経営課題にAIを!ビジネスインフォマティクス研究会 (第18回)
JSAI Special Interest Group on Business Informatics (SIG-BI #18)

個々の特徴的な因果関係を発見する技術の開発とマーケティングデータへの適用

Developing a Framework for Individual Causal Discovery and its Application to Real Marketing Data

小柳 佑介[1*]　上村 健人[1]　浅井 達哉[1]　金児 純司[1]　大堀 耕太郎[1]
Yusuke Koyanagi[1]　Kento Uemura[1]　Tatsuya Asai[1]　Junji Kaneko[1]　Kotaro Ohori[1]

[1] 株式会社富士通研究所 人工知能研究所
[1] Artificial Intelligence Laboratory, Fujitsu Laboratories Ltd.

Abstract: In this paper, we propose a framework for discovering characteristic causal relationships for each person or thing, such as each individual customer in marketing and each individual defective product in manufacturing. Our method enumerates all correlations from a given set of samples and discovers characteristic causalities by applying a causal discovery method to all subsets of samples identified by conditions derived from the correlations. After that, we can obtain a characteristic causality for each new sample by identifying a condition that the new sample satisfies. We also report experimental results on real marketing data to show the effectiveness of our method.
keywords: explainable AI, emerging pattern discovery, causal discovery, knowledge discovery

1　はじめに

近年，マーケティングや製造，医療などの様々な業務において，実問題解決のためにAIを活用して施策を立案することが増えている．解決したい問題の重要な要因を特定し，効果的な施策を立案するためには，従来のAIでよく用いられている相関関係だけでなく,「AだからBである」のように原因と結果の関係性まで表現した**因果関係 (causal relationships)** に注目する必要がある．

これまでデータ全体に対する因果関係を推定する技術が研究されている [11]．一方，多くの実問題解決のためには，個々のデータが持つ因果関係を推定することが必要である．たとえば，マーケティングの現場におけるプロモーションの場合，多くの顧客それぞれが購入につながる異なった特性をもっている．したがって，顧客一人ひとりに適切な施策を立案するためには，顧客全員に共通する原因ではなく，顧客一人ひとりにとっての原因を見つけることが必要である．

個々のデータに対する特徴的な因果関係を正確に求めるためには，対応する個人やモノに同じ条件のもとで異なる操作や作用を与えた結果を比較する必要がある．しかし，たとえば，一人の顧客に対して異なるプロモーション施策を実施した結果を同時に得ることは難しい．

我々は以下のアプローチに基づき，個々のデータに特徴的な因果関係を発見する手法を開発したので，本稿にて報告する．本手法では，与えられたサンプル集合に対して，顕在パターン発見技術 [2, 3] を用いて，目的変数との高い相関をもつ説明変数の組合せを探索し，それらを条件として得られたサンプルの部分集合に統計的因果探索技術 [11] を適用することにより特徴的な因果関係を網羅的に発見する．そして，得られた条件と因果関係の組を用いて，因果関係を知りたい新たなデータについて，特徴的な因果関係を特定する．実際のマーケティングデータに本手法を適用した結果についても報告する．

本稿の構成は以下のとおりである．2節では関連研究について述べる．3節では我々の提案手法について述べる．4節では，実際のマーケティングデータを用いた実験を報告する．5節で本稿をまとめる．

2　関連研究

データマイニングの分野では，頻出パターン [1] や顕在パターン [2] や最適パターン [6] の高速発見技術がさかんに研究されている．これらの技術では，データ

*連絡先：株式会社富士通研究所 人工知能研究所
〒 211-8588 神奈川県川崎市中原区上小田中 4-1-1
E-mail:koyanagi.yusuke@fujitsu.com

に出現するあらゆる説明変数の組合せを探索し，高頻度で共起する組合せや，教師ラベルとの相関が高い組合せを網羅的に発見する．共起や相関関係に基づき得られたパターンから，因果関係を成り立たせるパターンやルールを抽出する研究も行われている [4, 5, 9] が，条件ごとに異なる特徴的な因果関係を網羅的に見つける本研究とは位置づけが異なる．

一方，データから因果関係を推定する技術として，統計的因果探索 [11] の研究が行われている．因果探索では，データの背後にある因果構造をモデル化し，手元のデータが生成されたと考えられるモデルを推定することで，事前知識なしにデータから因果構造を推定する．これまで変数の種類や因果関係の関係式が異なる様々なモデルおよび各々の推定アルゴリズムが提案されてきている [8, 10, 12]．これらの技術は，与えられたデータに対して，その背後にある 1 つの因果構造を推定する．本研究では，先に求めた特徴的な因果関係を持つと思われるデータの部分集合それぞれに対して，因果探索技術を適用することで，データ全体ではなく個々に特徴的な因果関係を推定する．

3 提案手法

3.1 アイデア

提案手法では，個々のサンプルに対する因果関係を推定することを目的とする．しかし，対象の単一サンプルに異なる操作や作用を与えた結果を比較することは不可能であり，また，単一サンプルから統計的な因果探索技術により因果関係を推定することも困難である．

そこで提案手法では，過去のサンプル集合を用いて，説明変数が目的変数に対して特徴的な因果関係を持つための条件と，その条件下での因果関係の組をあらかじめ網羅的に求めておくアプローチをとる．条件とは，説明変数[1]の積項により記述される式とし，特徴的な因果関係を持つとは，目的変数に対して強い原因となる説明変数が存在することとする．このような原因となる説明変数を重要因子とよぶ．たとえば，条件 $x_1 \wedge x_2$ の下で x_3 が y の強い原因となる，のように，条件とその下での重要因子の組を網羅的に求め，保持しておく．興味の対象であるサンプルが新たに与えられたとき，そのサンプルに合致する条件を選択し，対応する重要因子を提示することで，個々のサンプルに対する因果関係の推定を試みる．

[1]本稿では説明変数および目的変数は二値として扱い，連続変数に対しては適当な二値化がなされているものとする．

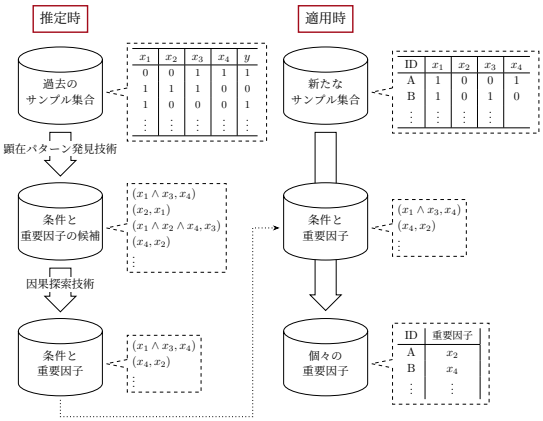

図 1: 提案手法のフレームワーク.

3.2 実現方法

本アイデアを単純に実行することは，計算量の観点から現実的ではない．ある条件が所望の性質を持つか，すなわちその条件下で重要因子となる説明変数が存在するかを判定するためには，その条件に当てはまる過去のサンプル集合に対して因果探索を実行した結果が必要となるためである．ゆえに，網羅的に条件と重要因子の組を求めておくためには，考えられるすべての条件候補に対して因果探索を実行する必要があり，これは変数の数がごく少数の場合を除き現実的ではない．

そこで提案手法では，条件の探索対象を因果関係から相関関係に緩和した問題を考えることで，因果探索すべき条件の数を効率的に絞り込む．一般に，変数 x が変数 y の原因であるとき，すなわち因果関係 $x \to y$ にあるとき，同時に x と y は相関関係にもある．これより，求める対象を「目的変数の原因となる説明変数が存在する条件」から,「目的変数と相関を持つ説明変数が存在する条件」へと緩和し，これらを先に求めておくことで，明らかに重要因子が存在しない条件を除外することができる．この緩和した条件は，顕在パターンや最適パターンの発見技術 [2, 6] により高速に求めることができることから，効率的に因果探索すべき条件候補の数を削減することが可能となる．

提案手法のフレームワークを図 1 にまとめる．まず，顕在パターン発見技術を用いて，過去のサンプル集合から，特定の条件下で目的変数と強い相関を持つ重要因子候補と，その時の条件の組を網羅的に求める．その後，求めた条件それぞれに対して，その条件下での重要因子候補が正しく重要因子であるかを判定する．具体的には，過去のサンプル集合のうち，条件を満たす部分集合に因果探索技術を適用し，目的変数の原因となる説明変数を推定することで，対象の重要因子候補が含まれているか否かを確認する．このようにして求め

た条件と重要因子をデータベースとして保持する[2]．適用時には，因果関係を知りたいサンプルに対して，そのサンプルが満たす条件をデータベースから選択し，対応する重要因子を提示する．

4 実験

4.1 実験設定

富士通のマーケティング部門が保有する顧客データに提案手法を適用し，顧客が優良顧客に変わる重要因子と，その時の条件の推定を試みる．顧客データは，優良顧客 38 件と，それ以外の顧客 468 件からなる計 506 件の顧客のサンプルで構成される．説明変数は 311 種類であり，顧客の過去の行動を示す行動属性と，所属部門や役職などの属性が含まれる．

提案手法における顕在パターン発見技術としては，岩下らの手法 [3] を利用し，また，因果探索技術としては，DirectLiNGAM[10] を用いた．本実験では，条件として探索する積項の長さは最大 3 とした．加えて，因果関係の推定精度の観点から，その条件下でのサンプル数が 100 以上かつ正例割合が 1 割以上となる条件のみを探索の対象とした．

4.2 結果

データ全体に対して求めた重要因子と，提案手法によって得られた条件と重要因子の組を示し，それぞれの結果の違いを確認する．

表 1 に，データ全体に対して因果探索技術を適用して得られた重要因子と因果効果を示す．因果効果とは，その重要因子を変化させたときの目的変数値の変化量を表し，正値の場合は優良顧客である方に影響を与え，負値の場合は逆の影響を与えることを意味する．データ全体に対する因果探索においては 5 個の重要因子が得られた．

次に，提案手法を適用した結果を示す．提案手法では，936 個の条件が抽出され，それぞれの条件における因果探索結果が得られた．本稿では，それらのうち，含まれる重要因子が少ないものと多いものの例として，以下の 2 つの条件における結果を，表 2 と表 3 にそれぞれ記載する．

条件 1 「所属組織から直近 1ヵ月で展示会 A に関するメールのクリックが少ない」

条件 2 「本人から直近 3ヵ月で製品関連のメールをクリックなし」 ∧ 「本人から展示会 A に関する WEB ページへのアクセスなし」∧ 「所属組織から直近 1ヵ月で実績紹介の WEB ページへのアクセスなし」

データ全体，条件 1，条件 2 それぞれの結果で得られた重要因子を比較する．条件 1 と条件 2 の両方において，データ全体では一番因果効果が高かった重要因子「本人が直近 3ヵ月でセミナーに申込もしくは参加が多い」よりも因果効果の高い重要因子がそれぞれ得られた．また，データ全体における因果効果の値は最大でも 0.2 程度であるのに対し，条件 1 と条件 2 の下では，ともにより因果効果の高い重要因子が得られた．条件 1 の結果においては，データ全体や条件 2 において重要因子とされた「所属組織から資料ダウンロードの WEB ページへのアクセスが多い」は，重要因子とはならなかった．一方で，条件 2 の結果においては，データ全体や条件 1 では重要因子とならなかった「所属組織から直近 1ヵ月でイベントの基調講演に申込もしくは参加なし」などが重要因子として出力された．

条件 1，条件 2 以外の提案手法によって抽出された条件においては,「所属組織から展示会 A に関する WEB ページにアクセスが少ない」「所属組織から直近 1ヵ月で実績紹介の WEB ページにアクセスが多い」「所属組織から資料ダウンロードのメールをクリックが多い」が一番因果効果の大きい重要因子として得る条件が存在した．

これらの結果から，提案手法は，データ全体に対する因果探索では得られない，特定条件における重要因子を抽出できることが確認できた．

5 むすび

本稿では，与えられたサンプル集合から特徴的な因果関係を導出する条件を網羅的に計算しておくことにより，後から与えられる個々のサンプルに対して，それぞれに特徴的な因果関係の発見を可能とする手法について述べた．また，実際のマーケティングデータを用いた実験結果について報告した．

今後は，人工データを用いた性能解析や，マーケティング，医療などの現場において個々の特徴的な因果関係の発見に関する実証評価を行う予定である．また，富士通株式会社の説明可能な AI (XAI) 技術である Wide Learning™[7][3]の説明性や納得性を強化すべく，本手法の実用化を推進していく．

[2]重要因子に限らず，各条件下での因果構造全体自体を保持することも可能である．

[3]Wide Learning™ 公式サイト「Hello, Wide Learning!」: http://widelearning.labs.fujitsu.com/

表 1: データ全体における重要因子と因果効果

重要因子	因果効果
本人の直近 3ヵ月でセミナーに申込もしくは参加が多い	0.183
所属組織から資料ダウンロードの WEB ページへのアクセスが多い	0.098
所属組織から資料ダウンロードのメールのクリックが多い	0.096
所属組織から直近 3ヵ月で展示会 A に関するメールのクリックが少ない	0.065
本人から展示会 B に関する WEB ページへのアクセスが多い	0.016

表 2: 条件 1 における重要因子と因果効果

重要因子	因果効果
本人から展示会 B に関する WEB ページへのアクセスが多い	0.866
所属企業から展示会 A に関する WEB ページへのアクセスが少ない	0.274
本人の直近 3ヵ月でのセミナーに申込もしくは参加が多い	0.206

表 3: 条件 2 における重要因子と因果効果

重要因子	因果効果
所属組織から資料ダウンロードの WEB ページへのアクセスが多い	0.440
所属組織から直近 1ヵ月でイベントの基調講演に申込もしくは参加なし	-0.434
本人が直近 3ヵ月でセミナーに申込もしくは参加が多い	0.383
所属組織からの直近 3ヵ月で展示会 A に関するメールのクリックが少ない	0.111
本人がイベントに申込もしくは参加が少ない	0.030
所属組織から直近 1ヵ月でセミナー関連の WEB ページへのアクセスが少ない	0.025
所属企業からの展示会 A に関する WEB ページへのアクセスが少ない	0.018
本人から展示会 B に関する WEB ページへのアクセスが多い	0.011
所属組織から直近 1ヵ月でイベントの展示デモに申込もしくは参加が少ない	0.011

参考文献

[1] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, Proc. 20th international conference on very large data bases (VLDB), pp. 487–499, 1994.

[2] G. Dong, J. Li, Efficient mining of emerging patterns: Discovering trends and differences, Proc. 5th ACM international conference on knowledge discovery and data mining (KDD), pp. 43–52, 1999.

[3] H. Iwashita, T. Takagi, H. Suzuki, K. Goto, K. Ohori, H. Arimura, Efficient constrained pattern mining using dynamic item ordering for explainable classification, arXiv,, CoRR abs/2004.08015, 2020.

[4] Z. Jin, J. Li, L. Liu, T. D. Le, B. Sun, R. Wang, Discovery of causal rules using partial association, Proc. IEEE 12th international conference on data mining (ICDM), pp. 309–318, 2012.

[5] J. Li, T. D. Le, L. Liu, J. Liu, Z. Jin, B. Sun, Mining causal association rules, Proc. 2013 IEEE 13th international conference data mining workshops (ICDMW), pp. 114-123, 2013.

[6] S. Morishita, J. Sese, Transversing itemset lattices with statistical metric pruning, Proc. ACM 19th SIGMOD-SIGACT-SIGART symposium on principles of database systems (PODS), pp. 226–236, 2000.

[7] 大堀耕太郎, 浅井達哉, 岩下洋哲, 後藤啓介, 重住淳一, 高木拓也, 中尾悠里, 穴井宏和: 知識発見によって信頼をつなぐ Wide Learning 技術, FUJITSU, **70(4)**, pp. 48–54, 2019.

[8] J. Peters, J. M. Mooij,D. Janzing, B. Schölkopf, Causal Discovery with Continuous Additive Noise Models, Journal of Machine Learning Research, vol. 15, pp. 2009–2053, 2014.

[9] C. Silverstein, S. Brin, R. Motwani, J. Ullman, Scalable techniques for mining causal structures, *Data Mining and Knowledge Discovery*, **4(2–3)**, pp. 163–194, 2000.

[10] S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio,

P. O. Hoyer, K. Bollen, DirectLiNGAM: A Direct Method for Learning a Linear Non-Gaussian Structural Equation Model, The Journal of Machine Learning Research, vol. 12, pp. 1225–1248, 2011.

[11] 清水昌平, 統計的因果探索, 機械学習プロフェッショナルシリーズ, 2017.

[12] K. Uemura, S. Shimizu, Estimation of Post-Nonlinear Causal Models Using Autoencoding Structure, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3312–3316, 2020.