3B1-2

不完全な知識に基づく帰納学習
Inductive Learning Based on Incomplete Knowledge

山本勝考　　栗原正仁
Katsutoshi Yamamoto　Masahito Kurihara

北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology, Hokkaido University

Inductive Logic Programming (ILP) is a research area formed at the intersection of Machine Learning and Logic Programming. Generation of regular rule in ILP needs generation of new intermediate predicates. In this paper, we propose how to generate them with extra learning and circumscription.

1. はじめに

帰納論理プログラミング (Inductive Logic Programming; ILP) は計算機上で帰納推論を行うための技法である．これは機械学習法としても知られ，背景知識を使用できることが特徴である．しかし，実用的時間で学習するためには使用する知識が完全であることを前提とする必要があり，事例が追加される場合には始めから学習し直さなければならない．また，背景知識が不完全な場合には，それに存在しない概念を使用する必要がある．このような概念を補助述語と呼び，正しい学習をする上で不可欠である．そこで，不完全な知識を使用する ILP について考え，これらの問題を解決する．

2. ILP と関連研究

ILP とは一階述語論理において事例集合 E を説明するための仮説集合 H を，背景知識 B と E とから帰納するものである．

新たな事例 e を追加したときの事例集合を E' とし，ILP によって B と E' とから得られる仮説集合を H' とする．このとき，一般には $H \neq H'$ であるが，既に存在する H を無駄にすべきではないので，H を修正して H' を得たい．そこで，底節交差汎化法 [伊藤 99] と極小限定を応用し，補助述語を適宜加えながら仮説集合を修正する手法を提案する．

2.1 底節交差汎化法

ILP の解集合を仮説空間と呼び，包摂という節間の順序関係に基づく束構造である．この仮説空間の最大元を特に底節と呼ぶ．

異なる二つの事例の底節をそれぞれ求め，それらの最小汎化をした節を一般化することにより，二つの事例を同時に背景知識の下で説明する仮説を求める手法を底節交差汎化法という．

リテラルの最小汎化は，同じ述語名で同符号の複数のリテラルについて，引数を比べたときに対応する項が異なる部分を共通変数に置き換えることである．節の最小汎化は，各リテラルの組に対して最小汎化を行うことである．

2.2 極小限定

極小限定 (Circumscription) とは，記述していないことは成り立たないという仮説を利用した推論方式である．具体的には，論理式の集合 A に出現する述語 p が成り立つ条件を A に記述しているもののみに限定する．

連絡先: 山本勝考，北海道大学大学院情報科学研究科コンピュータサイエンス専攻，札幌市北区北 14 条西 9 丁目北海道大学大学院情報科学研究科，011-706-6861，t010296a@main.ist.hokudai.ac.jp

3. 仮説修正

提案手法のアルゴリズムを以下に示す．E, B, H は節集合であり，それぞれ事例集合，背景知識，仮説集合を表す．ただし，事例が負事例の場合は失敗による否定 (Negation as Failure) を適用する．

step1 新たな事例 e を E に加える．

step2 B を用いて e をゴールとする反駁導出を行い，生成した底節の集合を底節集合 G とする．

step3 G の元と H の元の全ての組み合わせに対して修正を行い，G の全ての仮説を一般化して H に要素として加える．

以下では，仮説の修正方法について詳しく説明する．

3.1 底節交差汎化法の応用

新たな事例による底節 g と過去に生成した仮説 h について底節交差汎化法を行う．ただし，g の結論部と h の結論部の符号が逆の場合は，符号が同じであるとみなして最小汎化を行う．

g は新たな事例による ILP の仮説空間の最大元であり，h は過去に追加した事例による ILP の仮説空間の最大元である．g を最大元とする仮説空間と h を最大元とする仮説空間の共通部分は束構造であり，g と h の最小汎化で得られる節 q はこの束の最大元である．g の結論部と h の結論部の符号が同じ場合は，この束は過去の事例と新たな事例を併せた事例集合による ILP の仮説空間である．

例

節 g, h をそれぞれ

$$g : P \leftarrow R \wedge S \wedge T$$
$$h : P \leftarrow R \wedge S \wedge U$$

とすると，これらの最小汎化節 q は，

$$q : P \leftarrow R \wedge S$$

である．

g の結論部と h の結論部の符号が逆の場合は，お互いの仮説空間の共通する部分が矛盾を起こす．どちらか一方の仮説空間から共通部分を削除することで，この矛盾を避ける．矛盾を避けた仮説空間の和集合が，過去の事例と新たな事例を併せた事例による ILP の仮説空間である．以下では，極小限定を用いて仮説空間を分離することで，矛盾を避けた仮説空間を生成する．

3.2 極小限定の応用

例外となる条件を全てまとめて一つの述語で表し，極小限定を用いると，例外となる条件を一つも含まない仮説空間と，少なくとも一つは含んでいる仮説空間とに分かれる．このことにより，一つの束構造を正の仮説空間と負の仮説空間の二つの束構造に分割することができる．この例外条件をまとめた述語が補助述語に相当する．

g と h との最小汎化節を q とし，q の例外を表す述語を ab_q とする．g の結論部と h の結論部の符号は逆であり，q の結論部の符号が g の結論部と同じ場合の節を q_g，h の結論部と同じ場合の節を q_h とする．すると，q_g，q_h の例外を表す述語はそれぞれ ab_q_g，ab_q_h である．

例

節 g, h をそれぞれ

$$g : \neg P \leftarrow R \wedge S \wedge T$$
$$h : P \leftarrow R \wedge S$$

とすると，これらの最小汎化節 q_h は,

$$q_h : P \leftarrow R \wedge S$$

となる．節 g と h の符号の違いは T の有無に関係するので，T は h の例外であり，$ab\_q_h \leftarrow T$ という節を仮説に加えることができる．これに伴い，h を

$$h' : P \leftarrow R \wedge S \wedge \neg ab\_q_h$$

という節に書き換え，例外を除く．これが，例外条件を一つも含まない正の仮説空間である．また，g を

$$g' : \neg P \leftarrow R \wedge S \wedge ab\_q_h$$

という節に書き換える．例外条件 ab_q_h を必ず含むという条件を加えると，これは，例外条件を少なくとも一つは含む負の仮説空間である．

4. 提案手法の応用

本研究における提案手法は，事例集合が不完全な場合の ILP における問題を解決するためのものであった．しかし，背景知識が不完全な場合には，補助述語の特徴付けを十分に行えない可能性がある．そこで，背景知識が不完全な場合に他の知識集合から知識を補うための手法として提案手法を用いることを試みる．

4.1 実装

提案手法によって生成される補助述語を学習するために，再度 ILP を行う．しかし，補助述語が生成されるかどうかは予め分からないので，二回目以降の ILP の事例集合は必然的に不完全である．つまり，前章の提案手法を用いてこの問題を解決できる．そこで，事例集合 E，背景知識 B，仮説集合 H，それ以外の知識の集合 B' を用意し，ILP を以下のように設定する．

step1 提案手法の手順を踏む.

step2 極小限定によって補助述語が生成されたら，それに関する知識を集めて節集合 E' とし，そうでなければ終了する.

step3 E' を事例集合，B' を背景知識とする ILP を，step1 から行う.

4.2 例題

例として，うるう年の法則を導く．以下，$leap(X)$ は「西暦 X 年はうるう年である」ことを意味し，$multX(Y)$ は「Y は X の倍数である」ことを意味する．

事例集合と背景知識をそれぞれ

$$E = \{leap(4), \neg leap(100)\}$$
$$B = \{mult4(4), mult4(100)\}$$

とする．事例 $leap(4)$ を入力したときの仮説は,

$$H = \{leap(X) \leftarrow mult4(X)\}$$

である．続けて事例 $\neg leap(100)$ を入力すると，提案手法によって,

$$H = \{leap(X) \leftarrow mult4(X) \wedge \neg new(X),$$
$$\neg leap(X) \leftarrow mult4(X) \wedge new(X),$$
$$new(100)\}$$

という仮説集合を得る．ここで，述語 new は補助述語である．背景知識 B 以外の知識集合を

$$B' = \{mult100(100)\}$$

とし，これを背景知識とした ILP に事例集合 $new(100)$ を入力すると，新しい仮説として

$$new(X) \leftarrow mult100(X)$$

が得られる．この仮説と H とを併せると，うるう年の仮説として「4 の倍数の年はうるう年である．ただし，100 の倍数の年を除く」という法則が得られる．以下同様に事例 $leap(400)$ を入力すると，正しいうるう年の法則を導くことができる．

4.3 考察

提案手法では，一度に全ての事例を入力する必要が無いので，事例が無限に存在する場合にも学習可能である．また，ILP が間違った仮説を出力したときに，それを修正するための反例をユーザが考えて入力できるので，無駄な事例を入力せずに済む．さらに，分散して存在する背景知識において ILP を行っているので，メモリに収めきれないほど膨大な背景知識を用いて ILP を行う場合には，データベースを分割して使用することができる．つまり，メモリ上の背景知識だけでは仮説が不十分であるときのみ，分散された別の背景知識を使用するので，ハードディスクへのアクセスを減らすことができることを期待できる．

5. おわりに

本研究では事例集合が不完全な場合の ILP のために，底節交差汎化法と極小限定を応用した手法を提案した．また，提案手法で生成した補助述語を特徴付けるために，その手法が使えることを示した．

今後の課題として，不完全な知識を提案手法によって補うことができることを，論理的な定理を用いて証明することが挙げられる．

参考文献

[伊藤 99] 伊藤公人，山本章博: 底節の最小汎化に基づく仮説の発見手法，人工知能学会誌，Vol.14，No.4，pp.709-716 (1999).

[斎藤 06] 斎藤悠，井上克巳: 極小限定を用いた帰納推論，人工知能学会論文誌，Vol.21，No.2，pp.143-152 (2006).