

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年5月31日現在

機関番号：13501

研究種目：若手研究（B）

研究期間：2010～2013

課題番号：22700141

研究課題名（和文）システム生物学応用に向けた仮説枚挙システムの実現

研究課題名（英文）Hypothesis Enumeration for Knowledge Discovery in Systems Biology

研究代表者

山本 泰生（YAMAMOTO YOSHITAKA）

山梨大学・大学院医学工学総合研究部・助教

研究者番号：30550793

研究成果の概要（和文）：本研究では、従来の仮説推論システムが持っている2つの問題点：解探索の不完全性と解の単一性を解決することを目的とし、仮説構成手続きの簡略化による解探索空間の削減と仮説の逐次枚挙を可能とする仮説推論システムの提案及び開発を行った。また酵母菌の遺伝子制御ネットワークにおいて未だ確認されていない制御因子を仮説推論の枠組みで発見する課題に着手し、システム生物学における仮説推論の新たな応用展開に取り組んだ。

研究成果の概要（英文）：This research project aims at studying the methodology for finding inductive hypotheses in inductive logic programming (ILP), and applying it to systems biology. We first address two problems: incompleteness and uniqueness in terms of hypotheses obtained by the previously proposed ILP methods. We then have developed a new ILP engine that can overcome the two problems. Next, we have applied it to the completion problem in a biological system, called the glucose repression system of *Saccharomyces cerevisiae*, in order to find gene regulatory factors that are missed in the prior system.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,300,000	390,000	1,690,000
2011年度	800,000	240,000	1,040,000
2012年度	800,000	240,000	1,040,000
年度			
年度			
総計	2,900,000	870,000	3,770,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報額

キーワード：学習と知識獲得，仮説枚挙

## 1. 研究開始当初の背景

近年、仮説推論技術のシステム生物学応用が進められている。仮説推論のタスクは新しく観測した事例に対して、それを論理的に説明するような仮説を求めることである。言い換

えると、仮説推論は現在の知識に欠落しているファクトやルールを仮説として導出することができる。このような背景理論を伴う理論補完は決定木やベイズ法など分類学習を扱う従来の機械学習手法では難しい仮説推論特有の機能である。

細胞内の複雑な分子ネットワークの予測を目指すシステム生物学にあって、近年、仮説推論は、現在の生化学情報に欠落している因果規則を推定する新技術として注目されている。例えば、化学物質のもつ代謝系の阻害効果を推定する問題や遺伝子発現ネットワークを補完する問題に仮説推論が実際に適用された研究事例が報告されてきている。

しかし研究を開始した当初は、従来の仮説推論システムが持つ2つの問題点：**解探索の不完全性** および **解の単一性** から、生物学の専門家が積極的に利用されるまでには至っていなかった。そこで本研究では、これら2つの問題点を解決する新しい仮説推論システムを構築することにより、システム生物学の諸問題に対する新しいソリューションを提供することを目標とし、以下に述べるような研究課題に着手した。

## 2. 研究の目的

本研究において取り組んだ課題は以下の2つのトピックである。

- (1) 仮説枚挙システムの実現とその高度化
- (2) 帰納推論技術による酵母菌のグルコース抑制に関する知識発見

課題(1)では、従来システムがかかえていた解探索の不完全性と解の単一性の問題を解決することを前提とし、論理的に考えられる任意の仮説を完全に構成可能な仮説推論アルゴリズムの理論的検討および複数の解を同時に提示することが可能な仮説枚挙システムの開発の2つを目的とした。

課題(2)では、課題(1)で開発する仮説推論技術をシステム生物学における未解決問題に実際に適用しその可用性を確認するとともに、仮説推論の枠組みでシステム生物学における知識発見を誘導する応用研究をさらに推進させることを目的とした。

## 3. 研究の方法

課題(1)では、**逆伴意法 (IE)** と呼ばれる完全な仮説構成手続きに焦点をあて、IE法に基づいて、ユーザーの欲する全仮説を枚挙する新しい仮説推論システムを開発することとした。IE法は任意の仮説を構成できる反面、いくつかの非決定的な手続きから成り、自動的に解を枚挙することは困難とされていた。そこでこの問題を解決し、本研究課題の目的を達成するため、次のような2つの部分課題に取り組んだ。

- ① 逆伴意法の論理的簡略化
- ② 仮説枚挙システム開発とその効率化

課題(2)では、扱うべき未解決問題として、**酵母菌のグルコース抑制機構**に焦点をあてた。グルコース抑制は真核生物が共通して持つ必須の糖代謝機能であるが、遺伝子制御、シグナル伝達系などのさまざまなレベルの機構が複雑にクロストークしていることが知られている。近年、システム生物学的アプローチに基づき、これらの個別の系が一つの生体ネットワークとして再構築されるようになってきている。しかしその一方で、現行モデルにはない欠落しているインタラクションが存在していることが知られている。

そこで本研究では、以下のような3つの部分課題に取り組み、仮説推論の枠組みで、現行モデルにはない因果規則を、実験データから論理的に導出し、生物学の専門家(研究協力者: 岡山大・守屋准教授)とともに仮説の妥当性を確認 (corroboration) した。

- ① グルコース抑制に関するネットワークモデルの論理形式変換
- ② 仮説推論による隠れた遺伝子転写因子の自動発見
- ③ 導出した仮説の生物学的確認

## 4. 研究成果

### 課題(1)① 逆伴意法の論理的簡略化

従来の仮説構成手続きには逆伴意法 (IE) のほかに逆包摂法 (IS) と呼ばれる手法が存在する。IE法は任意の仮説を構成できる反面、いくつかの非決定的な手続きから成り、これまで自動的に解を枚挙することはできなかった。一方、IS法は不完全な手続きであるが、包摂束上の効率的な束探索手法を用いることが可能であり、IE法に比べると解探索は比較的容易に制御することができる。

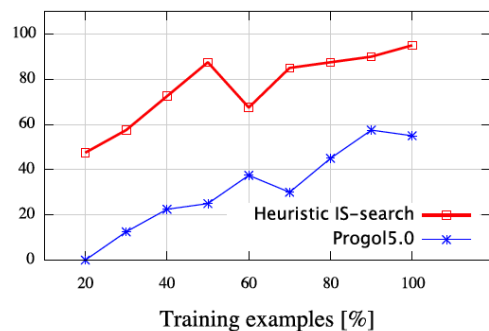
本研究課題では、IE法をIS法に論理的に変換する簡略化手法を構築し、IS法に基づく完全な仮説構成法を世界に先駆けて提案した。本研究成果は当該分野の国際学会であるILP2010において発表し、最優秀学生論文賞を受賞し、学術論文はMachine Learning誌に掲載された(論文②, 発表⑦)。

### 課題(1)② 仮説枚挙システム開発と効率化

IS法に基づき仮説を逐次枚挙できる新しい仮説推論システムを開発した。また仮説の発見的評価関数を設定し最良優先探索に基づくヒューリスティクス探索をIS法に導入した。これにより仮説探索にかかる計算コストを効率化した。また従来の代表的な仮説推論システムであるProgolに比べて予測精度の高い仮説が実際に生成されることを実験により確かめた(論文①, ③, 発表⑥)。

(図1参照、ただし本実験に用いた問題は小規模な問題にとどまっております、大規模問題を用いた検証は今後さらに行う必要がある)

図 1: IS 法により導出した仮説の予測精度



また IE 法 (ならびに提案している完全 IS 法) では、ブール式の双対化計算 (DNF-CNS 式変換) を必要としているが、これには非常に大きな計算コストがかかっていた。そこで本研究では、任意のブール式に対する一般双対化問題が、単調なブール式を扱う単調双対化問題に帰着できることを理論的に明らかにし、これまでに提案されている効率的な単調双対化アルゴリズムを用いて双対化計算を行うよう仮説推論システムを改善した (発表①, ⑤)。

#### 課題(2) 帰納推論技術による酵母菌のグルコース抑制に関する知識発見

酵母菌のグルコース抑制機構については、2009 年にシステム生物学的アプローチに基づいて再構築された生体ネットワークが存在する。このネットワークは、それまでに関係することが知られていた遺伝子制御系とシグナル伝達系の 77 個の頂点 (遺伝子 50, タンパク質 22, 細胞内代謝物 3, 細胞外代謝物 2) および 118 個のハイパーエッジから構成されるハイパーグラフとして表現されている。

しかしながら、現行のネットワークモデルを用いた遺伝子発現の予測精度は平均 55% 程度にとどまっており充分高いとはいえなかった。そこで本研究では現行ネットワークの予測精度を向上させる因果規則を仮説推論によって論理的に導出させる研究課題に取り組んだ。

一般の仮説探索空間は莫大なものとなる。このため、対象とする探索空間を生物学的に許容できる想定のもとで限定することとした。この限定した探索空間のもとで計算機実験を行った結果、予測精度を 90% 程度にまで向上させることが可能な仮説を実時間で発見することができた。なお、このような予測精度を最大にする仮説は複数存在しており、本課題ではそのすべての仮説を枚挙することに成功している (発表④)。

また導出した仮説集合を研究協力者である専門家に確認してもらった結果、酵母菌の遺伝子転写機構に関するデータベース

YEASTRACT において最近確認されている (このため現行モデルには含まれていなかった) 7 つの転写因子 GAL10, FBP1, PCK1, MLS1, SFC1, CAT2, IDP2 を実際に抽出できていることを確認した。さらに 2 値ではなく多値論理の枠組みを導入することで、より正確な定性解釈のもとで遺伝子発現データを離散化し、多値論理に基づき発現データと生体ネットワークモデルの整合性を再検証する取り組みも行っている (発表②)。

仮説推論が実問題に展開されていくなか、一つではなくある程度の複数解をユーザーへ提供できる柔軟なシステムが求められてきている。このような背景のもと、課題(1)では、ユーザーの求めたい仮説探索空間に属する解のみを逐次的に枚挙する仮説推論システムを実現した。またこのような仮説推論システムの解集合は従来システムの任意の解を包含するものであり、実際に本研究を通して、従来システムでは発見できなかった新仮説をユーザーに提供できたことを小規模問題により実証している。

近年、生化学データが急速に蓄積されるようになり、その学術利便性も向上しているなか、コンピュータサイエンス的なアプローチによって生命、生体、細胞に隠された知識発見を誘導させる取り組みの重要性はますます高まってきている。このような背景のもと課題(2)の取り組みを通して、酵母菌のグルコース抑制に関する知識発見を促すとともに、仮説推論技術の新しい応用領域を開拓することにも貢献した。

#### 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 3 件)

① Y. Yamamoto, K. Inoue, K. Iwanuma. Heuristic Inverse Subsumption in Full-clausal Theories. 22<sup>nd</sup> International Conference on Inductive Logic Programming (ILP2012) Revised Selected Papers, Lecture Notes in Artificial Intelligence, Vol. 7842, pp.241-256, 2013. 6

② Y. Yamamoto, K. Inoue, K. Iwanuma. Inverse Subsumption for Complete Explanatory Induction, Journal of Machine Learning, Vol.86, pp.115-139, 2011.6

③ Y. Yamamoto, K. Inoue, K. Iwanuma. Comparison of Upward and Downward Generalizations in CF-induction. 21<sup>st</sup>

International Conference on Inductive Logic Programming (ILP2011) Revised Selected Papers, Lecture Notes in Artificial Intelligence, Vol. 7207, pp.373-388, 2012. 3

[学会発表] (計 7 件)

① Y. Yamamoto, K. Iwanuma, K. Inoue. Late Breaking Papers of the 22<sup>nd</sup> International Conference on Inductive Logic Programming (ILP2012), CEUR, vol. 975, pp. 74-79, 2012.12

② 坂本悠, 山本泰生, 岩沼宏治. 多値論理を用いた生体ネットワークシステムのモデル検査. 情報処理学会バイオ情報学第29回研究会, 2012. 6

③ 宮城智輝, 山本泰生, 岩沼宏治. 時系列信号処理に基づくSAT解法: Wave-SATソルバの実現に向けて. 人工知能学会全国大会 (第26回), 2012. 6

④ 坂本悠, 山本泰生, 岩沼宏治. 論理モデルによるグルコース抑制機構のパスウェイ補完. 人工知能学会全国大会 (第25回), 2011. 6

⑤ 山本泰生, 鍋島英知, 岩沼宏治. 一般双対化問題における冗長節生成の抑制法とその評価. 人工知能学会全国大会 (第25回), 2011. 6

⑥ Y. Yamamoto. Incremental Hypothesis Enumeration with Inverse Subsumption, Imperial College-NII Learning Workshop, 2010.12

⑦ Y. Yamamoto, K. Inoue, K. Iwanuma. From Inverse Entailment to Inverse Subsumption, 20<sup>th</sup> Int. Conf. on Inductive Logic Programming (ILP 2010), 2010.10 [最優秀学生論文賞受賞]

[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>

## 6. 研究組織

(1) 研究代表者

山本 泰生 (YAMAMOTO YOSHITAKA)

山梨大学・大学院医学工学総合研究部・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)

岡山大学・異分野融合先端コア・准教授

研究者番号: 60500808