2L1-4

Markov logic network を事前分布に持つ潜在変数モデルの検討

Latent variable model with prior distribution of Markov logic network

秋元 康佑　　武石 直也　　堀 浩一　　矢入 健久
Kosuke Akimoto　　Naoya Takeishi　　Koichi Hori　　Takehisa Yairi

東京大学大学院工学系研究科
School of Engineering, The University of Tokyo

Inference and learning of a Markov logic network require actual truth values of evidence ground atoms, and it makes it difficult to apply Markov logic networks to problems in which these truth values cannot be determined manually in advance. Even in such situations, it may be possible to infer truth values of evidence ground atoms using other (possibly continuous) observations in data. In the proposed model, these unkown truth values are treated as binary latent variables of observation models, and a Markov logic network is used as a prior distribution for these models. In this paper, the authors show sampling algorithm and learning scheme for this model, and validate these methods through a simple setting experiment.

1. 背景

近年の機械学習分野における研究成果により、大量のデータから学習を行うことにより高い精度で分類や予測を行うことが可能になりつつある。しかし実際にこれらの技術を応用する場合には、分類や予測の根拠を提示したり、現場で蓄えられてきた専門家の知識を導入できることが望ましい。

例えばタクシー会社が客の数を予測したいという場合、その日の天気や曜日などの情報を基に予測器を学習することが考えられる。しかし単に入力から予測結果をブラックボックスに出力するだけの予測器では、人間側の予測と予測器の結果が食い違った場合に説得力を持って予測した値を提案することはできない。さらにタクシー会社は、例えば「雨の日には客が多い」や「休日で外出日和の日には観光客が多い」、「冬で暖かい日は外出日和の日である」といったような予測に有用な知識を既に持っており、これらの知識を学習により「再発見」することに意味は無く、むしろこれらの知識を用いてより少ないデータで学習を行えるような場合の方が望ましい。

以上のようにデータだけでなく人間の持つ知識を用いて推論や学習を行う手法への需要は大きいが、人間の持つ知識には以下のような扱いづらい性質を持つものが多く、論理推論のみによるアプローチを困難にしている。

不正確さ

不正確な知識とは、常に成り立つとは限らない知識である。
上の例では、「雨の日には客が増える」という知識が不正確な知識であり、雨が降っているのに客が少ない日もあることに対応している。

漠然さ

漠然とした知識とは、その知識に含まれる概念を明示的な記述を用いて定義できないものである。
上の例では、「天気がいい」という概念は個人差があり、具体的な日照や降水量、気温などが満たすべき条件として厳密に定義することはできないため、「天気がいい日には観光客が多い」という知識は漠然とした知識である。

連絡先: 連絡先：akimoto@ailab.t.u-tokyo.ac.jp

このような問題を解決するために、本研究では以下を実現するようなモデルを提案することを目的とした。

- **人間の持つ不正確で漠然とした知識を基に、確率的な推論とデータからの学習を行う**
  確率的な推論を目標とするのは、例えば予測や分類、異常検知といった確率的な要素を必要とする応用に用いるためである。
- **漠然とした概念に対応するデータ空間上のパターンを抽出する**
  漠然とした概念に対するパターンの抽出を行う目的は、推論結果の根拠を示す際に、漠然とした概念に対応するパターンを提示することで根拠の透明性を増すことである。また従来人間の暗黙知として扱われていた概念をデータ上のパターンとして明示的に得る知識発見の目的もある。

2. 関連研究

2.1 確率的な論理推論の拡張

従来の論理推論に確率的な要素を加えることで不正確な知識ベースを用いた確率的な推論を行おうとする試みは多く行われており、例えば P-log [Baral 09] や ProbLog [De Raedt 07]、distributional clause [Gutmann 11] などがある。しかしこれらの手法は知識ベースに含まれる具体的な確率値まで含めて既知であることを仮定しており、学習にはそれほど大きな比重が置かれていない。

一方異なるアプローチとして、論理的な知識ベースを確率モデルに変換して推論を行うような手法も提案されており、Bayesian logic programming [Kersting 07] や Markov logic network [Richardson 06] などがある。これらの手法では知識ベースを確率的なモデルに変換してから推論を行うため、他の確率的なモデルと同様にモデルに含まれるパラメータの学習を行うことができる。

2.2 Markov logiv network

今回の研究では Markov logic network (MLN) [Richardson 06] に注目した。

MLN は実数値により重みを付けられた一階述語論理の論理式からなる知識ベースを、マルコフ確率場に変換して推論や学習を行うモデルである。

具体的には知識ベースに含まれる述語 (predicate) および論理式 (formula) に対し、ある定数の集合を与えて変数を定数で置き換える grounding と呼ばれる処理を行い、それぞれ ground atom と ground formula の集合を得る。

この ground atom それぞれに対し、真か偽かの真理値に対応して $\{0,1\}$ の二値を取る離散変数 X_i を用意する。また ground atom(またそれに対応する離散変数) の取る値の組み合わせ (possible world) $\boldsymbol{X} = (X_1, ..., X_i) \in \mathcal{X}$ により、ある ground formula GF_j が真となるか偽となるかが変化するが、これに対応する以下のような feature 関数 $f_j(\boldsymbol{X})$ を定義する。

$$f_j(\boldsymbol{X}) = \begin{cases} 1 & (\mathrm{GF}_i\text{が真の時}) \\ 0 & (\text{そうでないとき}) \end{cases}$$

この feature 関数と、ground formula GF_j に対応する論理式の重み w_j を用いて、ある possible world $\boldsymbol{X}$ に対する確率が (1) のように定義される。

$$\begin{aligned} p(\boldsymbol{X} = x) &= \frac{1}{Z} \exp\left(\sum_j w_j f_j(x)\right) \\ Z &= \sum_{x \in \mathcal{X}} \exp\left(\sum_j w_j f_j(x)\right) \end{aligned} \tag{1}$$

MLN を用いた推論手法としては、エルゴート性と詳細釣り合い条件を満たしたマルコフ連鎖からサンプルを得る MC-SAT と呼ばれるアルゴリズムが提案されており [Poon 06]、MCMC による推論を行うことができる。またデータに対する対数尤度を最大化することにより重み w_j の学習も行うことができ、厳密な知識ベースの構築が難しい場合であってもデータを用いて重みの調整を行うことができる。

2.3 Markov logic network の連続値変数への拡張

MLN を用いて学習や推論を行う場合、evidence として用いる ground atom の真理値については予め定められなければならない。しかし 1. 章で指摘したように、人間の知識には必ずしも真偽を明確に定められない概念が存在しており、常に evidence として ground atom の真理値を与えられるとは限らない。

一方そのような場合であっても、その ground atom に対して何らかの観測値（連続値でありうる）によって間接的に evidence を与えることが可能な場合がある。その場合 ground atom を潜在変数と捉えて何らかの観測モデルを用意し、連続値の観測変数を与えて推論・学習を行うことが考えられる。しかし MLN は連続値変数を扱うことができないため、このような問題に対応することができない。

このような背景から MLN において連続値変数を扱えるように拡張した手法として Hybrid Markov logic network (Hybrid MLN) が提案されている [Wang 08]。

MLN と異なり、Hybrid MLN においては ground atom に対応する二値変数だけでなく連続値を取る確率変数が定義される。さらにこれらの変数に対し、MLN のような二値の feature 関数でなく任意の実数値を取れる feature 関数を定義することで、任意の指数型分布族を含むマルコフ確率場を定義することができる。

また MLN とは異なるアプローチであるが、連続値変数と知識ベースの両方を用いている手法として論理的制約を用いた RegBayes がある [Mei 14]。

論理的制約を用いた RegBayes においては潜在変数 $\boldsymbol{Z}$ を持つ確率モデルに対して、事後分布 $p(\boldsymbol{Z}|\boldsymbol{X})$ にできるだけ近く、かつ一階述語論理により記述された制約 $\phi_l(\boldsymbol{Z}, \boldsymbol{X})$ が満たされる期待値が指定した割合 γ_l に近いような分布 $q(\boldsymbol{Z})$ を最適化問題 (2) を解くことにより求められる。

$$\begin{aligned} \min_{q(\boldsymbol{Z}) \in \mathcal{P}, \boldsymbol{\xi} \in \mathbb{R}_+^L} & KL(q(\boldsymbol{Z}) || p(\boldsymbol{Z}|\boldsymbol{X})) + C \sum_l \xi_l \\ \text{s.t.} \quad & |\mathbb{E}_{q(\boldsymbol{Z})}[\phi_l(\boldsymbol{Z}, \boldsymbol{X}) - \gamma_l]| \le \epsilon + \xi_l \end{aligned} \tag{2}$$

3. 検討モデル

3.1 方針

前章で紹介したように、論理式により表現された知識から連続値を含むような問題に対して確率的な推論・学習を行う手法が既にいくつか提案されている。

しかし RegBayes においては MLN における MC-SAT のような効率的な推論手法が提案されておらず、知識ベースを基にした推論を行いたい今回の研究目的を満たしていない。

また今回は漠然とした概念に対するデータ空間上のパターンを明示的に抽出したいため、観測できない ground atom から観測されるデータへの観測モデルは Hybrid MLN のように無向グラフではなく、ground atom からデータへの有向グラフの形でモデル化することとした。

3.2 定式化

前節で述べたような方針から、今回は知識ベースを MLN の形で表現し、観測できない ground atom から観測変数に対して有向グラフの観測モデルをもつようなモデルを検討した。

検討モデルを一般的な形で表現すると (3) のようになる。ここで p_{MLN} は (4) で定義されるような MLN であり、$\tilde{p}$ は任意の観測モデルである。

$$p(\boldsymbol{X}, \boldsymbol{Z}) = p_{\mathrm{MLN}}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{z}_1, \boldsymbol{z}_2; \boldsymbol{w}) \tilde{p}(\boldsymbol{x}_3, \boldsymbol{z}_3 | \boldsymbol{x}_2, \boldsymbol{z}_2; \boldsymbol{\theta}) \tag{3}$$

$$p_{\mathrm{MLN}}(\boldsymbol{X}, \boldsymbol{Z}) = \frac{1}{Z} \exp\left(\sum_{i \in \mathrm{GF}} w_i f_i(\boldsymbol{X}, \boldsymbol{Z})\right) \tag{4}$$

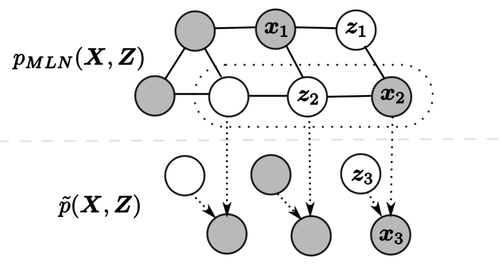

図 1: 検討モデルのグラフィカルモデル

図 1 はこれをグラフィカルモデルで表現したものである。$\boldsymbol{x}_1, \boldsymbol{z}_1$ は MLN に含まれる ground atom に対応する変数のう

ち、観測モデルが存在しないものである。一方 $\boldsymbol{x}_2, \boldsymbol{z}_2$ は同様に MLN に含まれる ground atom のうち観測モデルが存在するものであり、$\boldsymbol{x}_3, \boldsymbol{z}_3$ は観測される変数や観測モデルに含まれる潜在変数である。

前のタクシー会社の例では、$\boldsymbol{x}_1, \boldsymbol{z}_1$ は「今日は休日である」や「今日は外出日和の日である」といったような観測モデルが無い述語に対応する。また $\boldsymbol{x}_2, \boldsymbol{z}_2$ は観測モデルが作れるような「客が多い」や「暖かい」といったような述語に対応し、$\boldsymbol{x}_3, \boldsymbol{z}_3$ には実際の観測値である客の数や気温、または観測モデル中のベイズ的に扱うパラメータなどが対応する。

3.3 推論手法

一つのステップにおいて $\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3$ のそれぞれの潜在変数に対し、それぞれ異なる更新処理を行うマルコフ連鎖を用いて MCMC による推論を行う。

より具体的には、毎ステップ以下のような 3 つの処理をそれぞれ 0.5 の確率で $(1 \to 3) \to 2$ もしくは $2 \to (1 \to 3)$ のどちらかの手順を選び、実行することによってすべての潜在変数に対する更新を行う。(ここで括弧内 1 と 3 の処理については順序は自由に交換可能)

1. $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{z}_2$ を evidence として固定した MLN から $\boldsymbol{z}_1$ を MC-SAT により更新する
2. $\boldsymbol{z}_2$ に属する変数を Gibbs サンプリングによって更新
3. $p(\boldsymbol{x}_3, \boldsymbol{z}_3|\boldsymbol{x}_2, \boldsymbol{z}_2)$ について詳細釣り合い条件を満たすマルコフ連鎖の遷移確率によて $\boldsymbol{z}_3$ を更新する

このようなマルコフ連鎖は今回検討しているモデル (3) に対して詳細釣り合い条件を満たし、また $\boldsymbol{z}_2$ を含む hard constraint が存在しないことと、処理 3 において用いるマルコフ連鎖において任意の $\boldsymbol{z}_3$ に対する遷移確率が非ゼロであるという比較的緩い条件によりエルゴート性も満たすことを示すことができる [*1]。

3.4 学習法

今回検討しているモデル (3) における p_{MLN} の重み $\boldsymbol{w}$ および観測モデル $\tilde{p}$ のパラメータ $\boldsymbol{\theta}$ を学習する際には、E ステップを MCMC によって得られたサンプルによって近似した EM アルゴリズムを用いる [*2]。

E ステップで得られた L 個のサンプルを $\boldsymbol{Z}^{(l)}$ とすると、M ステップで最大化する関数は (5) のようになる。

ここで第一項については通常の MLN において対数尤度を最大化する処理に対応しており、既存の MLN に対する学習法を用いることができる。

また第二項については観測モデルの対数尤度を最大化する処理に対応している。観測モデルとしてガウス分布やポアソン分布といった単純なモデルを採用した場合には第二項の対数尤度の最大化は解析的に行うことができる。さらに観測モデルにニューラルネットなどを用いていて厳密な対数尤度の最大化が不可能な場合であっても、対数尤度の下限や近似を数値的に最大化することにより近似的に第二項の部分のパラメータ更新を行うことができる。

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{w}) &= \int p(\boldsymbol{Z}|\boldsymbol{X}; \boldsymbol{\theta}_{\mathrm{old}}, \boldsymbol{w}_{\mathrm{old}}) \log p(\boldsymbol{Z}, \boldsymbol{X}; \boldsymbol{\theta}, \boldsymbol{w}) d\boldsymbol{Z} \\
&\simeq \frac{1}{L} \sum_{l=1}^{L} \log p(\boldsymbol{Z}^{(l)}, \boldsymbol{X}; \boldsymbol{\theta}, \boldsymbol{w}) \\
&= \left(\frac{1}{L} \sum_{l=1}^{L} \log p_{\mathrm{MLN}}(\boldsymbol{X}, \boldsymbol{Z}^{(l)}; \boldsymbol{w}) \right) \\
&\qquad + \left(\frac{1}{L} \sum_{l=1}^{L} \log \tilde{p}(\boldsymbol{x}_3, \boldsymbol{z}_3^{(l)} | \boldsymbol{x}_2, \boldsymbol{z}_2^{(l)}; \boldsymbol{\theta}) \right)
\end{aligned} \tag{5}
$$

4. 実験

4.1 半人工データを用いた実験

まず最初に真理値を予め与えられない漠然とした概念を含む知識ベースを用いて、真理値ではなく観測値を evidence として与えることによる学習や推論が行えるかを実験により確かめる。

具体的には

- ある人物が沢山読書しているならば、その人物は合格する
- ある人物の親しい友人が沢山読書をしているならば、その人物も沢山読書をする

という知識ベースの重みを学習することを目的とした。今回の実験では学習時に合格したかの真理値のみ与え、沢山本を読んでいるかどうかの真理値、そして親しい友人かの真理値は直接与えず、それぞれ読書数とメールの送受信量の数値データのみを与えた。

実験の目的は真理値を直接与えずに数値の観測量のみ与えた場合であっても、真理値を直接与えた場合と同様の学習結果を得られるか確認することである。

実験に用いたデータは実データを基に必要なデータを生成した半人工的なものを用いた。具体的には友人関係については SNAP データセット [Jure 14] の facebook における友人関係のデータを用いた。そしてそれぞれの人物について「本を沢山読んでいる」かどうかと「合格した」かどうかについては適当に設定した [*3](図 2)。さらに読書数とメールのやり取りの量については、それぞれ「本を沢山読んでいる」かどうか、「友人関係にある」かどうかで異なる平均を持つポアソン分布に従って生成した (読書量については平均 10.0 と 30.0、メール量については平均 10.0 と 100.0)。

モデルの MLN には表 1 の知識ベースを与え、重みをすべて 0.0 で初期化した [*4]。また観測モデルについてはデータに含まれる N 人の人物に対して読書量 b_i とメールのやり取りの量 m_{ij} に対し (6),(7) のように混合ポアソン分布によりモデル化した。

$$
\tilde{p}(\boldsymbol{b}, \boldsymbol{m}|\boldsymbol{z}) = \prod_{i=1}^{N} p_{\mathrm{poi}}(b_n | z_{\mathtt{Reader}(i)}) \prod_{j=1}^{N} p_{\mathrm{poi}}(m_{ij} | z_{\mathtt{Friends}(i,j)}) \tag{6}
$$

*1 $\boldsymbol{z}_1, \boldsymbol{z}_2$ の双方に対し、$\tilde{p}$ の影響を適切に反映した MLN から MC-SAT を用いてサンプルすることで $\boldsymbol{z}_2$ を含む hard constraint の存在も許容できるようになるが、今回は実装の容易さから Gibbs サンプルを用いるアルゴリズムを用いた

*2 実際には後述する MLN の重みなどに関して厳密な最大化は不可能なので GEM アルゴリズムである

*3 具体的にはランダムに数人の読書習慣と合不合格を指定した後、上記知識ベースと MLN を用いて MC-SAT により残りの人物の読書習慣と合不合格をサンプルした

*4 MLN についてこの実験では述語 `Friends` を evidence とするような discriminative な定式化を用いた。これは `Friends` に対応する ground atom に無情報事前分布 $p(\boldsymbol{z}_{\mathtt{Friends}})$ を与え、モデル化の時点で $p_{\mathrm{MLN}}(\boldsymbol{X}, \boldsymbol{Z})$ を $p(\boldsymbol{z}_{\mathtt{Friends}}) p_{\mathrm{MLN}}(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{z}_{\mathtt{Friends}})$ に置き換えたことに相当する

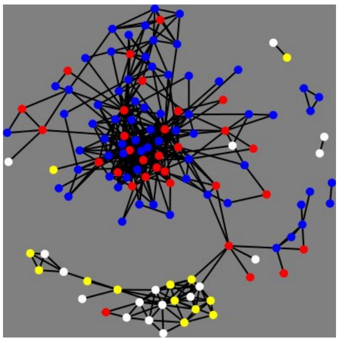

図 2: 友人関係のネットワークおよび読書習慣と合格状況
各エッジが友人関係に相当し、各ノードの色は青（読書多・合格）、赤（読書多・不合格）、黄（読書少・合格）、白（読書少・不合格）に相当している

$$p_{\mathrm{poi}}(x|z) = \left(\frac{\lambda_0^x \exp(-\lambda_0)}{x!}\right)^{1-z} \left(\frac{\lambda_1^x \exp(-\lambda_1)}{x!}\right)^{z} \quad (7)$$

表 1: 実験で与えた知識ベース

$\texttt{Reader}(p) \Rightarrow \texttt{PassExam}(p)$
$\texttt{Friends}(x,y) \wedge \texttt{Reader}(x) \Rightarrow \texttt{Reader}(y)$
$\texttt{Friends}(x,y) \wedge \texttt{Reader}(y) \Rightarrow \texttt{Reader}(x)$

学習の M ステップにおいて MLN の部分の対数尤度最大化には Alchemy [Kok 09] を用いた。またポアソン分布の部分の対数尤度最大化は解析的に行った。

学習によって得られた読書量、メール量の観測モデルのポアソン分布のパラメータ、および各論理式の重みは表 2 のようになった。さらに比較実験として同じ訓練データに対して「沢山本を読んでいるか」と「親しい友人かどうか」について正しい真理値を与えて学習した結果は表 3 のようになった。

これらの結果から今回検討しているモデルによって、観測値により間接的に evidence を与えた場合でも真理値を直接与えた場合とほぼ同じ学習結果が得られることが確認できた。さらに得られたポアソン分布のパラメータはデータを生成した際のポアソン分布のパラメータとほぼ等しく、観測モデルも適切に学習できていることが確認できた。

表 2: 学習結果(二列目が学習結果の重みまたはパラメータ)

読書量	9.45, 29.31
メール量	10.07, 99.15
$\texttt{Reader}(p) \Rightarrow \texttt{PassExam}(p)$	0.003
$\texttt{Friends}(x,y) \wedge \texttt{Reader}(x) \Rightarrow \texttt{Reader}(y)$	0.75
$\texttt{Friends}(x,y) \wedge \texttt{Reader}(y) \Rightarrow \texttt{Reader}(x)$	0.75

4.2 実データを用いた実験

本稿執筆時点において、実データを用いたより現実的な設定での実験を検討中である。

5. 今後の研究

今回提案したモデルにより、不正確で漠然とした知識ベースに対して数値的な観測値を evidence として与えることで直接真理値を与えることなく知識ベースを用いた推論や学習を行えることが確認できた。

今後の研究方針としては予測や異常検知などの実問題への応用を検討することや、事前分布として扱う MLN 由来のマルコフ確率場の推論を高速化してより大規模な知識ベースからの推論を実現することが考えられる。

表 3: 真理値を与えた学習結果(二列目が学習結果の重み)

$\texttt{Reader}(p) \Rightarrow \texttt{PassExam}(p)$	0.03
$\texttt{Friends}(x,y) \wedge \texttt{Reader}(x) \Rightarrow \texttt{Reader}(y)$	0.70
$\texttt{Friends}(x,y) \wedge \texttt{Reader}(y) \Rightarrow \texttt{Reader}(x)$	0.70

参考文献

[Baral 09] Baral, Chitta, Michael Gelfond, and Nelson Rushton. ”Probabilistic reasoning with answer sets.” Theory and Practice of Logic Programming 9.01 (2009): 57-144.

[De Raedt 07] De Raedt, Luc, Angelika Kimmig, and Hannu Toivonen. ”ProbLog: A Probabilistic Prolog and Its Application in Link Discovery.” IJCAI. Vol. 7. 2007.

[Gutmann 11] Gutmann, Bernd, et al. ”The magic of logical inference in probabilistic programming.” Theory and Practice of Logic Programming 11.4-5 (2011): 663-680.

[Jure 14] Leskovec, Jure, and Andrej Krevl. ”SNAP Datasets: Stanford Large Network Dataset Collection.” Jun. 2014, `snap.stanford.edu/data`.

[Kersting 07] Kersting, Kristian, and Luc De Raedt. ”1 Bayesian Logic Programming: Theory and Tool.” Statistical Relational Learning (2007): 291.

[Kok 09] Kok, Stanley, et al. ”The alchemy system for statistical relational AI.” (2009).

[Poon 06] Poon, Hoifung, and Pedro Domingos. ”Sound and efficient inference with probabilistic and deterministic dependencies.” AAAI. Vol. 6. 2006.

[Richardson 06] Richardson, Matthew, and Pedro Domingos. ”Markov logic networks.” Machine learning 62.1-2 (2006): 107-136.

[Mei 14] Mei, Shike, Jun Zhu, and Jerry Zhu. ”Robust RegBayes: Selectively Incorporating First-Order Logic Domain Knowledge into Bayesian Models.” ICML. 2014.

[Wang 08] Wang, Jue, and Pedro M. Domingos. ”Hybrid Markov Logic Networks.” AAAI. Vol. 8. 2008.