては、ベクトル自己相関モデル(VAR)[北川 05]や、構造時系列モデリングなどもここに該当しよう。

II. ii 影響変数が与えられており、関係が与えられているときに、関数やルールを推定する問題

$\langle GT, hT \rangle$ が与えられているもとで G が h を説明する f' を学習する。すなわち、動的に変化する影響変数に関連した関係を学習する。このような問題を扱う具体的な手法はあまり研究されていない。

1.2.3 記述的な問題および、その他の問題

ここでは関係を扱う手法のうち、1.2.2 で扱った問題に該当しないものを紹介する。一般的には影響変数が与えられていない関係データ G_s もしくは GT のみが与えられた状況下で、これらの特徴を抽出したり、また頻出パターンを抽出するといった問題がある。

社会ネットワーク分析における中心性の計算（ノードがもつリンクに関する特徴量）や、似たような関係をまとめるブロックモデルの作成や、ネットワークが閉じた集まりになっているものを抽出するクリーク抽出といったさまざまな課題や技法が存在する[安田 01]が、これらは一般的には影響変数があるわけではなく、ネットワークのもつ特徴を抽出することが多い。

このほかグラフマイニングのひとつである GBI など主に正例からの学習や頻出パターン抽出を行う場合も、この影響変数が与えられていない場合に該当する[Washio 03]。

ネットワーク時系列という問題もある。これはグラフ構造が時間的に変化していくもの、つまり GT がデータとして与えられたときに、その中から、ある時間に特徴的なパターンを学習したり、全体として頻出する特徴を抽出するという課題である[Tanya 06]。また時系列データマイニングの一種としてグラフを扱う場合もある。この場合、グラフの特徴から、その変化をよみとり、グラフ時系列の変化点を学習することが主な目的となる。具体的な手

法としては、グラフ同士の距離を計算して、その距離の変化を追っていくという方法がある[Pincombe 07]. また、ある時点でのグラフの特徴ベクトルを計算して、そのベクトルの変化をおう方法も開発されている[井手 04].

1.2.4 問題の解法

以上、さまざまな問題について解説をしてきた。これらの問題に関連する手法についても説明してきた。

ただし、これらの手法は、とりあげた問題以外の問題にも適用できる場合や、その利点や欠点がある。そこで本研究に関連する手法をいくつかとりあげ長所や短所ごとにまとめると以下の表1-2のようになる。

このような各手法の特徴を生かして、改良を加えて利用を検討することが本研究となる。各方法の詳細については、帰納論理プログラミングは2章および4章、自己相関ネットワークについては2章、グラフを扱う時系列データマイニングや、ネットワーク時系列分析については4章でのべる。

表1-2. 諸問題に関する解法とその特徴

方法	利用可能なタスク	長所	短所
帰納論理プログラミング (予測タイプ) ※述語論理形式 ※詳細は後述	主に影響変数がある タイプの問題。	複雑な構造が導入できる。 事前知識の導入。	探索的な場合は、計算コストがかかる。
グラフマイニング	影響変数がない問題。頻出 パターンの抽出。	頻出する関係パターンの 探索が高速にできる。	複雑な構造、たとえば、 制約を組み込むなどの柔軟性は少ない。
ネットワーク自己相関 $y = \alpha Wy + \varepsilon$ y: 説明変数 W: 関係行列	影響変数があり、関係が 1つだけ与えられている 問題。関係の効果の検証。 影響変数はノードの属性。	関係の効果について検定 ができる。	関係が未知である場合な どは適用できない。
時系列データマイニング (グラフ距離を使うもの)	影響変数がない問題。 時間情報のある関係デー タを扱い、変化点の抽出 を行う。	変化点の検出は可能。オ ンラインデータなどから の異常値検出などにも利 用可能。	詳細な構造の変化の分析 はできない。
ネットワーク時系列分析	影響変数がない問題。 時間情報のある関係 データ。頻出パターンの 抽出。	特定の構造では、頻出パ ターンをより早く検出で きる。	問題が限定されている。

1.3 本研究の課題の位置づけとアプローチ

次に具体的な課題と 1.2 でのべた問題群とを対応させ、それぞれについて解法を検討する。

1.1.1 でのべた 3 つの問題はそれぞれ以下のように対応づけられる。

- A. 関係の影響の検証については、すでに何らかの関係が与えられているうえでその成果を調べるため I . ii のいずれかの課題
- B. 未知の関係の獲得については、多くの場合影響変数が与えられているうえで関係を探ることが多いので I . i . i もしくは I . i . ii
- C. 関係と影響変数が動的に変化している特徴抽出の問題については、II . ii の時間情報があるデータの分析の課題

となる。

また、それぞれの課題を解く具体的な応用範囲としては認知、運動、社会といったそれぞれのスケールを想定して以下のように設定する。詳細な問題設定は 2 章以降の各章で行う。

- A. 信頼の認知：特に商品とその情報源に対する信頼の認知において、情報源としての依存関係によって影響していることを検証する。
- B. 運動スキルの学習：特に楽器演奏に関して筋肉の動きの協調関係を抽出する。この際、筋肉の活動パターン同士の時間関係を時区間論理という知識を用いる。
- C. 社会におけるコミュニケーション：特にコミュニケーションが時間ごとに変化するなかでその変化の特徴をノードの属性もふまえた上で抽出する。

続いて、それぞれの問題についての利用が可能な解法について検討する。

A については、I . ii のうち I . ii . ii の問題となりネットワーク自己相関が利用できる。この問題は、情報源としての依存関係についての情報がえられているうえで、影響変数を商品と各情報源の信頼、つまりノードの属性に設定するためである。また、関係の影響を検証できるという点もある。

B については、I . i . ii に該当し帰納論理プログラミングが利用できる。この手法を用いる理由は、あるノードの属性が与えられたうえで、関係に関する知識（たとえば、時間の前後関係を扱う時区間論理）を導入して、そのパターンを獲得する必要があったことによる。

C についても、帰納論理プログラミングが利用できる。この手法を用いる理由は、ノードの属性、リンクの属性、さらに時間に関する知識（たとえば、ノードとリンクの前後関係）や、関係についての知識など複雑な構造を有しているため、帰納論理プログラミングを用いることが妥当であると考えた。また、影響変数が外部データとして与えられているため、それらを帰納論理プログラミングにおける正負事例として利用できるという理由がある。

なお、この帰納論理プログラミングとネットワーク自己相関モデルの違いは、いずれも関係のみならず実体の属性を有してはいるが、学習しているモデルが異なっていること、背景知識が利用できるかできないかという2つの点が挙げられる。それぞれ具体例をあげる。

モデルの違いについては E , F , G を変数として E と F が G に関連ある場合は(i)式になる(α , β は係数)。一方 E や F を述語として、ある y がこれらに該当した場合、 G が成立することを記述した場合は(ii)式 (Prolog 形式で表現) になる。

$$G = \alpha E + \beta F + \varepsilon \quad \dots (i)$$

$$G(y) :- E(y), F(y) \quad \dots (ii)$$

このように、(ii) は命題論理レベルのルールを学習するが、(i) は α というように線形のモデルを学習することで重みの学習ができる。ネットワーク自己相関モデルは、目的となる変数がベクトル y 自体であるので、実際には $y = \alpha Wy + \varepsilon$ というモデルになっている。このモデルの詳細は2章で扱う。

つぎに背景知識の利用ができるかできないかという違いがある。背景知識は I.i.ii でのべたように、あるデータとあるデータの関連を事前に規定したもので、たとえば、B の課題ではある時間区間をもった運動 α と別の運動 β があつたとき、 α が β の前に行われている場合は、 $\text{before}(\alpha, \beta)$ と表現し、重なって起きた場合は $\text{overlap}(\alpha, \beta)$ という関係で表現することを取りきめておけるかおけないかということである。これは C の課題においても、ノードとノードの前後関係、リンクとリンクの前後関係、ノードとリンクの前後関係などを取り扱っている。

このように各具体的な問題に合わせて、それにふさわしい方法を利用し、また個別に改良を行うことで問題の解決を行える。これら、各課題とそのアプローチ法についてまとめたのが図 1-4 になる。

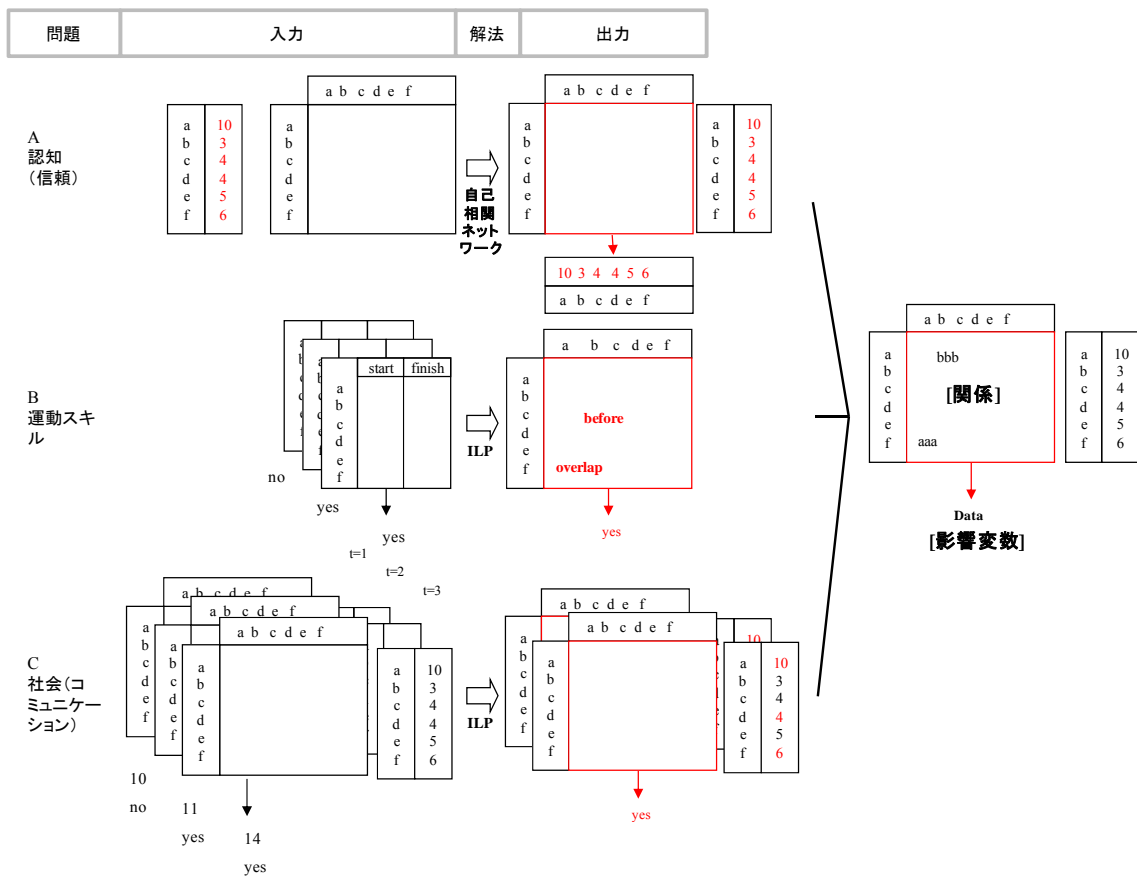


図1-4. 本研究の課題とそのアプローチ

1.4 本論文の成果

1.3 までに、関係に関するデータおよび、それと関連する影響変数があたえられている問題について整理を行った。そのうえで現在目的としている、A.関係の影響変数への効果の検証、B.影響変数を説明する未知の関係の獲得や、C.関係と影響変数が動的に変化しているときの特徴抽出といった本研究の課題がどこに該当するのかを検討した。その結果、AはI.ii.iiのノードと関係の関連をしらべる問題であることがわかった。Bは関係が事前に与えられていない場合の問題でI.i.iやI.i.iiあることがわかった。最後のCは動的な関係と影響変数との関連を扱ったII.iiの課題であることもわかった。これらをふまえた上で次章以降で行った各研究の成果を事前にのべる。

1. 関係がもつ内部の影響変数への効果について統計的に検証ができた。検証のためにネットワーク自己相関モデルを利用した。
具体的には、商品の信頼に対して情報源の信頼が寄与していることや、情報源同士の信頼も、情報源であるという関係に基づいて相互に依存していることがわかった。
2. 関係が明確でないものから、影響変数があるなかから背景知識を用いて新たに関係抽出ができた。その際、背景知識の利用のために帰納論理プログラミングを利用した。
具体的には、チェロ演奏において熟達者があるタスクを演奏する際の筋肉の協調関係を抽出することができた。これは既存の知見とも一致した結果であった。
3. 関係と影響変数があり、それらが動的に変化するなかで共通ルールをみつけるという新たな問題を提案した。さらにこの問題を解くことができた。方法としては背景知識を利用し、時間的な依存関係を扱うために帰納論理プログラミングを利用した。
具体的には、不況の際の景気認識のクチコミがどの時点で変化をしているのか、変化の前後でどのような層の間でコミュニケーションの変化が起きたかを特定することができた。

全体としての意義は、応用範囲の広い3つの課題について、「関係と影響変数との関連を扱う問題」という視点から捉えなおし、各課題が問題群のなかのどこにあたるか。そして、各問題について具体的な解法を示したことにある。数値情報が中心で関係の検証が必要な場合は自己相関ネットワークが有効であることを示した。また複雑な構造がある場合や定性的な情報が多い場合は帰納論理プログラミングが利用できることがわかった。また、そのうえでそれぞれの分野において一定の知見をもたらしたことも挙げられる。

派生的な意義としては、Bの課題において情報量基準を用いて、モデル選択を行い、時系列データのセグメンテーションをすることが有効であることがわかったこと。Cにおいて一

貫性制約の抽出という新たな問題領域を設定したことが挙げられる。

1.5 本論文の構成

本論文の構成について説明する。

2章では A の問題について認知、特にある商品の信頼の認知という分野に適用した分析結果とその汎用性の考察をしめす。

3章では B の問題について運動スキルの分野に適用した分析結果とその汎用性の考察を示す。

4章では C の問題の枠組みについて、社会におけるコミュニケーション過程を分析し結果とその汎用性の考察を示す。

5章では、以上1章での考察と、2、3、4章で実データを用いた検証をふまえて考察を行う。

6章では、本アプローチ全般をふまえて結論をのべる。

2章 商品および商品の情報源に対する信頼の統計的ネットワークモデル

本章では、関係についてのデータがすでに与えられている中で影響変数に対する関係の影響について検証する。

具体的には消費者による商品およびその情報源への信頼が“情報源という関係”によって相互に依存していることについて議論する。今日、消費者にとっての商品に対する信頼の確保は、大きな問題となっている。こうした商品に対する信頼について本章では特に、その情報源に対する信頼からの影響を検証する。加えて、これらの情報源の信頼の相互依存関係についても検証する。これらの問題に対する解として“信頼のネットワーク仮説”を提示し、この仮説を統計的に検証するためにネットワーク自己相関モデルを用いる。その結果、商品に対する信頼とその情報源の信頼とが相互に依存しあっていることを示す。特に食品に関しては製造源や販売元といった情報源に対する信頼の影響が強く、またさまざまな情報源に対する信頼形成の中心に、友人・知人への信頼があることを示す。

これらを通じて、Aで提示した問題に対して本章の提案が妥当であることを示す。

2.1 背景と問題設定

インターネットの普及により商品・サービスについての情報が氾濫する中、どのような情報源からのどのような情報を信頼していけばよいのかが消費者にとって大きな問題となっている。たとえば、ある1つの商品に対し良い評価と悪い評価の両方が存在するが、それぞれの情報がどれほど信頼できるかがわからない状況が多々ある。また食品の偽装問題などが大きな社会問題として報道されていることから、商品の製造元に対する信頼も揺らいでいることがうかがわれる。このような商品や情報源に対する信頼の揺らぎは、商品への信頼を獲得しようとする企業、商品が信頼できるか判断する消費者の情報探索のコストを増加させ社会に大きな影響を及ぼす。そこで、本章では信頼の形成メカニズムの解明とそれによる示唆をえるため、消費者にとっての商品やさまざまな情報源に対しての信頼のモデル化と検証を試みる。特に以下の4つの問題について検討する。

はじめに、商品に対する信頼を考える。商品が信頼を得るためには、商品がもつ属性に加えて、商品の情報源への信頼が重要であると考えられる。ここで、商品の属性とは、たとえば、品質を保障する外部機関のマークがついている、きれいに包装されているなど、商品を手にとったときに知覚される情報である。一方、商品の情報源とは、それ以外の手段による、商品に関するさまざまな伝聞である。多くの消費者は特定の商品についての十分な知識がない場合も多く、商品の属性だけから信頼を判断できるとは考えにくい。むしろ現実には、販売店やその商品を周りで紹介した友人など、商品についての情報を提供した「情報源」に対する信頼も、商品の信頼に同時に寄与しているだろう。いいかえると商品を取りまくさまざまな情報源の信頼により、信頼が形成されていることが示唆される。これまでの信頼の研究においても、ある対象に対して知覚される属性をもとにした信頼（皮相）と、対象に関する情報源によって形成される信頼（発信者や評判）といった区別は存在したが[加藤 06]、後者の情報源からの寄与について詳細な構造の解明は行われていなかった。これをふまえ、特に後者に焦点をあて第1、2の問題を次のように設定する。

問題1 商品に対する信頼は、情報源に対する信頼から影響をうけているか。

問題2 商品に対する信頼が情報源から影響をうけていた場合、どのような情報源に対する信頼から主に影響を受けて形成されているのか。

つづいて、情報源に対する信頼について考える。商品と同様に、ある情報源に対する信頼も、その情報源について紹介した別の情報源の信頼と、ある情報源がもつ属性から成り立っていると考えられる。たとえば、前者については、商品の情報源である「店や製造元への信頼」が、信頼のおける専門家や、友人などを介して紹介されることで成り立つことがある。このように、情報源が互いに他の情報源に関する情報を提供することで、情報源どうしを結ぶネットワークが作られ、相互の信頼に寄与していると想定される。しかし、

こうした情報源の相互依存関係に関する研究はなく、また相互依存の詳細な構造の解明は行われていない。そこで第3、4の問題を次のように設定する。

問題3 情報源に対する信頼は、相互に影響しあっているか。

問題4 情報源に対する信頼が相互に影響しあっていた場合、どのような情報源への信頼から主に信頼が形成されているのか。

以上のように本章は、商品と情報源の信頼の相互依存ネットワークの検証を課題の中心とする。

以降、上記4つの問題に対してそれぞれ仮説を示すとともに、これらを統合的に扱う「信頼ネットワーク仮説」を示す。信頼ネットワーク仮説は、ある対象者にとって商品や情報源に対する信頼が相互に依存しあうことを概念的に表現した認知的なモデルである。2.2では、本章における構成概念である信頼やネットワークなどを定義する。2.3では、本章の主題である信頼ネットワーク仮説について説明する。また、これを検証するため統計的なネットワークモデルについても概説する。2.4では、実際に調査データを用いて検証を行い、商品情報について信頼のネットワーク仮説が支持されることを提示する。2.5では、信頼やネットワーク、そしてマーケティングに関する既存研究を紹介しながら本章の位置づけを行う。最後の2.6でこれらの研究により得られた結果のまとめを行う。

2.2 構成概念の定義

準備として本研究の仮説で扱う概念である、対象者、商品、情報源、信頼、ネットワークの定義をのべる。

(1) 対象者

対象者とは、本研究である商品に対して信頼をおく主体で、情報源についてのネットワークを認識している主体である。検証の調査においては回答者にあたる。

(2) 商品

商品とは、個別の銘柄などの上位概念に位置するカテゴリで、たとえば食パンや牛肉、電子レンジなどのカテゴリの単位で構成される。

(3) 情報源

情報源とは、ある商品についての情報を発信する主体・組織といった要素のことである。本研究では、消費者との利害関係から情報源を大きく4つに分類し、さらにこれらを細分化した7つの要素を使用する。4つの分類は次の通りである。(a)商品を販売することで利益を得る情報源である「製造元」および「販売元」、(b)これらとの間接的な利害関係はあるマスメディアなどの情報源として「テレビ番組」および「新聞」、(c)利害関係は直接的にはないと考えられる「専門家」および「第3者が発信したインターネット上の情報」(以

下, 第3者のインターネット上の情報と表記), (d) 自分自身と直接関係がある「友人・知人(家族も含む)」である。より詳細な区分は可能であるが, まずは大きな要素をとりあげて分類し, より単純なモデルで検証するために要素は限定した。これらの要素で検証されれば拡張は容易に行うことが可能である。

(4) 信頼

信頼とは, 対象者が事前に予想する, ある対象に期待している機能と実際の機能とが一致している程度と定義する。対象の機能とは, 商品であればそのおいしさや味などの品質, 情報源であれば, その情報源から提供される情報の正確さなどをさす。

ただし信頼は多義的な性質を持つとされており[Ye 02], その定義は分野により多種多様である。本研究では, Oxford English Dictionary(1971)における信頼(Trust)の定義(「confidence in or reliance on some quality or attribute of a person or thing, or the truth of a statement」)を参考にしている。その理由は, 信頼のもつ機能に基づいていることから汎用性が高いと考えられるためである。

また, 操作的な定義としては, 各商品の品質や情報源からの情報などについて「非常に信頼している」「まあ信頼している」「どちらともいえない」「あまり信頼していない」「まったく信頼していない」のいずれかを対象者に解答させて, 各順に5, 4, 3, 2, 1の得点を割り振った値とした。

(5) ネットワーク

ネットワークとは, ノードとエッジの集合である。

ノードは, 7つの情報源および商品が対応する。エッジは, あるノードが別のノードについて情報源となっていることを対象者が認識している関係を表す。認識されていない関係は対象者にとっての信頼形成に寄与しないと考える。また各対象者ごとに各エッジに1(情報源と認識している), 0(認識していない)を割り当て, これを調査対象者全体について集計した値をエッジの重みとして扱う。詳細については, 2.3.1でのべる。

2.3 信頼のネットワーク仮説

信頼はどのように形成されているのか。これまでの議論をふまえて2.3.1では, 1章で提示した4つの問題に答えるべく, 仮説を提示する。そして, これらを統一的に扱うために商品に対する信頼のネットワーク仮説を提案する。2.3.2ではネットワーク仮説を検証するためにネットワーク自己相関モデルを提示し, 検証方法ものべる。

2.3.1 仮説

問題1, 2に対する仮説をそれぞれのべる。

仮説1「商品に対する信頼は, 情報源により影響を受けている。」

仮説2「商品に対する信頼は、情報源のうち製造元と販売元に対する信頼から主に影響を受け形成される。」

仮説1について、商品の信頼に情報源との関係が機能している事例をあげる。まず、対象者がある食品をはじめて購入するときには、食品のラベル表示のようなさまざまな属性情報を収集する。しかし、生鮮食品の場合など、商品に記載されている情報自体が少ない場合もある。この場合、店員のお勧めなど販売元の情報に頼ることが多くなる。つまり商品から得られる情報が少ないときは、商品そのものより情報源に対する信頼が、商品の信頼に影響すると考えられる。また、すべての商品の属性を調べたうえで比較検討するのはコストがかかるため、だれかの勧める商品を選択したり、信頼できる店であるため購入を行う場合がある。このとき、情報源への信頼がその商品の信頼に寄与していると考えられる。以上の2つの理由により、仮説1を導いた。

次にこれをふまえて、問題2に対する仮説をのべる。情報源の中では製造元と販売元の信頼が、商品に対する信頼に対し重要になると考えられる。一部の商品について友人・知人、マスコミなどが情報源になることはある。しかし、あらゆる商品について一つ一つ第三者の情報源に頼ることは探索コストが高くなる。このことが信頼が重要になるはじめの理由として挙げられる。たとえば、すべての商品選択において友人のクチコミを利用したり、インターネット上での評判を確認することはコストが高い。このため、通常ひとは信頼のおける店で購入をしたり、その商品に記載されているメーカーが信頼できるかどうかで判断することが多い。2つめの理由は、製造元や販売元は、友人・知人やそのほかのメディアとは異なり、商品の製造・販売を直接行っており、商品に対して大きな責任のある主体であるからである。3つめの理由は、商品に対する信頼ではないが、販売店と顧客の関係が継続購買などのロイヤリティ形成に大切であるという知見があることが挙げられる[Sirfeshmukh 02]。このロイヤリティの背後に商品・サービスへの信頼があるとすると、販売店との関係が重要であるといえる。以上の理由から、問題の2に対して仮説2を示した。続いて、問題3、4に対する仮説をのべる。

仮説3「さまざまな情報源に対する信頼は、相互に影響しあっている。」

仮説4「さまざまな情報源に対する信頼は、友人・知人の信頼が中心となって形成されている。」

前段では商品に対する信頼は、主に製造元や販売元といった情報源の信頼により形成されると仮定した。これら情報源に対する信頼はどこからくるのであろうか。一般には、Fogg & Tsengの研究でもあるように情報源の属性情報も重要であろう[Fogg 99]。ただし、こうした情報が少なく判断が難しい場合や信頼形成のために情報を調べるコストがかかる場合もある。このときその情報源の信頼は、どのような情報源を通じて「情報源についての情

報」がつたわっているかにも影響を受ける。たとえば、テレビや新聞などのマスメディアを信頼している人は、そのメディアがある製造元を好意的に取りあげていると、その製造元も信頼する。またテレビや新聞などへの信頼も、テレビ番組の情報を紹介した友人や知人に対する信頼に影響されることがある。このようにさまざまな情報源に対する信頼は、相互に影響しあっていると考えられる。以上より問題3に対して、仮説の3を提示した。

ではさまざまな情報源に対する信頼を考えた場合、対象者にとってのもっとも重要で、かつ他の情報源への影響力も高いものはなにか。その答えは友人・知人に対する信頼ではないかと考えられる。理由としては、まず友人・知人はあらゆる情報源についての情報源になることが可能で、すべての情報源の信頼に対して影響を与えることが可能だからである。ネットワーク理論でいいかえると出次数が多く中心性が高いのでほかに影響を与えることができる。たとえば、友人・知人からメディアの情報のみならず製造元、販売元の情報などあらゆる情報源の話の聞くことができる。2つめの理由としては、友人・知人と対象者は直接コミットしているため、裏切ったり誤った情報を与えた場合、友人自身の利害に影響するため一般に信頼は形成されやすいからである。さらに、消費に関する情報についてのメディアに対する信頼の中では友人・知人などの信頼がマスメディアより高いとされており、友人・知人が情報源として重要な機能があることが推測できる[宮木 09]。以上の考察をもとに問題4に対して仮説4を提示した。

以上のような4つの仮説が成り立つのであれば、商品の信頼は情報源の信頼に依存しており、加えてこれらの情報源同士の信頼も相互に依存しているという複雑な構造をもつ。たとえば、商品の信頼はその製造元や販売元の信頼により成立し、さらにこの二つの信頼も友人、知人の信頼により成立しているということである。そこで構成要素が複雑に関連していることをまとめて扱うため、これまでのべた構成要素、ネットワークの定義、4つの仮説を元にして以下のネットワーク仮説を提案する。本仮説は構成要素の関係を図示化した概念的な枠組みである。4つの仮説の検証は、具体的な商品を設定したうえで、この統合した仮説の要素をつかい、関連する質問紙調査を行い、4.2でのべる統計的モデルを用いた検定および比較によって行われる。商品により実際に推定されるモデルは異なる。

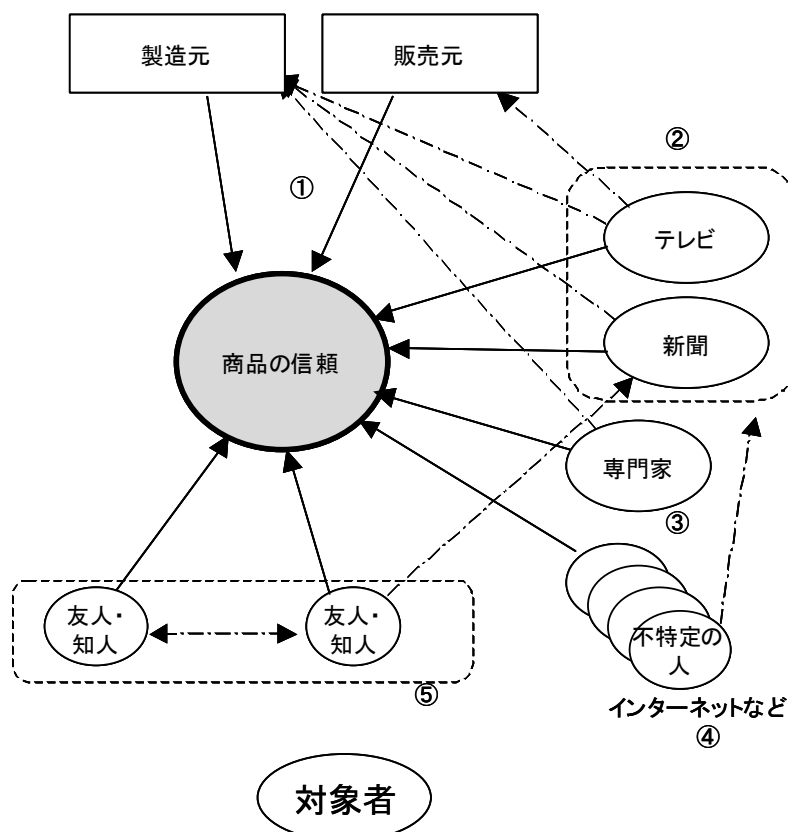


図 2-1. 信賴のネットワーク仮説

信賴のネットワーク仮説とは、対象者が商品、および情報源に対して信賴をもち、この値が「それぞれが互いに他の情報源になっているという依存関係」によって影響しあっていることを示す認知的枠組みである。仮説 1, 2 と関連する商品の信賴の情報源からの影響は、製造元や小売の信賴①、そしてテレビや新聞などのマスコミのマスメディアに対する信賴②と専門家に対する信賴③、そしてインターネット上にある情報など自分と直接関係のない人に対する信賴④、友人・知人（自分と直接的に関係があり、そこでコミットメントにより利害関係が形成されている）に対する信賴⑤で形成される（図 2-1：実線）。また情報源の信賴についても、相互に影響しあっている。①, ②, ③, ④, ⑤（図 2-1：点線）。これは仮説 3, 4 に関連する変数間関係をすべて表現したものである。依存関係のリンクは、矢印の元が A、先が B であるとすると、「A は B の情報源である」という関係を表している。

これらの解釈についてのである。たとえば、商品に対して製造元からリンクがある場合、「商品について製造元が情報源となっており、さらに商品に対する信賴は製造元への信賴によって影響を受ける」と解釈する。また、友人・知人からテレビへリンクがあるときは「商品に関するテレビ番組についての情報が、主に友人・知人を介して入手しており、友人・知人への信賴がテレビ番組への信賴にも影響を与えている」と解釈される。

このモデルを既存の研究と比較すると、加藤(2006)の研究における「評判や発信者」が本研究における情報源の部分に該当する[加藤 06]。また、原ら(2007)のモデルとも近いが、さまざまな情報源を扱い、その相互作用を想定しているという点で異なっている[原 07]。

2.3.2 統計モデルによる検証方法

では、仮説を検証するにはどうすればいいか。ここでは仮説検証に用いる統計的モデルと、仮説の検証方法についてのべる。

本研究では商品と情報源、情報源同士の信頼の依存関係を同時に分析するために、ネットワーク自己相関モデルを用いる。ネットワーク自己相関モデルとは、計量経済学で開発された、空間的自己相関モデルをもとにしたものである。空間的自己相関モデルは、隣接した地域が同じ傾向をしめすかどうかという影響をみるものである。このモデルは地理上で隣同士の依存関係を、 W という隣接行列(ある集合の要素同士の関係を隣であれば1、隣でなければ0で表した行列)で表現したうえで各地域がもつ属性値がこの隣接行列に依存しているかをみる。この空間的自己相関モデルと異なり、依存関係において地理的な距離などの制約をおかず、ネットワークを表現する隣接行列を W に利用したものが、ネットワーク自己相関モデルである[Leenders 02, Roge 02, 渡辺 05]。この W は本研究との対応では4.1のネットワーク仮説にある各エッジに該当する。それぞれの対象についてどこが情報源になっているかを調査し、 W を作成する。

ネットワーク自己相関モデルを用いる理由は、すでに与えられている依存関係 W の影響が判定できる、すなわちネットワークによる影響が判定できるという点にある。また信頼が、互い依存して相互作用があることを組み込んで推定できる点も挙げられる。線形回帰だけでは、このように相互作用を組み込むことは難しい。特に今回、ふたつの要素の信頼の依存や相関を単純にみるのではなく、その間に「情報源であるか否か」といった条件も考える必要があること。また、ある信頼を説明する別の信頼自体もさらに別の信頼から影響を受けており、説明に用いる変数間の作用を同時に捉える必要があることから相関分析や線形回帰は用いていない。

具体的にモデルを説明する。ネットワーク自己相関モデルには、二種類の外乱モデル(disturbance model)と効果モデル(effect model)がある。今回は、特に信頼の関係を直接的に調べるために、以下のような効果モデルを使用する。

$$y = \rho Wy + x\beta + \varepsilon$$

W は上述の依存関係、すなわちネットワークを表す隣接行列である。本研究ではネットワークのエッジの部分にあたる。 y は商品・情報源に対しての信頼である。 x は情報源の属性などである。 ρ は y が W により依存しているというネットワーク効果、 β は属性の効果、 ε は $N(0, \sigma^2)$ である。ただし、今回は情報源の信頼がネットワークに依存していることに焦点をあてるため、下記のような、単純なネットワーク効果 ρ のみがきいている一次のネッ

トワーク自己相関モデルを用いて検定を行う。これを用いた理由は、属性による効果を分離せず、属性の効果も含めた総合的なネットワーク効果（総合効果）を検証するという意味がある。

$$y = \rho Wy + \varepsilon$$

このモデルは、空間的自己相関モデルにおける一次空間的自己相関モデル(First-order autoregressive model)に相当する。

パラメタとなる、 ρ および σ^2 の推定方法についてのべる。最小二乗法で推定すると、バイアスが生じることが知られている。そこで、尤度関数の最大化によってパラメタを求める。また、近年ではベイズ推定をしているケースもある[Yang 03]。このネットワーク自己相関モデルの尤度関数は

$$L(y|\rho, \sigma^2) = \frac{1}{2\pi\sigma^2(n/2)} |I_n - \rho W| \exp \left\{ -\frac{1}{2\sigma^2} (y - \rho Wy)' (y - \rho Wy) \right\}$$

となり、これから対数尤度関数を導出し、最適化法を利用して最大化を行ってパラメタを推定する。このとき n はベクトルの数とする。本研究では、収束の早い quasi-Newton 法を用いる。

ところで、ネットワーク自己相関モデルで用いる W の計算・作成には、いくつかの方法が考えられる[Leender 02]。たとえば、リンク関係の有無をバイナリーデータとしてあつかったものや、類似関係を計算して行列化したものなどの利用が可能であり、さまざまな値を用いることができる。これら W の選択については Yang らの研究(2003)でも議論されており、ここでは複数の関係を同時に扱っている[Yang 03]。本研究では先ほどのべたようにそれぞれの対象に対し、どこが情報源になっているかを聞くことで W を作成し、また各対象に対する影響力を比較するために対象への影響の合計が同じになるようにする。すなわち、各行の合計が 1 になるように標準化を行う[Leender 02]。このような W の作成方法は、各行のスケールをそろえているため、各要素への影響をみるという点では妥当性は高い。質問項目の詳細については、2.4.1 の調査設計でのべる。

ネットワーク自己相関モデルをふまえて 4 つの仮説の検証方法をのべる。

まずネットワークの効果 ρ が有意であることが示されれば、具体的に既に得られている W という構造が有意に機能して、ノード同士は相互に依存しているということがわかる。つまり、ネットワーク仮説が支持される。これにより情報源の信頼と商品の信頼が依存しているという仮説 1 も支持される。

このネットワーク効果が認められると、仮に A が B の主な情報源である場合は、A の情報源としての信頼が高ければ、B の信頼も高くなるということを意味する。このネットワーク効果の検定がクリアしたことをふまえ、次に商品に対する情報の多くが、製造元や販売元からきていることを順位などで示せば仮説 2 が支持される。

仮説 3 が支持されるか否かは、仮説 1 と同様にネットワークの効果が示すことで判定できる。すなわち、 W という構造が有意であれば、情報源同士が相互に依存していることがわ

かる。

仮説4については、ネットワーク効果が示されたうえで、友人・知人が商品やすべての情報源に対する情報のもととなり、影響が大きいことを示せばよい。検証の方法としては、出次数中心性が高いこと、すなわち友人・知人がリンクを出している数が多いことを示す方法、また影響力をしめす基準化した W の値を各ノードごとに合計した値がもっとも高いものを選ぶという二つの方法がある。

2.4 検証

本節では、調査に基づいて仮説の検証をする。まず、調査設計をのべ、つぎに各商品と情報源の調査結果を順次のべたあとに、モデルの推定を行い、最後に考察を行う。

2.4.1 調査設計

全国の男女 632 名に対し、インターネットによる質問紙調査を行った。調査日時は 2008 年 10 月 20 日から 22 日である。男女比率は、全国比率に揃えた。被験者は以下の構成になった。

表 2-1. 被験者の構成

年齢	男性					女性					全体
	20-29	30-39	40-49	50-59	60-69	20-29	30-39	40-49	50-59	60-69	
人数	56	71	58	70	57	54	68	57	71	61	623
比率	9.0	11.4	9.3	11.2	9.1	8.7	10.9	9.1	11.4	9.8	100.0

次に質問項目についてのべる。

信頼を調べる対象としては食品を扱い、4 品目(食パン、牛乳、緑黄食野菜、牛肉)を選び、それぞれの品目に対して質問項目を回答させた。この商品は、一般的によく食する商品(加工食品)、以前に問題のあった商品(加工食品)、生鮮食品、以前に問題のあった生鮮食品という区分を行い、それらに該当する食品を 1 つずつ選んでいる。調査した主な項目は、仮説に基づき以下の表 2-2 のようになる。具体的な質問紙については付録に記述している。

表 2-2. 仮説と調査項目

問題	仮説	統合した仮説：検証のための概念モデル	概念	調査項目
問題 1	仮説 1「商品に対する信頼は、情報源により影響を受けている。」	信頼ネットワーク仮説 (図 2-1) 『対象者が商品に対しての信頼、さまざまな商品の情報源に対して信頼をもち、この値が「それぞれどこが情報源になっているかという依存関係」であるネットワークによって影響しあっている。』	信頼	商品に対する信頼①
問題 2	仮説 2「商品に対する信頼は、情報源のうち製造元と販売元に対する信頼から影響を受け形成される。」			食一般に関しての情報源(製造元, 販売元, マスメディア, 友人・知人・家族, インターネット等)に対する信頼②
問題 3	仮説 3「さまざまな情報源に対する信頼は、相互依存しあっている。」	利用する統計モデル： ネットワーク自己相関モデル	ネットワーク (情報源としての依存関係)	商品の情報源③
問題 4	仮説 4「さまざまな情報源に対する信頼は、友人・知人の信頼を中心に形成されている。」			情報源についての情報源④
その他			属性	デモグラフィック変数⑤

① 商品に対する信頼

「非常に信頼している」「まあ信頼している」「どちらともいえない」「あまり信頼していない」「まったく信頼していない」の 5 段階尺度で測定。(付録:Q1)

② 食一般に関しての情報源に対する信頼

(製造元, 販売元, マスメディア, 友人・知人(家族), インターネット等)

それぞれ「非常に信頼している」「まあ信頼している」「どちらともいえない」「あまり信頼していない」「まったく信頼していない」の 5 段階尺度で測定。(付録:Q2, Q3, Q5)

③ 商品の情報源

商品の情報源については個別の商品では、事前のテスト調査で想起が難しかったため、個別商品の上位カテゴリーで情報源を質問した。すなわち、食パンや牛乳の上位概念である「食品」を用いた。(付録:Q4)

④ 情報源についての情報源

信頼を形成するネットワークに用いる情報源同士の関係性を、ある対象がある対象の情報源となっているか否かを調べていくことで特定した。

(付録:Q6)

⑤ デモグラフィック変数

性別、年齢、都道府県、未既婚、子供の有無、世帯年収、個人年収、職業

2.4.2 結果

調査の結果についてのべる。各品目・情報源に対しての信頼の平均値を示したあとに、分析を容易にするためにこれらの中でもっとも信頼が高いものをえらぶ。つぎにこの品目とその情報源で構成されるネットワークを特定し、さらにネットワーク効果についての検定を行う。この品目に焦点をあてる理由には、信頼が高い商品はどのようにして信頼を高めているのかを検証する意味もある。信頼の高い品目についての結果をのべたあと、個別の商品についての検定結果をのべる。

はじめに、直近でみかけた各品目に対する信頼は以下の図2-2のようになった。それぞれ「非常に信頼している」「まあ信頼している」「どちらともいえない」「あまり信頼していない」「まったく信頼していない」に対し、5, 4, 3, 2, 1の得点をふって平均をとったものである。

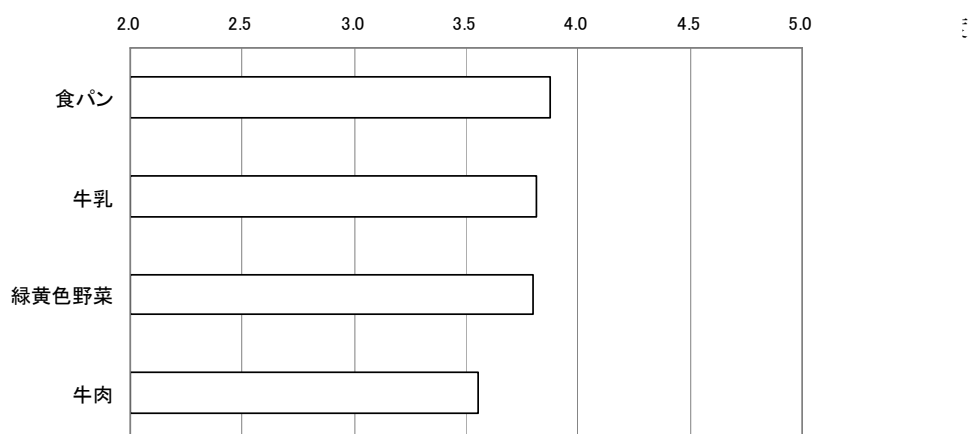


図2-2. 各食品の信頼

上記の結果をみると、各食品において大きな差はなかったが食パンで信頼が高く（牛乳，緑黄色野菜，牛肉に対して対応のある平均値の差の検定で有意水準 5%で有意差があった，それぞれ順に t 値 3.04, 3.63, 11.93, 自由度 622, p 値 0.0024, 0.0003, 1.00e-29; 多重比較の結果も有意であった），一方で牛肉など偽装問題になった商品に対する信頼が低くなっている．続いて，各情報源に対する信頼について記述する．

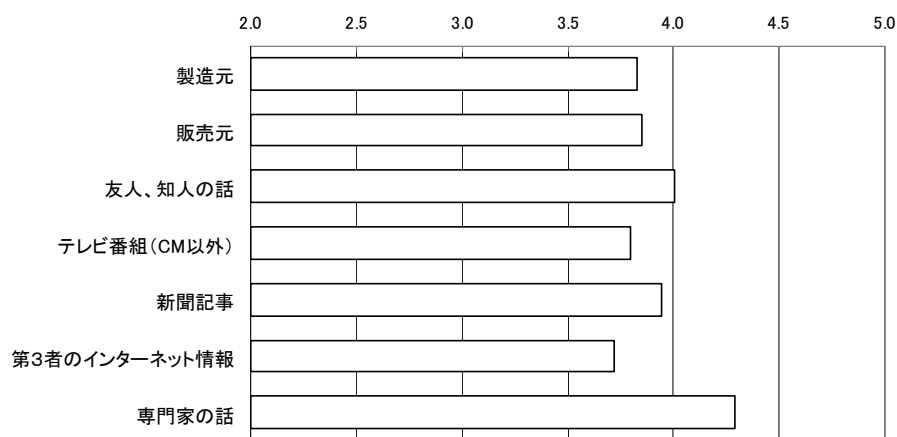


図 2-3. 食についての情報源に対する信頼

図 2-3 は情報源に対する信頼の平均値である．

情報源に対する信頼では，友人・知人のほなしといった直接人間関係がある場合が，信頼度が高く，さらに専門家の話も信頼度が高いことがわかった．

続いて，商品・情報源同士のネットワーク効果を調べる．まず，もっとも信頼の高かった食パンに焦点を当ててネットワーク自己相関モデルによる推計を行う．具体的には，商品と情報源の依存関係，情報源同士の依存関係，すなわち 2.4.1 の③と④についての質問項目の集計からネットワークを作成したのちに，個別の商品や情報源に対する信頼の平均値を用いて検証を行う．

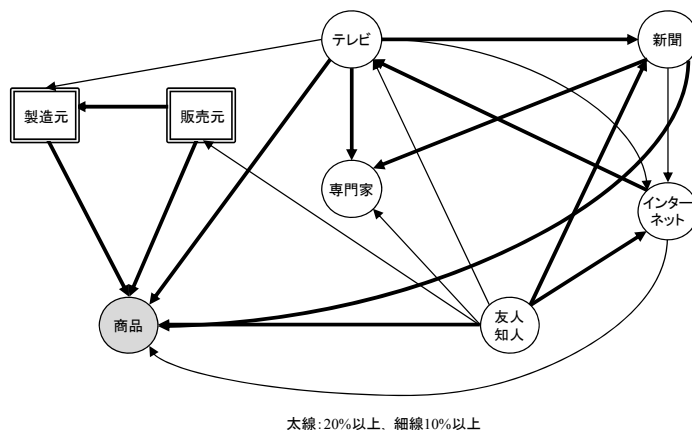


図 2-4. 情報源同士の情報源としての依存関係

食パンの情報源についてのネットワークを提示する。図2-4は、ある情報源を利用しているひとを母数としてその情報を入手するもととなっている情報源についての比率を計算し、図示化するために20%以上を太線、10%以上を細線で表したものである。たとえば、テレビから専門家へのリンクは専門家についての情報をテレビから入手するというを示している。

このような情報源の依存ネットワークを各行の合計が1になるように2.3.2で述べた方法で標準化したものをネットワーク自己相関モデルにおける W とした(付録2に記載)。そして、各情報源に対する信頼の平均値を説明変数および被説明変数に用いて、ネットワーク自己相関モデルを推計してネットワーク効果があるかどうかの検定を行った。まず食パンにおけるネットワーク効果とその誤差について検証を行う。

表2-5. 食パンにおけるネットワーク自己相関

変数	推定値	標準誤差	Z値	Pr(> z)
ρ	0.906	0.133	6.793	1.10e-11 ***
σ	1.471	0.149		

Residual standard error: 4.232 on 7 degrees of freedom

AIC: 37.88 BIC: 38.04

食パンにおける各係数の値は表2-5の通りである。 ρ がネットワーク効果、 σ は誤差の分散である。ネットワーク効果 ρ は、5%の水準で有意であった。このことから、商品・情報源に対する信頼は、相互にその情報源に依存していることがわかった。同様に、各食品についてもネットワーク効果の検証を行い、食パン以外の食品についても、ネットワーク効果が有意であることが確認できた。つまり食に関してさまざまな情報源同士が依存しており、またその影響が商品の信頼へと影響していることが確認できた。これにより仮説1, 3が確認できた。

表2-6. 各食品でのネットワーク効果

	変数	推定値	標準誤差	Z値	Pr(> z)
牛乳	ρ	0.9043	0.1336	6.770	1.29e-11 ***
緑黄色野菜	ρ	0.8991	0.1338	6.722	1.79e-11 ***
牛肉	ρ	0.8890	0.1350	6.587	4.48e-11 ***

再び食パンに焦点をあてて、個別の情報源の依存関係を図2-4でみる。食パンについては、その多くは製造元や販売元に依存しておることがわかった。実際、商品に対する情報源からの影響力を、標準化されたWを用いて集計してみると、図2-7のようになり、製造元や販売元の影響が強いことがわかる。このことから、仮説2が支持された。

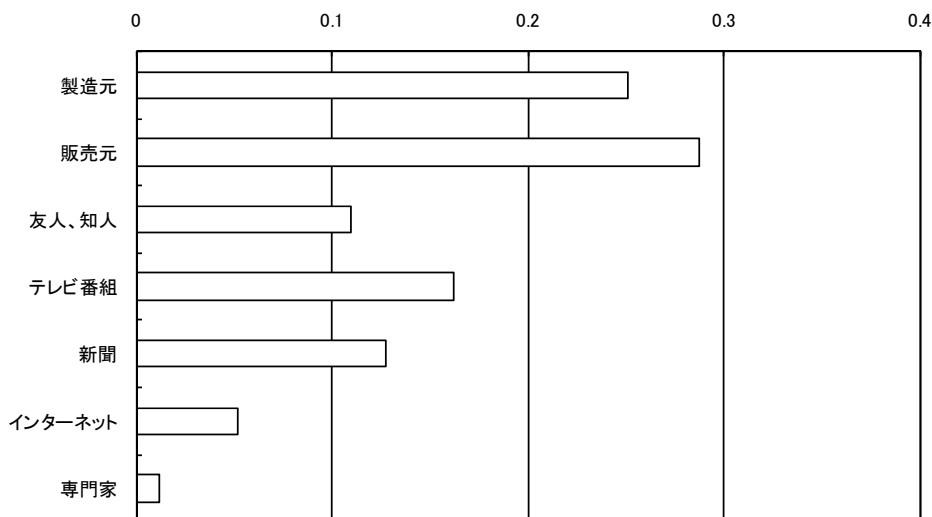


図2-7. 商品（食パン）への影響

また情報源としては友人・知人の話・テレビ・新聞から、エッジがでていることが多く、情報源に対しての信頼の鍵となっている。製造元や販売元へはテレビ番組や友人・知人からの影響があることがわかる。これについても、実際にネットワーク全体に与える影響度を標準化したWを各ノードごと（列ごと）に集計をしても以下のように友人・知人の影響度やマスメディアからの影響度が高いことが確認された（図2-8）。これにより仮説4が支持された。

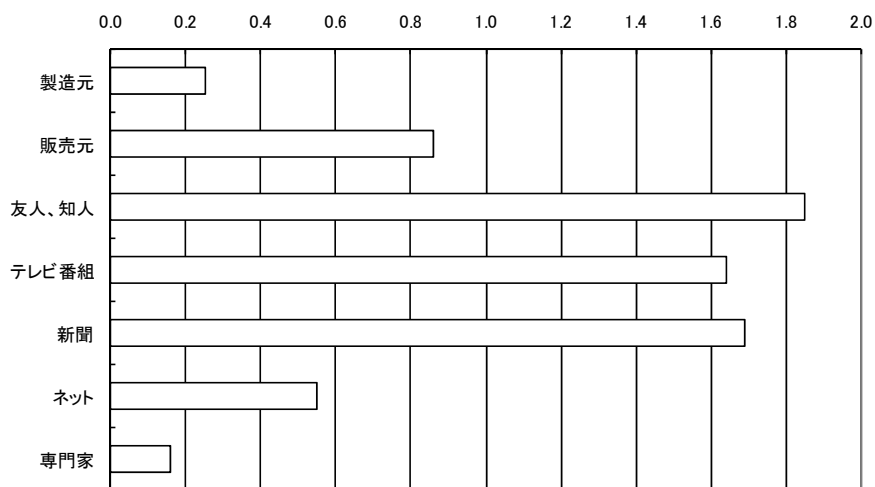


図2-8. 各情報源の全体への影響：各Wの合計値

2.4.3 考察

2.4.2 で行ってきた分析結果について詳細に考察を行う。

仮説 1 については支持された。ネットワーク自己相関モデルにおけるネットワーク効果が 4 つの品目すべてにおいて有意であり、商品の信頼と情報源の信頼との依存関係がみとめられる。そのため商品の信頼が情報源の信頼から影響していることが示されている。

仮説 2 も支持された。ネットワーク効果があり、また商品への影響を依存関係からみた結果、商品、特に食パンについての信頼には、図 2-7 のように製造元と販売元の信頼が重要であることがわかった。加えて、この製造元と販売元への信頼は他の情報源からの影響では、テレビや友人・知人からの影響があることもわかった。ただし、ほかに比べると低い。またその他の情報源からの商品への効果については、テレビや新聞、友人・知人という情報源の信頼を介して商品の信頼が高められていることもわかった。前者の商品への信頼が製造元や販売元から形成されるということは、これまでの知見とも一致するところであり、本研究の妥当性があるともいえよう。またその中で、製造元や販売元の信頼がさらにその情報源から影響を受け、高められているということは、これまであまり分析されていなかった新たな知見である。

つづいて、仮説 3 についても支持された。すなわち仮説 1 と同様にネットワーク効果が認められ、情報源同士の信頼が相互に依存していることが示されている。

仮説 4 についても支持された。仮説 3 が示され情報源同士の信頼が相互に影響していることがわかった。そのうえで、図 2-4 や図 2-8 にあるように友人・知人の信頼の他の信頼への影響力が高く、影響を与えている情報源の数が多いことがわかった。そのため友人・知人への信頼が重要であることが示された。加えて、信頼の方向ではなく情報の流れからみていった場合、つまりネットワーク仮説の矢印を逆からみていくと、多くの情報源に関する情報が友人・知人を介していることがわかる。つまり別の解釈をするとある種の信頼のフィルターとして機能していることも推測される。ただし、友人・知人に対する情報源としての信頼がどのようにして形成されているかは今後の課題として挙げられる。

このように商品そして商品の情報源への信頼はそれを取りまく情報源との関係性の中で構成されていることがわかった。

以上の知見の応用としては、製造元や販売元が個別の商品の信頼を獲得する方法への示唆などがある。具体的に信頼を獲得するためには、まずは製造元や販売元の信頼を得ていくことが重要である。この際、製造元の情報はテレビなどを通じて伝わっていくことが大切であるといえる。また、商品は、そのほかの情報源からも影響を受けていることから、信頼ある別の情報源を利用や活用していくことで、結果として商品の信頼性を高めることができる。特にこれらの信頼の中で重要な役割を果たしているのが友人・知人の信頼である。友人・知人はエッジの出次数（5本）からみてもさまざまな情報の中核、フィルターとなっていることがわかる。これを直接利用していくのは難しいが、信頼におけるクチヨミの

重要性を示唆している。仮に友人・知人に信頼を得たとき、ネットワーク全体の信頼が高まった結果、商品への信頼が形成されることが考えられる。

信頼をもった対象同士が相互に関連しあい、信頼を高めあうということはこれまでも研究されてきたが、今回統計的なモデルを導入することで確認できた。

2.5 関連研究

本研究の特徴は、信頼を情報源同士のネットワークから捉えて帰納的かつ形式的にモデリングを行ったことにある。特に、今日食品偽装事件などで社会的に問題となっており、かつ商品にある表示情報が少なく不確実性の高い、食品に対する信頼の問題に焦点をあてて分析および検証を行った。そのため本研究は、社会ネットワーク研究とマーケティング研究の境界に位置する。

そこで 2.5.1 で信頼のネットワークモデルに関する従来研究、2.5.2 でマーケティング分野における信頼・ネットワーク研究について、本研究との関連や違いをのべる。

2.5.1 信頼のネットワークモデルに関する関連研究

信頼にネットワークモデルを導入した関連研究には、人間関係のネットワークに焦点をあてたものが多い[山岸 98; Patnum 93]。ネットワークに人間以外を含む異種の要素を加えたものとして、アクターネットワーク理論がある[Latour 87]。本研究は人に限定したものではなく、この様な異種の要素があるネットワークを扱う。また、本研究で想定するような情報源を対象とした例として、原らによるウェブ信頼性の研究がある[原 07]。原らは、ウェブ上の情報の信頼を情報源に注目してモデル化し、権威型とコミュニケーション型の2つを設定した上で、さらに情報が伝わる構造により4タイプに分類している。ただし、これらの研究では定性的な分析法を適用しており、統計的な検証は行われていなかった。その点で本研究とは異なっている。

本研究と同様に信頼のネットワークモデルを形式的に記述し、数式を用いて分析を行った研究もある。たとえば、数理社会学関連の分野で、Audun&Ross らは、オブジェクト論理を用いて個人間の信頼に関する分析を行っている[Audun 06]。社会ネットワーク分析においては佐藤嘉倫らが、市場における信頼性について、人同士の関係をモデリングを行ったうえで信頼関係とコミットメント関係についてシミュレーションで検証を行っている[佐藤 05]。さらに、情報工学分野では、人のリンク関係に基づき、各人の信頼性もしくは重要性を計算するリンク解析の研究も多い[Matsuo 04]。前者の2つの研究では、定性的な調査やシミュレーションを適用して分析しているが、本研究では定量データに基づいて帰納的に検証できるようにしている点が異なる。また、既存研究はなんらかの実体間の関係をネッ

トワークを用いてモデル化するものであるが、本研究は、商品に対して信頼を寄せる対象者の認知的なモデルとしてネットワークを用いていることも違いとして挙げられる。

2.5.2 マーケティングにおける信頼とネットワーク研究

本研究と関連のあるマーケティングの分野でも、信頼は幅広く研究されている。マーケティングにおいては、生産者と流通者の間で信頼が必要であることに加え、販売者と購入者との間での信頼がロイヤリティ、継続購買などに影響することなどが示されている [Doney 97; Dwyer 87; Sirfeshmukh 02]。

マーケティング領域におけるネットワーク研究もある。久保田はマーケティングには組織と組織、組織と個人、組織内、個人と個人、個人の中のネットワークという5つの種類が存在するとしている [久保田 09; Burt 92; Rindfleisch 01]。個人と個人の例では、山本らが他者との関係性を組み込んだレコメンデーションシステムのモデルを作成して評価を行い、個人と個人のネットワークが購買に影響を与えることを示している [山本 07]。個人内のネットワークについては、ブランド連想に関するいくつかの概念に対して、ネットワーク分析を適用した研究がある。たとえば、ソーシャルマーケティングの分野で、リサイクルのイメージについてラダリング法を適用し、その結果に対してネットワーク分析を行った例がある [Bagozzi 96]。本研究はネットワークの分類ではこのような個人の中の認知ネットワークに位置している。

最後に、商品のうち特に食品に関する信頼は、リスク研究において注目されている。Frewer, L. J., Fischer, A. R., Van den Brink 他 (2008) は、食品リスクの低減において、食品が農家から消費者へと繋がっていくフードチェーンと、消費者が情報を得て食品の購買にいたるプロセスとの接点で考える必要があるとしている [Frewer 08]。David (1992) や Poortinga & Pidgeon ら (2005) は、後者の消費者側からのプロセスで、直接効果や間接的なものも含め消費者の食品や製造元に対する信頼が重要であることを示している [David 92; Poortinga 05]。このように、本研究であつかう信頼のモデリングは、リスク低減の観点からも重要な課題である。

こうした観点から考えると、本研究は主に消費者からみた個人内の認知的なネットワーク研究に位置しており、その目的は個人における商品やさまざまな情報源への信頼、そしてその相互関係の認知構造について明らかにすることにある。

2.6 本章のまとめ

商品をめぐる信頼の形成についてのネットワーク仮説を提唱し、さらにこれらを統計的モデルとしてとらえて、定量調査により実証することができた。

本章で行ったことはつぎの2つにまとめられる。

まず、商品に関する信頼のネットワークを形式的に定義したことである。もうひとつは統計的ネットワークモデルを信頼のモデリングに応用したということである。特に、ネットワーク自己相関モデルを活用して仮説の検証したことが挙げられる。この際、その他考えうる多くの変数を考慮せず、あえて単純なネットワーク効果のみが存在するモデルを用いた。このひとつの効果だけでも信頼に関しての説明力があつたことは意義があると考えられる。

また、詳細に情報源などの個別の信頼を追及していく際には、今回は分析の対象にしなかった、判断している人自身や商品、情報源の持つ属性の寄与、その他の情報源や情報源の細分化も考慮する必要があるだろう。すなわち、新たな要素を組み込んだモデルについては今後の研究の課題としたい。また拡張として、商品における購買プロセスの変化のなかでどのように信頼のネットワークが変化しているのかについて扱うことが挙げられる。

ただし、問題としてある商品の情報源としての関係と、情報源同士の情報源としての関係のスケールがやや異なっていることが挙げられる。前者は個別商品レベル、後者は食レベルでとらえている点がある。これは調査において具体的に取得できる限界との兼ね合いもあり今回はこのようなモデルおよびデータになったが、これらについては分離するかもしれない。もしくは別のモデルを想定するなど今後の課題として挙げておきたい。

加えて、本研究では情報源となっているかいないかという相互関係の有無のみをとらえたため、今後は情報源がどのような情報を発したかということも含めたさらなるモデリングが必要であろう。後者には既存の態度研究の知見[藤原 02]を組み入れていくことができる。

この章を通じた成果について、第1章との関連で述べると、Aの形式の問題に対する解法としてネットワーク自己相関モデルが適用できることを示したことにあつる。

ただし、アドホックな点としてはグラフにおいて個別のノードを信頼の程度としたところ、関係については情報源としての関係を用いたところがある。この部分を変更すれば、同じように「G(重み付き)が1つだけ与えられているときにリンクを介してノード同士がどのように影響しているかという問題」に対して適用できる。応用が可能なものとしては、組織を例にあげると、会社における人間関係を関係とし、個々の人の生産性を影響変数としてその関連を扱うことができるであろう。

ただし本手法の限界としては、関係があらかじめ既知であることが前提となっているため関係がデータとして与えられていない場合は推定が難しい。また関係が動的に変化していく場合や、そのノードに対する影響についての分析は直接扱えないことが挙げられる。5章で扱うネットワーク全体の指標と関係についての関連であれば扱うことは可能であるが、ノードへの影響をネットワーク自己相関という形式で取り扱いながら時間変化をみるということについては今後研究していく必要がある。

3章 帰納推論による時系列データからの関係構造の抽出

本章では、外部に影響変数があるなかで関係を学習する問題を扱う。特に、問題としては運動スキル、演奏スキルを扱い、影響変数としてはその熟達者のある演奏タスクというものに焦点をあて、それを説明する筋肉同士の協調関係を学習する。そのために筋電波形のような非定常な時系列データから関係パターンを抽出するための分析アルゴリズムを提案する。

まず、我々は与えられた多変量な計測データのそれぞれの時系列データに対して、局所的に定常なモデルを推定していく。次に、我々はこのように推定したものをクラスタリングし、それぞれのクラスタに対して記号を与え、結果として各時系列データを記号の系列データにする。つづいて、得られた系列における記号に開始時点と終了時点を加えて、時区間表現系 Time-Interval-Symbol (TIS)に変換する。最後に得られた TIS から帰納論理プログラミングを利用し、「時区間関係」のパターンを抽出する。

これを通じて、B の課題である外部に影響変数があるなかで関係を学習する問題に対する提案枠組みの有効性を検証する。

3.1 背景と問題設定

今日、多くの身体的な技術や暗黙的な技術が後継者に伝承されていないという状況下で、熟達者のもつ身体的な技術、すなわちスキルを解明するのは重要な課題である。こうした課題に対し、センサーから取得した情報の解析、人工知能やデータマイニングの手法の応用、認知科学の知見等を用いて、人間の身体の暗黙的な知識の言語化を目指しているスキル分析はひとつのトピックとなっている。この中で、身体に取り付けた複数のセンサーから得られる時系列データから、マイニングなどの技術を利用して言語的知識を獲得することは、スキルサイエンスにおける中心的な課題となっている。こうして獲得された知識は、伝承につかえるほか、新しいスキルの提案にも結びつく。

このようなセンサーと知識化・言語化との関係は、記号接地問題[Harnad 90]の記号の初期段階における実世界との関連や、記号がすなわち汎化であるとする議論[伊庭 96]とも関係しており、知能の解明という、より大きな枠組みの中で捉えることも可能である。

この技術は、身体のみならずさまざまな分野での応用が期待できる。たとえば、ユビキタス社会の到来により、社会生活のさまざまな場面でセンサー情報の取得と活用が急務となっているなか、大量の時系列データを処理し、情報を圧縮してまとめていく技術は多くの場面で必要とされるだろう。

本章は、分節化やルール獲得などの時系列データマイニング手法を主軸に、センサー情報からの高次の推論に向けた情報処理の開発を目指す。具体的な応用分野として、楽器演奏のスキル分析を取り上げ、各筋肉同士の協調動作を、筋電から自動的に抽出する問題に取り組む。

この様な問題を形式的に表現する。筋電の分析は、 X をデータ、 t を時間としたとき、非定常である時系列 (T) に複数のパターンがあり、このような系列が N 個存在する場合にどのように協調動作のルールを見つけるかという問題として定式化できる[古川 06]。

$$T = \{X_1, X_2, \dots, X_i, \dots, X_t\}$$

$$TS = \{T_1, T_2, \dots, T_n, \dots, T_N\}$$

これまでの研究と異なり、スキルサイエンスにおいて重要なのは、定常的な時系列を対象として相関を求めるのではなく、依存関係が非定常性を有していることを想定し、単純な相関だけでは取られない時間的に離れた依存関係を扱う、たとえばある系列が状態Aのあと、しばらくして別の系列で状態Bが発生するという関係を扱うということである。さらに外部から与えられる情報（記号データとして、楽譜情報や被験者情報など）を考慮する必要があるということもある。

従来の時系列解析としては、計量経済学ではベクトル自己相関モデルをはじめとした方法が用いられ、最近では相関のある時系列同士をグラフストリームと捉える研究[井手 07]などがあるが、これらの技術では相関を想定しているため上記のような複雑な依存関係を扱うことが難しい。

上記をふまえ本章では、非定常な時系列データ群に対し、まず系列ごとに局所的にモデルが変化していると考え、情報量規準を利用して推計を行う。またこのモデル同士が時区間をもった関係で表現できると考え、時区間データマイニングの方法を導入して分析していくことを試みる。またその際に、述語表現と帰納論理プログラミングを用いることで、時区間の関係のみならず、より柔軟な知識獲得を可能とするプラットフォームを提案する。

これにより、センサーデータから、情報量規準を利用したセグメンテーションを経由して、最終的にこれまで開発された記号処理の技術による推論への橋渡しが可能になると考えられる。

以降、3.2 で具体的な手法について説明を行う。3.3 では実験を行い、提案手法を実際に熟練者から取得した筋電時系列データに適用し、運動学的な観点から解釈可能な記号ルールを抽出する。これにより、センサー情報から高度な推論への結びつけの可能性を示す。3.4 では本手法の特徴をのべ関連研究との比較を行い、最後に 3.5 でまとめを行う。

3.2 提案手法

3.2.1 提案手法の概要

本章における提案手法の概要をのべる。入力となるデータについて再度提示する。
N 系列の時系列データ群

$$TS = \{T_1, T_2, \dots, T_n, \dots, T_N\}$$

T_n は X をデータとした t 個の時系列データ

$$T_n = \{X_{n1}, X_{n2}, \dots, X_{ni}, \dots, X_{nt}\}$$

とする。

入力をこの TS として、出力は述語論理により時区間とともに記述した時系列モデルの表現系を a sequence of Time-Interval-Symbol(TIS)で記述を行った、モデル同士の時間的な関係構造(Temporal relational pattern(TRP)となる。TIS や TRP の詳細については 3.2.4 でのべる。これらをふまえ提案手法は、以下の 3 つのステップから構成される (図 3-1)。

Step 1. 再帰分割推定アルゴリズムによる分節化：

入力した時系列群 TS から T_n を取り出し、情報量規準を利用した再分割推定アルゴリズムを適用し、局所的に最適な時系列モデルを推定し、 T_n を時系列モデル集合 best_models_n で表現する。

Step 2. 時系列モデルのクラスタリングによる記号化：Step 1 で得られた時系列モデル (best_models_n) をクラスタリングした結果、クラスタ集合 C_n をえる (ただし、元となるクラスタにインデックス I をふった： C_{ni})

Step 3. 帰納論理プログラミングによる時区間関係の抽出：各クラスタ C_{ni} に記号を割り当て、時系列データを述語論理による表現系 TIS に変換する。記号同士の時間的な関係

構造 TRP を時区間論理を組み込んだ帰納論理プログラミングを用いて抽出する。
以下, Step 1 から Step 3 について, 具体的な手法をのべる。

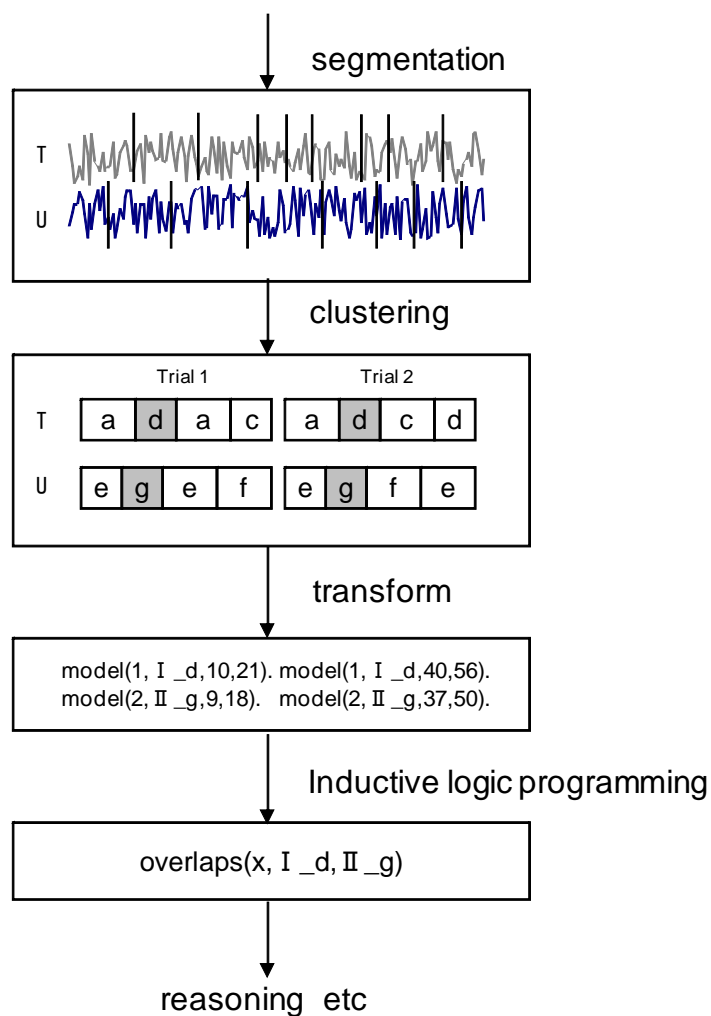


図 3 - 1. 提案手法

3.2.2 再帰分割推定アルゴリズムによる分節化

まず、Step 1 を説明する。ここでの目的は、時系列データを時間軸にそって分割し、各領域に対して、パラメタ数や値を最適化した（異なる）時系列モデルを対応させることである。この際の領域分割方法および領域ごとの時系列モデルを自動的に定めるために、本章では再帰分割推定アルゴリズムと呼ぶ手法を提案する。再起分割推定アルゴリズムは、情報量規準でモデルの当てはまりを評価することで時系列データの自動的な分節化を可能にするもので、一定の条件が満足されるまで、領域を最適に二分割する操作を再帰的に繰り返す。全体の流れは以下の通りになる。

1. データのあてはめに用いる時系列モデルの種類を選択する。
2. 時系列群 TS より 1 つの時系列 T_n を選び、全体へのモデルを当てはめて (estimate)、最小となる情報量基準(AIC)[Akaike 73]の値 (one_model_aic) を計算する。
3. 次に同じ系列を任意の時点 i 時点 ($1 < i < t$) で分割し、分割点の左右の領域に対して、あるパラメタ数内 (paranum) で時系列モデルを推定し、それぞれの最小の AIC 値を計算する (two_segment)。
4. 左右のモデルの最小 AIC 値の合計 (two_sum_minaic) が最小になる時点で最適分割点 (segpoint) として、時系列データ T_n を分割する。
5. 分割後、最小 AIC (minsum_aic) と分割前の AIC の値(one_model_aic)を比較し、改善されている場合、さらに分割した左右の領域に対して、1 以下のプロセスを繰り返す。
6. 時系列モデル群 (best_models) を出力する。

図 3-2 に簡略化したアルゴリズムを載せる。再帰分割推定アルゴリズム全体を regseg、一時系列を一回だけ二分割するアルゴリズムを two_segment として、a_model は推定したパラメタの値を示している。モデルのパラメタ数がデータ数を上回ってしまう場合は計算を停止するような条件を課している。(paranum は今回任意に近似精度は高かめに 5 とした。なお、2, 3 でも高い精度は確保された)

```

[one_model_aic,a_model]=estimate(T)
Algorithm best_models=regseg(T, one_model_aic,a_model)
for i=paranum to length(T)-paranum
    [two_minaic,models]=two_segment(T,i)
    sumaics=list(sumaics,[two_minaic,models])
end
minsumaic=min(sumaics)
[segpoint,two_minaic,models] =search(minsumaic,sumaics)
if minsumaic<one_model_aic
    models_right
        =regseg(T[1:segpoint],two_minaic_left,a_model_left)
    models_left
        =regseg(T[segpoint+1:length(T)],two_minaic_right,a_model_right)
    best_models=list(models_right,models_left)
else
    best_models=a_model
end
end

```

図 3-2. 再起分割推定アルゴリズム

データのあてはめに用いる時系列モデルとして、本章では、筋電の特徴の分析に用いられる自己回帰モデル（以下 **Aute Regression** モデル）を使用する。AR モデルは、筋肉の一活動状態に対応している[Loudon 92]。X が AR モデルに従う時系列データとすると次のモデルで表現できる。

ただし、 a_i は係数、 ε は平均 0、分散 σ^2 の誤差項、 p は次数である。

$$X_t = \sum_{i=1}^p a_i X_{t-1} + \varepsilon$$

推定方法は、計算の効率性を考えて、ユールウォーカー法を採用した。さらに AIC の計算方法は、L を尤度関数、 Σ を分散共分散行列、N をベクトル数、 θ を AR モデルのパラメタ $(a_1, \dots, a_p, \sigma^2)$ 、M をパラメタ数として、以下の式になる。

$$AIC = -\log L(\hat{\theta}) + 2(M + 1)$$

$$L(\hat{\theta}) = (2\pi)^{-N/2} |\Sigma|^{-1/2} \exp \left[-\frac{1}{2} X^T \Sigma^{-1} X \right]$$

提案する再起分割推定アルゴリズムは、時系列データ分析の分野で開発されてきた手法を応用している。たとえば、局所定常自己回帰モデル[北川 05]、ノンパラメトリック回帰における再帰分割線形モデル[Hastie02]の応用などの手法が提案されてきた。こうした方法を用いてトップダウンで分節化をしていく方法となっている。この方法は時系列モデルを導入することで情報量規準を用いて自動的にセグメントを抽出できるほか、トップダウン型の分節化により正確に変化点を抽出できるという特徴がある。このような本手法を用いたメリットや他の手法との比較などの詳細については 3.4 の関連研究で述べる。

3.2.3 時系列モデルのクラスタリングによる記号化

次に Step 2 を説明する. Step 2 では, 一つの時系列データ T_n から得られた時系列モデル群 `best_models` に対して `k-means` などの従来手法を用いてクラスタリングし, T_n ごとに部分集合 C を元とする集合族 $C_i (1 \leq i \leq m)$ をえる. これにより, 同じ系列で似た特徴を持つ部分を抽出することが出来るのみならず, クラスタに記号を割り振ることで記号処理に持ち込む前処理が出来る.

まずクラスタリングをするために時系列モデル同士の距離を定義する. 今回の筋電の特徴分析に用いた AR モデルについて, Picco [Picco 89] は, 二つの系列の距離を AR 係数をベースにしたユークリッド距離で定義している. これ以外にも周波数解析の結果を用いて距離を測る方法やカルバックライブラー情報量を用いた方法などが考えられる. 本章では, 計算上の簡便性からこの AR 係数をもとにした距離を採用する.

時系列モデルのクラスタリングについて, ここではクラスタ数 m を決定するための判定基準が開発されていることから, `k-means` を使用した. 判定基準があるものでは混合分布によるクラスタリングもあるが, 得られているデータ数では分布の仮定を検定することができないため用いなかった. クラスタ数は, Krzanowski らの提案 [Krzanowski 85] にしたがって, 次式による KL 値が最大となるものに決定した.

$$KL(k) = \left| \frac{DIFF(k)}{DIFF(k+1)} \right|$$

$$DIFF(k) = (k-1)^{2/p} W_{k-1} - k^{2/p} W_k$$

ここで, p はパラメタベクトルの次元数, W は群内の平均平方和の平均値であり, KL 値は, 群内平方和の平均の変化を表す. すなわち, 当てはまりの良さが大きく改善されているクラスタ数 k で最大となり, これをクラスタ数 m に用いる. これらは筋肉の活動状態数を表現していると考えられる. ただし, クラスタ数が多い場合は解釈が困難になるため, あらかじめ m は多くともデータ数の $1/2$ 以下とした. $1/2$ 以上であれば必ず状態が 1 つとなるクラスタが生成され, こうしたことが増えることを避けるためである.

最後に, 3.2.2 と 3.2.3 での提案手法の妥当性を検証するため, 局所的に定常な時系列データを用意し, その推定や分割点の特定を行った. 非定常な時系列は下記のような二つの AR モデル A, B に従うものでその組み合わせで 3 種類用意した. e は $N(1,0)$, AR 過程の初期値はいずれの場合も 0 とした. テストデータ 2 のデータは以下になる.

$$X(t+1) = 0.1X(t) + e \quad 1 \leq t \leq 200 \quad (A)$$

$$X(t+1) = 0.7X(t) + e \quad 201 \leq t \leq 400 \quad (B)$$

データ 3 は, このデータのあとに再び A のデータが来たもの, データ 4 は, データ 2 が 2 回繰り返したデータとする. このデータに本アルゴリズムを各 10 回以上適用し, セグメントの位置関係と個数が特定され, かつ同じクラス (class) に属するセグメントを同じも

のと特定しているかをテストした結果、以下のようになった。正答率はその比率、なおセグメントの特定は誤差の平均値、そのほかはそれぞれの推定値である。

表 3-1. テスト結果

テストデータ	クラスの特 定 (正答率)	セグメントの 特定	パラメータ class:a	パラメータ class:b
2(class2,t=200)×20	0.9	0±5.6	0.127±0.092	0.687±0.068
3(class2,t=200)×10	0.85	0±8.4	0.122±0.050	0.689±0.056
4(class2,t=200)×10	0.83	0±9.1	0.094±0.081	0.663±0.074

時系列データを正確に分割し、また AR の係数やその数も推定出来た (表 3-1)。クラスの特定については、正答率が高かった。ただし、セグメントの位置やパラメタについてはクラスタ数が増加するにつれ、精度が悪くなることなどがわかった。結果として、それぞれの時系列データの表現として時系列モデルの系列を得ることが出来る。以下はデータ 4 の推定およびクラスタリングの結果である (図 3-3)。

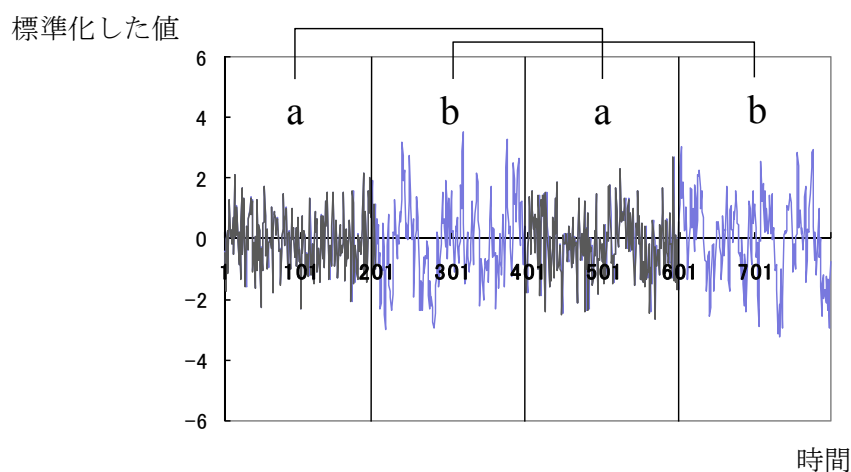


図 3-3. テスト結果図

3.2.4 帰納論理プログラミングによる時区間関係の抽出

最後に Step 3 について説明する。Step 3 の目的は、3.2.3 で抽出した特徴間から記号間に成り立つ関係構造を抽出することである。たとえば、「あるモデルがあるモデルの前にある」や「あるモデルがあるモデルと重なっている」などの時区間関係、ある特徴点と特徴

点の前後関係のルールを自動獲得することである。ここでは、帰納論理プログラミング (ILP)[Muggleton95, 古川 01]を用いて特に時区間関係を抽出する手法を提案する。ILP は、正事例、負事例（すなわちルールとは反する事例、たとえばある拍子でのルールを獲得する場合はそれ以外の拍子を負事例とする）そして背景知識を与えて関係概念を獲得する一階述語論理に基づく学習器である。

本研究では Aleph[Srinivasan 99]をもとに時区間関係を獲得するシステムを構築した。基本的には、まず、時系列モデルを区間を持った記号列として表現する。次に、それを一階述語論理で表現し、さらに、時区間に関する背景知識を ILP に与え、最後に推論により時区間パターン TRP を抽出する。

1. 時系列モデルの記号列への変換：Step 1 の分節化により得られた *best_models* に対して、時系列 T_n に対応する筋肉名を *Muscle*, Step 2 のクラスタリングで得た *C* の元となる部分集合 C_{ni} に任意につけた記号を *cluster* として、記号 *Muscle_cluster* を割り当てる。

2. 一階述語論理による表現：時系列モデルを述語論理で表現するために、1 で得られた記号に試行を区別する *id* を加え、時系列モデルの開始時点 T_{start} 、終了時点 T_{end} を対応させて以下の形式に変換する。

model(id, Muscle_cluster, T_start, T_end).

本章では以下、このように記号化した表現形式のことを TIS と呼ぶ。なお、異常値など1時点での特徴の場合でも T_{start} 、 T_{end} が一致するようにすると TIS で表現可能である。

3. 時区間論理による背景知識の導入：区間を持ったもの同士の関係構造を扱うために時区間論理[Allen 94]を導入する。時区間論理とは、時間的な区間を持った対象同士の関係を記述するための体系である。この体系の中では、二つの事象の時区間の関係を13種類に分類して表現している（表3-2）。たとえば、もしある試行 *id* において *X* が *Y* の前に起きたのであれば、“before(*id*, *X*, *Y*)”と表現する。表をみてもわかるように、対称性があるため結果としては全7種類となる。以下は、上記をルールとして ILP に加える場合の記述例である。記号化された2つの時区間 *X* と *Y* の間に前後関係が存在することを記述している。

before(*id*, *X*, *Y*):-

model(id, X, T_start1, T_end1),

model(id, Y, T_start2, T_end2), T_end1 < T_start2.

4. 帰納論理プログラミングによる TRP の抽出：3 でのべた時区間論理による体系を ILP

の背景知識に加え、さらに観測された時区間データの TIS を加えて、頻出するパターンや分類ルールを学習することで、最終的に時区間パターン TRP を抽出する。

表 3-2. 時区間論理

before(<i>id,X,Y</i>)	after (<i>id,Y,X</i>)	$X \longleftrightarrow$ $Y \longleftarrow \longrightarrow$
meets(<i>id,X,Y</i>)	met_by (<i>id,Y,X</i>)	$X \longleftrightarrow$ $Y \longleftarrow \longrightarrow$
overlaps(<i>id,X,Y</i>)	overlapped_by (<i>id,Y,X</i>)	$X \longleftrightarrow$ $Y \longleftarrow \longrightarrow$
starts (<i>id,X,Y</i>)	started_by (<i>id,Y,X</i>)	$X \longleftrightarrow$ $Y \longleftarrow \longrightarrow$
during(<i>id,X,Y</i>)	contains (<i>id,Y,X</i>)	$X \longleftrightarrow$ $Y \longleftarrow \longrightarrow$
Finishes(<i>id,X,Y</i>)	finished_by (<i>id,Y,X</i>)	$X \longleftrightarrow$ $Y \longleftarrow \longrightarrow$
equals (<i>id,X,Y</i>)		$X \longleftrightarrow$ $Y \longleftrightarrow$

このような ILP を用いた本研究の特徴は、システムの検証を ILP システムの健全性と背景知識の健全性におくことができるというシステムの検証性や、様々な知識の導入やデータの追加などが容易にできるという拡張性にある。これらの特徴については 3.4 章で詳細にのべる。

3.3 実験

本手法を評価するため、チェロにおける熟達者の演奏の分析を行った。まず単純なタスクを用いて演奏データを分析し、次にチェロとバイオリンの演奏の比較を行い、最後に楽譜にしたがった複雑なタスクについて演奏データを解析した。実験では、熟達した被験者 1 名に対し、前腕筋 (forearm), 上腕二頭筋(bicepsbrachii), 上腕三頭筋(tricepsbrachii), 三角筋-前(deltoid-f), 三角筋-後(deltoid-b), 脊柱起立筋-右(spinae-right), 脊柱起立筋-左(spinae-left)の筋電位を計測した。計測機器としては、PolymateAP1000(TEACCorpolation)を用いて、600Hz でハイパスフィルタ(高周波をカットする)をかけた。サンプリングレートは 200Hz とした。以下、運動学的観点からの考察も踏まえながら、それぞれについて説明する。なお、運動学的視点からの考察については、得られたルールを同時に計測したビデオ

オ、および楽譜と照会しながら、演奏に詳しい専門家との議論の上で行った。

3.3.1 単純なタスクからの個人スキルの抽出

まず、拍子の違いによる演奏の変化を調べるために8、4、2、1拍子でD線を演奏してもらった。それぞれの拍子で、16、8、4、2回の繰り返し動作を行い、拍子によらず演奏時間を一定とした。演奏回数は4とした。すなわち全体としては $16+8+4+2=30$ 、 $30 \times 4=120$ 試行となる。

上記のようにして得られた7個の筋電位の時系列について、前節でのべた方法による解析を行った。まず、それぞれの筋肉について3.2.2の再帰分割推定アルゴリズムを適用し、時系列モデル群を抽出した。次に、3.2.3のクラスタリングを適用した。表3-3にKL値の変化を参考に決定した各筋肉のクラスタ数を示す(ただし表中の「脊柱筋」は「脊柱起立筋」の略)。

表 3-3. 各筋肉のクラスタ数

前腕筋	上腕二頭筋	上腕三頭筋	脊柱筋左	三角筋前	三角筋後	脊柱筋右
4	4	3	3	3	4	3

ここで、KL値を用いたクラスタ数の決定について、前腕筋における詳細なデータをみる。前腕筋におけるクラスタ数2から5における評価値は表3-4の通りであり、結果として最適クラスタ数は4となる。モデル数の半分以上のクラスタ数については、状態の解釈が困難になるため考慮しなかった。前腕筋時系列データの分節化により得られた時系列モデルの数は全部で11個で、得られた4個のクラスタに含まれる時系列モデルの数はそれぞれ3、4、1、3となった。

表 3-4. 前腕筋のKL値

クラスタ数	2	3	4	5
KL値	0.959	0.994	9.501	0.107

前腕筋のデータから得られた4つのクラスタに記号A、B、C、Dを割り振って時系列データを記号の系列データにしたものは以下の表3-5のようになる。

表 3-5. 記号系列

順番	1 2 3 4 5 6 7 8 9 10 11
記号系列	B D A C A A B B D B D

最後に、こうして得られた時系列モデルの記号をもとに、3.2.4の記号処理を行って時区間パタンの抽出を行う。今回は、それぞれの拍子における全試行を正事例とし、残りを負事例とすることでルールを抽出した。たとえば、8拍子の場合には正事例が64、負事例が56となる。ルールを抽出する規準としては、分類精度が高いもの、すなわち、精度(=100×(被覆された正事例の数)/(被覆された正負例の数))が80%以上かつ最も高いものを採用し、さらに順に節数を増やし、より記述長が短くなるものを選択した。アルゴリズムは集合被覆アルゴリズムを用いている。最終的に帰納論理プログラミングを実行し結果として得られた全てのルールを示す。参考までに χ^2 乗検定(イエーツの補正)を行った結果もあわせて示す。

8拍子におけるルール ($Z = 4.8 > \chi^2(1, 0.05)$)

$\text{beat}(X) :- \text{during}(X, \text{forearm_A}, \text{tricepsbrachii_B}).$

$\text{beat}(X) :- \text{overlaps}(X, \text{forearm_A}, \text{bicepsbrachii_C}).$

1,2拍子におけるルール ($Z = 4.8 > \chi^2(1, 0.05)$)

$\text{beat}(X) :- \text{before}(X, \text{forearm_B}, \text{bicepsbrachii_C})$

次に解釈を行う。8拍子では「上腕三頭筋のB活動が起きている間に前腕筋のA活動がおきている」「前腕筋がA活動を開始した後に、重なって上腕二頭筋のC活動が起きている」、1, 2拍子では「前腕筋のB活動が起きた後に、上腕三頭筋のC活動が起きている」と解釈できる。これらを筋電の生データで見てみる。

図3-4は8拍子で抽出された $\text{beat}(X) :- \text{before}(X, \text{forearm_B}, \text{bicepsbrachii_C}).$ を示す事例の一部である。縦軸は筋電位、横軸は時間軸(1/200Sec)である。それぞれ活動度が前後に比べて下がったことがルールとして検出されていることがわかる。このルールは、前腕の活動が下がった後に上腕二頭筋の活動が抑えられるということを示唆している。これは1, 2拍子で演奏を行う場合は、腕の端から順に力を抜いて演奏を行っていることを示す。

一方、それ以外の8拍子におけるルールでは、前腕の活性が長期に渡っており、その間で上腕三頭筋の活性が見られることが観測された。このルールは、拍子を上げて演奏をする際に各筋肉のインピーダンスを調整し、腕を伸ばして振り子のような演奏をしている[古川05]ことをデータで示しており既存の研究と対応がつく。たとえば、八分音符に対する腕の振り子長の0.25mを実現するために、肘を支点とした振り子運動で運弓動作を実現する

ことを考えた場合、バイオリンでは可能であるが、チェロではそれが出来ない。それは、バイオリンでは弦の方向と前腕の方向がほぼ平行になるが、チェロでは弦の方向に平行になるのは肩と手首を結んだ線となるからである。そのため、全体のインピーダンスを調整し、上腕に力を入れ腕を伸ばして演奏をしている。これは図3-4の下図において上腕二頭筋の外側にあたる上腕三頭筋を固めていることになる。

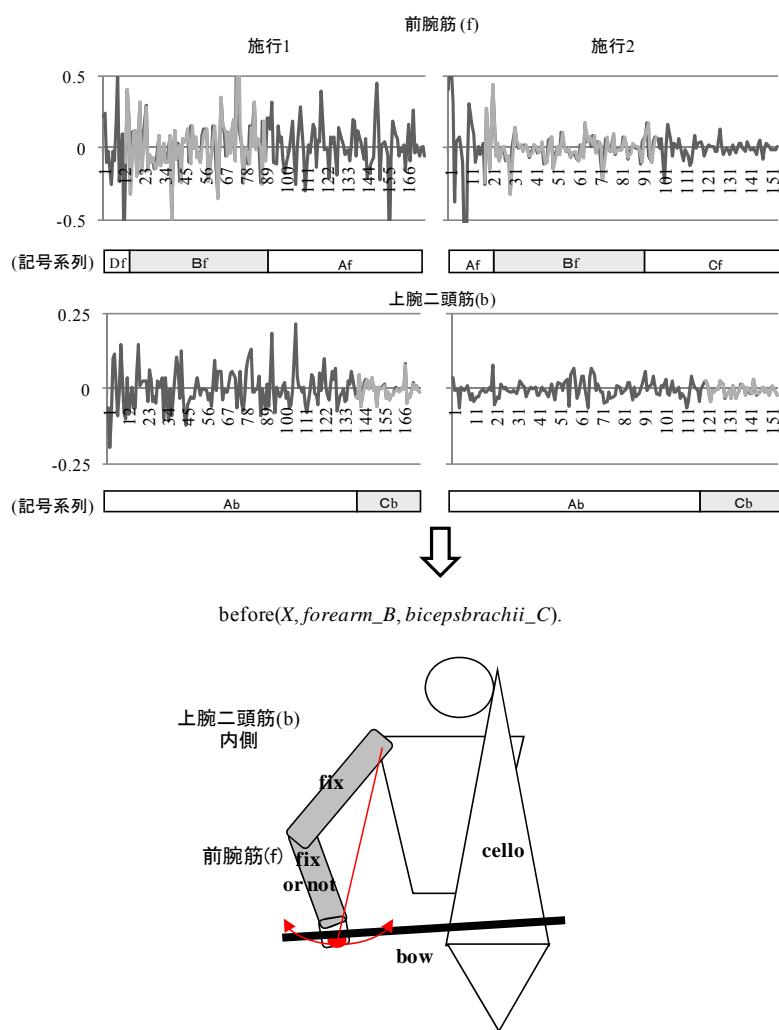


図 3-4. 筋電波形と動作図

3.3.2 バイオリン演奏とチェロ演奏のスキル比較

次に、バイオリンの演奏の熟達者 1 名に対し、3.3.1 と同様にさまざまな拍子において、どのように筋肉を使っているのかを調べる実験を行った。試行回数および試行内での繰り返し数も同数にそろえ、全体では 120 試行とした。そこから、バイオリンとチェロにおける分類ルールを学習を行うことで拍子ごとに両者のスキルの違いを調べた。この場合、正事例はバイオリンの演奏全体となり、事例数は前回の倍となる。バイオリン演奏について

抽出されたすべてのルールを以下に示す.

バイオリンにおけるルール

$$(Z = 13.9 > \chi^2(1, 0.05))$$

$\text{beat}(X) \text{ :- before}(X, \text{spinae-right_B}, \text{spinae-right_D})$

$\text{beat}(X) \text{ :- overlaps}(X, \text{trapezius_A}, \text{bicepsbrachii_C})$

第一のルールでは、脊柱起立筋の中での二つの活動パターンが検出されており、「脊柱起立筋・右の B 活動の後に、脊柱起立筋・右の D 活動がみられる」と解釈できる。これは活動が抑えられている状態の後に、活動が活発になるという単純なパターンが検出されたものである。特に演奏の最初の段階で検出がされている。第二のルールは、「僧帽筋の A 活動のあとに重なりながら、上腕二頭筋の C 活動が観察される」と解釈ができる。僧帽筋と上腕二頭筋の活動パターンで、それぞれ活動が落ち着いているときの時系列データが検出されている。すなわち、僧帽筋の力を抜いた後に上腕二頭筋の力を抜くというスキルである。奏者は背中から徐々に力を抜いており、腕のインピーダンス調整をしていないことから、3.3.1 でのべたバイオリン演奏とチェロ演奏[古川 05]の違いを示唆しているといえる。

3.3.3 複合タスクからの個人スキルの抽出

最後に複雑なチェロ演奏における筋電の個人スキルの抽出を行った。実験では、ルイジ・ボッケリーニのロンド(Boccherini Rondo)を4試行、熟達者に行わせた。そして3.3.1や3.3.2の知見を利用し、筋電部位で協調性があり、さらに運動学的にも意味のある筋肉部位の対5組を選んだ。そして、他の箇所とは異なる特徴の弾き方をしているルールを抽出するため、注目する1組を正事例、残りの4組を負事例として学習を行った。これにより正事例4、負例16の計20試行(4×5)が得られる。このタスクより得られた3つすべてのルールを以下に示す。

複合タスクにおけるルール

$\text{trial}(X) \text{ :- before}(X, \text{deltoid-f_A}, \text{deltoid-f_B})$

$$(Z = 4.6 > \chi^2(1, 0.05))$$

$\text{trial}(X) \text{ :- during}(X, \text{tricepsbrachii_B}, \text{bicepsbrachii_B})$

$$(Z = 5.1 > \chi^2(1, 0.05))$$

$\text{trial}(X) \text{ :- overlaps}(X, \text{spinae-right_C}, \text{spinae-left_C})$

$$(Z = 9.5 > \chi^2(1, 0.05))$$

以上の結果を解釈すると、順次「三角筋肉前の A 活動が三角筋肉前の B 活動の前でおきる」「上腕三頭筋の B 活動が上腕二頭筋の B 活動の間で起きている」「脊柱起立筋右の C 活

動のみられた後に、重なって脊柱起立筋左の C 活動がみられる」となる。はじめのルールにおける三角筋前の活動は、開始の部分ではじめの4つの音を出すための準備動作を捉えている。次のルールにおける上腕二頭筋と上腕三頭筋の活動は、スラスタッカートをしている部分に相当する。これは、上腕を強く活性化してインピーダンスを急激に増大させる必要があるからである。スラスタッカートは技術的に困難な課題であり、その奏法を解析した例は知られていない。本ルールはそこでの上腕のインピーダンス調節の重要性を示唆しており、練習法の具体的な提案にもつながっている。これは、科学的な発見と考えてよい。最後のルールでは、背筋においては脊柱起立筋右の活性化後に脊柱起立筋の左が活性化するという現象が交互に見られた。この部分は素早い弓の切り返しを必要とするところ（図3-5）で、そのために脊柱起立筋を活用していると考えられる。こうすることで、左右の脊椎起立筋を一体化させて動かす場合に比べて、同期運動に対する体の周期を約2倍に上げられる。高速運動を実現するためのこのような体のたくみな使い方が自動的にルールとして抽出できた点は、特筆するに価する。

以上、3つのルールについて楽譜と必要な動きとの考察を行った。



図3-5. 該当する楽譜部分

3.3.4 実験結果のまとめ

実験では、熟達者によるチェロ演奏において遅い拍子で演奏を行う場合、腕の端から順に力を抜いて演奏を行っていること、それ以外の早い拍子におけるルールでは、前腕の活性が長期に渡っており、その間で上腕三頭筋の活性が見られることが発見された。これらは、事前に得られていた知識である振り子演奏とも対応関係がつく。また上腕二頭筋と上腕三頭筋の活動でそれぞれ活動が持続している場合があるが、これはスラスタッカートのときに、インピーダンスの調節を行って腕全体で引いている場面を観測している。バイオリンの解析においてはマクロなパターンとして、演奏が早くなればなるほど徐々に脊柱起立筋の活動が増えるということが観測された。また、チェロ演奏とは異なり、常に一定の活動を伴う、すなわち筋肉・活動を固定した動きはあまり行わないといえよう。

3.4 関連研究

ここでは本研究で用いた主要な提案手法である、時系列データを分節化する方法および、ILP をもちいた時区間関係の抽出方法についての特徴と、それに関連する研究についてのべる。

時系列データを分節化する手法の特徴は以下の3つになる。

- i. Piecewise aggregate approximation など時系列データマイニングにおける従来の記号化の手法[Last 04] での記号の数値範囲を規定するパラメタ，既存研究におけるセグメンテーションアルゴリズム[植野 04;Keogh 03]の最小誤差などのパラメタを設定する必要がなく，Hoppner も同様のセグメンテーションおよび，時区間関係の抽出の研究において，任意の閾値や教師なしクラスタリングを用いているのに対して[Höppner 01]，明示的に情報量規準というという統計量を組み込んだこと。
- ii. 時系列データ含まれる誤差の情報を想定してモデルの導入をしていること。従来の研究でもいわゆる変化点予測を用いた時系列データマイニング法や[Guralnik 99]，モデル選択の際にクロスバリゼーションを用いる手法などが数多く存在するが，AR モデルを用いてモデル選択について情報量規準を利用して分割を繰り返しているという点が異なる。
- iii. ボトムアップ型のセグメンテーションアルゴリズムとは異なり，精確な分節点を測定できる[Last 04].ただし，その一方で計算量が多くなるというデメリットもある。

ただし上記では，時系列データが局所的に定常であることを仮定して AR モデルを適用して分析したが，時系列データのトレンドが変化するという非定常な時系列においても，提案手法の適用が可能である。たとえば，複雑なトレンドに対しては線形回帰を区分的に推定して，後で差し引くことで定常な系列を得ることも出来る。また，モデルの推定値からの距離を計算することで異常値の検出も可能となる。ただし，時系列データの分節化において，ある状態が別の状態にすぐさま移ることを想定している。これについては，時変係数時系列モデルを用いることや，局所的に周波数解析を用いるなどの対応法が考えられるが記号化との接点が明確でないため，今後の課題とする。

次に ILP を用いた時区間データのマイニング手法の具体的な特徴とメリットを挙げる。時区間データのマイニングの手法としてはすでにアプリアリなどを用いて頻出する時区間パターンを抽出する手法がある。こうした手法はインターバルデータマイニングとも呼ばれており，いくつかの手法が存在している[Höppner 01; Ho 07]。これらとは異なる特徴として次の2点が挙げられる。

- i. システムの検証性。まずメリットとして，システムの検証を ILP システムの健全性と背景知識の健全性に求めることが出来る。スキルサイエンスにおいては，マイニングを行うだけではなく，専門家の知識の検証作業も重要な要素となっており，仮説検証や新たな知識獲得，推論のプラットフォームとして必要な機能である。
- ii. 情報統合と拡張性。提案手法の大きなメリットとして，拡張性をあげることができ

る。本手法では新しいルールやデータの追加が容易にできるほか、同じデータを用いて、別の関係構造の抽出が可能となることなど、分野に特化したマイニングとは異なり、複雑なデータ構造の分析への対応が容易であり、アルゴリズムの開発時間やコストの低減をするラピッドプロトタイピングに向いている。本研究では、時区間関係の獲得に特化した。それ以外のルールの記述も容易であり、たとえば点と点同士、点と区間、誤差に対応し拡張した時区間論理などの関係構造も考えられる。別のルールの獲得という点で拡張した研究として、すでに本枠組みと TIS を用いて動作一貫性制約(MIC)を抽出したものなど[澤井 08]が存在する。

また、新しいデータとして、音楽データなど他の種類のデータも追加可能であり、本研究でも拍子などの情報を一部組み込んでおり、こうした情報統合の側面もある。

最後に、ILP を用いて時区間パターンを抽出していることから、得られた関係構造の高次の発想推論[小林 07]等への再活用が容易で、これらを一つのシステムとして実現することが可能である。

ただし、デメリットとしては、柔軟性を重視したことで計算が増えることがあげられるが ILP とグラフマイニング手法の精度が比較されているように必ずしも精度は劣ってはいない[Ketkar 06]。

本研究では、前述のように専門家による検証作業も重視していることから、獲得対象を時区間関係に限定できるような背景知識を ILP システムに与えることで、探索対象となるリテラルの組み合わせを制限し、探索空間を減少させている、計算量の問題に対処している。そのため、他の分野へ適用する場合の計算は探索性をどこまで必要するかということと、分野における背景知識に依存する。

3.5 本章のまとめ

本章では、局所的にパターンの変化する複数の時系列データを対象として、特徴同士の関係をルールの形で抽出する方法を提案した。提案手法を具体的にチェロ演奏およびバイオリン演奏の解析に適用して、事前に予想された知識の検証や、さらに新たな知見を獲得することができた。

この研究は、3.1 でのべているような記号と実世界に関連する問題に対する対応策のひとつとして特に、「情報量規準」を利用してセンサー入力を自動的に記号化し、その記号間の関係を帰納論理を用いて獲得する一連の処理の流れを示したことにも意義がある。

ただし、本研究の限界としてはどの関係が重要であるかは抽出できたが、得られた重要な関係のうちさらにどのような関係が重要なのかといった問題には適用できないことがある。具体的に述べると、本研究において時区間関係からある影響変数への重みについては学習できていないため、どの「関係がもっとも重要であるか」といった比較を直接行うことは難しい。

また、本章における固有の課題としては個人の特徴的なパタンの抽出を行ったが、複数の被験者に適用することで演奏についての一般的なルールの獲得をしていく必要もあろう。

なお応用としては、企業組織に対して当てはめると個々人の属性とその組み合わせ関係と、その企業の生産性・売上などの関連をあつかうことができる。さらに今回扱った時系列データを扱う方法については人間に関するデータ以外でも利用できる。たとえば、ユビキタス社会におけるセンサーデータとそこでおきたほかの事象との関連（センサーのパターンと事故）をみていくことができる。この応用については、本研究では筋電の分析に焦点をあて AR モデルを採用したことから、モーションキャプチャーのデータやユビキタスなどの技術で得られるセンサーデータなどさまざまな時系列データにそのままの形で適用することはできない。提案手法を適用するためには、分析する時系列データの特徴をうまくとらえる時系列モデルの選択がポイントであり、個々の分野に関する既存の知見や研究を応用して、適切なモデルはなにかを体系化していくことが必要であろう。

最後に今後の発展の可能性としては、帰納推論の結果として抽出したルールを発想推論やプランニングと組み合わせていくことで、新たな演奏法の提案や解釈を支援するシステムなどが期待できる[小林 07]。

以上、この章の研究を通じて本論文全体との対応関係をみると、第 1 章でのべた **B** の問題にあたる「 h_s と $G_s' (= G_s - E)$ が与えられているもとの、 E を獲得しつつ h_s を説明する f' を学習する」という問題に対する解法を提示し、さらに有用な知見がえられることを示した。ただし、今回、時区間関係の知識、時系列データ、記号系列データからの学習に特化した。これを変更することにより属性データが付随しているノード V とその影響変数 h_s が与えられている問題であれば本手法を適用できる汎用的な方法である。

4章 調査データに基づく社会構造変化の抽出

本章では、影響変数を伴う関係データがそれぞれ時間情報を持ち、変化するとき分析を行う問題を扱う。

特に動的に変化する関係を抽出する手法を提案する。具体的には影響変数があり、さらに関係の属性やノードの属性が存在する関係データを扱い、影響変数が急激に変化した前後での関係の違いについて抽出を行う。

本手法の有効性について、本章ではコミュニケーションの分析を扱う。

具体的には集団間のコミュニケーション・ネットワークの変化というものを想定し、それを調査を用いて推定する方法を示した。次に抽出した集団の動的なネットワークから、ある時間前後での関係の変化や一貫性制約条件の変化について、帰納論理プログラミングを用いて時間的な前後関係の知識を組み込みこんだ上で抽出を行う。

これらを通じて、Cの課題である影響変数を伴う関係データが時間的に変化するなかから有効な構造を抽出するという問題に、提案した手法が有効かを検証する。

4.1 背景と問題設定

1.1.1でも述べたように、関係に注目した研究が盛んに行われている。その背景として、ネットワークの構造的特性が計算機シミュレーションにより簡単に調べられるようになったこと[Newman 03]や、ブログなどの情報環境の発達によりネットワーク構造を持った大規模なデータが手に入り易くなったことなどが挙げられる。

ネットワーク構造を持ったデータの分析法として、社会ネットワーク分析[安田 01]やリンクマイニング[Getoor 03]などが発展している。これらの研究では一般には、関係をリンクの有無に対応させ、ネットワーク内の共通構造を発見・抽出するというアプローチが用いられる。その理由としては、関係というものを抽象化することで、より一般的な知識や特徴を得ようという問題意識があると考えられる。

しかし、単なる関係の有無だけでは十分な分析が行えない場合もある。たとえば、1章のCの例で述べたように企業における組織構造を変えた結果、生産性が向上したという例を取りあげよう。このような現象の場合、どのような属性（役職や能力）をもった人が、別の人とどのような関係（同等か上司・部下関係か）をもっており、その構造がどのように変化したかを考える必要がある。そこで本章では、次の三点を考慮した分析の枠組みを提案する。

- (1) 関係の属性：関係にはさまざまな種類や属性がある。
- (2) ノードの属性情報：個々のノードの持つ属性情報は重要な意味をもっている。
- (3) 時間情報：関係は動的に変化する。

たとえば、社会ネットワークであれば、(1)として親子関係や職場関係などの種別や、関係の強さなどの属性があげられる。また(2)については、性別、年齢や身長などが考えられる。(3)では、会社組織における組織図の変化やウェブログ同士のリンクの変化などがあげられる。従来(1)や(2)の属性情報などを含めた統合的な立場からの分析はあまり行われていなかった。また、(3)については、シミュレーションなどで扱われることは多かったが、データ分析という帰納的な面からのアプローチは必ずしも多くはなかった。これに対して本章では、特に帰納論理プログラミングを用いて、(1)~(3)を考慮したデータ分析を行うを試みる。

加えて、本研究の主題でもあるように、ネットワーク構造はある外部の変数や指標と関係している。先ほどの企業の例を用いると、組織のネットワーク構造は生産性と結びついている。他にも、市場のネットワーク構造と全体の魅力度[佐藤 05]など、ある種の外部の影響変数を通じてネットワーク全体を評価している事例は多い。このような場合、外部の影響変数の値の急激な変化は、ネットワークの構造変化を示すと考えられる。

これらをふまえ、本章では外部の影響変数を伴う動的なネットワークを対象に、関係や

その変化に加え、関係の種類やノードの属性情報も考慮した新たな知識獲得の枠組みを提案する。また本章では、応用として、質問紙調査の結果より得られる集団間の動的ネットワークを分析の対象とする。ブログや SNS を用いた社会ネットワーク分析では分析対象が参加者に限定されるが、質問紙調査を使った本方法では調査主体が自由に調査対象や設問を決められるため、集団という限定はあるがより柔軟な分析が可能になる。

はじめに本研究で対象とする問題の形式的な定義を与える。

第 1 章の 1.2.2 の定義を再び用いるとある時点 $t(1 \leq t \leq n)$ におけるネットワークを

$$G_t = \{V_t, E_t, \Phi_{V_t}, \Phi_{E_t}\}$$

と表記する。なお $G(i:j)$ で、集合 $\{G_x | i \leq x < j\}$ を表す。

一方、ネットワーク G_t における外部の影響変数を $h_t \in R$ と表す。以上の準備のもと、本研究で対象とする動的ネットワークは、

$$I = \langle GT, hT \rangle = \{ \langle G_1, h_1 \rangle, \langle G_2, h_2 \rangle, \dots, \langle G_t, h_t \rangle, \dots, \langle G_n, h_n \rangle \}$$

と表現する。

本研究では、上記の動的ネットワーク I に対し、(1) ネットワークに大きな質的变化が起きた時点、及び (2) その背後の変化した構造、すなわちノード・リンクの有無やノード・リンクの属性を抽出することを目的とする。

従来の動的なネットワークに関する分析では、そこからある集団（クラスター）を抽出し、その変化を追うという研究[Tantipathananandh 07]があった。しかし本研究は、外部の影響変数を通じて構造の変化を追うという点で問題設定が異なっている。

本章の構成を以下に示す。まず 4.2 では問題設定を行い、本研究の問題の形式化を行う。続いて 4.3 では提案手法について具体的に説明する。4.4 でシミュレーションデータと実データに本手法を適用する。4.5 では既存の研究との比較とサーベイをし、4.6 でまとめを行う。

4.2 提案手法

4.1 でのべた問題に対し、本研究で提案する枠組みの概要をのべる。

まず 4.1 の(1)の目的に対し、外部の影響変数 h_t が G_t と関連していることを仮定し、入力データ I から外部の影響変数を取り出し、その変化に基づいて、大きな変化が起きた時点 ch を抽出する(4.2.1 節)。すなわち、 h_1, \dots, h_n の値に着目し、大きな変化がみられる複数の時点、 $\{h_{ch_1}, \dots, h_{ch_m}\} (1 < ch_1 < \dots < ch_m < n)$ を抽出する。また便宜上 $ch_0 = 1$, $ch_{m+1} = n$ とする。またこれらの集合を CH とする。

次いで、4.1 の(2)の目的である変化の要因や構造の変化を抽出するために、上記で抽出した CH を利用して、述語論理ベースの学習器である帰納論理プログラミング(Inductive Logic

Programming)[Mugglten 94; 古川 01]を用いて比較することで、ノード属性も含めた構造変化の抽出、及び一貫性制約の抽出を行う。ここで ILP の論理設定をのべると、

Given: $E = E^+ \cup E^-$, BK

Find: H s.t. $BK \cup H \models E^+$, $BK \cup H \cup E^- \not\models \square$

ここでBKは背景知識、 H は仮説とする。今回は E として G_t を利用することとし、 E^+ を変化点 ch 以後のネットワーク集合、 E^- を ch より前のネットワーク集合とする。構造変化とは、外部の影響変数の質的な違いを反映するリンクや属性の違いのことをさす。出力は述語形式で表現される。また、一貫性制約とは、「あるデータが違反してはいけないルール」のことを指し、述語形式を用い、 $false:- A(x), B(x)$ の形をしたルールとして表現される。本研究では、変化点前後での違反してはいけないルールの違いに着目し、変化以降でなくなった制約の抽出を行う。詳細な定義は 4.2.2 で扱う。

また詳細な分析を行うために、この作業の前処理としてあらかじめ、既に得られている属性や社会ネットワーク分析の結果などの属性をノードに追加した上で、 G_t を述語表現に変換をする。同様に、あらかじめ時間的な依存関係についての背景知識も導入する。

図 4-1 は以上のプロセスを簡略的に表現したものである。図 4-1 においてはネットワークの述語表現についても簡略したものを採用している。例ではリンク $a-c$ は、 $G(ch0:ch1)$ 中のグラフには存在せず、 $G(ch1:ch2)$ 中のグラフにのみ存在するので構造変化として抽出される。一方、リンク $a-b$ は、 $G(ch1:ch2)$ 中のいくつかのグラフには存在するが、 $G(ch0:ch1)$ 中のグラフにはまったく存在しない。よって、 $G(ch0:ch1)$ には存在したが、 $G(ch1:ch2)$ では存在しない制約として抽出される。

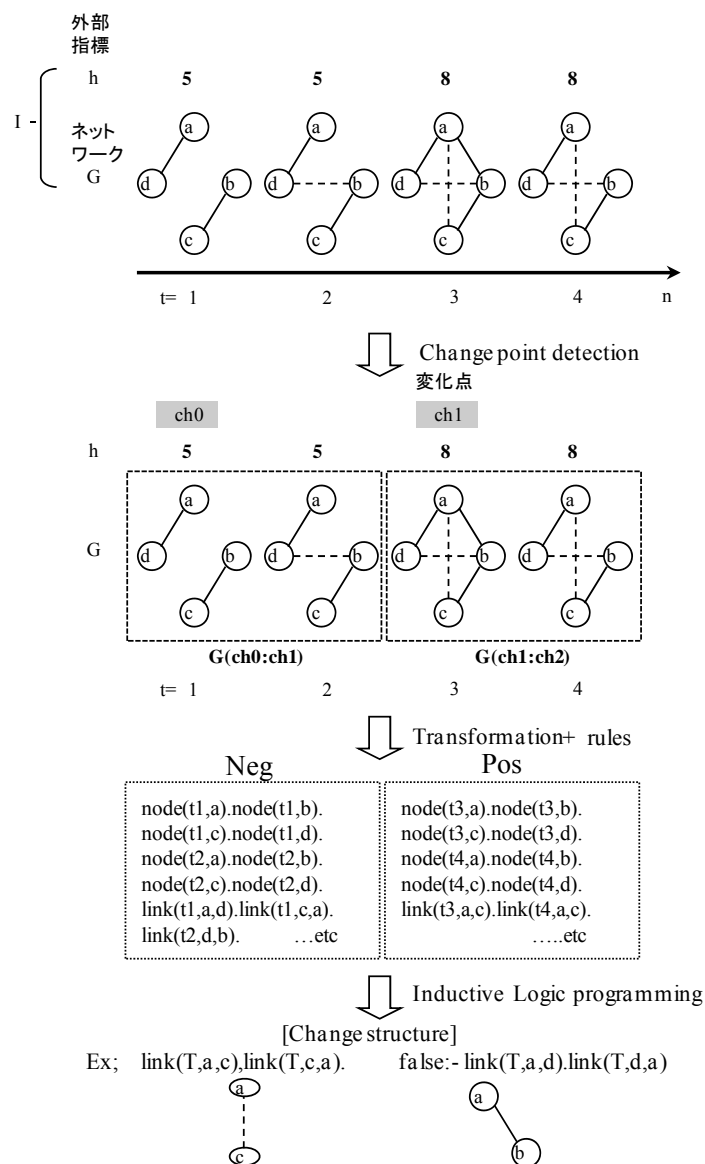


図 4-1. 提案手法の概要

本研究で提案する動的ネットワーク分析のための枠組みの利点は、

- (1) 動的に変化するネットワークからの構造変化の抽出、
- (2) 一貫性制約の抽出、
- (3) 属性や複雑な関係構造の導入、

を実現できることにある。(3)については、ILPを使用して述語表現を利用したほか、背景知識の導入をすることで実現ができた。また他の利点として、得られる結果の解釈が容易であることも挙げられる。

一方、欠点としては、既存の社会ネットワーク分析の手法と異なり、中心性やクラスタの抽出のように関係を抽象化した情報を容易には扱えないことや、述語の組み合わせを考

えるため計算コストが高くなることが挙げられる。しかし、前者に対してはマクロな情報を属性として組み込んで分析を行う、後者に対しては仮説の深度を浅くするなどの対応策が考えられる。また、本枠組みでは、ネットワーク構造の質的な変化が外部の影響変数 h の大きな変化に対応していることを前提としており、大きな変化点のみを抽出するので、ネットワークやその指標が常に漸進的に変化をしている場合には適用できない。

以下では提案手法の詳細についてのべる。4.2.1 節で変化点の抽出、4.2.2 節で ILP を用いた構造変化の抽出法について扱う。

4.2.1 変化点の検出

はじめに、入力データを I とし、そこから h を抜き出して構造が変化した点 CH を調べる。

ネットワーク構造の変化点を調べる際には、大きく二つの場合がある。外的指標がある場合と外的指標がない場合である。本章では、問題設定でのべたように前者を想定しており、外的指標の二時点の変化（ベクトルの場合は、各時点同士の距離）をもとにある規準値 α 以上となる点を変化点として抽出する。すなわち

$$CH = \{ t \mid |h_t - h_{t-1}| \geq \alpha \} (2 < t < n)$$

となる。変化点の検出方法自体は本研究の主題ではないため、ここでは、もっとも簡便な方法を用いたが、このほか、時系列分析の分野で用いられている数多くの変化点検出法を用いることができる。

一方、本章では直接対象とはしないが、原理的には、外部の影響変数が与えられていない場合においても、変化点の抽出は可能である。たとえば隣接行列から各ノードのもつ中心性などの特徴量を計算した後に距離を計算して変化を調べる、ネットワーク全体の特徴量（平均経路長など）やネットワーク同士の距離の変化を探るなどの方法が考えられる。

4.2.2 帰納論理プログラミングによる構造変化の抽出

次に本研究では、4.2.1 節で検出した変化点 CH を利用して正事例・負事例を作り、ILP 上に組み込んで構造変化を獲得する。詳細について説明する。変化点が1つであった場合を考えよう。入力データ $I = \{ \langle G_1, h_1 \rangle, \langle G_2, h_2 \rangle, \dots, \langle G_t, h_t \rangle, \dots, \langle G_n, h_n \rangle \}$ と変化点 ch_1 に対し、 $G(ch_1: ch_{m+1})$ を正集合、 $G(ch_0: ch_1)$ を負集合として、ILP を適用する。このとき、ILP に与える正事例、負事例は時間 t を index とした各時点の G_t の述語 $G(t)$ をヘッダーとして以下で構成される。

正事例 E^+ は $G(ch_1). \dots G(ch_{m+1})$.

負事例 E^- は $G(ch_0). \dots G(ch_1 - 1)$.

この判別により、どのような関係が変化したかがわかり、その変化により全体としての変化が起きたと特定できる。ただし、 ch は複数存在する場合もあり、たとえば ch_1, ch_2, ch_3 が

変化点のとき $G(ch_0: ch_1)$ と $G(ch_1: ch_2)$, $G(ch_2: ch_3)$, $G(ch_3: ch_{m+1})$ のそれぞれ隣接する組, つまり負事例 $G(ch_0: ch_1)$ と正事例 $G(ch_1: ch_2)$, 負事例 $G(ch_1: ch_2)$ と正事例 $G(ch_2: ch_3)$, 負事例 $G(ch_2: ch_3)$ と正事例 $G(ch_3: ch_{m+1})$ という組み合わせを作成して構造変化を抽出する. このため, 極端に変化点の間の事例数が少ない場合, 本手法は適用が困難である.

本研究では, 上記の ILP の前処理として①まず元データ G_t から必要な属性を選択・生成を行い述語形式に変換する. ②その後, 構造に関する知識, 特にノードとリンクに関する時間的関係を背景知識BKに導入する. これらの処理をした後, 変化した構造そのものの違いを抽出する. また構造変化の抽出に加えて, ③変化前後における一貫性制約条件の違いも抽出する. 詳しくは後述するが, 一貫性制約も構造に関する知見である. 本章では, 構造変化及び一貫性制約という2つの構造の抽出を試みる.

以下では具体的な前処理と一貫性制約の抽出法についてのべる.

① データの属性選択・生成と変換

入力 I のうち G (必要に応じて h)を対象に, データの属性選択・生成および変換を行う. データの属性選択・生成とは, 必要に応じ, ノード V に付随する変数 (属性) 集合 D_1 を選択・生成する操作である. これにより, ルールをえる際に探索空間を縮小できるメリットがある.

属性選択とは, ノードに付随する必要な属性 D_1 を選ぶ操作である. ある程度調べる対象が限定されている場合は, 属性選択の必要はなく, すべての属性を利用すればよい. ただし, 変数が多い場合は, 探索空間が広がるため外部の影響変数の変化と属性の変化の相関を利用して, 必要な属性を絞り込むことができる. たとえば, ある社会でのネットワーク全体の取引量などが存在した場合, その時間変化(2時点間変化)と属性の時間変化の相関の高さにより選定ができる.

属性生成とは, 必要に応じてノードに新たな属性を D_1 に追加する操作である. たとえばリンク集合により計算されるノードについてのネットワークの特微量, 中心性などを計算するという操作である. また属性が多い場合に主成分分析などを行い属性をまとめることで新たな属性を生成することもできる.

以上の作業後, ネットワーク G_t を述語形式に変換する. V_t と Φv_t , E_t と Φe_t は, インデックス(ID)と, ノード属性(NodeAttribute), リンク属性(LinkAttribute)及び時間(T)の3種類の要素で構成され, 述語形式として

$node(T, ID, NodeAttribute).$

$link(T, ID1, ID2, LinkAttribute).$

と表現される. ここで前者は, 「時刻Tのグラフにおいて, ノードIDは, 属性NodeAttributeを持つ」ことを表す. また後者は, 「時刻Tのグラフにおいて, ノードID2からノードID1への属性LinkAttributeを持つ有向リンクが存在する」ことを表す. なお, 述語論理と関係データ

についての関連研究は 4.4 でのべる。

② 構造に関する知識の導入

構造変化の抽出の前段階として、構造に関する知識をBKに導入する。利用するデータや獲得すべきルールに関し、何の前提知識もない場合は、グラフマイニングのように頻出するパターンやルールを全て抽出することは有効である。しかし、ある程度の知識がある場合には、それを背景知識として組み込むことにより、探索空間を制限し、より詳細な知識の獲得に焦点をあてることが可能となる。たとえば、リンクに関しての時間的關係の知識を取り入れる場合と、ノードの属性に関しての時間的關係の知識の例を記述した。

以下に、リンク及びノード属性に関する時間關係の知識の記述例を示す。

linkBefore(T1, X, Y, Z, La, Lb): –

link(T1, X, Y, La), link(T2, X, Z, Lb), T2 ::= T1 – 1.

(ノードX, Zの間にリンクLbがあったあとにX, YにリンクLaが存在する)

nodeAttBefore(T1, X, A, B): –

node(T1, X, A), node(T2, X, B), T2 ::= T1 – 1.

(ノードXにおいて属性Bのあとに属性Aを有する)

beforeNodeLink(T1, X, Z, La, B): –

link(T1, X, Z, La), node(T2, Z, B), T2 ::= T1 – 1.

(ノードZに属性Bができたあとに、ノードX, ZにリンクLaが存在する)

以上のような情報を仮説に加えることでlinkや属性に関して1区間のみの前後關係のみを調べることも可能となる(上2つの知識)。さらに、ノードとノードの關係、ノードとリンクの關係、リンクとリンクの關係、その他属性間の關係などの記述も可能である。たとえば、3つめのルールでは、あるノードの属性がBであり、その次の時点で別のあるノードに対してAというリンクをはるというものになる。このような時間的な關係を扱うことで、 G_t 間の時間的な獨立性の仮定を取り去ることができる。

以上の①と②の設定により、ILPを用いて変化点ch前後の判別ルールとして構造変化を与える。

③ 一貫性制約の変化の抽出

次に、一貫性制約の抽出について説明する。ある構造が時点CHにて変化した要因を制約条件の変化として捉える。一貫性制約とはもともとはデータベース分野の用語であり、あ

るデータベースにおいてデータベースが従う規則をあらかじめ記述しておくことで、データベースの一貫性を保つという役割を果たすものである[日本工業規格 95]。これに基づいて、本研究では一貫性制約を「あるデータが違反をしてはいけないルール」と定義する。また、一貫性制約の記述には大きくは二つの方法が考えられる。与えられたデータについて真のルールで表現する制約と、偽のルールで表現する制約がある。前者はたとえば、関数従属性のようにある属性 X, Y の間で $X \rightarrow Y$ が成立しているなどの否定表現を用いないルールのことをさす。後者については、たとえば年齢のデータベースであれば、「人間の年齢は 0 歳未満や 200 歳以上になってはいけない。」など主に事実を反した否定で表現でき、常に偽となるルール[佐藤 08]にあたる。本研究ではこの後者を扱う。このような制約は特に

false: $\neg A$.

という形式で表現する。頭部の **false** は節部分 A が一貫性制約を示したもので、 A は事実としては成立しないものである。これは集合で考えると、与えられているデータの補集合の一般化とほぼおなじものとなる。

この定義を採用する理由として、通常扱われる与えられている集合ではなく、補集合のようなデータの外枠となる条件の変化も、構造の変化に対して重要である可能性があるからである。制約が変化した結果として、さまざまな事象があらわれていることが往々にしてある。たとえば社会科学で扱われる制度（制度論）としての視点からも重要であり、法律など違反してはいけないというルール自体が変化することによって、その内部の変数の自由度が変わり、ネットワーク構造や外部の影響変数が変化することがある。このように一貫性制約を扱った研究にはたとえば高橋による Web コンテンツの管理において一貫性制約の中で包含従属性を抽出するという研究がある[高橋 08]。ただし、正事例が与えられている中で、その内部での包含関係を扱う課題であり、むしろ負としてルールを抽出する本研究とは異なる。

具体的な制約の方法は、本研究では制約が変化したかどうかをみるために、正事例・負事例それぞれに該当するデータに対して存在しないデータ、たとえば隣接行列に焦点をあてるのであれば 0 となっている部分のデータ（補グラフ）を生成し、通常の構造変化の抽出と同じように ch の前後を判別するルールを得ることで抽出した。この方法により、多く制約のなかからより意味のある制約に絞ることが可能となる。正事例のみが与えられている場合は、予め設定した制約が成り立つかどうかをチェックしていき、すべてのデータに対して成り立たない場合、一貫性制約として抽出するという手法がある[Srinivasan 99]。しかし、今回は正・負事例を扱い、重要な制約の変化を抽出するというタスクのためこの手法は適当ではない。ほかにも事例ごとに一貫性制約を抽出し、その上で差分を取る方法も考えられよう。しかし、この場合、節同士の一般/特殊性を考慮して抽出する必要もあり複雑になるため用いていない。

今回採用した手法を図 4-2 を用いて示す。図 4-2 の+と-はそれぞれ正事例、負事例である。このとき、一般的に判別ルールとしては、正事例にすべてあり、負事例にはなり

リンクとして b と d の間のリンク，波線が抽出される．一方，負事例すべてにおいて存在しなかった，すなわち負事例においては制約があったが，正事例では存在している，すなわち制約が解消された差分データとして a と b のリンク， b と d のリンクがある．つまり波線のみならず太線のリンクもある．このような，仮に正事例において多くの事例を説明するものではないとしても，負事例には無かった新しい事例（制約差分）を抽出する．通常の学習により得られるルールの否定が得られる可能性もあるが，一貫性を保つためにすべての事例に成り立っていないとはいけないという条件を課していることもあり，図 4-2 のように別のルールが得られることが多い．

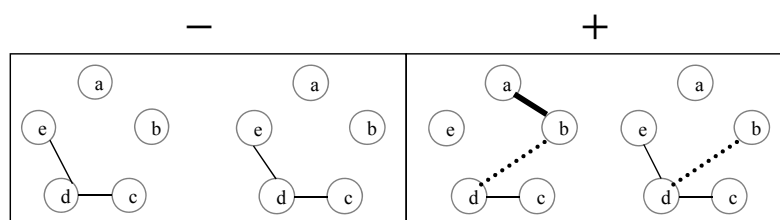


図 4-2. 本研究の一貫性制約の事例

ただし，何の条件も与えない場合，存在しないデータの組み合わせのため数多くの制約が得られることが予想される．そのため，あらかじめ考える重要なデータのみを焦点をあてるといった制限も必要である．たとえば，図 4-2 の例にあるようにリンクについての制約が重要であるならば，存在しないリンクの組み合わせについて考えていく必要がある．また，中心性などのネットワークの特徴量を利用してあらかじめ重要なノードを特定しておき，このノードを中心とした制約を特定するという方法も考えられよう．このような種，分析者にとって獲得したい制約が限定されている場合に本手法は有効であろう．

4.3 実験

提案手法の妥当性を検証するために，集団間ネットワークと呼ばれるネットワーク，特に集団間のクチコミを対象に，実データを用いた実験を行った．以下では，まず 4.3.1 で集団間ネットワークを導入し，その後，4.3.2 で集団間ネットワークに関するシミュレーションデータを用いた実験結果及び 4.3.3 で実データを用いた実験結果を示す．なお，本研究では ILP システムとして Aleph[Srinivasan 99]を利用した．

4.3.1 集団間のネットワークの作成方法

まず，集団間ネットワークについて説明する．はじめはネットワークのノード集合 V の作成である．集団間ネットワークにおいて V を構成する集合として，ある集合（母集団） O を

考える．このとき， O を何らかの基準で p 個の排他的な集合 X_1, X_2, \dots, X_p ($O = \cup_{i=1, \dots, p} X_i, \forall i \neq j, X_i \cap X_j = \phi$) に分割する．ここで， X_i ($i = 1, \dots, P$) を集団間ネットワークのノードとする．本研究では，世代間の情報伝達，すなわちクチコミに関する集団間ネットワークを考えている．具体的には，母集団 O として人間の集合を考え，それを世代（年齢）によって $X_i = \{o | o \in O, o \text{ の年齢は } (i * 10) \text{ 歳代}\}$ に分割することで得られる各集団 X_i ($i = 1, \dots, P$) をノードとする．

つづいてリンク E における各リンク e の作成について説明する．これは集団（属性） X_i と集団（属性） X_j の関係を把握することで可能となる．本研究では，集団同士のクチコミの程度をリンクと考える．このとき，どの属性（集団）の人がどの属性（集団）の人と会話を行ったかを調査し，集計し，複数の集団間のリンク指標を作成して，それに閾値を使うことで作成する．

具体的に，集団間のネットワークにおけるリンクの作成についてのべる．まず X_i 集団と別の集団 X_j があるとしたときに，これらの母集団間におけるリンク指標は Z_{ij} を，以下のように定義する．

$$Z_{ij} = C_{ij} / (|X_i| \times |X_j|)$$

集団で生成可能なリンクを分母として， X_i と X_j の間のクチコミの総数 C_{ij} を分子とした比率として作成する．

なお，全母集団を調査することが出来ない場合は，一部のサンプルからこれらを推測することになる．ただし，上記の定義ではもとの集団の数 M が多くなるにつれて，分母が $O(n^2)$ で増加するためより多くの C を必要とする値になっている．そのため，サンプルを用いる場合はサンプリング比率 (= サンプル数 / M ; ただし母集団における集団の構成比とサンプルにおける集団の構成比はほぼ同じでサンプリングの比率はほぼ一定とする) で調整を行い，次の Z' を用いる．

$$Z'_{ij} = c_{ij} / (m_i \times m_j \times (1/p))$$

c は回答数， m はサンプル数， p は想定しているサンプリング比率である．上記のような値を用いる理由は，全母集団によるネットワーク全体の可能なリンクに対してどれだけ値が大きいかを示すためである．

次にこうして作成したリンク指標 Z' に対してある閾値 σ を設定することで σ 以上の場合に 1，そうでないときは 0 として隣接行列 E を作成する．また，全体の平均との差がプラスかマイナスか，リンクがあるか否かをみることで {high, low, 0} という 3 値の関係行列を作成することもできる．

以上でのべた方法で準備した頂点集合 V と辺集合 E をもとにネットワーク G を作成する．ただし，推定や精度などに関しては，今後の研究としたい．なお，集団間のネットワークは，従来からブロックモデルと表現されているものに近い [金光 03; Doreian 86]．ブロックモデルについては，前規定する方法もあるが，多くの場合データから帰納的に集団のネットワークをつくる．本手法の場合はあらかじめ定義を与えているほか，母集団を考慮してい

ることが違いとしてあげられる。

4.3.2 シミュレーションデータを用いた実験

ここでは、提案手法の予測精度の検証を目的に、シミュレーションデータを用い、情報伝達、クチコミに関する集団間ネットワークからの構造変化及び制約の抽出を試みる。また、これにより事前の情報共有の割合に比べてどの程度の割合で構造が変化した場合に、その変化の特定が可能かを検証する。まず、第一の実験として、外部の影響変数に大きな変化をもたらしたリンク構造の変化の特定を考える。このタスクのために図4-3に示すプロセスに従い、「一定の割合で情報伝達が行われていたが、ある時点である集団でより多くの情報伝達が発生することで全体の認知数が向上した」という現象をモデル化したシミュレーションを行った。

-
1. 500人で構成される集団を作成する。
 2. このうち50人×10個の集団を作る。
 3. まず全体500人のうち、ある確率 p_0 でランダムに情報を伝える。
 4. あらかじめ情報が伝達されたひとから確率でランダムに、不特定の別のひとへ情報が伝達される。この過程4を5回繰り返す。
 5. 続いて、500人の間である確率で情報伝達をさせ、それに加えて特定集団50人においてさらにある確率で情報伝達をさせる。
なお、同様にあらかじめ情報伝達がされている人からのみ情報伝達が行なわれるとする。
この過程を同様に5回繰り返す。
 6. 以上の過程での情報の認知数の時系列、および500人での情報のやり取りから作成された10個の集団同士で情報伝達をあらゆるリンク関係 G の時系列データを出力する。
-

図4-3. シミュレーションプロセス

具体的には500人で構成される集団を用意し、このうち全体での情報の認知数を h とする。またネットワーク G のノード V は各50人の10個の集団を用意し、リンク E はこの500人同士で情報伝達を行わせ、4.3.1に従って計算される Z_{ij} が全時間における全 Z の平均よりも高ければリンクを作成する。シミュレーションにおいて以上の方法で構成されるネットワーク系列を提案手法に対する入力 I とする。この入力データ I に対し、認知率 h が大きな変化を起こした背後の構造として、予め設定した途中から活動が活発になる集団同士のリンクを特定できるかどうかがこの問題のタスクとなる。

本研究では全体の集団における初期のリンク率 $p_0 = 0.05$ ，リンク生成の割合 p_1 は 0.005 に特定し，計 10 回のテストを行った．

実験例として認知数の時系列変化（ただし， $p_0 = 0.05$ ， $p_1 = 0.005$ ， $p_2 = 0.05$ とする）は 図 4-4 のようになる．縦軸は 500 ノード中の認知者で，横軸は時間変化である．これを見ると，任意にリンクを増加させた時間 $t=5$ から $t=6$ の間で変化が起きていることがわかる．

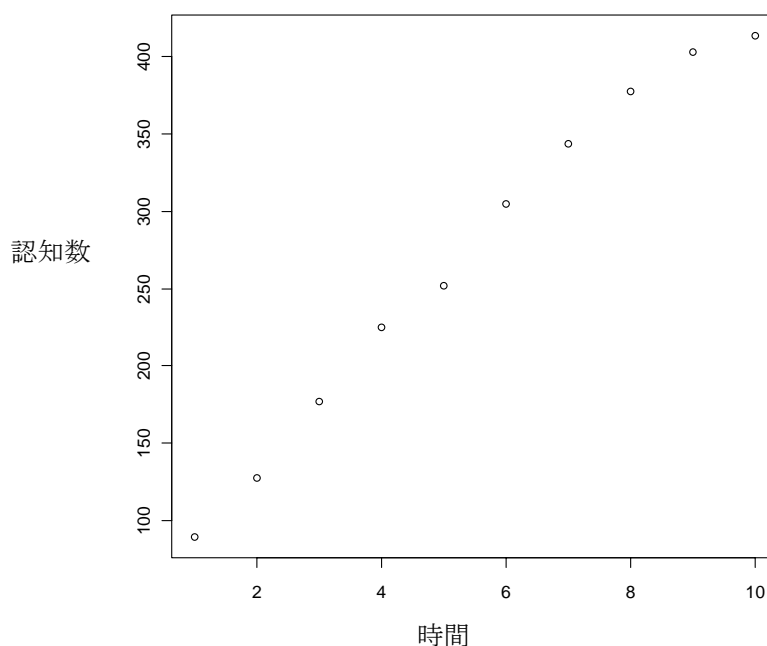


図 4-4. 認知数の時系列変化

このようにして得られたデータに本手法を適用し，得られたルールのなかで実際にある集団がクチコミを行っている様子を特定できた割合，特定率(=抽出したルールがある集団の情報伝達を特定した数/テスト数)をみる．たとえば，20 代から他の世代への情報交換を活性化させた場合において，20 代から他の世代へのリンクが抽出できるか，つまり得られたルールが $link(T, ID1, ID2, LinkAttribute)$ のときID1が 20 代となっているかどうかを確認した．表 4-1 は結果として得られたある集団の情報伝達を活性化させる割合 p_2 と，特定率である．なお，抽出ルール数は各テストで1つであった．

表 4-1. 特定結果

特定集団の情報伝達確率 p_2	0.005	0.010	0.001	0.050	0.100
特定率	0.300	0.500	0.400	0.900	0.900

結果をみてもわかるように，確率値があがるにつれて，特定する率はあがっている．特

に 5.0%以上であればほぼ確実に関係を抽出することができる。この値は本手法において、事前の情報伝達の割合に比べてどのくらいの割合(生成確率)で情報伝達の構造が変化した場合、特定(特定率)が可能かというひとつの目安となっている。つまり平均よりも追加で 5.0%以上情報伝達をする比率が高ければ特定がされやすいことを示す。それぞれの生成確率での各試行において抽出されたルールはいずれも 1 つであった。

続いて、第二の実験として一貫性制約に関するシミュレーションを行い、制約の変化の特定というタスクを考える。そのために、図 4-5 に示すプロセスに従いシミュレーションデータを用意した。このシミュレーションは特定の集団のネットワークが活性化するネットワークとは逆に、「特定のネットワークにおいて情報伝達が行われなかったが途中でその制約がなくなる、すなわち情報伝達はされるものの、自らの情報伝達はない特定の集団を用意し、ある時点から情報伝達を行わないという制約をとる」という状況をモデル化したものである。

-
1. 500人で構成される集団を作成する。
 2. このうち50人×10個の集団を作る。
 3. まず全体500人のうち、ある確率 p_0 でランダムに情報を伝える。
 4. あらかじめ情報が伝達されたひとからランダムに確率で、不特定の別のひとへ情報が伝達される。ただし、ある50人の集団では情報伝達ができないと制約をおく。
この過程4を5回繰り返す。
 5. 続いて、500人の間でランダムにある確率 p_1 で情報伝達をさせ、同時に4でおいた特定の50人の集団での制約を除外する。
この過程を同様に5回繰り返す。
 6. 以上の過程での情報の認知数の時系列、および500人での情報のやり取りから作成された10個の集団同士で情報伝達をあらゆるリンク関係 G の時系列データを出力する。
-

図 4-5. シミュレーションプロセス 2

具体的には、図 4-3 のシミュレーションの第 4 の段階で、ある集団からのリンクの出力を禁止し、第 5 の段階ではその禁止というルールを除外するというものである。この除外という操作を行った結果として、認知数の向上がみられるということを想定している。出力するデータ I は情報の認知数 h と、集団間のネットワーク G であり、構造変化のシミュレーションと同じ設定とする。これにより得られた I に対し、認知率 h が大きな変化を起こした背後の構造として、集団同士のネットワークにおける制約の変化を特定できるかどうかがこの問題のタスクとなる。

本研究では、全体の集団における初期のリンク率 p_0 を0.05に特定し、全体が5.0%, 1.0%, 0.5%の確率で情報交換を行っていくなか、制約を特定できるのかについてそれぞれで5回テストを行った。このとき特定できているか否かは、一貫性制約の中に今回情報を伝達するの制約を行った集団(20代)が含まれているか否かを調べた。たとえば、得られたルールがfalse: $-\text{link}(T, ID1, ID2, \text{LinkAttribute})$.のときID1もしくはID2に不活性集団が含まれているか否かを確認する。

その結果、5.0%, 1.0%の時の抽出した制約数はそれぞれテストで2つずつで、特定率は100% (また再現率の平均はそれぞれ0.70, 0.90)であった。しかし0.5%の場合は抽出した制約数は2つであり、制約の特定が60% (再現率は0.80)まで低下した。このことは、0.5%以下での少ない情報伝達のなかでは制約変化の特定が難しいことを示している。この主な原因は、ネットワーク全体で活性化するリンクが少なければ少ないほど、ネットワーク全体は疎な状態になり、ある一定時間で完全にリンクのない関係が、多く生じてしまうということであろう。

4.3.3 実データを用いた実験

つづいて、実際の調査で得られたデータに対して本手法の適用を行う。景気悪化の認識が時間的に変化していく過程で、ある属性とある属性の間の情報交換がどれくらい変化していったのかを調べることを目的として質問紙調査を行い、集団間のネットワークを作成した。

まず、実データでの検証に向けて、調査設計についてのべる。調査方法は、インターネット調査を用いた。時期は、2008年8月14日～8月19日で、調査対象者は日本に住む20・69歳男女個人とし、サンプル数は1,034名とし、性別年代別の男女の比率を母集団比率にあわせて抽出した。

質問としては、まず景気悪化認識について調査を行い、つづいて、そのような認識にいたった時期と人びとと情報交換を行った時期について調査をおこなった。同時にその個人の属性と景気の悪化について情報交換を行った相手の属性を調査し、さらに情報交換を行った時期についての調査も同時に行った。情報交換を行った相手については、今回最も多く話した人を特定させ、単一回答してもらったが、一般には複数回答してもらってもよい。また、誰とも情報交換をしていない人は、「誰もいない」と回答させた。今回上記のように想起により時期を特定したが、本来は継続的に調査を行っていくことが望ましい。これについては今後の研究としたい。

景気悪化認識の時期についての時系列変化を追い、変化点を特定し、つづいてその質的な変化の背景にある集団間の情報交換のネットワークの変化について、抽出された結果を表示し、その解釈についてのべる。

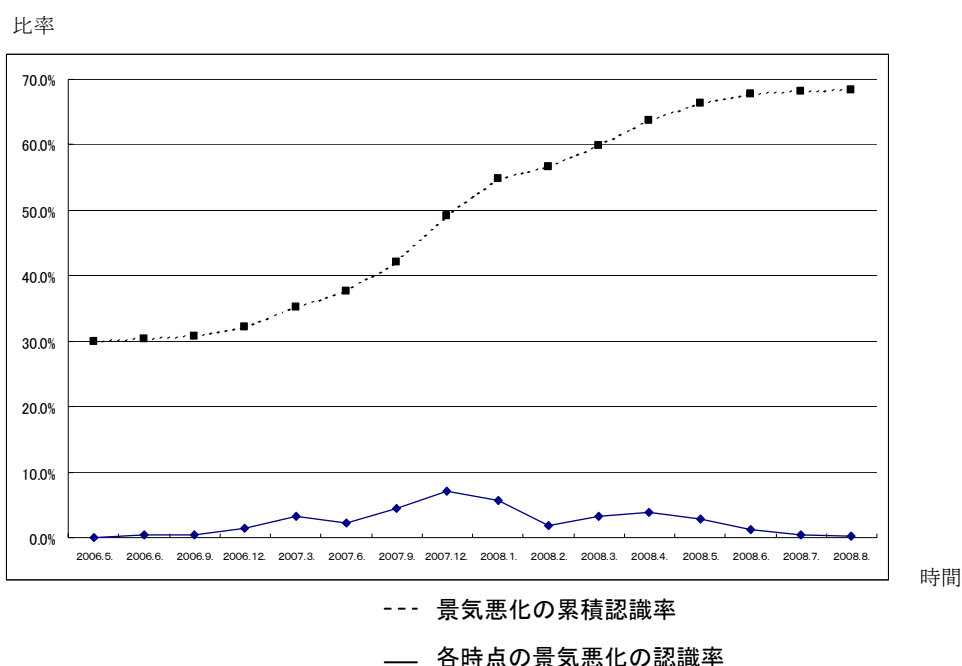


図 4-6. 景気悪化認識率の推移

図 4-6 に景気悪化の累積認識率と、その時点ごとの景気悪化の認識率を示す。この認識率が今回指標 h となる。

これらの結果、2007 年の 12 月時点と 2008 年 3 月時点で大きな変化があることがわかった。2007 年の 12 月の前の段階では景気の悪化認識が一部の人に限定されていたが、中間期では伸びが非常に高くなっており、最後の段階ではある程度、安定的な伸びに変化している。2007 年 3 月でも変化は認められるが、この時点ではサブプライムローン問題などの今回の景気悪化の原因となる要素が顕在化していない時点のため誤差と考えて除いた(同年 6 月 22 日に米大手証券ベアスターン傘下のヘッジファンドのサブプライムローンが破たんし、翌月以降株価など経済混乱が生じた)。

つづいて調査データを用いて 4.3.1 で解説した方法により、性別および年代も考慮した集団間ネットワーク G を作成した。具体的には集団間ネットワークのノード V にあたる集団としては性年代(seinen), 男・女の 20 代, 30 代, 40 代, 50 代, 60 代の 10 集団を用意した。加えて、男性を m , 女性を f とし, 20 代男性ならば $m20$ と以下記述する。また, リンク E については Z が全ての Z データの平均より高ければ $high$, 低ければ low を付加して, リンクとその属性を作成した。たとえば性年代という関係が強ければ $seinen_high$ となる。ノードの属性として, このリンクについての情報を用いて隣接行列を作成したうえで, 各集団の中心性 (主要な指標である次数中心性と媒介中心性 [安田 01]; それぞれ $degree$, $between$ と表記) を使用し, 平均以上 ($high$), 平均より低い値 (low) の 2 値に離散化したものを付加した。たとえば, 次数中心性が高い場合は $degree_high$ という属性になる。また, 時間的依存関係の知識については, リンクの前後関係を表す背景知識とノードの前

後関係、ノードからリンクへ関係を表す背景知識について組み込んだ。リンクからノードへ関係を取らなかった理由は、リンクがあった結果として何らかのノード変化が起きるという因果の解釈が困難であったためである。

以上の変化時点、および集団ネットワークGを利用して構造の変化を特定した結果、7個のルールが得られた。以下得られた結果すべてに対してその解釈を行う。ただし、一貫性制約のルールについては、否定の情報を使っているため、表層的なルールの否定として逆に解釈していく必要がある。

I. 2007年12月期以降

まず2007年の12月期前後での変化した構造の抽出を行う。なお正事例・負事例はそれぞれ4事例存在する。またルールが説明した正事例数をPosとする。Accuracyはルールが説明する正事例/(正事例+負事例)の値である。

構造変化ではカイ2乗値による独立性の検定も行った。

【構造変化】

G(T) : -link(T, 60m, 40m, seinen_low).

(Pos=2, Accuracy=1, $\chi^2 = 4.0$, df = 1, p = 0.045 < 0.05)

【一貫性制約の変化】

false : -link(T, 20m, 20m, seinen_high),

link(T, 60m, 40m, seinen_high),

link(T, 60m, 60m, seinen_high)

(Pos=4, Accuracy = 1)

false: -beforeNodeLink(T, 50m, 40m, seinen_high, degree_high),

beforeNodeLink(T, 50m, 50f, seinen_high, degree_high).

(Pos=4, Accuracy = 1)

それぞれのルールの解釈を行う。

構造変化では、まず初期段階では40代男性から60代男性への情報交換が低い程度であるがみられる。これらは60代の層は主に資産運用をしている層でもある*1ことから資産に対する不安などで情報交換を行ったものと推測できる。

一貫性制約条件をみると、以前の2007年12月より前の時期は20代の男性同士でこのような情報交換が一切行われず、60代男性同士および60代男性から40代男性への情報交換も同時に高いというわけではなかったことが伺える。また、beforeNodeLinkはある時点である属性が高い(high)ノードから次の時点で別のノードにリンクがはられるという述語であ

るので、40代男性は高い次数であったにも関わらず、50代男性へとリンクがなかったこと、そして50代男性から50代女性への伝達も低かったことがわかる。以前の時点では、40代、50代男性を中心とした人は情報は持っていたにも関わらず、それを伝達することは一部に限られており、若年や女性に対する伝達を含めあまり周辺への伝達は行っていなかったが、つづく2007年12月の段階ではこうした情報交換が起きたことがわかる。

II. 2008年3月以降

つづいて、2008年3月期以降の時点での構造の抽出を行った。結果として以下のルールが得られた。

【構造変化】

G (T): -

link(T, 40m, 40m, seinen_low),
link(T, 60m, 50m, seinen_high).

(Pos=2, Accuracy =1, $\chi^2 =4.0$, df =1, p =0.045<0.05)

【一貫性制約の変化】

false : -beforeNodeLink(T, 50m, 50m, seinen_high, between_high).

false : -beforeNodeLink(T, 20m, 50m, seinen_high, degree_high),
beforeNodeLink(T, 20m, 60m, seinen_high, degree_high).

(Pos=4, Accuracy = 1)

false : -link(T, 30m, 40m, seinen_high),
link(T, 50m, 50m, seinen_high),
link(T, 60f, 50m, seinen_high).

(Pos=4, Accuracy = 0.75)

2008年3月以降をみると、40代男性同士での情報交換が内部で起きているほか、50代男性から60代男性へと別の層に対して多くの情報交換を行っていることがわかる。

また一貫性制約条件の変化をみてみると、50代男性が媒介中心性という指標が高くなり、その後同じ50代への高い情報交換がなされたことがわかる。これは会社の重役など景気や経済動向に注意をしていると思われる50代がその内部で情報共有していることを示唆する。また、次数中心性が高く活発な50代男性から20代男性、60代男性から20代男性への徐々にこれまでなかった高い年齢層から低年齢層への情報交換が起きている。

また、この時期から40代男性から30代男性、50代同士、50代男性から60代女性への高い情報交換も生成している。

これらのことから、この時期におもに鍵になっているのは50代内部での情報交換の活発

化，そしてその結果として多くの年代へと伝達を行っていったことが最終的に大きな構造の変化をもたらしたということが示唆される。

4.4 関連研究

これまでに，動的なグラフを対象とした多くの研究が行われている。本章では主に，①時系列ネットワークモデル，②リンク予測，③時系列データマイニングの観点から関連研究について言及し，提案手法のメリット，デメリットや相違点について考察を行う。

① 時系列ネットワークモデル

時系列ネットワークモデルを扱ったものとしては，生成モデルを Newman らがサーベイなどを行っている [Newman 03] ほか，Carley が DyNet を作成し，Dark network などのモデリングした研究もある [Carley 03]。

こうした研究は，時系列に変化する構造を構成的に，もしくは演繹的にシミュレーションをするアプローチをとっている。しかし，本研究では，時系列変化するネットワーク形式のデータが与えられたうえで帰納的に分析を行って構造変化やその要因を考えるという点で異なる。

社会ネットワーク分析における時系列ネットワークモデルとしては，Wasserman らのように生成モデルに近いもの [Wasserman 80] や，さまざまな経験則や統計的な検証から，人間関係における関係構造のパターンを分類した研究などがある。Heider は，バランス理論を提唱して，符号の間に構造定理が存在することを仮定して分析を行っているが，この際構造の変化も暗に前提としている [Heider 44]。また，Katz, Procter らは微分方程式でネットワークの構造を表現している [Katz 59]。Holland, Leinhardt らは確率過程としてのモデル化し，局所構造の分布（ネットワークセンサス）などの導入も行っている [Holland 77]。このような社会ネットワークにまつわる，個々の文脈や経験，社会心理学などに基づいたさまざまな理論や仮説は，今回のたとえば事前知識として，帰納論理プログラミングの中に組み込んで検証が可能であろう。ただし，これらもある程度演繹的にモデルを構築して実証するアプローチを採用しているのに対し，本研究ではこれらの実証に関して統計的かつ帰納的にアプローチしている点が大きく異なる。

② リンク予測

続いて，リンク予測モデルといわれる，データマイニングの一分野での手法との比較をする。リンク予測とは，与えられたネットワーク中に陽に現れていないリンクを予測する問題である。たとえば，Liben-Nowell らは，中心性の高いもの同士が繋がりがやすいなど，ネットワーク指標に基づき，リンクを予測する方法を考案している [Liben-Nowell 04]。ほかには，ノード情報と構造情報を属性とする教師付き（あり）学習問題としてリンク予測

問題を定式化する研究もある[Hassan 05]. 別の観点からは、イベントベースのネットワークデータに対する、学習による予測法も提案されている[O'Madadhain 05]. そのほか時系列ネットワークを元にしたモデルを提案し、リンク予測を実現したもの[鹿島 07]などがある.

このように、ネットワークに関する属性やイベントなどを取り扱った技術の例はある. 時間要素を取り入れた手法もあるが[Pavlov 07], その数は多くない. また、リンク予測に焦点をあてていることから明らかなように、与えられるネットワークは必ずしも完全であることを前提としていない. 一方、提案手法では、完全なネットワークデータが与えられることを前提としている. また、リンク情報を利用するだけではなく、属性を含めた関係構造、複数の関係ハイパーグラフなども扱うことが出来るという点、背景知識の導入が出来るという点も挙げられる[Blokeel 07;Ketkar06].

③ 時系列データマイニング

時系列データマイニングの分野においても、グラフを取り扱った研究がある. 時系列データマイニングとは、主に時系列データにデータマイニングを適用する方法であるが、これをグラフ時系列に適用した枠組みがグラフを扱う時系列データマイニングである. 特にBurke らがやっている方法では、グラフ間の距離の定義を行い、動的に変化するグラフ構造から変化点や異常値などを探っている[Burke 04]. これらの方法は、3.1節において説明した、外部の影響変数がない場合の方法に使用できる. この方法を適用するために距離の定義もさまざまなものが提案されており、グラフ編集距離やグラフスペクトルを用いた距離などが存在する. 距離の時系列変化については、平均からの乖離や距離に対してARMA(AutoRegressive Moving Average)などの時系列モデルなどを想定する方法もある[Pincombe 07]. 同様な方法として、ネットワークデータをあらゆる隣接行列に対する特異値分解を利用して、各ノードの固有の活動度ベクトルを抽出してネットワークの情報を圧縮し、このベクトルの時系列変化を追うことで異常値を検出するという方法も提案されている[井手 04]. これらの関連研究では、グラフ間の距離を用いて変化をみているが、本研究では、グラフ構造そのものではなく、グラフが持っている外部の影響変数に着目して、それに対応した構造の違いをみるという点で異なる.

以上をまとめると、既存研究に対する本研究のメリットと相違点としては、

- (1) 時系列で変化する構造を帰納的に分析している点
 - (2) 複雑な構造や前提となるモデル、制約などを容易に導入してデータで検証できる点
 - (3) 外部の影響変数を利用し、時間変化前後で判別する構造を抽出するという点
- が挙げられる. またこうした情報は例えば今回の集団間のネットワークにおいて、単純に集団を項目としたクロス集計やその比較による差の抽出では得られない複雑な構造を有している.

一方、デメリットとしては、計算量が多くなることが挙げられる。ただし、3.2節でのべたように、重要な特徴などを選定することによりノードや属性を絞り込み、その上で複雑な構造を仮説に組み込んでILPを利用することで計算を抑えることができよう。

4.5 本章のまとめ

本章では、動的に変化するネットワーク構造、すなわち関係データを分析する手法を提案した。本手法の有効性について、具体的に集団間のネットワークの変化というものを想定し、それを調査を用いて抽出する方法を提案した。このように抽出した集団の動的な関係構造から、構造の変化や制約の変化を帰納論理プログラミングを用いることにより特定した。本手法のメリットは、主に時間変化前後で、変化した構造を抽出するという点や、複雑な構造を自由に導入できるという点である。

本章で扱った枠組みは、本研究のCの課題である、「影響変数があり時間のある関係データの分析、つまり $I = \langle GT, hT \rangle$ が与えられているもとで hT を説明するような f を学習する問題」に適用することができる。ただし、リンクとノードについての時間を含めた関係の設定、つまり前期のリンクとその次期のリンク、前期のノードとその次期のノードなどの設定はこの手法に限って用いたものであるため、この部分を必要に応じて修正すればより広い範囲に適用できる汎用的なものである。

応用としてはさまざまなセンサーデータ、調査データの分析にも用いることができ、特に集団間のネットワーク構造の変化の特定に利用が可能である。たとえば、運動についてはあるスキルが熟達するにつれてどのような変遷をたどっていったかがわかる、組織においては全体の生産性とその組織構造の変化の分析にも適用できるであろう。他にも、同様に構造を持ったデータが動的に変化する、たとえば産業連関表のような経済データに対しても利用することが可能であると考えられる。

拡張としては外部の影響変数の変化のパターンをクラスタリングし、ネットワークデータと関連付ける方法も考えられる。具体的には、急な上昇や横ばいなどのときにどのようなルールがあるかを分類するということである。

ただし限界として、判別するためのデータが少ない場合は本手法の適用は難しい。たとえば、正事例や負事例のいずれかが極めて少ない場合は判別が困難なため、変化事例の抽出は困難である。

そのほかの限界としては、変化点があることを設定していたが、 hT が漸増する場合には適用できない。これについては hT が直接数値データであれば、その値を推定するような関数を設定することも可能であろう。この場合は、今回用いた方法とは異なり、正事例と負事例をもちいていないため正事例の学習を利用する必要がある。

今後の発展としては、今回分割点はあらかじめ検出したが外部の影響変数とそれに付随する関係を同時にモデリングし、分節点前後でモデルを情報量基準などで評価をしながら

自動的に最適な分割点を探索することも必要であろう。また Allen の時区間論理[Allen 94]のように時間的な依存関係についての一つ体系を構築することにより、知識の獲得が容易になるようにしたい。加えて関係データモデル[Kemp 06]として、集団およびその属性をモデリングできる可能性もあろう。こうした関係データモデルの利用についても今後を行っていきたい。

5. 考察

影響変数が存在する関係の分析の応用として人や集団についての課題を事例に分析に取り組んだ。これらの主要な課題には、A. そもそも関係が影響しているのかという、関係の影響変数への効果の検証、B. 関係が与えられていないなかから、影響変数に応じた未知の関係の獲得、C. 関係と影響変数が動的に変化している場合の影響変数に応じた特徴抽出という3つが存在し、本研究では各課題に応じて具体的な解法を示し、検証を行った。それぞれの新規性、汎用性、限界を述べる。

まずAの課題では、商品に対する信頼と新聞や友人などの情報源に対する信頼を影響変数として、これらが「XはYの情報源である」という関係に依存していることを検証した。統計的な検定を可能にするためにネットワーク自己相関モデルを利用した。その結果、商品に対する信頼に情報源に対する信頼が寄与していることや、情報源同士の信頼も、互いに情報源であるという関係に基づいて相互に依存していることがわかった。すなわち、関係と影響変数の相互作用について統計的に検証することができた。

これにより「Gが1つだけ与えられている状況で関係がノードの属性へとおよぼす影響を調べるという問題」の形式であれば、適用が可能な方法を示した。今回扱った事例以外に考えられる応用としては、組織において人間関係を関係とし、個人の属性値（生産性など）を影響変数とした場合の影響変数への関係の効果の検証などが考えられる。

ただし、この方法の限界としてはまず関係が既知である必要があり、事前に関係がデータとして与えられていない場合は利用できないという点がある。また関係が動的に変化する問題についても扱うことができないという点も挙げられる。具体的に扱った商品に対する信頼についての課題としては、個々の人の信頼のネットワークを作成することが残されている。また、総合効果、すなわち属性などの条件も含めた効果を扱うことに焦点をあてたため、各ノードの属性効果の分離を行っていなかったことが挙げられる。構成要素についても、情報源についてさらに詳細に分離し、より多くの関係のなかからネットワークが及ぼす影響を特定する必要がある。

つぎにBの課題については、熟達者によるチェロ演奏のタスクを影響変数として与え、それを実現する筋肉の協調関係を抽出する問題を扱った。この課題でははじめに、筋肉の動きの特徴を抽出するために、時系列モデルと情報量規準を利用した筋電波形のセグメンテーションを行った。次に抽出された筋肉の動きの時区間関係を扱うために帰納論理プログラミングを利用した。実際に獲得した協調関係は、チェロの演奏に関する既存の知見とも一致した結果となった。これにより、関係が明確でないものから、背景知識を用いて新たに影響変数を説明する関係の抽出する手法を示すことができた。

これにより、「影響変数 hs とリンクの情報がないグラフ $Gs' (= Gs - E)$ が与えられているもとで、 E を獲得しつつ hs を説明する f' を学習する問題」に適用可能な方法を示した。この際、関係に関する制約などの事前の背景知識がある場合は、帰納論理プログラミングを用いた本手法が有効であることがわかった。これは2章で問題となっていた、関係が事前に与えられていない場合に関係そのものを探索するということへの対応のひとつとなっている。ただし、時系列データ、記号系列データからの学習に特化したことはこの問題特有のものである。ノード V の属性データと hs があれば基本的にこの手法が適用できるため応用する問題に応じてこの部分を修正する必要がある。応用としては、会社組織を例とすると、個人属性とその組み合わせ関係と、その組織全体の生産性・売上などの関連を扱うことが可能である。

しかし、時区間関係の影響変数に対する説明する重要度（重み）については学習できていないため、得られた関係のうちどの関係がもっとも重要であるかの詳細な判定は難しい点はこの手法の限界である。また、2章の課題と同様に関係や影響変数が動的に変化していく、スキルの熟達を例にあげると熟達するにつれてどのように筋肉の協調関係が変化していくのかを扱うことは難しい。加えて、チェロ演奏のスキルの分析における残された今後の課題としてはより多くの人に対して実験をすることで、一般的なルールを抽出すること、こうして獲得したルールを実際に初心者へフィードバックさせるシステムなどが必要であろう。

最後にCの課題については、不況の際の景気認識を影響変数、集団同士でのクチコミを関係として、景気認識の変化に対応したクチコミ関係の抽出を行った。方法としてはクチコミのコミュニケーションが大きく変化した時点特定し、次に帰納論理プログラミングを利用してリンクやノードの時間的な依存関係を扱う背景知識を導入したうえで、その前後の差分を抽出した。その結果、どの時点でクチコミが大きく変化をしているのか、変化の前後でどのような層の間でコミュニケーションの変化が起きたかといったことがわかった。これらは既存の手法では得られなかった知見である。

この研究を通じて、Cの課題つまり「 $\langle GT, hT \rangle$ が与えられているもとで hT を説明するような f' を学習する問題」に適用できる方法を示せた。その際、関係に関する制約など複雑な条件が事前にある場合は、本手法が有効であることがわかった。この問題は2章や3章で問題となっていた、関係や影響変数が時間的に変化する場合に、影響変数を説明する関係をどのようにして抽出するかという問題へのアプローチのひとつである。ただし、リンクとノードについての時間を含めた関係の設定、つまり前期のリンクとその次期のリンク、前期のノードとその次期のノードなどの設定はこの問題に限って用いたものである。そのため、この部分を問題に応じて修正すればより広い範囲に適用できる。応用としては、スキルサイエンスにおいて、スキルが熟達するにつれてどのような変遷をたどっていったか

や、会社組織においては組織の生産性と、その組織構造の変化について調べることができよう。

本手法の限界としては、判別するためのデータが少ない場合や、 hT が漸増的な場合には適用できないことが挙げられる。また **B** で提案した手法と同様に獲得した関係ごとの影響変数への寄与を示す重みの計算ができないことも挙げられる。

今後の研究としては、時間的な依存関係について、論理的なルールの形式で取り扱ったが、前後の相関を含めた重みも学習できる統計的なモデリングも考えていく必要がある。また一貫性制約の抽出問題についての形式化や問題の整理が残されている。

以上の各章における結果についてまとめたのが次の表 5 である。

また 3 つの課題とその解法について扱ったが、これらの方法の補完関係についても述べる。**B** は、**A** で課題となっていた、関係が与えられていない状態で関係を獲得するという課題へのひとつの対応となっているため前処理として利用できる可能性があるほか、**C** は、**A** や **B** で問題となっていた時間的に関係と影響変数が変化する際にどのように分析をするかということへのひとつのアプローチとなっている。その為、これらを組み合わせることで、たとえば **A** や **B** の課題で得られた関係のパターンを時系列変化で取得することが可能であれば **C** で扱った方法を用いて一つの側面から分析することが可能である。具体例をあげると、3 章においてはある時点のある熟達者のスキルの抽出を扱ったが、ある演奏者が徐々に熟達するのに伴いどのようにスキルが変化するのかという問題について、演奏の評価を影響変数とし、それに対応する筋肉の協調関係を入力として 4 章で述べた方法を利用することが可能であろう。このように本研究で述べた方法を相互に利用や補完をしあっていくことも可能である。

表 5. 本研究における各問題の新規性と今後の発展

	問題設定	具体事例	関係		既存研究の問題	本研究のアプローチ	結果	汎用性	限界	補完関係
			関係	影響変数						
課題 A 2章	関係が与えられている状況で、ノードが関係を通じて、ノードに与えている影響を調べる。	・認知：消費者が商品および商品の情報源に対して持つ信頼が、相互に依存しあっていることを検証。 ・調査データ	情報源としての依存関係	商品および各情報源の信頼	これまでの信頼研究において、ネットワークによる効果を、統計的に検証する方法が提案されていない。	統計的な検証のために、ネットワーク自己相関モデルを利用。	・商品に対する信頼に対して情報源の信頼が寄与、情報源同士の信頼も相互に、関係に基づいて依存していることが検証。	【汎用性】同じようにG(重みつき)が1つだけ与えられている場合に、関係の効果を調べたい場合に適用可能。 ただし属性の効果の分離はしていない。 応用例：会社組織という視点でみると、職場における人間関係と個々の人の生産性	・関係が既知。 ・ある一時点での効果を検証に特化、関係の時間的な変化やその影響は扱えない。	・関係が未知である場合は、3章で述べた手法を利用することができる。 ・時間的な変化を扱えない。 ・仮説の検証。
課題 B 3章	関係が事前に与えられていない状況で、影響変数を説明する関係を設定する、という問題の設定。	・運動：チェロ演奏において、熟達者がある演奏を実現するに必要となる筋肉の協調関係。 ・センサーデータ	複数の筋肉の動き同士の時空間関係	・熟達者の各拍子の演奏	これまで、時空間関係など複雑な構造をこれまでのネットワーク分析などでは使えない。	時空間論理や、外部の影響変数として楽譜の情報などを容易に扱えるように、帰納論理プログラミングを利用。	・チェロ演奏において熟達者があるタスクを演奏する際の筋肉の協調関係を抽出。 ・これらは既存の知見とも一致。	【汎用性】 h_s と $G_s^i(G_s-E)$ を問題に合わせて設定することで、Eを獲得しつつ h_s を説明するFを学習可能。 ただし、時空間論理はアドホック。 応用例：個々人の属性とその組み合わせ関係と企業全体の売上	・一時点でのルールの学習のため時間変化は扱えない。 ・論理形式の学習のため、関係から影響変数への重み学習が難しい	・時間的な変化を扱えないが、 σ により扱うことができる。 ・仮説を探索的に生成することができる。
課題 C 4章	時間により変化する影響変数と関係データが与えられているもとで、影響変数を説明する関係を学習問題。	・集団、コミュニケーション：不況時にある集団同士の関係変化。 ・調査データ	ある集団同士でコミュニケーション(クチャコミ)	景気認識における複数の相	これまでの統計やネットワークの手法は、リンクやノードの時間的な関係や関係の種類を扱うことができなかった。	時間的な関係や外部の影響変数を容易に扱うために帰納論理プログラミングを利用。 またあらかじめ探索空間が小さくなるように属性選択も行う。	・クチャコミが変化の前後でどのような層の間でコミュニケーションの変化が起きたかを特定。	【汎用性】 $\langle G, ht \rangle$ が与えられているもとで ht を説明するようなFを学習する問題に適用することができる。 ただし時間関係の設定はアドホック。 応用例：スキルサイエンスにおいて、あるスキルが熟達の際の筋肉の協調運動の変化など。	・変化点が存在していることが仮定。 htが線形トレンドで変化する場合には適用できない ・ネットのマクロな情報は組み込み。	・仮説を探索的に生成することができる。

6. 結論

「関係」というものは、単独で存在しているのではなく、関係によってつながっているものや、その関係の外部の変数に対して影響を及ぼしていることが多い。本研究の成果は、このように影響変数を伴う関係データの分析という問題と解法を整理し、人や社会における関係に関する課題に応用可能な方法を提案したことにある。

具体的にのべると、明示的に「関係が、主体や関係そして外部にある変数に対して影響を与えている」という統一的な視点から統計やネットワーク分析、機械学習で開発されてきた様々な問題やその解法をみていくことで、その特徴や限界、そして応用可能性を示唆することができた。そのうえで人や社会においてみられる関係に関する共通の課題として、A. 関係が影響しているのかという、関係の影響変数への影響の検証、B. 関係が与えられていない中から、影響変数に応じた未知の関係の獲得や、C. 関係と影響変数が動的に変化している場合の影響変数に応じた特徴抽出という3つに着目し、具体的な解法を示して知見を得た。

提案した解法はそれぞれ時間的な変化を扱えないなどの限界はあるが、取り組む課題に合わせて適切にノードやリンクを再設定することで汎用化が可能な手法である。特に、ノードやリンク同士の時間を伴ったメタ的な関係などのように従来と比べて複雑な関係構造を持っている場合には、様々な情報を統合しやすく、ラピッドプロトタイピングといった観点からも述語論理表現を用いた帰納論理プログラミングが有効であることを示した。また、提案手法の比較により、課題で用いた手法が相互に補完関係があることも示した。

このように、関係と影響変数の関連を分析することが可能になれば、新たな関係を発見すること、影響変数に関連した関係はなにかを特定すること、関係にもとづいて影響変数が今後どのようなようになるかといった予測することなどにも応用できるであろう。

今後の研究としては、まず、3章の研究の延長として発想論理プログラミングとの組み合わせによるスキルの提案など、本研究で提案した手法により学習した知識やルールの再利用をしていくことが挙げられる。次に、手法に関してはネットワーク自己相関モデルとILPを中心に扱ったが、後者については獲得した関係のなかでの重要度の比較をどう扱うかという問題に取り組む必要もある。それ以外に近年発展している統計的な関係データモデルなどをいかに利用するかということも挙げられる。この場合、複雑なデータ構造をどこまでモデリングできるかについては課題になると考えられる。最後に応用としては、提案した方法を今回扱わなかった人や組織に関連したそのほかの事象に対しても適用していくことが挙げられよう。これを通じて、関係の影響という側面から人に関して新しい知見が得られることを期待する。

謝辞

本論文は多くの方々のお力添えのもとに完成いたしました。

はじめに指導教官でアドバイザーでありました国立情報学研究所相澤彰子教授におきましては本研究全般に関して多くの御指導を賜りました。論文の記述なども含めて細かい指導をしていただいたほか、研究の過程で様々なアイデアをいただきました。また時間の都合で休日にご連絡することもありましたが、それにも対応していただくなど多くの配慮をしていただきました。そして本論文の作成にあたり、幾度も激励していただくとともにデータの収集も含めて様々な領域に分散しがちな研究であったにも関わらず、我慢強くご支援くださったことなど先生の御指導なくしては本研究は完成しなかったと思います。心より御礼申し上げます。

また副指導教官であり、本論文の主査でありました国立情報学研究所市瀬龍太郎准教授には研究の当初より機械学習の研究など様々なご指摘やアドバイスをいただき、本論文全体をまとめるのに重要な指導をいただきましたことを感謝します。

同じく副指導教官でありました国立情報学研究所井上克己教授にも定式化や定義など多くのアドバイスをいただけたことを感謝します。論文をまとめる最終段階で帰納や仮説等の対応関係など重要かつ根本的な議論をすることができたことは多くの実りがありました。

各章において、2章については、社会調査をはじめとおこなったこともあり不慣れなかでも根気強く指導していただいた国立情報学研究所小林哲郎助教の協力があったからこそまとめることができましたことを感謝します。3章については、修士時代からお世話になっています、慶應義塾大学古川康一名誉教授にメインの技術である帰納論理プログラミングなどをご指導していただいたことに感謝しています。4章については、神戸大学尾崎知伸助教に問題の形式化という点からはじまり、文体の指導までこと細かくしていただき、完成させることができたことを深く御礼申し上げます。

本論文全般をまとめるにあたっては論文審査委員である、孫媛教授、大山敬三教授の先生方からも統計的な処理の問題や、論文全体の統一性やそこからくる示唆や考察についての問題をご指摘いただくなど、多大な御指導と御鞭撻を賜りましたことを感謝します。

そして、発表の練習も含め問題点などの指摘に協力していただいた相澤研究室のメンバーに感謝します。特に鈴木崇史特任研究員には研究の方向性も含めて多くのアドバイスをいただいたことに重ねて御礼申し上げます。

総合研究大学院大学での研究に協力とご理解をいただき、時間の都合も含めて協力していただいた株式会社 JMR 生活総合研究所の方々に感謝します。特に、現福島工業高等専門学校松本行真講師には発表の練習も含めて、論文の細かな問題やその不足のご指摘、ご指導いただいたことに深く感謝します。

最後に長らく支援していただいた家族、友人知人に感謝します。

参考文献

- [Agrawal 95] Agrawal, R., Lin, K.I.H., Sawhney, S., Shim, K.: Fast similarity search in the presence of noise, scaling, and translation in Time-Series Databases, Proceedings of the 21th VLDB Conference, pp.490-501 (1995)
- [Akaike 73] Akaike, H.: Information theory and an extension of the maximum likelihood principle, 2nd International Symposium on Information Theory, Petrov and csakieds, Akademiai Kiado, Budapest, pp.267-281 (1973)
- [Allen 94] Allen, J.F., Ferguson, G.: Actions and events in interval temporal logic, Journal of Logic and Computation, pp.531-579 (1994)
- [Anselin 88] Anselin, L.: Spatial Econometrics: Methods and Models, Springer (1988)
- [Audun 06] Audun, J., Ross, H., Simon, P.: Trust network analysis with subjective logic, ACM International Conference Proceeding Series; Vol.171, Proceedings of the 29th Australasian Computer Science Conference, 48, pp.85-94 (2006)
- [Bunke 04] Bunke, H., Kraetzl, M.: Classification and detection of anormal events in time series of Graphs , (Mark Last, Abraham Kandel & Horst Bunke “data mining in time series databases”), chapter6, World Scientific, pp.123-144 (2004)
- [Bagozzi 96] Bagozzi, R.P., Geraldine, H., Pratubha, A.D., Dawn, I.: Network analysis of hierarchical cognitive connections between concrete abstract goals, In Dawn, I(Eds.), Network in marketing: SAGE publications, pp.367-383 (1996)
- [Bagnall 04] Bagnall, A.J., Janacek, G.J.: Clustering time series from ARMA models with clipped data, Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.49-58 (2004)
- [Blockeel 07] Blockeel, H., Witsenburg, Kok, T.J.: Graphs, hypergraphs, and inductive logic programming, In 5th International Workshop on Mining and Learning with Graphs (MLG07), pp.93-96 (2007)
- [Burt 92] Burt, R.S.: Structural holes the social structure of competition, Harvard University Press (1992)(安田雪訳, 『競争の社会的構造-構造的空間の理論, 新曜社, 2006)
- [Carley 03] Carley, K.M., Remminga, J., Kamneva, N.: Destabilizing terrorist networks, Proceedings of the 8th International Command and Control Research and Technology Symposium, the National Defense War College, Washington DC., Track 3, Electronic Publication (2003)
- [Chen 76] Chen, P.: The Entity-Relationship Model: toward a unified view of data, ACM transactions on database systems, 1(1), pp.9-36 (1976)
- [David 92] David, P.: Relationships between trust and risk perceptions: Literature,

- Focus Groups, and Methodological Issues, Nevada Nuclear Waste Policy Office. 1991-1992 (1992)
- [Deacon 97] Deacon, T.: Symbolic Species: The Coevolution of language and the Brain, W.W.Norton and Company Inc. (1997)
- [Dekker 03] Dekker, D., Krackhardt, D., Snijders, T.A.B.: Multicollinearity robust QAP for multiple regression, CASOS Working Paper, Carnegie Mellon University (2003)
- [Doerschuk 83] Doerschuk, P.C., Gustafson, W.: Upper extremity limb function discrimination using EMG signal analysis, IEEE Transactions on Biomedical Engineering, pp.18-38 (1983)
- [Domingos 04] Domingos, P., Richardson, M.: Markov Logic: a unifying framework for statistical relational learning, Proceedings of the ICML-2004 Workshop on Statistical Relational Learning and its Connections to Other Fields, pp.49-54 (2004)
- [Doney 97] Doney, P.M., Canon, J.P.: An examination of the nature of trust in buyer-seller relationships, Journal of Marketing, 61(2), pp.35-51 (1997)
- [Doreian 86] Doreian, P., Batagelj, V: Generalized blockmodeling, unpublished manuscript (1986)
- [Dwyer 67] Dwyer, R.F., Schurr, P.H., Oh, S.: Developing buyer-seller relationships, Journal of Marketing, 51(3), pp.22-27 (1987)
- [Codd 70] Codd, E.F.: A relational model of data for large shared data banks, Communications of the ACM, 13(6), pp.377-387 (1970)
- [Fogg 99] Fogg, B.J., Tseng, H.: The elements of computer credibility, Proceedings of CHI 99, pp.80-87 (1999)
- [Frewer 08] Frewer, L.J., Fischer, A.R., Van den Brink, P.J., Byrne, P., Theo Brock, T., Brown, C., Crocker, J., Goerlitz, G., Hart, H., Scholderer, J., Solomon, K.: Optimising the communication of probabilistic risk assessments to end-users and decision-makers, Health and Ecotoxicological Risk Assessment, 14, pp.166-178 (2008)
- [Getoor 03] Getoor, L.: Link mining: a new data mining challenge, ACM SIGKDD Explorations Newsletter archive, 5(1), pp.84-89 (2003)
- [Guralnik 99] Guralnik, V., Srivastava, J.: Event Detection from time series data, in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp.33-42 (1999)
- [Harnad 90] Harnad, S.: Symbol grounding problem, Physica D, 42, pp.335-346 (1990)
- [Hasan 05] Hasan, M.A., Chaoji, V., Salem, S., Zaki, M.: Link prediction using supervised learning, In Proceedings of SDM 06: SIAM Conference on Data Mining Workshop on Link Discovery Counterterrorism and Security (2005)
- [Hastie 04] Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning,

Springer (2002)

[Heider 44] Heider, F.: Social perception and phenomenal organization, *Psychological Review* 51, pp.358-374 (1944)

[Ho 07] Ho, T.B., Nguyen, C.H., Kawasaki, S., Le, S.Q., Takabayashi, K.: Temporal relations extraction in mining hepatitis data, *Journal of New Generation Computing*, 25(3), pp.247-262 (2007)

[Holland 77] Holland, P.W., Leinhardt, S.: A Dynamic model for social networks, *Journal of mathematical sociology*, 5, pp.5-20 (1977)

[Höppner 01] Höppner, F., Klawonn, F.: Finding Informative rules in interval sequences, *Advances in Intelligent Data Analysis, Proceedings of the 4th International Symposium, Lecture Notes in Computer Sciences* 2189, Springer, pp.123-132 (2001)

[Horvath 06] Horvath, T., Bringmann, B., L.De Raedt.: Frequent hypergraph mining, *Inductive Logic Programming, 16th International Conference on Inductive logic programming*, pp.244-259 (2006)

[Katz 59] Katz, L., Procter, C.H.: The concept of configuration of interpersonal relations in a group as a time-dependent stochastic process, *Psychometrika*, 24(4), pp.317-327 (1959)

[Kalpakis 00] Kalpakis, K., Gada, D., Puttagunta, V.: Distance measures for effective clustering of ARIMA time series, *Proceedings of the 2001 IEEE International Conference on Data Mining*, pp.273-280 (2000)

[Kemp 06] Kemp, C., Tenenbaum, J., Griffiths, T., Yamada, T., Ueda, N.: Learning systems of concepts with an Infinite Relational Model, *Proceedings of the 21st National Conference on Artificial Intelligence*, pp.381-388 (2006)

[Keogh 03] Keogh, E., Kasetty, S.: On the need for time series data mining benchmarks: A survey and empirical demonstration, *Data Mining and Knowledge Discovery*, Springer, pp.102-111(2003)

[Ketkar 06] Ketkar, N.S., Holder, L.B., Cook, D.J.: Comparison of graph-based and logic-based multi-relational data mining, *ACM SIGKDD Explorations Special Issue on Link Mining*, 7(2), pp.64-71(2006)

[Krackhardt 88] Krackhardt, D.: Predicting with networks: nonparametric multiple regression analyses of dyadic data, *Social Networks*, 10, pp.359-382 (1988)

[Krzanowski 85] Krzanowski, W.J., Lai, Y.T.A.: Criterion for determining the number of groups in a data set using sum of squares clustering, *Biometrika*, 44, pp.23-34 (1985)

[Last 04] Last, M., Kandel, A., Bunke, H.: *Data mining in time series databases*, World Scientific (2004)

[Latour 87] Latour, B.: *Science in action: how to follow scientists and engineers through*

- society. Harvard University Press(1987) (川崎勝他 訳,『科学が作られているとき』, 産業図書, 1999)
- [Leenders 02] Leenders, R.T.A.J.: Modeling Social influence through network autocorrelation: Constructing the Weight Matrix, *Social Networks*, 24(1), pp. 21-47 (2002)
- [Liben-Nowell 04] Liben-Nowell, D., Kleinberg, J.: The link prediction problem for social networks, *Proceedings of the Twelfth International Conference on Information and Knowledge Management (CIKM)*, pp.556-559 (2004)
- [Loudon 92] Loudon, G.H., Jones, N.B., Sehmi, A.S.: New signal processing techniques for the decomposition of EMG signals, *Medical and Biological Engineering and Computing*, 30(6), pp.591-599 (1992)
- [Matsuo 04] Matsuo, Y., Tomobe, H., Hasida, K., Ishizuka, M.: Finding social network for trust calculation, *Proceedings 16th European Conf. on Artificial Intelligence (ECAI2004)*, pp.510-514 (2004)
- [Money 98] Money, R.B., Gilly, M.C., Graham, J.L.: Explorations of national culture and word-of-mouth referral behavior in the purchase of industrial services in the United States and Japan, *Journal of Marketing*, 62, pp.76-87 (1998)
- [Mooney 04] Mooney, R.J., Melville, P., Tang, L.R., Shavlik, J., Dutra, I., Page, D., Kargupta, H., Joshi, A., Sivakumar, K., Yesha, Y.: Relational data mining with inductive logic programming for link discovery, *Data Mining: Next Generation Challenges and Future Directions*, AAAI Press, pp.239-254 (2004)
- [Muggleton 94] Muggleton, S., Raedt, L.D.: Inductive logic programming: Theory and methods, *Journal of Logic Programming*, Vol.19, pp.629-679 (1994)
- [Newman 03] Newman, M.E.J.: The structure and function of complex networks, *SIAM Review*, 45(2), pp.167-256 (2003)
- [O'Madadhain 05] O'Madadhain, J., Hutchins, J., Smyth, P.: Prediction and ranking algorithms for event-based network data, *ACM Special Interest Group on Knowledge Discovery and Data Mining Explorations*, 7(2), pp.23-30 (2005)
- [Oxford 71] Oxford. U. P.: *Oxford English dictionary: the compact edition*, Oxford University Press (1971)
- [Pavlov 07] Pavlov, M., Ichise, R.: Finding experts by link prediction in co-authorship networks, *Proceedings of the Workshop on Finding Experts on the Web with Semantics (FEWS2007)*, pp.42-55(2007)
- [Putnam 92] Putnam, R.D.: *Making democracy work:Civic Traditions in Modern Italy*, Princeton University Press (1992)(河田潤(翻訳)「哲学する民主主義－伝統と改革の市民的構造」NTT 出版, 2001)

- [Picco 89] Picco, D.: On a measure of dissimilarity between ARIMA models, Proceedings of the ASA Meetings, Business and Economic Stat. Section, Washington, DC, pp. 231-236 (1989)
- [Pincombe 07] Pincombe, B.: Detecting changes in time series of network graph using minimum mean squared error and cumulative summation, Australia and New Zealand Industrial Applied Mathematics Journal, CTAC2006, 48, pp.450-473 (2007)
- [Rindfleisch 01] Rindfleisch, Moorman, A.C.: The acquisition and utilization of information in new product information in new product alliances: A strength-ties perspective, Journal of Marketing, 65, pp.1-18 (2001)
- [Rissanen 78] Rissanen, J.: Modeling by shortest description, Automatica, 14, pp. 456-471 (1978)
- [Saigo 08] Saigo, H., Kramer, N., Tsuda, K.: Partial least squares regression for graph mining, Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.578-586 (2008)
- [Sirfeshmukh 02] Sirfeshmukh, D., Singh, J., Sabol, B: Consumer trust, value, and loyalty in relational exchanges, Journal of Marketing, 66(1), pp.15-37 (2002)
- [Srinivasan 99] Srinivasan, A.: The Aleph Manual
<http://www.comlab.ox.ac.uk/oucl/research/areas/machlearn/Aleph/> (1999)
- [Tantipathananandh 07] Tantipathananandh, C., Wolf, B.T., Kempe, D.: A framework for community identification in dynamic social networks, International Conference on Knowledge Discovery and Data Mining, Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.717-726 (2007)
- [Tanya 06] Tanya, Y.berger-wolf, Saia, J.: A framework for analysis of dynamic social network, International Conference on Knowledge Discovery and Data Mining, Proceedings of the 12th ACM SIGKDD International conference on Knowledge discovery and data mining, pp.523-528 (2006)
- [Wasserman 80] Wasserman, S.: Analyzing social networks as stochastic process, Journal of the American Statistical Association, 75(370), pp.280-294 (1980)
- [Washio 03] Washio, T., Motoda, H.: State of the art of graph-based data mining, ACM SIGKDD Explorations Newsletter archive, 5(1), pp.59-68 (2003)
- [Poortinga 05] Poortinga, W., Pidgeon, N.F.: Trust in Risk regulation: Cause or Consequence of the Acceptability of GM Food?, Risk analysis, 25(1), pp.197-209 (2005)
- [Xifeng 02] Yan, X., Han, J.: gSpan: Graph-based Substructure Pattern Mining, Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM02), pp.721-724 (2002)
- [Yang 03] Yang, S.: Modeling Interdependent consumer preferences, Journal of

- Marketing Research, 40, pp.282-294 (2003)
- [Ye 02] Ye, D.W., Emurian, H.H.: An overview of online trust: Concepts, elements, and implications, Computers in Human Behavior, 21, pp.105-125 (2002)
- [Zan 05] Huang, Z., Li, X., Chen, H.: Link prediction approach to collaborative filtering, International Conference on Digital Libraries, Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, pp.141-142 (2005)
- [井手 04] 井手剛, 鹿島久嗣:固有空間におけるコンピュータシステムの障害検知, The 18th annual conference of the Japanese society for artificial intelligence, 3F3-05 (2004)
- [井手 07] 井手剛: 近傍保存原理による異常検知, 2007 workshop on Information-Based Induction Sciences (IBIS2007), pp.212-217 (2007)
- [伊庭 96] 伊庭幸人: 基礎的問題から見た情報統合<「情報統合への視点」>, 人工知能学会誌, 11(2), pp.193-200 (1996)
- [猪口 04] 猪口明博: グラフマイニングと ILP システムの比較考察, The 18th annual conference of the Japanese society for artificial intelligence (2004)
- [植野 04] 植野研, 古川康一: 定性モデリングによる演奏スキルデータの個人差比較解析, The 18th annual conference of the Japanese society for artificial intelligence, 3D2-02 (2004)
- [尾崎 07] 尾崎知伸, 渡沼智己, 大川剛直: 多次元構造データからの分類知識の獲得 (Mining Classification Rules from Multidimensional Structured Databases), 人工知能学会論文誌, 22(2), pp.173-182 (2007)
- [鹿島 06] 鹿島久嗣, 津村直史, 井手剛, 野ヶ山尊秀, 平出涼, 江藤博明, 福田剛志: ネットワークデータを用いた分散システムにおける異常検出, 電子情報通信学会論文誌, J89-D, No. 2, pp.183-198 (2006)
- [鹿島 07] 鹿島久嗣, 安倍直樹: ネットワーク構造の確率的な時変モデルに基づく教師ありリンク予測, 人工知能学会論文誌, 22(2), pp.209-217 (2007)
- [金光 03] 金光淳: 『社会ネットワーク分析の基礎』, 勁草書房 (2003)
- [加藤 06] 加藤義清, 黒橋禎夫, 江本浩: 情報コンテンツの信頼性とその評価技術, 人工知能学会第 14 回セマンティックウェブとオントロジー研究会 (2006)
- [北川 05] 北川源四郎: 『時系列解析入門』, 岩波書店 (2005)
- [久保田 09] 久保田進彦, 芳賀康治: マーケティング研究におけるネットワーク・パースペクティブ, Japan Marketing Journal, 108, pp.4-19 (2009)
- [小林 07] 小林郁夫, 古川康一: 発想制約論理プログラミングによる演奏スキルの発想支援, The 21 annual conference of the Japanese society for Artificial Intelligence (2007)
- [佐藤 05] 佐藤嘉倫:市場における信頼関係とコミットメント関係, 『ネットワーク・ダイナミクス—社会ネットワークと合理的選択 (数理社会学シリーズ)』(佐藤嘉倫, 平松闊編, 勁草書房), pp.53-69 (2005)

- [佐藤 08] 佐藤健:論理コンピューティング, *Journal of Japanese Society for Artificial Intelligence*, 23(5), pp.677-686 (2008)
- [澤井 08] 澤井啓吾: 帰納論理プログラミングを利用した生体データからの動作一貫性制約の抽出, 慶応義塾大学政策メディア研究科修士論文 (2008)
- [高橋 08] 高橋公海, 森嶋博行, 松本亜希子, 杉本重雄, 北川博之: Web コンテンツ管理のための一貫性制約発見手法, *DBSJ Journal*, 7(3), pp.25-30 (2008)
- [中村 03] 中村仁彦:『ロボットの脳を創る 脳科学から知能の構成へ』, 岩波書店 (2003)
- [日本工業規格 95] 日本工業規格: データベース言語 SQL JISX3005-1995, pp.504 (1995)
- [日本リスク学会 00] 日本リスク研究学会編:『リスク学事典』, TBS ブリタニカ (2000)
- [野口 07] 野口廣:『トポロジー 基礎と方法』, 筑摩書房 (2007)
- [原 07] 原由祐子, 長谷川敦司: 情報信頼メカニズムにおけるトラストモデルの検討 (TA・TP), *The First Workshop on Information Credibility on the Web (WICOW)* (2007)
- [藤原 02] 藤原武弘:『社会的態度の理論・測定・応用』, 関西学院大学出版会 (2002)
- [古川 01] 古川康一, 植野研, 尾崎知伸:『帰納論理プログラミング』, 共立出版 (2001)
- [古川 05] 古川康一, 尾崎知伸, 植野研: 身体知研究の潮流: 身体知の解明に向けて, *the 19th annual conference of the Japanese society for Artificial Intelligence* (2005)
- [古川 06] 古川康一:スキルサイエンスの展望, *the 20th annual conference of the Japanese society for Artificial Intelligence* (2006)
- [宮木 09] 宮木由貴子:消費に関する情報の受発信と信頼性—商品・サービス購入にかかわるコミュニケーション調査より—, *Life design report ライフデザインレポート*, 第一生命経済研究所ライフデザイン研究本部, pp.24-31 (2009)
- [安田 01] 安田雪:『実践ネットワーク分析』新曜社 (2001)
- [安田09] 安田雪: ネットワーク分析の本質, *一橋ビジネスレビュー*, 57(2), pp.24-35 (2009)
- [山岸 98] 山岸俊男:『信頼の構造—こころと社会の進化ゲーム』東京大学出版会 (1998)
- [山本 07] 山本晶, 阿部誠: 消費者間ネットワークを利用したレコメンデーションエージェント, 『Web マーケティングの科学』, 日本マーケティング・サイエンス学会, 千倉書房, pp.165-191 (2007)
- [渡辺 05] 渡辺理, 樋口洋一郎: ネットワーク自己相関分析:モバイル IT システム利用行動における連携利用パタンの把握と活用に関する考察, *情報処理学会論文誌*, 46(5), pp.1222 - 1232 (2005)

研究業績

学術論文（査読付き）

金城敬太, 相澤彰子, 小林哲郎: 属性と関係構造に基づく商品信頼のネットワーク構造モデル", 経営行動科学会誌, Vol.23, No.1, pp.37-51, 2010

金城敬太, 相澤彰子, 尾崎知伸: 調査データに基づく社会構造変化の抽出, 人工知能学会論文誌, Vol.25, No.3, pp.452-463, 2010

金城敬太, 相澤彰子, 古川康一: 帰納推論による時系列データからの関係構造の抽出・スキル解析に向けたプラットフォーム, 知能と情報, Vol.21, No.4, pp.587-597, 2009

国際会議

Keita Kinjo, Aizawa Akiko: "Change detection in structured data using Inductive Logic Programming", Proceedings of Joint Meeting of 4th World Conference of the IASC and 6th Conference of the Asian Regional Section of the IASC on Computational Statistics & Data Analysis (IASC2008), pp.880-887, 2008

Keita Kinjo, Sawai Keigo, Furukawa Koichi: Time series data mining using local modeling and temporal interval logic, International Symposium on Skill Science, 2007

Furukawa Koichi, Saori Yoshinaga, Keigo Sawai, Satoshi Shimizu, Keita Kinjo: Movement in cello playing by whip motion, Joint 2005 Proceedings of the 3rd European Medical and Biological Engineering Conference (EMBEC), 2005

国内会議

金城敬太, 相澤彰子: 属性と関係構造に基づく商品信頼のネットワーク構造モデル, マーケティング・サイエンス学会 消費者の相互作用とダイナミズムに関する研究部会, 2009

金城敬太, 相澤彰子: 多重なネットワークの調査とシミュレーション, 人工知能学会第 81 回 知識ベースシステム研究会 (SIG-KBS), Vol.81, pp.23-29, 2008

金城敬太, 相澤彰子, 古川康一: 帰納論理によるネットワークデータからの知識獲得, The 22th Annual Conference of the Japanese Society for Artificial Intelligence, 2008

金城敬太, 澤井啓吾, 古川康一: 局所モデリング時系列データマイニングと帰納論理による知識獲得, The 21th Annual Conference of the Japanese Society for Artificial Intelligence, 2007

金城敬太, 澤井啓吾, 古川康一: 局所モデリングと時区間論理を用いた時系列データマイ

ニング, 人工知能基本問題研究会, 第 64 回, 2006

金城敬太, 古川康一: ILP を利用した社会ネットワーク分析, 社会情報学フェア, 情報処理学会 EIP 研究会, 2005

金城敬太, 古川康一, 尾崎友伸, 澤井啓吾: 相関ルールとネットワーク分析を用いた時系列データから知識獲得, The 19th Annual Conference of the Japanese Society for Artificial Intelligence, 2005

澤井啓吾, 金城敬太, 古川康一: 非利き手書字行為におけるスキル熟達過程からの自己組織化マップを用いた知識獲得, The 20th Annual Conference of the Japanese Society for Artificial Intelligence, 2006

西山武繁, 金城敬太, 古川康一: ボクシングにおけるスキル習熟過程について, The 20th Annual Conference of the Japanese Society for Artificial Intelligence, 2006

その他の研究・研究発表

金城敬太, 相澤彰子: 関係データと付随する変数との関連についての研究, 2009 年度新領域融合プロジェクトによる研究会 「大規模データ・リンケージ, データマイニングと統計手法」 予稿集, 2009

金城敬太, 古川康一, 相澤彰子: 帰納論理プログラミングによる定性ネットワーク分析, 2007 年度新領域融合プロジェクトによる研究会 「大規模データ・リンケージ, データマイニングと統計手法」 予稿集, 2008

金城敬太: ネットワークデータの分析とその応用, 『消費経済レビュー2007 vol.7』, 2007, JMR 生活総合研究所

金城敬太: 多重なネットワークの調査とシミュレーション, 『消費経済レビュー2007 vol.8』, 2007, JMR 生活総合研究所

金城敬太: データフュージョンの新展開, 『消費経済レビュー2008 vol.9』, 2008, JMR 生活総合研究所

金城敬太: 景気認識のミクロ的構造, 『消費経済レビュー2008 vol.10』, 2008, JMR 生活総合研究所

金城敬太: ネットワーク分析と帰納論理プログラミング, 『消費経済レビュー2008 vol.10』, 2008, JMR 生活総合研究所

金城敬太: 動的に変化する関係構造の分析についての研究, 『消費経済レビュー2009 vol.11』, 2009, JMR 生活総合研究所

付録

【付録. 1】 2章の調査で用いた質問項目

以下、本研究の調査に用いた関連する質問項目について記載する。

『次にあげた財について、それぞれ直近でみかけたものをひとつ思い浮かべながら各質問についてお答えください。』

食品（食パン）、食品（牛乳）、食品（緑黄色野菜（サラダを含む））、食品（牛肉）

Q1. 以下の商品について、お伺いします。もっともあてはまるものをお答えください。あまり深く考えずにお答えください。（ひとつだけ）※商品の品質や機能などを含めて総合的に判断してください

A.食品（食パン） B.食品（牛乳） C.食品（緑黄色野菜（サラダを含む）） D.食品（牛肉）

1. 非常に信頼している
2. まあ信頼している
3. どちらでもない
4. あまり信頼していない
5. まったく信頼していない

Q2. 以下の商品についての製造元について、お伺いします。もっともあてはまるものをお答えください。（それぞれひとつだけ）

※思い浮かべていただいた各商品の製造元ひとつについてお伺いします。

A.食品（食パン） B.食品（牛乳） C.食品（緑黄色野菜（サラダを含む）） D.食品（牛肉）

1. 非常に信頼している
2. まあ信頼している
3. どちらでもない
4. あまり信頼していない
5. まったく信頼していない
6. 製造元を知らない

Q2s1. 前問でお答えの、「商品の製造元に関する情報」はどこからお知りになりましたか。（それぞれいくつでも）

※思い浮かべていただいた各商品の製造元ひとつについてお伺いします。"

A.食品（食パン） B.食品（牛乳） C.食品（緑黄色野菜（サラダを含む）） D.食品（牛肉）

1. 友人, 知人(家族, 会社関係のひとも含む)の話
2. テレビ番組・ラジオ番組 (CM 以外)
3. 新聞記事・雑誌記事・書籍
4. ブログ・口コミサイトなど第三者のインターネット情報
5. 医師や政府の専門機関など専門家の話
6. その他【 】
7. 覚えていない

Q3. それぞれの商品を購入したかたは, 直近で購入したお店についてお伺いします. それ以外のかたは, その商品を販売している店のなかで, ふだん最もご利用になっているお店についてお伺いします. もっともあてはまるものをお答えください. (それぞれひとつだけ)

A.食品 (食パン) B.食品 (牛乳) C.食品 (緑黄色野菜 (サラダを含む)) D.食品 (牛肉)

1. 非常に信頼している
2. まあ信頼している
3. どちらでもない
4. あまり信頼していない
5. まったく信頼していない
6. お店を利用したことはない

Q3s1.前問でお答えの「商品を販売するお店の情報」について、普段どこから情報を入手しますか。(それぞれいくつでも)

A.食品 (食パン) B.食品 (牛乳) C.食品 (緑黄色野菜 (サラダを含む)) D.食品 (牛肉)

1. 友人, 知人(家族, 会社関係のひとも含む)の話
2. テレビ番組・ラジオ番組 (CM 以外)
3. 新聞記事・雑誌記事・書籍
4. ブログ・口コミサイトなど第三者のインターネット情報
5. 医師や政府の専門機関など専門家の話
6. その他【 】
7. 覚えていない

Q4. 食品について普段 (ここ半年以内) 情報源としているものをお答えください. (いくつでも)

1. 友人, 知人(家族, 会社関係のひとも含む)の話

2. テレビ番組・ラジオ番組 (CM 以外)
3. 新聞記事・雑誌記事・書籍
4. ブログ・口コミサイトなど第三者のインターネット情報
5. 医師や政府の専門機関など専門家の話
6. その他【 】

Q5. 上記の質問でお選びになった、食品の情報源についてお伺いします。もっともあてはまるものをお答えください。(それぞれひとつだけ)

- A.友人, 知人(家族, 会社関係のひとも含む)の話
- B.テレビ番組・ラジオ番組 (CM 以外)
- C.新聞記事・雑誌記事・書籍
- D.ブログ・口コミサイトなど第三者のインターネット情報
- E.医師や政府の専門機関など専門家の話

1. 非常に信頼している
2. まあ信頼している
3. どちらでもない
4. あまり信頼していない
5. まったく信頼していない

Q6. 上記の質問でお答えになった食品についての情報源の情報を、普段どこから得ていますか。それぞれについてお答えください。(それぞれいくつでも) ※たとえば、以下の項目について、友人から聞いている場合は、「4.友人, 知人(家族, 会社関係のひとも含む)の話」をお選びください。

- A.テレビ番組・ラジオ番組 (CM 以外)
- B.新聞記事・雑誌記事・書籍
- C.ブログ・口コミサイトなど第三者のインターネット情報
- D.医師や政府の専門機関など専門家の話

1. 新聞記事・雑誌記事・書籍
2. インターネット上の広告や記事
3. テレビ番組・ラジオ番組 (CM 以外)
4. 友人, 知人(家族, 会社関係のひとも含む)の話
5. ブログ・口コミサイトなど第三者のインターネット情報 (検索エンジンも含める)
6. 医師や政府の専門機関など専門家の話
7. その他/偶然みかけるなど

8. 覚えていない

【付録.2】標準化した W の計算値(行列)

	製造元	販売元	友人、 知人	テレビ・ ラジオ 番組	新聞	ネット	専門家
商品	0.25	0.29	0.11	0.16	0.13	0.05	0.01
製造元	0.00	0.57	0.10	0.15	0.12	0.03	0.02
販売元	0.00	0.00	0.35	0.28	0.23	0.12	0.02
友人、知人	0	0	0	0	0	0	0
テレビ・ラジオ番組	0.00	0.00	0.16	0.53	0.26	0.05	0.00
新聞	0.00	0.00	0.17	0.22	0.53	0.06	0.02
ネット	0.00	0.00	0.22	0.13	0.13	0.48	0.03
専門家	0.00	0.00	0.11	0.22	0.28	0.08	0.31