

Belief Network を用いた語義判別規則の学習

○ 阿部修也
茨城大学大学院理工学研究科
システム工学専攻

新納浩幸
茨城大学工学部システム工学科

1 はじめに

本論文では Belief Network を用いた語義判別規則の学習を試みる。

自然言語処理の個々の問題を分類問題として定式化し、帰納学習の手法により解決するというアプローチは大きな成功をおさめている。そこで利用される帰納学習手法は主に確率統計的な手法である。特に Naive Bayes は単純なモデルであるにもかかわらず、比較的よい規則を学習できるために広く利用されている [4]。Naive Bayes は変数間の独立性を仮定しているために計算が容易であるという長所をもつが、その反面変数間の関係を記述することができず、ある種の情報を落しているとも考えられる。一方、Belief Network は確率推論技術の一種であり、利用するモデルを柔軟に設定できることから、近年、活発に研究が行われている [3]。そこで本論文では、多義語の曖昧性解消問題を題材に、Naive Bayes 型のモデルに変数間の関係を追加したモデルを設定し、Belief Network による学習を試みる。

また Belief Network は学習の訓練データの中に背景知識を容易に組み込めるという特徴も有している。これは Naive Bayes などの確率統計的な手法にはない大きな長所である。本論文では背景知識を導入した場合の実験も行う。

実験では SENSEVAL2 の日本語辞書タスクを用いる。Naive Bayes との正解率の比較により本手法を評価する。

2 Belief Network による語義判別

2.1 Naive Bayes による語義判別

ある事例 x が素性のリストとして、以下のように表現されたとする。

$$x = (f_1, f_2, \dots, f_n) \quad (1)$$

x の分類先のクラスの集合を $C = \{c_1, c_2, \dots, c_m\}$ と置く。分類問題は $P(c|x)$ の分布を推定することで解決できる。実際に、 x のクラス c_x は以下の式で求

まる。

$$c_x = \arg \max_{c \in C} P(c|x) \quad (2)$$

ベイズの定理を用いると、

$$P(c|x) = \frac{P(c)P(x|c)}{P(x)}$$

なので、結局、以下が成立する。

$$c_x = \arg \max_{c \in C} P(c)P(x|c)$$

ここで、 $P(c)$ は比較的簡単に推定できる。問題は、 $P(x|c)$ の推定だが、これは現実的に難しい。Naive Bayes 型のモデルは、素性間の独立性を仮定して、以下の変形により $P(x|c)$ の推定を行う。

$$P(x|c) = \prod_{i=1}^n P(f_i|c) \quad (3)$$

2.2 Belief Network への拡張

Naive Bayes は素性間の独立性を仮定している。確率モデルの観点でみると素性は確率変数に対応しており、確率変数間の独立性を仮定していると見なせる。しかし現実には確率変数間には何らかの依存関係がある場合が多く、より自然なモデルを扱うためには Naive Bayes は適していない。そこでここでは素性間の依存関係を考慮したモデルを扱える Belief Network を用いる。

Belief Network が扱うモデルは以下の条件を満たす有向非循環グラフ (Directed Acyclic Graph; DAG) として表される [5]。

- 確率変数の集合がネットワークのノードを形成する。
- リンクまたは矢印の集合がノード対を結び、ノード間の因果的確率関係を表す。
- 各ノードは親ノードがそのノードへ及ぼす影響を定量化した条件付き確率表を持つ。
- グラフは矢印の方向にサイクルを持たない。

図1は天候から傘が必要かを判断する例である。雨ならば傘が必要で、晴天ならば傘は必要ない。しかし、目的地までの距離が近い場合や地下鉄で移動できる場合は雨が降っていても傘は必要ない。紫外線に気を配る人は晴天でも傘が必要になる。曇りの日は傘が必要かどうかわからない。図1はこのような天候と傘に関する条件付き確率の関係を表すモデルとして捉えられる。そして天候の確率は表1、ある天候のとき傘が必要な条件付き確率は表2の形にまとめられる。

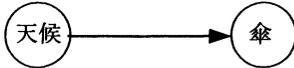


図1: 傘が必要かどうかのモデル

表1: 天候の確率

天候	$P(\text{天候})$
晴天	0.3
曇り	0.5
雨天	0.2

表2: ある天候の元で、傘が必要な確率

天候	傘が必要か?	$P(\text{傘} \text{天候})$
晴天	必要	0.1
	必要なし	0.8
曇り	必要	0.5
	必要なし	0.5
雨天	必要	0.9
	必要なし	0.1

しかし、現実はまだ少し複雑であり、季節や天気予報というファクターが天候や傘に影響を与えている。例えば、季節によって雨が多く降ることもある。また、夏に日傘をさす人は多いが、日差しが強くて冬はあまり日傘をささない。さらに、実際の天候によって天気予報は変化し、天気予報によって傘が必要かどうか判断する人は多い。このような因果関係を先のモデルで表した場合、図2のようになる。

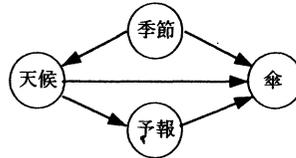


図2: 現実的な傘が必要かどうかのモデル

図2のモデルは図1を拡張し、より現実世界を反映しているが、図1と比べて条件付き確率を計算するのが困難になっている。Belief Networkは図2のようなモデルに対して、効率よく条件付き確率を計算できる学習手法である。

2.3 モデルへの素性の導入

語義判別の手がかりとなる属性として以下のものを設定した。

- e1 直前の単語
- e2 直後の単語
- e3 前方の内容語2つまで
- e4 後方内容語2つまで

例えば、語義判別対象の単語を「出す」として、以下の事例を考える（形態素解析され各単語は原型に戻されているとする）。

短い/コメント/を/出す/に/とどまる/た/。

この場合、「出す」の直前、直後の単語は「を」と「に」なので、「e1=を」、「e2=に」となる。次に、「出す」の前方の内容語は「短い」と「コメント」なので、「e3=短い」、「e3=コメント」の2つが作られる。またここでは句読点も内容語に設定しているので、「出す」の後方内容語は「とどまる」「。」となり、「e4=とどまる」、「e4=。」が作られる。

結果として、上記の例文に対しては以下の6つの素性が得られる。

- e1=を, e2=に, e3=短い, e3=コメント,
- e4=とどまる, e4=。

図3は、素性間の独立性を仮定したNaive Bayesモデルである。直前の単語(e1)、直後の単語(e2)、前方の単語(e3)、後方の単語(e4)それぞれの間にリンクがなく、単語間の関係は考慮されていない。

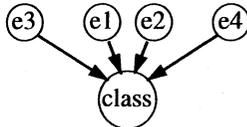


図 3: Naive Bayes 型モデル

図 4 は、前方の単語 (e3) から直前の単語 (e1) への影響、後方の単語 (e4) から直後の単語 (e2) への影響を追加したモデルである。

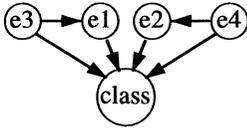


図 4: Belief Network 型モデル

図 3 では素性同士の関係が独立であるが、前方の単語と直前の単語がなんの関係も持たないとは考え難い。そこで、前方の単語から直前の単語への関係を追加したのが図 4 である (同様に後方の単語から直後の単語への関係も追加した)。

本論文では、背景知識としてラベルなしデータを用いる。ラベルなしデータとはクラスが付与されていない事例の集合である。一般に、ラベル付きデータを作成するのはコストが高く、十分な量を用意することができない。一方、背景知識として用いるラベルなしデータは、ラベル付けする必要がないことから簡単に作成することができる。そのため、ラベルなしデータを用いて素性間の関係を学習することで、高い正解率が期待できる。本論文ではラベルなしデータから、前方の単語から直前の単語への関係 (e3 → e1) と、後方の単語から直後の単語への関係 (e4 → e2) を学習する。

3 実験

本手法の有効性を確認するために、SENSEVAL2[2] の日本語辞書タスクの課題に対して語義判別を試みた。

SENSEVAL2 の日本語辞書タスクは、単純な語義判別問題である。対象単語は名詞 50 単語、動詞 50 単語の計 100 単語である。ラベル付きの訓練データは 1 単語平均して名詞は 177.4 事例、動詞は 172.7 事例用意されている。またテストデータは各単語に

対して 100 問のテスト事例が用意されている。つまり、名詞に対しては計 5000 問、動詞に対しても計 5000 問のテストが行える。ただし、

ラベルなし訓練データは SENSEVAL2 から提供されていない。RWC テキストデータベース第 2 版に納められた毎日新聞 95 年度版の 1 年分の記事を利用して、ラベルなし訓練データを収集した。収集できたラベルなし訓練データの数は 1 単語平均して名詞は 7585.5 事例、動詞は 6571.9 事例である。

実験結果を表 3 に示す。表中の NB は Naive Bayes 型モデルを、BN は Belief Network 型モデルを、BN+ は Belief Network 型モデルに背景知識を用いたものを表している。

表 3: 実験結果

	NB	BN	BN+
名詞	75.85 %	76.00 %	75.74 %
動詞	76.77 %	76.85 %	76.77 %

4 考察

Naive Bayes 型モデルの正解率は、名詞 75.85%、動詞 76.77%。Belief Network 型モデルの正解率は、名詞 76.00%、動詞 76.85%。名詞、動詞共に Naive Bayes 型モデルを適用させた場合に比べ Belief Network 型モデルを適用させた方が正解率が高かった。Naive Bayes 型モデルに、素性同士の関係を追加することでより適切なモデルを構築することができたと考えられる。

しかし Belief Network 型モデルに背景知識を適用させたときの正解率は、名詞 75.74%、動詞 76.77% であった。背景知識を利用した場合、動詞は Naive Bayes 型モデルと同じ正解率にまで下り、名詞は Naive Bayes 型モデルよりも正解率が下がった。背景知識を利用した場合は期待した効果が得られなかった。

背景知識が上手く利用されなかった原因を考察する。

未知の値を証拠として与えられたときの条件付き確率値の値、つまり事後確率を適切に設定することができなかつたため、背景知識を上手く利用することができなかつたと考えられる。

本実験では Belief Network を扱う手法として、接合木 (Junction Tree, Join Tree) を用いた [1]。本実験で用いたモデルは、変数から変数への経路が複数あるため通常的手法で計算することができない。そ

ここで、Belief Network から接合木という木構造を構築して、その木構造を利用して計算する手法を用いた。

接合木は3つの変数一つのクラスタ (cluster, clique) として扱い、クラスタの集合が木構造を成している。計算の際は、クラスタ単位で条件付き確率から、証拠をもとに事後確率を求める。その後、メッセージ伝播により木構造全体に事後確率を伝える。証拠とは変数に与えられる値のことで、変数から変数への条件付き確率に証拠が与えられると、変数から変数への事後確率が決定する。

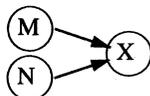


図5: $M \rightarrow X$ と $N \rightarrow X$ で構成されるクラスタ

接合木ではクラスタ単位で事後確率を求めるが、図5のクラスタにおいて変数 X に証拠 E_X , M に証拠 E_M , N に証拠 E_N が与えられて全ての証拠が満たされたとき、 $M(E_M) \rightarrow X(E_X)$ の事後確率が決定するが、 $N(E_N) \rightarrow X(E_X)$ の事後確率がわからない場合がある¹。この場合、クラスタの事後確率は0になり、 $M(E_M) \rightarrow X(E_X)$ の条件付き確率は無視されたことになる²。これは、このクラスタにおいて証拠 E_X , E_M , E_N が与えられなかった場合と等しい結果である³。これは証拠を与えられずに結果を導き出すことと等しい。

この現象は背景知識が与えられたとき頻繁に起きている。Naive Bayes 型モデルの場合は、名詞91回、動詞138回。Belief Network 型モデルの場合は、名詞107回、動詞147回。Belief Network 型モデルに背景知識を用いた場合は、名詞166回、動詞174回。Belief Network 型モデルと背景知識を用いた場合を比較すると、名詞は発生回数が59回増えて、動詞は発生回数が27回増えている。

背景知識を用いたときにこの問題が頻発する理

¹ 本実験のアルゴリズムは基本的に文献 [1] の通りであるが、本実験と異なり、文献 [1] では変数の中身を True または False であるとしているため、本実験で問題となった現象が文献 [1] では起らない。

² $N(E_N) \rightarrow X(E_X)$ の証拠が満たされない (N の証拠が満たされない) ように振舞う方法も試みたが本実験の結果よりも更に悪い結果であった。

³ 本実験では、条件付き確率が0になったクラスタの条件付き確率が他のクラスタに伝播しないように対策を行っている。

由は次の通りである。背景知識を用いると変数が知っている値を増やすことができる。そのため、変数が証拠を知っている回数は増え、図5における変数 M, N, X が証拠が満たす回数が増える。しかし、 $M(E_M) \rightarrow X(E_X)$ の事後確率が存在し、 $N(E_N) \rightarrow X(E_X)$ の事後確率が存在する回数はあまり増加していない。このため、前述の現象が発生する回数が増え、結果として証拠が生かされない回数が増える。

5 おわりに

本論文では Belief Network を用いた語義判別規則の学習を試みた。

実験では語義判別問題として SENSEVAL2 の日本語辞書タスクを用いた。Naive Bayes 型モデルを適用させた場合に比べ Belief Network 型モデルを適用させた場合の方が正解率が高かった。しかし、Belief Network 型モデルに背景知識を利用したところ、動詞は Naive Bayes 型モデルと同じ正解率にまで下り、名詞は Naive Bayes 型モデルよりも正解率が下った。背景知識を利用した場合は期待した効果が得られなかった。

背景知識が上手く利用されなかった原因として、未知の値を証拠として与えられたとき、事後確率を適切に設定することができなくなったためだと思われる。一般には未知の値を証拠として与えられた場合、事後確率に小さな確率を設定する。しかし、適切な小さな確率を決定することは難しく、本実験ではこの小さな確率を0としている。この問題を含めて背景知識を有効に活用することが今後の課題である。

参考文献

- [1] Cecil Huang, Adnan Darwiche.: Inference in Belief Networks: A Procedural Guide, *International Journal of Approximate Reasoning*, 11:1-158 (1994).
- [2] 黒橋禎夫, 白井清昭: SENSEVAL-2 日本語タスク, 信学技報, Vol101, No.351, pp1-8 (2001).
- [3] 佐藤泰介, 櫻井彰人: 特集「ベイジアンネット」, 人工知能学会誌, 17 巻, 5 号, pp538-565 (2002).
- [4] 新納浩幸, 佐々木稔: EM アルゴリズムの最適ループ回数の予測を用いた語義判別規則の教師なし学習, 情報処理学会自然言語処理研究会, 151-8, pp.51-58 (2002).
- [5] 古川康一監訳: エージェントアプローチ人工知能, 共立出版 (1997).