Computer Security Symposium 2020
26 - 29 October 2020

主観評価を加えたベイズ分類器の改良

池田 大地[1,a)]　森田 光[1]

概要：情報セキュリティの認証・認識の推定には機械学習で用いられる深層学習や統計的推論などが用いられることがある．推論の精度を上げるには，多数のデータを用いる．けれども，わずかなデータしか得られないケースでは，精度の上げようがなかった．統計的因果推論では，確率的グラフィカルモデル (以下，PGM) により，確率変数相互の関係を精査できることが知られている．著者らは，既存データに主観評価データを付け加えることで，精度を向上する手法を提案する．つまり，主観評価データを新たな確率変数として付け加えるのである．本稿では，主観評価の確率変数を付加するベイズ分類器の拡張式を導出するとともに考察を行う．

キーワード：単純ベイズ分類器，主観評価，確率的グラフィカルモデル，ベイジアンネットワーク

Extended Bayesian Classifier with Subjective Evaluation

Daichi Ikeda[1,a)]　Hikaru Morita[1]

Abstract: Deep learning and statistical inference, which are used in machine learning, are used to estimate authentication and recognition of information security. A large amount of data is used to improve the accuracy of inference. However, the accuracy could not be improved in the case where only a small amount of data was obtained. In statistical causal inference, it is known that the mutual relations of random variables can be closely examined by a probabilistic graphical model (hereinafter, PGM). The authors propose a method for improving accuracy by adding subjective evaluation data to existing data. Subjective evaluation data is added as a new random variable. In this paper, we derive and consider an extended formula of Bayesian classifier that adds a random variable of subjective evaluation.

Keywords: Naive Bayesean classifier, subjective evaluation, probablistic graphical model, Bayesian network

1. はじめに

情報セキュリティの認証・認識の推定には機械学習で用いられる深層学習や統計的推論などが用いられることがある．推論の精度を上げるには，多数のデータを用いる．けれども，わずかなデータしか得られないケースでは，精度の上げようがなかった．また，深層学習でもデータを増やして精度の向上を狙うアプローチになることから，少ないデータでは効率的な精度向上が見込めない．

そこで，著者らはデータに，主観評価データを付け加えることで，精度向上する手法を提案する．主観評価は入手できるデータとは異なる高度な知識背景のある人による学習効果が内包されていると期待できるからである．

本稿では，統計的因果推論を適用した分類問題の 1 つとも考えられる NBC の改良を試み，新たな確率変数を付け加えて判定問題を扱う．主観評価の確率変数を付加方法を導出するとともにその有効性について考察する．

2. 先行研究

2.1 NBC を用いたスパムメールの判定

NBC を用いたスパムメールの判定では，スパムメールから確率変数の bag of words(D) を取得し，カテゴリー (C) に分類を行う．そのために，事後確率である $P(C|D)$ の条件付き確率をベイズの定理を用いて求める．

[1] 神奈川大学大学院
Graduate School of Kanagawa University
[a)] r201970069wk@jindai.jp

$$P(C|D) = \frac{P(D|C)P(C)}{P(D)} \quad (1)$$

このとき，$P(D|C)$，$P(C)$，$P(D)$ は事前確率であるため，観測データから求めることができる．

通常 D はベクトルであり，$D_1, \cdots, D_n$ に分解できる．さらに，条件付き独立性を用いて $P(C|D_1, \cdots, D_n)$ は次のようになる．

$$P(C|D_1, \cdots, D_n) = \frac{P(D_1|C) \cdots P(D_n|C)P(C)}{P(D_1, \cdots, D_n)} \quad (2)$$

bag of words のデータ $(D_1, \cdots, D_n)$ から通常メールかスパムメールの 2 つのカテゴリー (c_1, c_2) を求めることができる．そのため，$P(c_1|D_1, \cdots, D_n)$ と $P(c_2|D_1, \cdots, D_n)$ を求めることができ，2 つの値の比を用いて判定する．このとき，NBC では分母の値が同じ値になるため，比で推定結果を導くことができる．

3. 問題点

3.1 NBC に主観評価を加える困難性

本稿では事例を少なくても，事例とは別の確率変数である主観評価データを新たに加えることで推定結果の向上を試みる．しかし，NBC には的する拡張法がない．

一例として単語のみで NBC を用いてスパムメールの判定を行う際，図 1 のように単語 (W) からスパムメールの判定結果 (Y) の推定を行う．そこで，主観評価 (U) を加える際，U から Y の推定することが考えられるため，図 2 のような関係性と仮定できる．この時，ベイズ定理を用いて条件付き確率 $P(Y|W,U)$ を推定する際，事前確率として $P(W,U|Y)$ を求める必要がある．しかし，NBC では W と U は独立か従属関係であるか判断できないため，条件付き独立性を用いて推定することが妥当であるか判断することができない．このことから，NBC では主観評価データを新たに加えたて推定することが困難であり，正当であるか判断することができない．

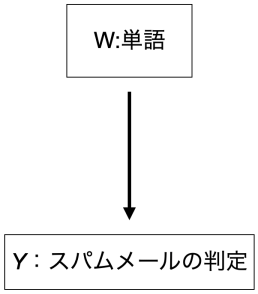

図 1　W と Y の関係性

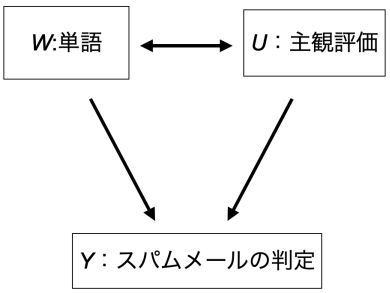

図 2　U を考慮した時の関係性

4. 事前知識

4.1 ベイジアンネットワーク [1], [2], [4]

ベイジアンネットワークは，因果関係を確率により記述する PGM の一種で，因果関係の推論を非循環有向グラフ (以降，DAG) により表し，変数間の関係を条件付き確率で表す確率推論のモデルである．

ベイジアンネットワークの推論では $P(y|x)$ に要約される．X は観測値の集合であり，Y は予測あるいは診断を行う際に重要であると考えられる変数集合である．同時分布 P が与えられた時，$P(y|x)$ の計算は概念的には明白であり，ベイズの規則を直接適用することによって以下の式を導き出せる．

$$P(y|x) = \frac{\sum_s P(y,x,s)}{\sum_{y,s} P(y,x,s)} \quad (3)$$

ここで,S は X と Y を除くすべての変数からなる集合を表す．ベイジアンネットワークに対して同時分布 P が定義されることで，DAG と条件付き確率を用いて $P(y|x)$ を計算できる．

4.2 d 分離 [2]

d 分離はノード同士を結ぶ経路が存在し，グラフによって生成されるすべてのデータセットに共通の関連性を予測することができるプロセスの 1 つである．2 つのノードが d 分離である場合，それらの変数は独立であることを意味する．また，2 つのノードが d 連結である場合，それらの変数は従属である可能性がある．以下に本稿で扱う d 分離の一部定義を示す．

定義 (d 分離)

道 p がノードの集合 Z によりブロックされていることは以下と同値である．

(1) 分離 $A \leftarrow B \rightarrow C$ を含み，中央のノード B が Z に含まれる．

Z がノード X と Y の間のすべての道をブロックするとき，Z が与えられた下で X と Y は d 分離されている．すなわち，Z が与えられた下で X と Y は条件付き独立である．

このことから，$A \leftarrow B \rightarrow C$ のモデルで B が条件付けされていないときは，A と C は従属関係であり，道が存在すると判断できる．そのため，B を条件付けしないモデルに変更した場合でも，A と C にはリンクが存在すると判断できる．

5. 提案方法

5.1 提案方法の概要

本稿では，主観評価データを加えて推定結果の向上を試みるため，複数の観測データとの関係性を判断し，推定方法の確立を行う．しかし，従来の NBC では複数のデータから関係性を把握することは困難であることから，本稿では PGM の一種であるベイジアンネットワークを用いて推定の改良を試みる．

ベイジアンネットワークでは，確率変数の関係性を仮定したグラフと d 分離プロセスを用いて検証が可能である．そのため，主観評価データの実数値を用いず既存の観測データとの関係性を検証できモデルを確立できる．また，ベイジアンネットワークの推論では，新たな確率変数である主観評価データを周辺化することで，既存のデータとの関係性を考慮して推論できる．そのため，従来の NBC では困難であった推論の手法を導き出すことができる．

本稿では PGM を用いて主観評価を加えたスパムメールの判定を行いグラフの仮定を行う．その後，d 分離を用いて局所的にデータの関係性の検証を行い，推論方法の確立を行う．

5.2 PGM によるグラフの仮定

新たな確率変数として主観評価 (U) を加えたスパムメールの判定を行う上で，図 3 のグラフを仮定する．スパムメールの判定を行う上で，文章 (S) はスパムメールを送信する人 (X) から強い影響を受け，単語 (W)，スパムメールの判定 (Y)，主観評価 (U) に影響を与えると判断したためである．しかし，X は人によって異なるため未知数のデータである．そのため，本稿では X を除外した図 4 のモデルから d 分離のプロセスを用いて検証を行う．

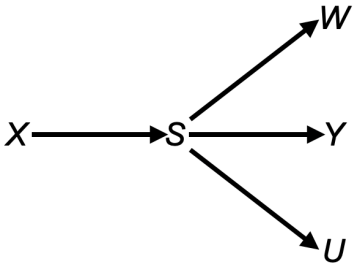

図 3 主観評価を付与したモデル

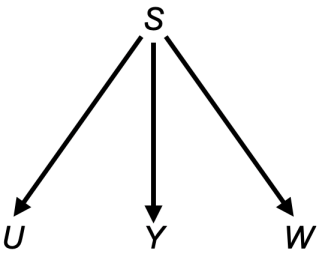

図 4 X を除外し主観評価を付与したモデル

5.3 d 分離によるグラフの検証

本稿では，図 4 を d 分離を用いて検証を行う．図 4 では，文章 (S) は主観評価 (U), スパムメールの判定結果 (Y), 単語 (W) に影響を与えていることから従属関係にあると判断できる．しかし，U と Y，U と W，W と Y は独立か従属関係か判断することが図 4 だけではできない．

そこで，d 分離から各々の関係性の判断を行う．本稿でも NBC のように S は考慮せずに推定を行うため，S は条件付けしていないと判断する．そのため，d 分離のプロセスから U と Y，U と W，W と Y は従属関係である．また本稿では，W と U から Y を推論することを目的としていることから，W と U は Y に影響を与えていると判断でき，U は通常 S から影響を受けるが考慮しないモデルでは，S から取得できる W から影響を受けていると判断する．そのため，d 分離の検証から図 5 のようなモデルを仮定した．

d 分離のプロセスから U と W は従属関係と判断できたので，主観評価データを付与したモデルでは，NBC と同じように U と W を条件付き独立性と判断し推論することは d 分離の観点から正当でないと判断できる．

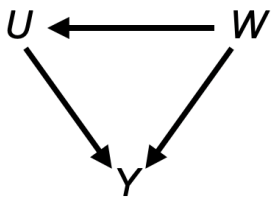

図 5 d 分離を用いて検証したモデル

5.4 ベイジアンネットワークによる推論

ベイジアンネットワークで推論を行う際，式 (3) を用いて行う．そのため，本稿では主観評価データを付与したモデルと式 (3) から推論手法の確立を行う．本稿では，単語 (W) は観測値の集合であり，スパムメールの判定 (Y) は予測を行う際重要である集合と判断する．また，主観評価 (U) は新たな確率変数として付与したことから，W と Y を除く変数からなる集合と判断する．

W は NBC と同様に，$W_1 \cdots W_n$ に分解することができる．また，推定を行う際特定の w_j を取得できない場合が

あるため，$W_j = \{w_j, \overline{w_j}\}$ と定義する．U は 0〜5 段階に分類を行うため，$U = \{u_0, u_1, u_2, u_3, u_4, u_5\}$ と定義する．u_0 は通常のメールであり，$u_1 \sim u_5$ はスパムメールとする．段階が上がるにつれ，スパムメールの度合いが強いことを示している．判定結果である Y は $Y = \{y, \overline{y}\}$ とする．これらにより以下のベイジアンネットワークの推論の式が確立する．

$$P(y|w) = \frac{\sum_U P(y,u,w)}{\sum_{U,Y} P(y,u,w)} \tag{4}$$

同時確率である $P(y,u,w)$ は NBC の改良という観点から乗法定理を用いて因数分解を行う．

$$P(y|w) = \frac{\sum_U P(w_1,\cdots,w_n|y,u_i)P(u_i|y)P(y)}{P(w_1,\cdots,w_n)} \tag{5}$$

式 (5) の $W_1,\cdots,W_n$ は NBC でも条件付き独立性と判断していたため，本稿でも条件付き独立性を用いる．

$$\sum_U P(w_1|y,u_i)\cdots P(w_n|y,u_i)P(u_i|y)P(y) \tag{6}$$

以上の式を用いて $P(y|w)$ と $P(\overline{y}|w)$ の比を用いて判定を行うことができる．ベイジアンネットワークの推論から，新たな確率変数を付与した場合でも判定の手法を確立することができる．

6. 考察

d 分離の検証では，NBC では判断できなかった確率変数の関係性を検証することができた．その結果，主観評価 (U) と単語 (W) は条件付き独立性が適切なのか不適切なのか判断することができた．また，U と W は従属関係だと判断でき，それぞれの確率変数の背景から影響を受けている流れも仮定することができた．

ベイジアンネットワークの推論では，U を周辺化することで，既存データである，W と Y と区別した状態で推論する手法を確立した．また，周辺化を行ったことで，複雑な計算式にはなったが，加法が加わったことで NBC では問題となっていたゼロ頻度問題の改善が期待できる．そのため，スムージングを利用せずに推定結果を導くことができると考える．

7. まとめ

本稿では，主観評価データを加えたスパムメールの判定を PGM を用いて行った．d 分離を用いてモデル検証を行い，ベイジアンネットワークを用いて推論手法を考察した．主観評価データを導入したベイズ分類器の拡張式を導き，その考察も行った．

参考文献

[1] Daphne Koller, Daphne Koller, “Probabilistic Graphical Models: Principles and Techniques”, The MIT Press, 2009

[2] Judea Pearl, “Causal Inference in Statistics: A Primer”, Wiley, 2016

[3] 植野真臣, “ベイジアンネットワーク”，コロナ社, 2013

[4] Judea Pearl, “Causality: Modles, Reasoning, and Inference”, Cambridge University Press, 2001