

因果情報を用いた経済数値予測

Forecast of Economic Indicators using Causal Information

和泉 潔¹ 坂地 泰紀¹ 佐野 仁美¹

Kiyoshi Izumi¹, Hiroki Sakaji¹, and Hitomi Sano¹

¹ 東京大学大学院工学系研究科 システム創成学専攻

¹Department of Systems Innovation, School of Engineering, the University of Tokyo

Abstract: In this paper, we propose a method that uses causal information extracted from economic texts to predict numerical indicators related to economic and financial fields, such as macroeconomic indicators and stock prices. The proposed method automatically detects whether each sentence in the economic text contains causal information or not, and if it does, it identifies the cause and effect expressions and stores them in our economic causality database. Furthermore, the proposed method calculates the similarity between the result expression of causal information contained in the economic causal database and the causal expression of another causal information, and generates causal chains from the given text data. Causal chains are used to predict how the numerical values of economic indicators will change in the future due to spillover effects.

1. はじめに

経済現象は人々の因果によって動く。つまり、その現象に関係する人たちが、自分たちが持っている因果の認識に基づいて、将来を予想し行動を決定する。その行動の集積の結果、経済システム全体の挙動が決まる。例えば「少子高齢化」を起点とした因果系列(因果チェーン)を考えてみよう。「少子高齢化→労働力低下→経済縮小」のように予想する人が株式市場で多くなり、株価の下落要因に考えられる銘柄もある。逆に「少子高齢化→自炊回数の減少→お弁当やお惣菜の利用増」のように予想され、株価の上昇要因に考えられる銘柄もある。このように、経済現象の将来予測には、人々が持っている因果の認識を分析することが必須となる。

しかし経済現象のように人間の行動が関係する事象の因果関係を、数値データだけから統計的に解析することは難しい。なぜなら、人間が原因事象をどのように認識して、それに対してどのような行動を取るのかという行動ルールが因果関係の鍵となっているからである。時と場合によって因果関係が変化する。自然科学的現象のように数値データの統計分析で、客観的かつ普遍的な因果系列を取り出すことはほぼ不可能である。

そこで、本研究では人間が認識した因果関係が含まれていると思われる経済テキストデータを解析し、経済分野に関わる因果関係のデータベースを構築す

る。さらに、特定の事象を表すフレーズから派生する因果系列を構築し、波及効果または潜在的な原因に関連する数値指標を提示するアルゴリズムを提案する。本手法を用いて、テキスト表現された因果情報を基に、ある出来事から他の出来事への波及効果や要因候補を探索し、出来事と数値指標との関係性を用いて数値の将来変化を予測したり見逃していた数値の変化を発見したりすることを行う。

本研究の主な貢献は、次の通りである。

- 自然言語処理で構築された因果系列(非構造化データ)と、数値指標(構造化データ)を統合する新たな手法を開発した。
- 本手法を用いて、関心ある事象を表すテキストから因果系列をたどって、数値指標の変化を予測する方法を新たに提案した。
- 提案手法による予測結果は、人間にも理解しやすい形での提示が可能となる。

2. 関連研究

近年、自然言語で記載された因果情報の抽出については、ニューラルネットワークベースのものが多い。Dasgupta et al.[Dasgupta 2018]はLSTMを用いて原因と結果を表す表現を抽出する手法を提案している。

さらに、英語の因果情報抽出に関しては、Coling

2020 のワークショップである Financial Narrative Processing Workshops(FNP) の Shared Task FinCausal2020^{*1} に含まれていることから、様々な手法がワークショップにおいて提案された。タスクには因果情報を含む文を抽出するタスクと、原因と結果で構成される因果情報を抽出するタスクが設定された。因果情報を含む文を抽出するタスクで提案された手法の多くが Bidirectional Encoder Representations from Transformers(BERT)に基づくもの[Ionescu 2020, Gordeev 2020, Gupta 2020]であり、高い性能が報告されている。また、原因と結果で構成される因果情報を抽出するタスクにおいても、BERTに基づく手法が提案されている[Imoto 2020]。

因果チェーンの構築に関する研究としては、Ishii et al.[Ishii 2012]や Alashri et al.[Alashri 2018], Zhao et al.[Zhao 2017]などが存在する。Ishii らは、新聞記事から因果関係を抽出し、それを SVO と上位下位関係辞書に基づいて結合する手法を提案している。Alshri らは、コンセプトに基づく因果チェーン構築手法を提案している。Zhao らは、cause-to-effect と effect-to-cause のパスを考慮した手法により、因果チェーンを構築する手法を提案している。

他にも、因果は因果チェーンを構築する以外にも様々な応用が期待されている。例えば、ロボティクスの世界では、因果を考慮した新たなベンチマークである CausalWorld[Ahmed 2020]が提案されている。

上記で述べた既存の研究では、因果抽出や因果チェーン構築だけにとどまっており、それがどのような事象や具体的な数値に関係するかが分からない。そのため、実際の経済分析に利用するには、数ステップ必要となる。それに対して本研究は、因果チェーンを数値指標に結び付け、その数値の変化予測ができる点が新しい。

3. 経済的因果関係の抽出とデータベース構築

最初に、人間が認知した因果関係について記述されたテキストデータを解析し、そこから因果関係を抽出する技術が必要となる。本システムでは、決算短信という上場企業が業績や財務状況を開示するために定期的に発行しているテキストから、手がかり表現を用いた手法[坂地 2015]により因果関係を抽出した。

- 使用したテキストデータ：2012年10月から2018年5月に約2,300社が発行した約2万個の決算短信テキスト
 - 抽出した因果関係：1,078,542個
- 抽出された因果関係は、決算短信の発行された日付、決算短信を発行した銘柄などと一緒にデータベースに保存される。

4. 因果チェーンの構築アルゴリズム

次に、上述の経済因果データベースから、特定の語句から関連する因果系列を構築する手法[和泉 2019][Izumi 2020]には、[西村 2018][Nishimura 2018]を基に改良した手法を用いた。具体的には以下の4ステップからなる(図1)。

1. ユーザーがフレーズを入力する。最初は終端ノードを入力テキストとする。
2. 終端ノードの表現と一定の類似性がある因果関係を経済因果データベースから選択し、ネットワーク追加候補とする。
3. ステップ2で抽出したネットワーク追加候補と終端ノードの組み合わせについて因果関係ノード間の類似度を計算する。因果関係ノード間の類似度が閾値 α 以上あるときにノードをネットワークに追加してネットワークを拡張する。
4. ステップ3でネットワークに追加したノードを終端ノードとしてステップ2,3を繰り返す。

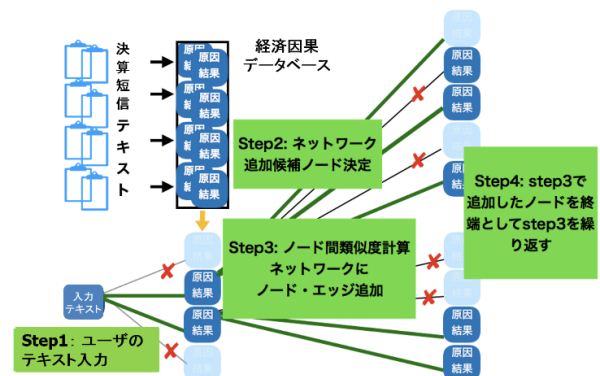


図1. 因果チェーン構築のアルゴリズム
([西村 2018] [Nishimura 2018]を基に作成)

本手法を用いて、ユーザーの入力テキストを最初の原因表現として原因-結果の順に波及効果を探索する場合を前向きの因果チェーン検索と呼ぶ。逆に入力テキストを最後の結果表現として結果-原因の

*1 <http://wp.lancs.ac.uk/cfie/fincausal2020/>

順で潜在的な原因に遡って探索する場合を後ろ向きの因果チェーン検索と呼ぶ。

5. 因果情報を用いた経済数値予測

ユーザーの入力テキストから、前節のアルゴリズムを用いて因果チェーンを構築し、次の手法により波及効果や潜在的な要因として変動が予想される数値を推定する(図 2)。

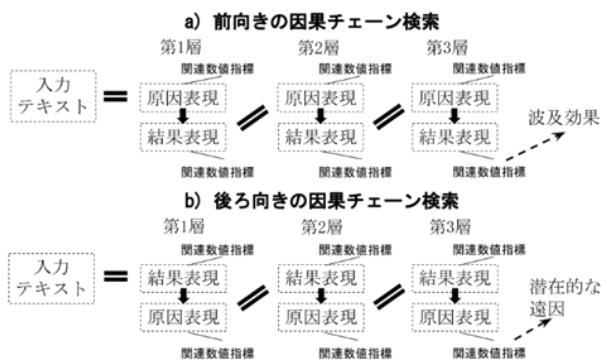


図 2. 因果チェーンによる関連数値指標の提示

前向きの因果チェーン(波及効果)検索の場合(図 2a) :

1. ある特定の現象を表すテキストからの波及効果を表す因果チェーンを構築する。
2. 因果チェーンに出現した原因表現や結果表現に含まれるテキストが、最初の特定現象からの波及効果に関連する表現となる。
3. 波及効果に関連するテキストが与えられたときに、事前に行った数値指標とそれに関連するテキスト集合の組み合わせを使用した学習結果により、テキストから関連数値指標を提示する。
4. ユーザーが与えたテキストデータに対して、結果表現に関連する数値指標を波及効果として変化しそうな指標として予測結果として提示する。

後ろ向きの因果チェーン(潜在的な要因)検索の場合(図 2b) :

1. ある特定の現象を表すテキストからの潜在的な原因を表す因果チェーンを構築する。
2. 因果チェーンに出現した原因表現や結果表現に含まれるテキストが、最初の特定現象をもたらす潜在的な要因に関連する表現となる。
3. 潜在的な要因に関連するテキストが与えられたときに、事前に行った数値指標とそれに関

連するテキスト集合の組み合わせを使用した学習結果により、テキストから関連数値指標を提示する。

4. ユーザーが与えたテキストデータに対して、原因表現に関連する数値指標を潜在的な要因として変化しそうな指標として予測結果として提示する。

6. 経済数値予測の実験結果

因果チェーン検索の結果表現から関連数値指標および数値を予測する実験を行った。本実験では、数値指標の特定に(株)Deep Data Research のオルタナティブデータを利用している。本データは上場企業のマンスリーレポート(自社サイトで開示する IR や PR 等)を収集し構造化したものであり、指標や数値のほか、報告関連の特記事項(以下「記事」という)を含む。本実験では、当該データ過去 6 年分(2015 年 1 月～2021 年 2 月)約 47 万件のうち、記事が付された約 15 万件を利用した。具体的な実験手順および結果を以下に示す。

6.1 実験手順

1. 因果チェーン検索に任意のテキストを与えて前向きの因果チェーン検索を行い、結果表現を取得する。
2. オルタナティブデータの記事をテキストマイニングし、記事に結果表現を含むデータの数値指標を「関連数値指標の候補」として取得する。
3. 上記候補と結果表現との関連度(S)を計算し、結果が上位の指標を「関連数値指標」として特定する。本研究では、この関連度について、利用データを「対象の数値指標を含むデータ群」と「全データ群」の 2 群に分割し双方の修正オッズ比により評価する。具体的には、以下の式により結果表現と数値指標との関連度を計算する。

$$S = \frac{(P+0.5) * (1-Q+0.5)}{(1-P+0.5) * (Q+0.5)}$$

ただし、

$$P = \frac{\text{対象の数値指標かつ記事に結果表現を含むデータ数}}{\text{対象の数値指標の全データ数}}$$

$$Q = \frac{\text{対象の数値指標かつ記事に結果表現を含むデータ数}}{\text{対象の数値指標の全データ数}}$$

6.2 実験結果

提案手法を用いて経済指標数値を予測した。本実験は、因果チェーン検索の波及効果を示す結果表現を得る際に、テキストのみを与えた場合と、テキストと検索対象期間を与えた場合の各事例で行った。前者はテキスト（「感染症」、「米国大統領選挙」および「地球温暖化」）を与えた3事例であり、後者はテキスト（「消費税」と2つの期間を因果チェーン検索に与えた。これらの結果表現から、変化が予想される数値指標をそれぞれ特定した。

「感染症」からの波及効果予測

因果チェーン検索にテキスト（「感染症」）を与え、波及効果を表す結果表現を取得し、当該結果表現に関連する経済数値指標を予測した。因果チェーン検索における2階層目の結果表現と強い関連度が予測された経済数値指標の上位5位を図3に示す。併せて、予測された数値指標のうち「客室稼働率」の数値変化を図4に示す。

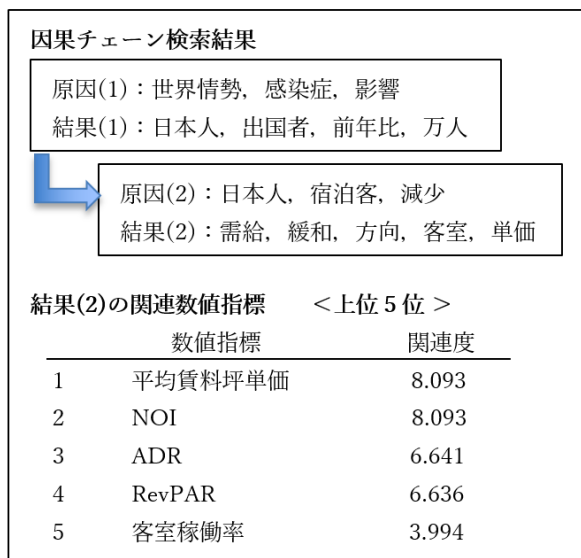


図3. 「感染症」から波及した関連数値指標

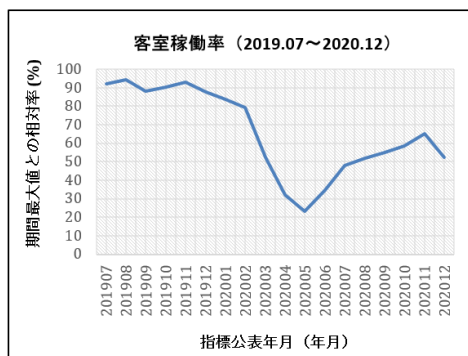


図4. 「客室稼働率」の推移

図4を参照すると、関連数値指標「客室稼働率」の推移は、国内の新型コロナウイルス感染拡大に伴い全国に緊急事態宣言が発令された2020年4月前後に急降下している。平時に比べ変動幅が大きいことから、予測された関連数値指標は「感染症」から波及した結果表現と関連性が高いと考えられる。

「米国大統領選挙」からの波及効果予測

因果チェーン検索にテキスト（「米国大統領選挙」）を与えた場合の実験結果を示す。因果チェーン検索では2階層目の結果表現に「円安」、「株高」、「景気」、「好転」等が得られた。これらの結果表現と強い関連度が予測された経済数値指標の上位3位を図5に示す。併せて、予測された数値指標のうち「外国為替取引高」の数値変化を図6に示す。

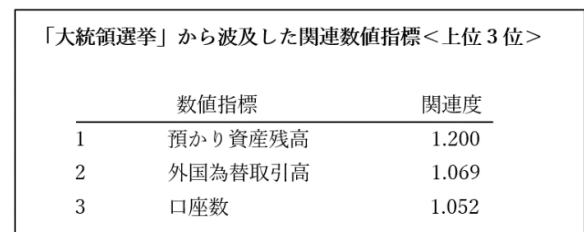


図5. 「米国大統領選挙」から波及した関連数値指標

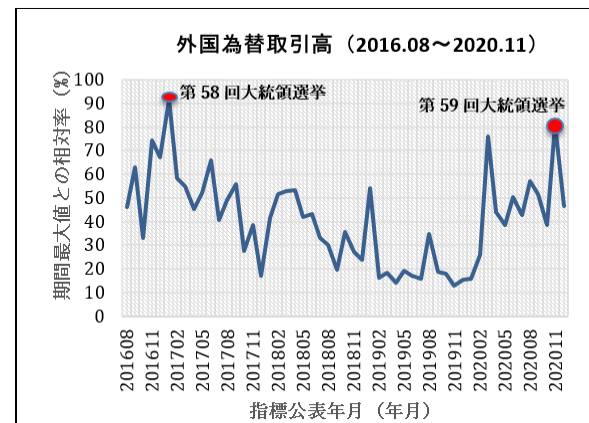


図6. 「外国為替取引高」の推移

図6を参照すると、2016年11月および2020年11月の双方の大統領選挙の時期に外国為替取引高が大幅に上昇している。2020年2月前後にコロナ禍の不安材料に伴う取引増加が見られるが、相対的には「米国大統領選挙」から波及した関連数値指標の数値が当該タイミングで変動した事実が確認できる。

「地球温暖化」からの波及効果予測

因果チェーン検索にテキスト（「地球温暖化」）を与えた場合の実験結果を示す。因果チェーン検索では1階層目の結果表現に「温室効果ガス」、「排出」、「削減」、「目標」等が得られた。これらの結果表現

と強い関連度が予測された経済数値指標の上位3位を図7に示す。併せて、予測された数値指標のうち「CO2削減量」の数値変化を図8に示す。

「地球温暖化」から波及した関連数値指標<上位3位>		
	数値指標	関連度
1	木材購入料	8.112
2	CO2削減量	1.933
3	発電量	1.045

図7. 「地球温暖化」から波及した関連数値指標

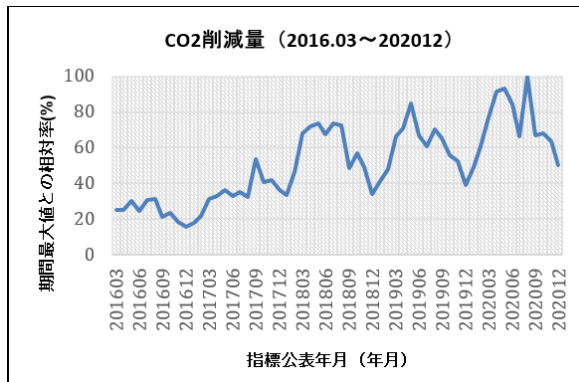


図8. 「CO2削減量」の推移

「地球温暖化」から波及した結果表現「温室ガスの削減」は世界共通の目標として進捗中であり、関連数値指標の「CO2削減量」(図8)は上昇傾向である。以上から、結果表現と関連性の高い経済指標および数値が得られたといえる。

「消費税」からの波及効果予測

因果チェーン検索にテキスト(「消費税」)および2種の対象期間を与え、提案手法を用いて経済指標を予測した。

まず、対象期間に消費税が8%に増税された2014年4月を含む「2013年4月～2014年9月」を与えた場合、因果チェーンの結果表現には「駆け込み需要」、「反動」、「物価上昇」および「懸念」等が得られた。これらは、2014年4月に実施された消費税率引き上げの前の駆け込み需要とその後の反動が波及効果に現れたと考えられる。実際に住宅市場でも駆け込み需要が起き、その後は新築を中心に低迷が続いている。関連する経済指標としては「平均賃料坪単価」、「新規契約賃料」、「更新期を迎えた既存テナントの更新割合」および「更新契約賃料」が予測された。

一方、対象期間に消費税が10%に増税された2019年10月を含む「2018年10月～2020年3月」を与えた場合、因果チェーンの結果表現には「薬価改定」、

「売上原価率」、「臨床試験」および「販売促進費」等が得られた。これらは、厚生労働省により、消費税率10%への引き上げに伴う2019年度診療報酬改定が告示されたことが波及効果に現れたと考えられる。そして、関連する経済指標には「客単価」、「店舗数」および「客数」が予測された。

このように、因果チェーン検索は、与えられたテキストと対象期間の組み合わせにより、ターゲットから波及する因果情報を鎖状に拡大し、対象期間の事象の関わりを可視化する。従って、本提案手法の適用により、因果チェーン検索を通じて、一般的には繋がりの見えにくい事象から関連する経済指標を取得し、動向を予測することが可能となる。

7. まとめ

本研究では、経済・金融に関わる事象の因果関係を、経済テキストデータから抽出したデータベースから連鎖的に検索し、波及効果または潜在的な原因に関連する数値指標を提示するアルゴリズムを提案した。本手法を用いて、具体的にいくつかの事象から人々が認識している因果関係から考えられる波及効果とそれに関連する数値指標を提示した。提示された数値指標の時系列変化を確認し、それらの変化のタイミングが、最初に与えられた事象と関連するタイミングであったことを見ることができた。

今後は因果系列の途中で出現したテキストを極性分析する手法を追加する。これにより、因果系列から推定した波及効果や潜在的な要因が関連する数値指標を増加させる方向の変化か減少させる方向の変化のどちらの影響を与えるのかを予測できる。

参考文献

- [Dasgupta 2018] T. Dasgupta, R. Saha, L. Dey, A. Naskar: Automatic Extraction of Causal Relations from Text using Linguistically Informed Deep Neural Networks, In Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue, pp. 306-316, 2018.
- [Ionescu 2020] Ionescu, M., Avram, A.-M., Dima, G.-A., Cercel, D.-C., and Dascalu, M.: UPB at FinCausal-2020, Tasks 1 & 2: Causality Analysis in Financial Documents using Pretrained Language Models, in Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation, pp.55-59, 2020.
- [Gordeev 2020] Gordeev, D., Davletov, A., Rey, A., and Arefiev, N.: LIORI at the FinCausal 2020 Shared task, in Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial

- Summarisation, pp.45–49, 2020.
- [Gupta 2020] Gupta, S.: FiNLP at FinCausal 2020 Task 1: Mixture of BERTs for Causal Sentence Identification in Financial Texts, in Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation, pp.74–79, 2020.
- [Imoto 2020] Imoto, T. and Ito, T.: JDD @ FinCausal 2020, Task 2: Financial Document Causality Detection, in Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation, pp.50–54, 2020.
- [Ishii 2012] Ishii, H., Ma, Q., and Yoshikawa, M.: Incremental construction of causal network from news articles, Journal of information processing, Vol. 20, No. 1, pp.207–215, 2012.
- [Alashri 2018] Alashri, S., Tsai, J.-Y., Koppela, A. R., and Davulcu, H.: Snowball: extracting causal chains from climate change text corpora, in 2018 1st International Conference on Data Intelligence and Security (ICDIS), pp.234–241, 2018.
- [Zhao 2017] Zhao, S., Wang, Q., Massung, S., Qin, B., Liu, T., Wang, B., and Zhai, C.: Constructing and Embedding Abstract Causality Networks from Text Snippets, in Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, p.335–344, 2017.
- [Ahmed 2020] O. Ahmed, F. Träuble, A. Goyal, A. Neitz, M. Wüthrich, Y. Bengio, B. Schölkopf, M. Wüthrich, S. Bauer: CausalWorld: A Robotic Manipulation Benchmark for Causal Structure and Transfer Learning, arXiv, 2010.04296, 2020.
- [坂地 2015] 坂地 泰紀, 酒井 浩之, 増山 繁: 決算短信 pdf からの原因・結果表現の抽出, 電子情報通信学会論文誌 D, vol. 98, no.5, pp.811-822, 2015.
- [西村 2018] 西村 弘平, 坂地 泰紀, 和泉 潔: ベクトル表現を用いた因果関係連鎖の抽出, 人工知能学会 第 20 回金融情報学研究会, pp.50-53, 2018 年 3 月.
- [Nishimura 2018] K. Nishimura, H. Sakaji, K. Izumi: Creation of Causal Relation Network using Semantic Similarity, 第 32 回人工知能学会全国大会, p. 1P104, 2018 年 6 月.
- [和泉 2019] 和泉 潔, 坂地 泰紀: 経済因果チェーン検索のシステム紹介と応用, 人工知能学会 第 22 回金融情報学研究会, 2019 年 3 月.
- [Izumi 2020] K. Izumi, H. Sakaji: Economic Causal-Chain Search Using Text Mining Technology, In: El Fallah Seghrouchni A., Same D. (eds) Artificial Intelligence. IJCAI 2019 International Workshops. IJCAI 2019. Lecture Notes in Computer Science, vol 12158. Springer, 2020.
- [坂地 2018] 坂地 泰紀, 和泉 潔, 松島 裕康: 金融テキストマイニングの基づいた投資家支援プラットフォームの開発, 人工知能学会 第 21 回金融情報学研究会, 2018 年 10 月.
- [坂地 2019] 坂地 泰紀, 和泉 潔, 松島 裕康, 川瀬 和哉, 林 寛, 接触履歴を用いた地域景況インデックスの自動生成, 日本知能情報ファジィ学会誌, vol. 21, no. 2, 2019.