ベイジアンネットワークに対する効率的な更新・問合せ手法

佐藤 亮 † 川島 英之 †† 北川 博之 ††
† 筑波大学情報学類 †† 筑波大学大学院システム情報工学研究科

1 はじめに

本研究の目指す所は，実世界状況を連続的に監視するアプリケーション (例：侵入者検知監視システム) を支援する基盤技術の開発にある．そのようなアプリケーションには，不確実な状況を扱う手法と，連続的にデータを処理する手法の2つが求められる．そこで本研究では，不確な実状況を取扱可能な一手法であるベイジアンネットワークを取り上げ，同手法に対するストリームデータ処理モデルおよび，その効率的な実現方法を提案する．我々の知る限り，本研究はベイジアンネットワークに時間的要素を導入した最初の研究である．

2 準備

2.1 ベイジアンネットワーク

不確実な状況を表現するためにベイジアンネットワークが広く用いられている．その例には状況推定[1]，麻酔行為表現，電子メールエージェント，遺伝子発現関係の抽出等がある．

ベイジアンネットワークは，数理的には確率変数をノードとするグラフ構造と，各ノードに割り当てられた条件付き確率分布群によってモデルが定義される．各変数の条件付き確率分布は，本研究で対象とする離散的な確率変数の場合には，親ノードと子ノードの各変数がとる具体的な値ごとに割り当てられた条件付き確率を離散的な条件付き確率表として表現される．

ベイジアンネットワークのアーキテクチャにおいては，確率値を読出す処理と確率値を更新する処理が必要である．本論文ではそれらを各々**reader**，**writer**と表記する．即ち，reader はベイジアンネットワークの読出を担い，その内容はユーザの指定条件を満たすノードに関する ID または確率値の取得である．そして，writer はベイジアンネットワークの更新を担い，その内容はイベントが生起したノードの確率値を1へ設定及び，他ノードへの確率伝播である．

2.2 ストリームデータ処理技術

ストリームデータ処理技術とは，システムに絶えまなく到着するストリームデータを連続的に処理する技術である．ベイジアンネットワークではイベントが生起したノードの情報がストリームデータとして writer に送られる．その情報をもとに確率伝播を行い各ノードでの確率値が計算される．その確率値を reader を通してユーザーに返す．

3 ベイジアンネットワークに対するストリームデータ処理モデルと問合せ

ベイジアンネットワークに対するストリームデータ処理モデルを構築するために，2.1節で定義したベイジアンネットワークのモデルに，次の拡張を加える．

master: master は時間幅のパラメタにより決定される．時間幅は計算を行いユーザーに送信するタイミングをを指定する．master の設定はユーザにより行なわれるものとする．

lifespan: ノード V で事象が発生した時，その発生の有効期間を lifespan により表す．lifespan の設定は本システムの管理者により与えられるものとする．

master は問合せに関して時間的要素を導入し，lifespan はイベントに関して時間的要素を導入する．

図1の問合せの例は,id が4未満のノードの情報を

```
1: MASTER 10sec
2: SELECT bn.getnodebyid(id<4).generate()
3: FROM table.bn
```

図 1: 問合せ例

１０秒間隔で要求している問合せである．ユーザーは問合せの頻度を master 節を使用し設定するが，このとき低頻度の要求 (1時間ごとに確率値が欲しい) を設定すると，ユーザーに結果を送信しない時もストリームデータが writer に送られるたびに計算をすることになる．それだと無駄な計算が増加するので，このシステムではユーザーに情報を送信する直前に確率伝播を実行し，計算している．しかし，ノード数や生起イベントが膨大になるとともに確率計算を実行する処理時間も増加するという欠点もある．その欠点を解決するために4節で問合せ処理の効率化を提案している．

Efficient Update and Query Methods on Bayesian Networks

† Ryo SATO(punisiro@kde.cs.tsukuba.ac.jp)
†† Hideyuki KAWASHIMA(kawasima@cs.tsukuba.ac.jp)
‡ Hiroyuki KITAGAWA(kitagawa@kde.cs.tsukuba.ac.jp)
College of Information Science, University of Tsukuba, University of Tsukuba (†)
Graduated School of Information and Systems Engineering, University of Tsukuba (††)
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

4 問合せ処理の効率化

問合せは次のように処理される．図 2 の単純法で

```
1: foreach event in window {
2:   writer は event_i の確率値を 1 に設定;
3:   writer は確率値伝播を実行;
4: }
5: foreach reader
6:   問合せを実行;
```

図 2: アルゴリズム 1: 単純法

は，reader は全ての event の更新処理が終了する迄，実行を待たなければならない．event 数が多い場合やノード数が多くなるほど writer の処理時間は長くなり，reader の待ち時間も長くなる．

そこで，reader の待ち時間を減らすアルゴリズムを提案する．これを図 3 に示す．図 3 の効率化法を使

```
1: foreach reader
2:   if (reader_i の対象は event のみ)
3:     reader_i を実行
4: foreach event in window {
5:   writer は event_i の確率値を 1 に設定;
6:   writer は確率値伝播を実行;
7: }
8: foreach reader
9:   if (reader_i の対象は event のみでない)
10:    問合せを実行;
```

図 3: アルゴリズム 2: 効率化法

えば,$reader_i$ が event のみを対象にしている場合には，writer の実行前にその処理を終えられる．すなわち，上記で問題としていた待ち時間を零にできる．

ただし，reader が event 以外を対象に含む場合には，待ち時間は従来アルゴリズムと同じになる．

提案手法を評価する為に，アルゴリズム 1 と 2 を Java 言語で実装した．PC(Athlon Dual-Core 2.00GHz CPU, 1982MB RAM, WindowsXP OS) 上で得た実験結果を図 4 に示す．図 4 の横軸はベイジアンネットワークのノード数を表し，縦軸は問合せ処理時間を表す．X%は，全ノード中の X%においてイベントが発生した割合を表す．問合せは図 1 の問合せを用い,lifespan は 1 秒に設定されている．また，各記録点は 10 回の実験結果の平均値である．

図 4 より，イベント発生率が 50%の場合には，提案手法は単純法よりも処理時間が最大で 19.8%短くなり，性能が向上することがわかる．一方，イベント発生率が 10%の場合には，提案手法は単純法よりも性能が若干劣る．この理由は図 3 の行 1〜3 が余計なオーバーヘッドになるからだと考えられる．

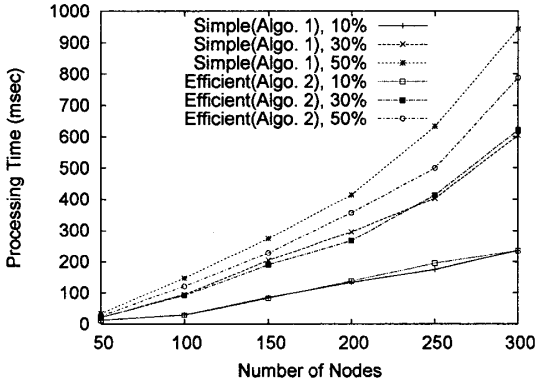

図 4: 効率化に関する実験結果

5 結論

実世界状況を連続的に監視するシステムを支援する基盤技術の開発を目指し，本研究では研究課題として，確率的表現能力と連続的問合せモデルを有する技術の開発を取り上げた．そこで本研究ではベイジアンネットワークに対するストリームデータ処理モデルを提案し，同モデル上における効率的な問合せ処理を提案した．提案モデルでは，ベイジアンネットワークに master と lifespan を導入したことで，問合せとイベントに時間的要素を表現可能にした．効率的な問合せ処理は，master で設定された時間内に生起しているイベントの数が多い場合には有効に働き，本研究における最大の場合には 19.8%の性能向上が観察された．以上より，本研究ではベイジアンネットワークに時間的要素を記述可能に，最大 19.8%の性能向上を得る効率化手法を提案できたと結論する．

謝辞

本研究の一部は科学研究費補助金基盤研究 (A)(# 18200005)，若手研究 (B)(# 18700096) による．

参考文献

[1] Masaya Kadota, Hiroto Aida, Jin Nakazawa, and Hideyuki Tokuda. D-jenga: A parallel distributed bayesian inference mechanism on wireless sensor nodes. In *Proc. of the International Conference on Networked Sensing Systems*, 2006.