

事故サーベイランスシステムからの知識獲得 - テキスト情報からの確率的因果構造のモデル化 - Knowledge Acquisition by Injury Surveillance System - Modeling of Probabilistic Causal Structure by Text Information -

三浦 未生^{*1} 本村 陽一^{*2} 柴田 康徳^{*3} 西田 佳史^{*2} 山本 哲也^{*1}
Miki MIURA Yoichi MOTOMURA Yasunori SHIBATA Yoshifumi NISHIDA Tetsuya YAMAMOTO

^{*1}東京都立産業技術高等専門学校
Tokyo Metropolitan College of Industrial Technology

^{*2}独立行政法人産業技術総合研究所
National Institute of Advanced Industrial and Technology

^{*3}日本電気株式会社
NEC Corporation

A method that acquires probabilistic causal information from text injury surveillance data is proposed. In order to convert text data to discrete random variables, an ontology and thesaurus about childhood accidents are introduced. We apply the proposed method to 2305 accident record, and construct Bayesian networks to evaluate probability of accident given particular behavior and situation.

1. 緒論

日本における子供の死亡原因の上位には、必ず不慮の事故がある。不慮の事故とは、“車のドアに指を挟んだ”や“ズボンの裾を踏んで転倒”など日常生活の中で起こる事故である。

しかし、これまで日本ではこのような事故に対しての予防策が確立されていなかった。そこで、筆者らは事故情報を知り、その事故の原因を推測することで子供の事故予防に役立つのではないかと考えた。この考えに基づいて、事故情報の収集及び事故の原因を推測できるように“事故サーベイランスシステム (ISS)”を開発し、実際に病院で運用することで事故情報が収集可能となった [Motomura 06]。ISS によって収集された事故データから事故時の状況を把握するための手法として、確率的因果構造モデルを構築できるベイジアンネットワークを用いている。現在、事故に関係したモノを用いて作成した確率的因果構造モデルから、事故の原因と考えられるモノを特定することが可能となっている [Shibata 07]。しかし、このモデルでは豊富な情報量を有する自由記述テキストが導入されていなかった。従って、自由記述テキストをベイジアンネットワークモデルに導入し、事故状況を把握することを目的とする。本稿では、次の手法を提案する (1) 自由記述テキストの構文解析及び品詞タグ付け (2) 行動項目のシソーラスを作成し (3) シソーラスを用いて構文解析結果から特徴的な情報を抽出、その情報を離散確率変数化し (4) 離散確率変数化した行動を従来モデルに導入し、確率的因果構造モデルを構築する。

2. 事故サーベイランスシステムに関する研究

“事故サーベイランスシステム (ISS)” は子供の事故予防のために、事故情報の収集及び事故情報を把握するために開発され、実際に病院で運用されている。ISS で収集される事故データの主な調査項目は次の通りである : (1) 受診年月日 (2) 性別 (3) 生年月日 (4) 年齢 (5) 発達段階 (6) 体重 (7) 身長 (8) 事故発生日 (9) 事故発生時間 (10) 事故の種類 (11) 怪我の種類 (12) 怪我の部位 (13) 事故に関連したモノ (14) 事故が起きた場所 (15) 事故直前の当事者の行動 (16) 一緒に

にいた人の事故直前の行動 (17) 事故の経過 (18) 事故の詳細。筆者らの研究では、これまでに ISS で収集された事故データを用いてモノデータベースやライフログを開発し、それらを用いて事故の分析を行ってきた。また、事故情報の知識化の手法としてベイジアンネットワークを利用し、事故データを用いて確率的因果構造モデルを構築し、確率推論により事故状況を把握してきた [Shibata 07]。

3. 事故直前の行動データの加工

本稿では、自由記述テキストの構文解析及び品詞タグ付けし、特徴的な情報を抽出し、離散確率変数化することを目的として、事故直前の行動をコード化した。

3.1 自由記述テキスト情報の離散確率変数化

自由記述テキストをベイジアンネットワークでモデル化するためには、データを離散確率変数として表す必要がある。そのために、構文解析及び品詞タグ付けと事故直前の行動のコード化を行った。使用した事故データは、成育医療センターから収集されたもので、総件数は 2305 件である。事故の種類として、転落、転倒、衝突、やけど、誤飲、はさむ、交通事故を使用した。

3.1.1 構文解析

最初に、事故直前の行動項目と事故の詳細項目それぞれについて構文解析を行い、品詞タグを付けた。この解析を行うために、Text Mining Studio [TMS] を使用した。解析結果は、行 ID、文章 ID、単語 ID、見出し語 (テキストの構文解析結果)、原形 (見出し語の原形)、置換語 (辞書によって変換された単語)、品詞などのよう出力される。このように構文解析することで、特定の品詞について表記ゆれを解消した単語の獲得が可能となった。しかし、このデータは離散データではないので、このままベイジアンネットワークでモデル化することはできない。そこで、今回は自由記述テキストである事故直前の行動に注目し、これを比較的ノードの少ない離散変数に変換するためのコード化を行った。

3.1.2 事故直前の行動のコード化

自由記述テキストが持つ表記ゆれや、ターミノロジーの問題を解決し、離散確率変数にするために事故直前の行動をコード

連絡先: 三浦未生, 都立産技高専, 東京都品川区東大井 1-10-40, 03-3471-6331, terrestrial.globe.of.sunca@gmail.com

化した。また、シソーラスにし、更に階層構造に組織化した。作成したシソーラスは次の通りである：(1) 全身動作 - 全身を動かす動作全般 (2) 頭部動作 - 頭部を動かす動作全般 (3) 上半身動作 - 上半身を主に動かす動作全般 (4) 下半身動作 - 下半身を主に動かす動作全般 (5) モノに関する動作 - モノに関しての動作全般 (6) 上下動作 - 上下で動くような動作全般、(7) 水平動作 - 水平に動く動作全般 (8) 静止動作 - 移動しないで動ける動作全般 (9) エネルギー大動作 - エネルギー量が比較的大きな動作全般 (10) エネルギー中動作 - 日常生活で消費する程度のエネルギー量を持つ動作全般 (11) エネルギー小動作 - エネルギー量が比較的小さな動作全般。

4. 確率的因果構造モデルの構築

4.1 事故データの確率的因果構造モデル

事故データから事故状況を把握するための手法として、確率的因果構造モデルを構築できるベイジアンネットワークを用いた。ベイジアンネットワークでは、確率変数をノード、因果関係をノード間の有向リンクで表す。このモデルを用いることにより (1) 依存関係の強い変数の間にリンクを張り (2) 確率推論によって、観測された事象から観測されていない変数の確率分布を予測・推定すること (3) “与えられた事実のもとで仮説が成り立つ可能性” を評価することが可能となる。モデル構築には BayoNet を用いた [Motomura 03]。

これまで行われてきた研究で確率的因果構造モデル構築の際に使用されていた項目は下記の通りである：(1) 事故発生時間帯 (2) 性別 (3) 年齢 (4) 事故に関係したモノ (5) 事故の種類 (6) 怪我の種類 (7) 怪我の部位。本稿では (6)(7) の代わりに事故直前の行動を導入し、確率的因果構造モデルを探索した。

4.2 オントロジーの導入

全てのノード間でリンクを張ると、その関係を理解することが困難になる。そこで、本稿では図 1 のような因果関係のオントロジーを導入する。これにより、モデルの構造探索速度を大幅に高速化し、ノード間の関係の理解を深める。

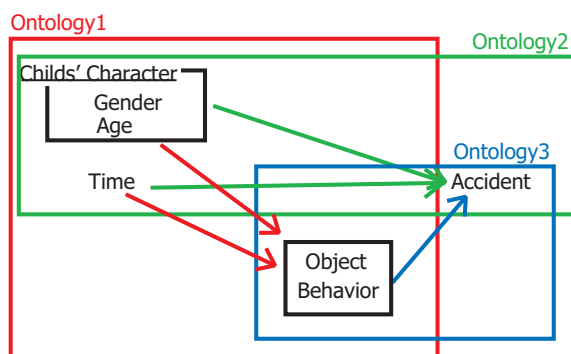


図 1: 因果関係のオントロジー

図 1 の関係に従って親ノードから子ノードへの有向リンクに制約条件をつけた。制約条件に基づき親ノードの仮説候補を制約した上で BayoNet により確率的因果構造の探索を行った。評価基準には AIC (赤池情報量基準) を使用した。

その結果、図 2 に示すようなベイジアンネットワークモデルが構築できた。この図から、年齢よりも事故の時間帯のほうが事故に関係するモノや事故直前の行動にリンクが張られてい

ることから、時間帯と事故に対して因果関係が強いことがわかる。また、ノード間の特徴的な関係は次の通りであった：(1) 1歳以下で乗れるモノ、静止動作の場合には転落事故の確率が高い (2) 熱いもの、下半身動作の場合にはやけど事故の確率が高い (3) はさみやすいモノであればさむ事故の確率が高い (4) 口に入るモノ、朝のそれぞれの場合で誤飲事故の確率が高い (5) 静止動作であればモノに関する動作の確率が高い、(6) 全身動作であればエネルギー大動作の確率が高い (7) 夕方乗れるモノであればモノの高さの確率が高い (8) 全身動作、エネルギー中動作の場合には上下動作の確率が高い。

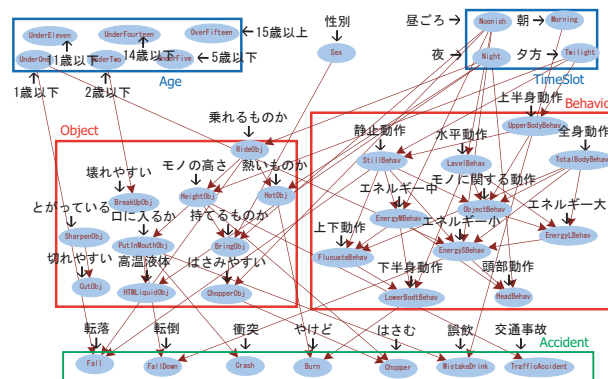


図 2: 確率的因果構造モデル

5. 結論

本稿では、自由記述テキストをオントロジーとシソーラスを用いてベイジアンネットワークモデルの変数に変換するための手法の提案を行った。その手法は次の通りである：(1) 自由記述データを構文解析及び品詞タグ付けし (2) 行動項目のシソーラスを作成し (3) シソーラスを用いて構文解析結果から特徴的な情報を抽出、その情報を離散確率変数化し (4) 離散確率変数化した行動を従来モデルに導入し、オントロジーを適用する。

この手法を用いて、事故データ 2305 件の自由記述テキストからベイジアンネットワークモデルが構築可能となった。

今後の課題としては、それぞれの項目についてシソーラスの階層構造の組織化を見直し、新しい項目 (場所、一緒にいた人など) の導入を考えている。

参考文献

[Motomura 06] 本村, 西田, 北村, 金子, 柴田, 溝口: 知識循環型事故サーベイランスシステム, 統計数理 (2006) .
 [Shibata 07] 柴田, 本村, 西田, 山中, 溝口: 事故サーベイランスシステムに関する研究 - 事故データベースと日常生活データベースの標準化による情報統合 -, 日本ロボット学会 (2007) .
 [TMS] (株) 数理システム: TextMiningStudio, (<http://www.msi.co.jp/tmstudio/>) (2007) .
 [Motomura 03] 本村: 計測と制御 - ベイジアンネットワークウェア BayoNet -, 計測自動制御学会, Vol.42, No.8, pp.693-694 (2003) .