

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 5 日現在

機関番号：14401

研究種目：挑戦的萌芽研究

研究期間：2012～2013

課題番号：24650069

研究課題名(和文) 希少・特殊条件における事象・シナリオ生起の確率的シミュレーションモデルの学習

研究課題名(英文) Learning Probabilistic Simulation Models for Rare Event/Condition Occurrence

研究代表者

鷲尾 隆 (WASHIO, Takashi)

大阪大学・産業科学研究所・教授

研究者番号：00192815

交付決定額(研究期間全体)：(直接経費) 2,900,000円、(間接経費) 870,000円

研究成果の概要(和文)：従来、データや背景知識を反映する確率モデル学習の研究が盛んに行われてきたが、希少・特殊な条件下での条件付き確率モデル学習法の研究は非常に少なかった。本研究では、既知のデータ・背景知識から希少または特殊な条件下での事象やシナリオ生起の条件付き確率シミュレーションモデルを効率的かつ高精度に学習する原理の確立を行った。更に稀な巨大災害発生シナリオなどのシミュレーションへの実験的適用を通じ、希少・特殊な事象やシナリオの確率的推定・予測の新たな方法論を提示した。

研究成果の概要(英文)：Various approaches for learning probabilistic models from given data and background knowledge have been studied in the past, however, studies on the probabilistic model learning for rare/special conditions have been very limited. In this study, we developed an efficient and accurate approach to learn probabilistic simulation models for the rare/special conditions by using a given data set and its associated background knowledge. Moreover, we demonstrated a novel framework for the probabilistic estimation and prediction of rare/special events and scenarios through its applications to a rare and large scale natural disaster.

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：知識発見 データマイニング 確率モデル 機械学習 希少事象 災害

1. 研究開始当初の背景

近年、ベイズ推定に基づく学習理論など、データと背景知識を反映する確率モデル学習の研究が盛んである。しかし、DNA 2重螺旋やタンパク質折り畳み構造形成など非常に特殊だが膨大な熱的試行により高確率で進む化学反応や、一度起これば甚大な被害をもたらす稀な災害など、所与のデータや背景知識では十分にカバーされない希少・特殊な事象、シナリオ生起の確率的シミュレーションモデル学習のニーズが高まっている。

従来、モンテカルロ計算では、マルチカノニカル法[1]、レプリカ交換法[2]、クロスエントロピー法 [3]など所与の確率モデル下での希少事象シミュレーション研究が進んできた。また、化学分野でも所与の確率モデルを用いた高分子の特殊な状態遷移シナリオの選択的シミュレーション方法が研究されてきた[4]。更に、多変数が複雑に影響し合う大規模状態遷移確率モデルを観測データから学習する研究も行われてきた[5]。しかし、背景知識と観測データを交えて希少・特殊な条件下の事象、シナリオ生起の確率的シミュレーションモデルを高効率、高精度に学習する手法の研究はほぼ手つかずである。

希少・特殊な条件下での事象やシナリオ生起の確率分布は、ほとんどの場合に一般的な条件下での分布と大きく異なる。そのため、所与のデータと背景知識から学習した確率モデルで膨大なシミュレーションを行っても、対象条件に合致する事象やシナリオを集めて統計的に十分な分布を推定することは困難である。従って、

(1) 対象条件下での確率的シミュレーションモデルの所与データと背景知識からの高精度学習原理、

(2) 同モデルからの対象とする条件付き確率分布の高効率シミュレーション原理、

(3) これら原理の統合的実現手法が必要となる。更に、

(4) これを実例題に適用し、新たなシミュレーションやリスク評価の方法論の可能性を提示することも求められる。

2. 研究の目的

以上のように従来、データや背景知識を反映する確率モデル学習の研究が盛んに行われてきた。一方、希少・特殊な条件下での条件付き確率モデル学習法の研究は非常に少ない。本研究では、上記で述べた(1)~(3)の理論的・技術的内容の実現を通じ、既知のデータ・背景知識から希少または特殊な条件下での事象やシナリオ生起の条件付き確率シミュレーションモデルを高効率かつ高精度に学習する原理の確立を目指す。更に(4)について、稀な巨大災害発生シナリオなどのシミュレーションへの実験的適用を通じ、希少・特殊な事象やシナリオの確率的推定・予測の新たな方法論を提示する。我々は、高次

元データからの確率モデル推定[5,6]や希少事象シナリオの確率的シミュレーション[1,2]で研究実績を有し、それを背景に本提案にて新たな研究分野の発展に取り組むこととした。

希少・特殊な条件付き確率モデルの高精度、高効率な学習・シミュレーションの統合研究は進んでおらず、本研究によって、確率モデル学習(機械学習・データ同化)、確率的シミュレーション(モンテカルロ計算)何れの観点からも新しい手法と研究分野が提示されると考える。更に本研究は、生化学反応のように特殊だが重要な事象シナリオや稀有な災害時の事象生起の確率評価など、現在喫緊の科学的、社会的課題に大きく応えることができる。

3. 研究の方法

本研究では、データマイニング及び高次元データからの確率モデル学習に研究実績を有する研究代表者に、希少事象の確率的シミュレーション手法研究の研究実績を持つ統計数理研究所の伊庭幸人教授を連携研究者に迎えた。更に研究代表者の研究室に所属する若手研究者2名を含め、理論と実践の両面で研究遂行力の高い研究組織を構成した。2年間の研究計画とし、平成24年度は、(1)希少・特殊条件付き確率シミュレーションモデル学習及び(2)正確な確率分布の効率的シミュレーションという2原理の確立に取り組んだ。平成25年度には、(3)両原理を統合し希少・特殊条件下でシミュレーションと学習を同時進行する手法開発を行った。また並行して両年度とも、(4)それらの数値実験検証、希少大規模災害などの実例題での検証、改良を重ねた。

課題(1)では、所与のデータと背景知識から対象とする希少・特殊な条件に相対的に関係がある部分を確率的に抽出し、それに基づく統計的不確定性を減らす人工的標本データ生成(一種のブートストラップ)による学習を行う原理を探索した。

課題(2)では、希少・特殊な条件時に有意な頻度分布を持つ標本の中で、元の確率的シミュレーションモデルの持つ相対的分布が正確に反映されるように標本生成する原理を探索した。これら原理の検証用例題として、人工的に確率的シミュレーションモデル及び希少・特殊な条件を準備し、種々のパラメータを変えて各提案原理の精度と効率に関する計算機実験評価を行った。その結果に基づいて、問題点の洗い出し、更なる精度や効率向上のための改良点の抽出を行い、各原理に反映させた。

(2)の原理により対象とする希少・特殊な条件下で有意な頻度の標本を正しい確率分布で効率的に生成でき、そのブートストラップ標本と(1)の原理からより正確に希少・特殊な条件下での確率的シミュレーションモデルのパラメータ学習ができる。これにより、

課題(3)では高い精度と効率を持つ希少または特殊な条件下での事象やシナリオ生起の条件付き確率シミュレーションモデルの学習を実現した。

更に、課題(4)で大規模河川洪水に関する仮想的な状況下での確率的シミュレーション例題を用意し、既存データと背景知識から与えられるモデルに基づいて、希少・特殊な条件下での確率的シミュレーションを実施する。そして、その妥当性や効率の評価を通じて方法論の有効性を提示した。

4. 研究成果

(1) 開発手法

今対象とするシステム S の時間発展 Y に影響を与える条件の集合 X をシナリオと呼ぶことにする。そして、 $P(X)$ を X の確率密度関数とする。さらに、 E を Y に基づいて定義される対象希少事象とし、 $P(E|X)$ を与えられた X の下での E の条件付き確率とする。我々の問題は、対象希少事象 E をもたらす条件集合 X の確率分布 $P(X|E)$ を以下のベイズの公式を用いて計算することである。

$$P(X|E) = P(E|X)P(X)/P(E) \quad (1)$$

$P(X)$ は過去に経験的に得られた X のデータセット $D(X)$ から推定する。一方、 $P(E|X)$ は条件 X を対象システム S のモデル M に入力して時間発展 Y の下で希少事象 E が起きるか否かをチェックすることで、以下のように得られる。

$$P(E|X) = \int P(E|X, Y)P(Y|X)dY \quad (2)$$

特にモデル M が決定的なものである場合には、 Y による積分無しに $P(E|X)$ を直接計算可能である。

一般には $P(E|X)$ は解析的ではないので、ここでは $P(E|X)$ をその標本分布

$$D(X|E) = \{X | X \text{ is drawn from } P(X|E)\}$$

から推定する。この標本分布は原理的には式(1)及び(2)から計算可能であるが、 E が与えられた条件分布 $P(X)$ の下で極めて希少な事象であるため、 $P(X)$ から E を引き起こす X を得ることは殆どできない。そこで、ここでは希少事象標本化技術の1つであるレプリカ交換モンテカルロ法(REM)[7]を適用することとする。一度 $D(X|E)$ が得られれば、 E の生起確率のみならず、それに関連して E を起こす期待条件 X や E が起こる期待時刻など種々の統計量を計算することが可能である。

(2) 河川洪水解析への応用

本開発手法を九州北部の筑後川の洪水解析に適用した。以下に示す河川モデルとシナリオを用いて、希少な洪水シナリオ確率とその性質の評価を行った。

河川モデル

河川流域による降雨量や保水容量の違いを反映するため、筑後川流域を8つのエリアに分割してモデリングを行った。各流域の保水挙動はタンクモデルで表す[8]。これは質量保存、流水圧動特性、表層と地下の流水特性によって表されるモデルである。8つのエ

リア中の4箇所で大規模ダムが存在するが、これらも最大容量まで降水を貯めこむタンクとしてモデル化された。これらの流域エリアとダムは河川本流・支流によって繋がっており、各本流・支流にはその幅と堤防高で決まる最大流量が知られている。もし流量がこの限界を超えれば、その部分で洪水が発生する。このモデルではあるシナリオが起これば洪水は必ず起きるので、モデルは決定的である。

降水量の確率分布モデル

洪水を起こし得る1つのシナリオ X は、4つのダム何れかにおける雨の降り始めから7日間の毎日の降水量によって表される。従って、1シナリオ X は各要素があるダムの日降水量である28次元ベクトルで表される。過去10年間の4つのダムの降水量記録データから、すべての X の情報を収集してデータセット $D(X)$ を得た。この $D(X)$ に対してパラメトリック確率モデルを当てはめ、希少な X も同様な分布に従うと仮定して希少シナリオまで外挿した確率モデル $P(X)$ を得た。具体的には、気象学の知見に基づき[9]、7日間の4つのダムの降水量総計は指数分布に従い、これらが28要素間にディリクレ分布に従って分配されると仮定した。従って、 $P(X)$ はこれら2つの分布の積で表され、そのパラメータは $D(X)$ に対する最尤推定によって決められた。

解析結果

図1は提案手法で得られた降水継続日数と洪水生起確率の関係を示す。これによると、殆どの洪水は降雨開始から24時間以内に発生すること、及び約100回の主要な降雨に一回の割合で洪水が発生することが分かる。降雨継続7日目にはじめて洪水が発生する確率は、1日目に発生する確率の1%に過ぎない。このことから、殆どの洪水は短時間の異常集中豪雨によって起こることが分かる。

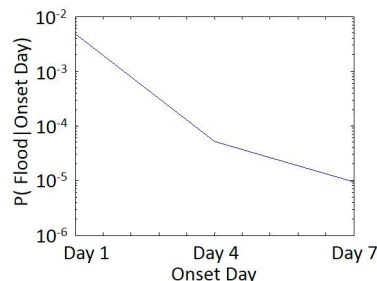
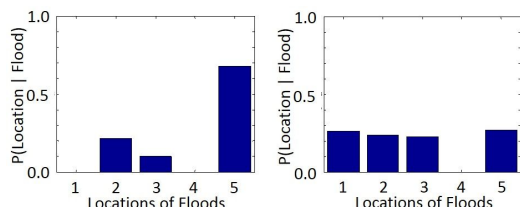


図1 各降雨継続日数と洪水確率の関係

図2(a), (b)は、洪水が生起する場所の瓶度分布である。上流から河川上の位置に番号が振ってあり、1番から5番までそれぞれ2, 4, 5, 7, 8つの流域エリアの降水が通過する。(a)より異常な集中豪雨の際には殆どの洪水は下流域で発生するのに対して、(b)

より7日間に亘る長雨の場合には4番の位置以外で満遍なく洪水が起きることが分かる。これは短期の集中豪雨の場合にはダムの保水によって上流での洪水が防止されるのに対して、雨が長期に亘るとダムが満水を防ぐために放水せざるを得ず、上流での洪水も防げなくなるためである。



(a)降水初日の洪水 (b)降水7日目の洪水
図3 洪水箇所分布

(3) まとめ

以上の研究成果より、既知のデータ・背景知識から希少または特殊な条件下での事象やシナリオ生起の条件付き確率シミュレーションモデルを効率的かつ高精度に学習する原理の確立と、更に稀な巨大災害発生シナリオなどのシミュレーションへの実験的適用を通じ、希少・特殊な事象やシナリオの確率的推定・予測の新たな方法論を提示できた。

参考文献

- [1] Iba, Y. and Hukushima, K.: J. Physical Soc. of Japan, 77, p.103801 (2008).
- [2] Yanagita, T. and Iba, Y.: J. Statistical Mechanics, p.P02043 (2009).
- [3] Rubinstein, R.Y.; J. Operation Research, 99, pp.89-112 (1997).
- [4] Bolhuis P.G. et al.; Annu. Rev. Phys. Chem., 53, pp.291-318 (2002).
- [5] Nguyen V.P., Washio T. and Higuchi T.; Int. J. Knowledge Eng. and Soft Data Paradigms, 2(4), pp.284-311 (2010).
- [6] Ting K.M., Washio T., Wells J. and Liu T.; Proc. IEEE ICDM2011 (Accepted) (2011).
- [7] Hukushima K. and Nemoto K., E J. Phys. Soc. Jpn., 65, pp.1604-160 (1996).
- [8] Ishihara Y. and Kobatake S., Bull. D.P.R.I., Kyoto Univ., 29(260), pp.27-43, 1979.
- [9] Suzuki A. and Kikuchi H., Weather, 31, 179-189, 1984 (in Japanese).

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 2件)

Y. Sogawa, T. Ueno, Y. Kawahara, & T. Washio, Active learning for noisy oracle via density power divergence, Neural Networks, 査読有, Vol.46,

(2013) pp.133-143.

DOI:dx.doi.org/10.1016/j.neunet.2013.05.007

十河康弘, 植野剛, 河原吉伸, 鷲尾隆, Density Power Divergence を用いたロバスト能動回帰学習, 人工知能学会論文誌, 査読有, Vol.28, No.1, (2013) pp.13-2. DOI: doi.org/10.1527/tjsai.28.13

[学会発表](計 3件)

T. Washio and Y. Iba, Rare Flood Scenario Analysis Using Observed Rain Fall Data, Proc. of JSST 2013; International Conference on Simulation Technology, 査読有, Tokyo, (2013年9月12日).

T. Ueno, K. Hayashi, T. Washio & Y. Kawahara, Weighted Likelihood Policy Search with Model Selection, in Advances in Neural Information Processing Systems, Vol.25 (Proc. of NIPS'12), 査読有, Lake Tahoe, Nevada, USA (2012年12月5日) pp.2366-2374.

Y. Sogawa, T. Ueno, Y. Kawahara & T. Washio, Robust active learning for linear regression via density power divergence, in Neural Information Processing (Lecture Notes in Computer Science, 査読有, Vol.7665, Doha, Qatar, (2012年11月14日) pp.594-602.

6. 研究組織

(1)研究代表者

鷲尾 隆 (WASHIO, Takashi)
大阪大学・産業科学研究所・教授
研究者番号: 00192815

(2)研究分担者

該当無し

(3)連携研究者

伊庭 幸人 (IBA, Yukito)
統計数理研究所・モデリング研究系・准教授
研究者番号: 30213200

清水 昌平 (SHIMIZU, Shohei)
大阪大学・産業科学研究所・准教授
研究者番号: 10509871

河原 吉伸 (KAWAHARA, Yoshinobu)
大阪大学・産業科学研究所・准教授
研究者番号: 00514796