

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 15 日現在

機関番号：13501

研究種目：若手研究(B)

研究期間：2013～2015

課題番号：25730133

研究課題名(和文) 逆包摂法に基づく仮説推論の効率化とシステム生物学への応用

研究課題名(英文) Improving logic-based hypothesis-finding methods with inverse subsumption and its applications to systems biology

研究代表者

山本 泰生 (YAMAMOTO, Yoshitaka)

山梨大学・総合研究部・助教

研究者番号：30550793

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本研究では、仮説発見問題について、逆包摂法と呼ばれる解法に着目し、実用的規模の問題を効率的に解くための技術を確立する。逆包摂法は、解探索の完全性を保証する手法であり、従来法が発見困難な仮説を導き得る仮説推論法として注目されている。他方、逆包摂法のスケーラビリティはまだ低く改善の余地がある。そこで本研究では、逆包摂法においてボトルネックとなる各種手続きの効率化をはかる。さらに細胞内の分子ネットワークのモデル記述言語であるSGBNを述語論理形式に翻訳する手法を開発する。これらの手法を通して、SGBNに基づく分子ネットワークモデルに仮説推論を適用し、モデルに欠落している新知識を導出する。

研究成果の概要(英文)：This research aims at improving logic-based hypothesis-finding methods and furthermore prompting to apply them to real problems in systems biology. First, we focus on so-called Inverse Subsumption (IS), which is a novel approach for finding hypotheses from observations with the background theory. Recently, it has been growing interests in IS to find such hypotheses that cannot be inherently obtained by the previously proposed approach. IS however has yet to achieve sufficient scalability in real problems. We consider to improve the two procedures of IS (dualization and subsumption-lattice search) in this research. Next, we focus on so-called SGBN, which is the standard markup language to describe molecular networks in systems biology. We establish an efficient way to translate SGBN into first-order logic (FOL). Together with SGBN-FOL translation, we apply hypothesis-finding methods to derive new knowledge in real SGBN-based molecular networks of cells.

研究分野：総合領域

キーワード：学習と知識獲得 システム生物学 逆包摂法 SGBN 転写因子ネットワーク 仮説推論

1. 研究開始当初の背景

仮説発見問題 (Hypothesis Finding Problem, HFP) とは、ある観測事例 E と背景知識 B が与えられたとき、 B とともに E を論理的に説明する仮説 H を求める問題である。HFP は、(1) 「述語」を必要とする高度な関係表現に対応できる点、(2) 背景知識を伴う仮説推論により欠落知識 (ミッシングリンク) を補完できる点など、決定木やベイズ法、カーネル法などの分類問題を扱う他の機械学習法にない特徴を持つ。

これらの特徴から近年 HFP をシステム生物学分野に応用する研究が進められている。システム生物学の目的は、細胞内の遺伝子制御、代謝、シグナル伝達といった個別の系を一つのシステムとしてモデル化することである。モデルの大規模・複雑化が急速に進むなかで、HFP を用いて生体モデルに隠れた相互作用や未知のモジュールを発見する取り組みが進められるようになってきている。

このような研究動向を受け、研究代表者らは、HFP の従来システムが持つ問題点: 解探索の不完全性と解の一意性に焦点をあて、これらを解決するための推論技法の確立と HFP のシステム生物学応用に関する研究に課題を推進してきた。

特に推論技法については、完全な仮説構成法である逆伴意法 (Inverse Entailment, IE) に着目し、IE を論理的に簡略化した逆包摂法 (Inverse Subsumption, IS) を提案している。IS の特徴は、IE の解探索の完全性を保証しながら、その探索空間をある包摂束に絞り込める点にある。束構造を取るデータ探索は人工知能とデータマイニング分野で広く研究されており、伴意に基づく煩雑な一般化操作を扱う IE に比べ、IS では効率的に仮説を探索することができる。

2. 研究の目的

そこで本研究では、HFP の実規模問題への応用を推進するため、IS に基づく仮説推論エンジンの効率化を行い、そのスケーラビリティを向上させる。さらに HFP の実際の応用として、生体ネットワークの知識発見にも取り組む。

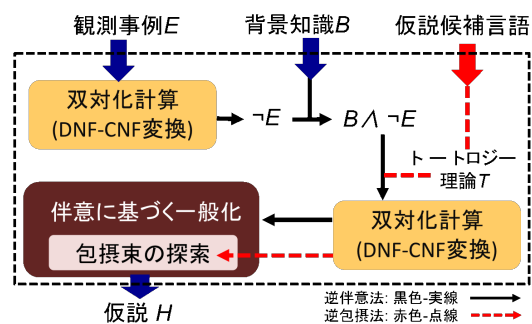
生体ネットワークのモデル設計は SBGN (Systems Biology Graphical Notation) と呼ばれる Markup 言語が標準化されており、統一規格に基づく設計が可能となっている。他方、従来の HFP の応用研究では、対象とする生体機構各々を解釈し、論理式に変換しなければならなかった。この作業は、生物と情報の専門知識を必要とする高度な技術が必要であり、HFP 応用を妨げる要因となっている。本研究では、SBGN 規格を述語論理に基づく関係表現に翻訳する課題に取り組む。これにより、SBGN 規格に準拠した任意の生体ネットワークに対して HFP を利用できる

ようになり、仮説推論エンジンによる応用研究が進捗することが期待できる。また、細胞の複雑な生体ネットワークに関する理解にも寄与できる。

3. 研究の方法

以上のような背景・目的のもと、本研究では次の2つの研究課題に取り組んだ。

(1) IS 法のスケーラビリティ向上: 下図は、IE 法と IS 法の手続きの流れを示したものである。IE と IS はともに、双対化と一般化と呼ばれる2つの手続きから構成される。



(非単調)双対化計算の効率化

双対化は、否定式などの DNF 式を CNF 式に変換する操作である。IS 法では、入力である B と $\neg E$ に加え、仮説候補言語から構成可能なトートロジー節集合 T をあわせてなる式の否定を双対化によって CNF 形式に変換する。追加されるトートロジー節の個数によって出力サイズは指数的に増加するため、圧縮表現を用いてコンパクトかつ高速に出力を求めることが本課題内容である。

包摂束上の探索の効率化

双対化による出力をボトム理論と呼ぶとき、IS 法では、ボトム理論を下限にもつ包摂束に探索空間を絞り、仮説を導く。しかし、この仮説導出は組み合わせ爆発を引き起こす難しさを持っている。本課題では、この包摂束上の探索を効率化するための各種のアルゴリズム技法を検討していく。

(2) SBGN ネットワーク上の知識発見:

SBGN ネットワークの翻訳

SBGN では、モデル化する分子間作用の抽象度に応じて、3つのクラス (PD, ER, AF) に分類されるが、各クラスの制約を一階述語論理形式 (First-order logic, FOL) に変換することが本課題内容である。

SBGN ネットワーク上の知識発見

上記の SBGN-FOL 変換を通して、SBGN ネットワークを論理式に変換し、HFP の枠組みを用いて分子ネットワーク上の未知の相互作用を発見する課題に取り組む。

知識発見を誘導するネットワーク可視化
知識発見の主体者はユーザー（実験生物学者）であり、有益な発見につなげるには、仮説をさまざまな観点から分析・検証することが求められる。仮説検証を含めて対話的にユーザーの知識発見をサポートする可視化インタフェースを作成する。

4. 研究成果

(1) 双対化に関しては、BDD/ZDD を用いた非単調双対化計算法を検討した。この中で、BDD/ZDD を用いた単調双対化計算法を非単調双対化問題に適用できるかどうか、帰納推論の実問題を用いて評価実験を行った。この結果、従来の列挙木をベースとするものに比べ、数倍程度の速度改善を得ることがわかった。本研究成果は、単調双対化計算に基づく非単調双対化法の提案アルゴリズムとともに 1st Int. Workshop on Learning and Nonmonotonic Reasoning (LNMR2) にて発表した（論文）。

(1) 包摂束上の探索問題の最も基礎的な共通問題はトランザクションデータから頻出アイテム集合を列挙する問題である。この問題は、組み合わせ爆発現象を扱うデータマイニング分野でよく知られた問題である。本研究では、仮説探索を高速化するための重要な基盤技術と位置づけ、これを解く新しいオンライン型近似手法を開発した。提案法は、与えられた計算資源（リソース）のもとで高速かつ省メモリに逐次処理を行うことができ、大量の突発的イベント（例：地震やハリケーン）を含む大規模トランザクションに対しても遅延時間やメモリリークを発生させず処理が可能である（耐バースト性）。本成果はデータマイニングならびにデータベース分野のトップ会議である SIGMOD2014 に採択されている（論文）。またこの「リソース指向型」近似手法と仮説推論を融合する取り組みも進めている（発表）。

(2) SBGN の3つのクラスのなかで最も抽象度の高い AF クラスについて、各種の制約を論理プログラム（論文）と FOL（図書）に帰着する手法をそれぞれ提案した。本成果は、SBFN ネットワーク上の推論について研究を進めている C. Froidevaux 教授（フランス LRI）らの研究グループとの共同研究により得られた。また、この成果をふまえて、2013年と2014年に日仏合同シンポジウムを LRI において開催し、GPCR 調節機構の解明にアドレスする研究プロジェクトを推進した（本課題は平成26年度国立情報学研究所戦略公募型共同研究事業として採択されている）。

(2) 真核細胞のモデル生物である出芽酵母のグルコース抑制機構（GRS）に注目し、HFP の枠組みを用いて、GRS に関与する遺伝子の発現変化量（マイクロアレイ）から GRS

に欠落する制御因子を発見する課題に取り組んだ。この中で、GRS の先行モデルに欠けていた Mig1p タンパク質から YGL157W 遺伝子への発現抑制を発見することに成功している。本成果は、当該分野の国際会議である FMMB2014 にて発表している（論文）。さらに、前述した日仏共同研究のなかで、FSHR 調節機構を取り上げ、FSHR の先行モデルに欠落している転写制御因子を発見する取り組みも進めた。FSHR は性腺刺激ホルモン受容体の一つであり、性腺機能低下症などの疾患に関与する GPCR である。この中で、先行モデルでは欠落していた、G タンパク質と PI3K の2つの調節系をクロストークさせる制御因子を導出することに成功している（発表）。

システム生物学においては、さまざまなパスウェイをまとめて全体の系を構成するが、上述した2つのケーススタディを通して、これらのパスウェイのクロストークに関する知識発見に HFP が効果的であるという感触を得ている。

(2) 出芽酵母の転写調節の制御ネットワークモデルとマイクロアレイデータから、モデル上の各転写因子の活性度を推定する手法を提案し（発表）、推定された転写活性全体を可視化するツールを開発するとともに（発表）、連携研究者の協力のもとツールの可用性についてヒアリングを実施した。プロトタイプ版のため、操作性については改善を要するものの、可視化によって、転写制御全体の様子を俯瞰できるようになり、細胞の表現型の動的変化を引き起こすメカニズムを理解する助けになるのではないかという感触を得ている。今後も引き続き、(2-B) と連動させながら本課題に取り組み、データ指向型の知識発見支援ツールとしてシステム開発を進めていく予定である。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕(計 4 件)

Yoshitaka Yamamoto, Adrien Rougny, Hidetomo Nabeshima, Katsumi Inoue, Hisao Moriya, Christine Froidevaux and Koji Iwanuma. (査読あり)

Completing SBGN-AF Networks by Logic-Based Hypothesis Finding. Proceedings of the 1st International Conference on Formal Methods in Macro-Biology (FMMB2014), Lecture Notes in Bioinformatics, volume 8738, pages 165-179, 2014.

Yoshitaka Yamamoto, Koji Iwanuma and Shoshi Fukuda.

Resource-oriented Approximation for Frequent Itemset Mining from Bursty Data

Streams. Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data (SIGMOD'14), pages 205-216, 2014. (査読あり)

Adrien Rougny, Chrisine Froidevaux, Yoshitaka Yamamoto and Katsumi Inoue. Translating the SBGN-AF Language into Logics to Analyze Signalling Networks. Post-Proceedings of the 1st International Workshop on Learning and Nonmonotonic Reasoning (LNMR2013), CoRR, volume 975, pages 43-54, 2013. (査読あり)

Yoshitaka Yamamoto, Koji Iwanuma and Hidetomo Nabeshima. Practically Fast Non-monotone Dualization based on Monotone Dualization. Post-Proceedings of the 1st International Workshop on Learning and Nonmonotonic Reasoning (LNMR2013), CoRR, volume 975, pages 55-66, 2013. (査読あり)

〔学会発表〕(計 5 件)

平沼 悠人, 山本 泰生, 守屋 央朗, 宋 剛秀, 岩沼 宏治. 分子ネットワーク上の状態推定とその可視化による知識発見支援. 情報処理学会第 45 回バイオ情報学研究会 (SIG BIO), 2016. 3 月 19 日, 北陸先端科学技術大学院大学, 石川県能美市

Yoshitaka Yamamoto and Koji Iwanuma. Online Pattern Mining for High-Dimensional Data Streams. Proceedings of IEEE BigData2015, pages 2615-2617, 2015 poster presentation. 10 月 30 日-11 月 1 日, サンタクララ (米国) (査読あり)

Yoshitaka Yamamoto, Resource-oriented Online Approach for Itemset-Mining and Hypothesis-Finding. Presentation at The ILP NII-Satellite Meeting, 2015. 8 月 26 日, 国立情報学研究所, 東京都千代田区

Adrien Rougny, Yoshitaka Yamamoto, Hidetomo Nabeshima, Gauvain Bourgne, Anne Poupon, Katsumi Inoue, and Christine Froidevaux.

Completing Signaling Networks by Abductive Reasoning with Perturbation Experiments. Proceedings of the 25st International Conference on Inductive Logic Programming (ILP2015), 2015.8 月 20 日-22 日, 京都大学楽友会館, 京都府左京区 (査読あり)

平沼 祐人, 山本 泰生, 岩沼 宏治. 遺伝子発現データを用いた転写因子束縛ネットワークの状態推定. 人工知能学会全国大会 2014, 5 月 15 日, ひめぎんホール別館第 15 会議室, 愛媛県松山市

〔図書〕(計 1 件)

Adrien Rougny, Christine Froidevaux, Yoshitaka Yamamoto, Katsumi Inoue. Analyzing SBGN-AF Networks using Normal Logic Programs. In handbook "Logical Modeling of Biological Systems", Katsumi Inoue, Luis Farinas (eds.) IStE-Ltd, ISBN: 978-1-84821-680-8, 2014.

〔その他〕

(1) 仮説推論システム:

<http://www.iwlab.org/our-lab/our-staff/yy/software>

(2) 研究代表者 HP

<http://www.iwlab.org/our-lab/our-staff/yy>

(3) 日仏ワークショップ HP

<https://www.lri.fr/~rougny/workshop2014.html>

6. 研究組織

(1) 研究代表者

山本 泰生 (YAMAMOTO YOSHITAKA)

山梨大学・大学院総合研究部・助教

研究者番号: 30550793

(2) 研究分担者

なし

(3) 連携研究者

なし

(4) 研究協力者

守屋 央朗 (MORIYA HISAO)