The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

多クラス分類半自動化エージェントにおける ロバストな確信度算出を目指した擬似学習データ生成手法

Dummy training data generation method towards robust estimation of confidence value of semi-automatic agents for multi-class classification

川口英俊　　中谷裕一
Hidetoshi Kawaguchi　　Yuichi Nakatani

NTT ネットワークサービスシステム研究所
NTT Network Service Systems Laboratories

In this paper, we propose a dummy training data generation method towards robust estimation confidence value of semi-automatic agents for multi-class classification for the purpose of improving the performance. In the case where machine-learning is used for calculating the confidence value, there is a problem that can calculate accurately only with existing training data. Our approach is using simple random values. As a result of experiments of actual data, we confirmed the improvement of the performance.

1. はじめに

本稿では，多クラス分類問題を半自動化するエージェントの性能を向上させる要素技術を提案する．また，IPS(Intrusion Prevention System) のシグネチャ分類作業という実問題への適用実験を通し提案技術の有効性を検証する．

機械学習技術を産業的に活用する需要は高まるばかりであり，人々のデータ分類作業をエージェントに置き換え自動化しようという動きも見られる．本稿では，タスクはデータの分類作業を指すものとし，タスクを実行する人のことをオペレータと呼称する．また，エージェントとはオペレータのタスクを代替する稼働削減を目的としたソフトウェアとする．

しかしながら，分類作業の自動化が精度不足により進んでいない分野も数多く存在する．自動化の基本的なアプローチは，教師あり学習などでオペレータのタスクのパターンを導出し，それをエージェントとして自動的に分類することである．本研究では，要求される精度をモデル設計等の一般的な精度向上の取り組みでは実現が現実的ではない場合を想定している．例えば，100% に極めて近い精度を要求される場合などである．

そこで本研究では，教師あり学習によるパターン単体での分類の自動化を目指すのではなく，オペレータとエージェントがタスクを効果的に分担する半自動化の実用化を目指している．

エージェントは以下の動作により半自動化を実現する．まずエージェントが分類を行うと同時に，その分類に関して正解の確率が高いか，低いかを推定する．正解の確率が高いと判断することを『自信あり』，低いと判断することを『自信なし』と呼称する．『自信あり』とされた分類については，それをそのまま自動分類の結果として用い，その分オペレータの稼働を削減したことになる．『自信なし』とされた分類については，オペレータにそのデータを仲介する．オペレータはその受け取ったデータに対して通常通りにタスクを実行することになる．この場合，オペレータの稼働は削減されない．

2. 先行研究

筆者らが先行研究にて提案した，上述したエージェントの動作を実現するための構成を図 1 に示す [Kawaguchi 18-1]．エージェントは「分類推定機能」「確信度算出機能」「確信度判断機能」の 3 つの機能で構成される．

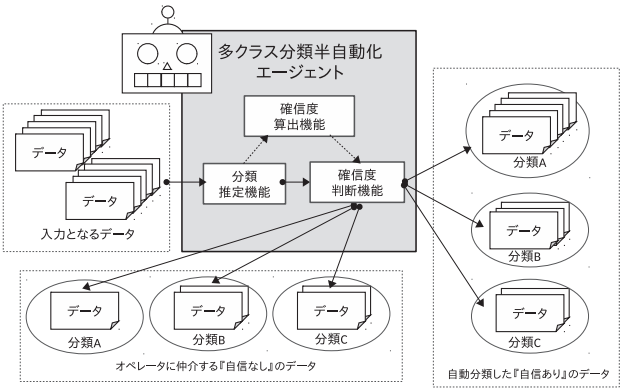

図 1: 前提となるエージェントのアーキテクチャ

分類推定機能は入力となるデータを純粋に分類する機能である．本稿では，オペレータの過去の分類結果を学習データとした教師あり学習により導出された人工ニューラルネットワーク (以下 ANN) 等のパターンを想定している．

確信度算出機能は，分類推定機能が分類を算出する過程を観測し，確信度 [Yamasaki 15] を算出する機能である．確信度とは実際の正解率と正の相関が高いほど望ましい指標であり，上述した『自信あり』『自信なし』を判断するために用いる．確信度に関する先行研究 [Yamasaki 15, Yamasaki 12, Maeda 14] は，教師あり学習の精度向上を目指したものである．

確信度判断機能は，確信度の閾値判定を行い『自信あり』か『自信なし』を判定する機能である．この閾値を本稿では α とおく．この閾値 α はエージェントのハイパーパラメータである．確信度が閾値 α 以上であれば『自信あり』を，未満であれば『自信なし』を分類に付与する．

確信度算出機能の最もシンプルな例は，推定した全クラスの所属確率のうち最も高い値を用いる手法であるが，より精度を高めるための試みとして教師あり学習を用いる手法も存在する．まず分類推定機能を構成するために用いた全学習データをその分類推定機能に再度入力し，各クラスの所属確率等を含んだ分類を推定する過程で算出される学習用の観測データを取得する．それと同時にその観測データに対応する正誤のリストを

連絡先: 川口英俊，NTT 武蔵野研究開発センタ，
〒 180-8585 東京都武蔵野市緑町 3-9-11，
E-mail: hidetoshi.kawaguchi.my@hco.ntt.co.jp

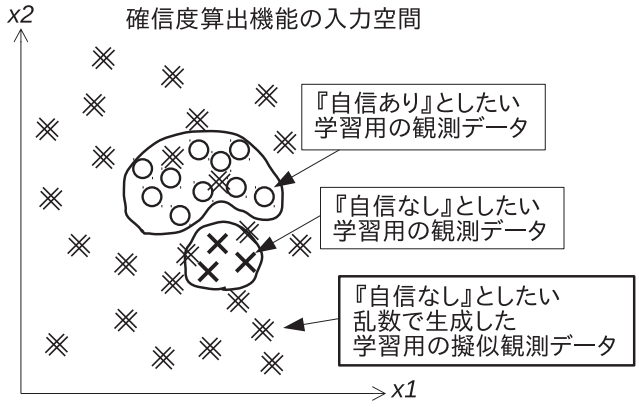

図 2: 提案手法のイメージ

取得する．正解であれば 1, 誤りであれば 0 とする．学習用観測データが入力，その正誤 (1 もしくは 0) が出力となるように教師あり学習を行い確信度算出機能を構成する．

3. 提案手法

上述した教師あり学習を用いて確信度算出機能を構成する場合，既存の学習データに対してのみ『自信あり』『自信なし』の判定の性能向上しか見込めないという問題がある．エージェントの基本的な方針として，学習データのうち正解することができるデータに対しては『自信あり』とすることが望ましく，それ以外の場合は『自信なし』としたいが,『自信なし』としたいデータには以下の 2 種類が存在する．

- 学習データに含まれるが誤って学習したデータ
- 学習データに含まれない未知のデータ

これらのデータは『自信なし』としてオペレータに仲介するほうが望ましいが，既存手法では前者の場合しか考慮にいれることができていない．

そこでこの問題を解決するために，確信度算出機能の入力空間に関して以下の 2 点を仮定した．

- 学習データからの入力は局所化している
- 未知のデータは学習データからの入力以外の領域に広く分布する

この仮定に基づき，確信度算出機能の学習用に乱数を使った擬似観測データを生成する手法を提案する．イメージを図 2 に示す．この手法により生成した擬似観測データを学習用観測データに加え，誤りとして確信度算出機能を学習する．乱数で一様に擬似的に観測データを生成したとしても，全体としては未知のデータに関して精度よく『自信なし』とする，つまり確信度を低めに出力することを狙う．

4. 適用実験

IPS のシグネチャとそれを分類するタスクを対象に，提案手法を用いたもの 1 種類と，比較手法 2 種類の計 3 種類の確信度算出機能を実装し，比較により提案手法の効果を検証する．

4.1 実験データ

対象となる実験データは，実際に運用されている IPS に用いられるシグネチャとその評価の実データセットである．そのシグネチャは IPS のセキュリティベンダより送信されてくるものであり，悪性・異常通信パターンに関する情報である．IPS はこれらのシグネチャと通信を比較し，マッチするものに対して「ロギング」「通知」「遮断」などのアクションを行う．

オペレータは IPS シグネチャの情報を確認しながらひとつずつシグネチャの分類を行っている．分類結果にはあらかじめ定義された IPS の設定情報が対応づいているため，この分類により設定が決定される．

学習用データは 3996 件，テスト用データは 372 件である．学習用データのクラスの内訳は 3437, 252, 111, 1 の 4 分類であり，テストデータは 301, 47, 24 の 3 分類である．

4.2 特徴ベクトルへの変換

本実験データのシグネチャは，1 つの自然言語と 13 個の名義特徴量，3 つの記号で表される順序特徴量の属性 1 つの合計 15 個で構成される．

これらの特徴を変換する手法はいくつか存在するが，先行研究 [Kawaguchi 18-2] にて最良だった特徴ベクトルへの変換手法を用いる．名義特徴量に対しては One-hot エンコーディングを，自然言語の属性には TF-IDF を用いて変換する．

4.3 分類推定機能の実装

自動分類を実行する分類推定機能については 3 層の ANN を誤差逆伝播法により学習して導出する．中間層ノード数は 100, 出力層ノード数は 4 とした．実装には Python の機械学習ライブラリである scikit-learn のバージョン 0.18.1 の MLPClassifier クラスを用いた．上述した設定以外は当ライブラリにおけるデフォルトの値を用いて学習している．

4.4 提案手法と比較手法

学習した分類推定機能ごとに，本実験では PRO, CMP, TOP の 3 つの確信度算出機能を実装して効果を検証する．

PRO は確信度算出機能に 3 層の ANN を用い，提案手法により生成した擬似的な観測データを組み込み誤差逆伝播法により学習して導出する．入力ベクトルは分類推定機能の各クラスの推定した所属確率とし，入力層のノード数は 4 とする．中間層のノード数は 100，出力層のノード数は 2 とした．分類推定機能と同様のライブラリを用いて実装する．提案手法によるサンプリング数は，学習データ数の 100 倍の 399600 件とした．また，所属確率の合計値は 1.0 となるため，その制約を満たす乱数を生成した．

CMP と TOP は比較のための手法であり，[Yamasaki 15] の考えを適用したものである．CMP は提案手法を組み込まないこと以外は PRO と同一の条件で確信度算出機能を構成する．つまり，CMP では分類推定機能の学習に用いたデータのみを用いて学習する．TOP は推定した各クラスの所属確率のうち最も高い値を確信度として用いる手法である．

4.5 性能評価指標

性能評価は確信度判断機能の閾値 α の値毎にテストデータを用いて以下の 2 つの指標を用いて行うものとする．

$$ACC_{co} = \frac{TC + FR + TR}{TC + TR + FR + FC} \quad (1)$$

$$EFF_{co} = \frac{TC + FC}{TC + TR + FR + FC} \quad (2)$$

TC は『自信あり』かつ分類が正解だった分類データ数である．TR は『自信なし』として，本当に不正解だった数である．FR

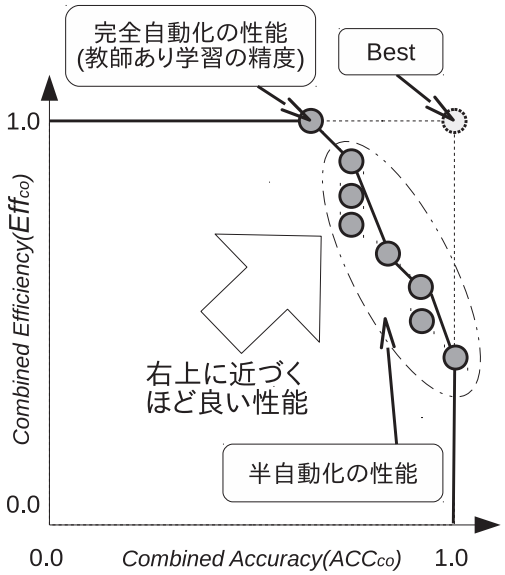

図 3: 共同分類精度と自動分類率のトレードオフ曲線

は『自信なし』としつつ，仮に『自信あり』とした場合，実際には正解だった分類データ数である．FC は『自信あり』として自動分類したものの，実際には不正解だった数である．

ACC_{co} は共同分類精度である．オペレータとエージェントの 2 つの主体で共同で分類した場合の全データに対する分類の正解率を表している．

EFF_{co} は自動分類率である．全データのうち自動で分類することができた割合を表している．つまり，エージェントが『自信あり』と分類した割合である．オペレータの稼働削減量を測るための単純な指標である．

以上の ACC_{co} と EFF_{co} はエージェントに設定する閾値 α により決定することができ，これらはトレードオフの関係にある．共同分類精度 ACC_{co} と自動分類率 EFF_{co} をそれぞれ横軸，縦軸におき α 毎にプロットし，各 ACC_{co} で最も EFF_{co} の高い点を線で繋いだトレードオフ曲線のイメージを図 3 に示す．α を増加させると共同分類精度 ACC_{co} は増加する可能性があるが，自動分類率 EFF_{co} は下降する．

4.6 実験結果と考察

図 4 に実験結果を示す．横軸は視認性の観点から，最小値を 0.90 に設定している．本実験では，上述した実験設定を 100 回試行し，1 試行毎にトレードオフ曲線を算出し，確信度算出機能の 3 種類それぞれについて，全 ACC_{co} ごとに EFF_{co} の平均を実験結果として算出した．1 試行ごとに分類推定機能を学習し，それについて上述した 3 種類の確信度算出機能を実装・学習した．また，すべての有意な閾値 α について共同分類精度 ACC_{co} と自動分類率 EFF_{co} を算出した．テストデータのサンプル数は 372 件であり，現実的に算出できる個数である．

$ACC_{co} < 1.0$ の場合おいて，提案手法を組み込んだ PRO の EFF_{co} が最良の結果となった．提案手法による擬似的に乱数で生成した学習データが有効に働いたと考えられる．特に CMP との差は顕著であり，$0.96 \leq ACC_{co} < 1.0$ の範囲内ではいずれもおよそ 20% の EFF_{co} の上昇を確認できた．

しかしながら，$ACC_{co} = 1.0$ の場合において TOP に劣る結果となった．提案手法による乱数のサンプリングにて，元々の学習データに干渉し，本来『自信あり』とするべきところも『自信なし』としてしまった可能性がある．

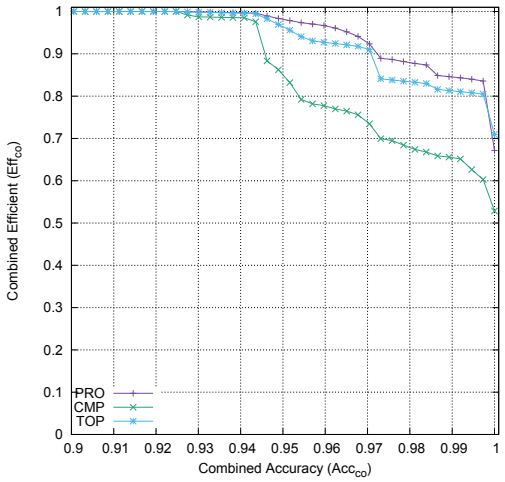

図 4: 実験結果

5. おわりに

本稿では，多クラス分類を半自動化するエージェントの構成要素のうち，確信度算出機能を教師あり学習で構成する際の擬似的な観測データを乱数を用いて生成する手法を提案し，効果を検証した．結果として，提案手法を行わない教師あり学習と比較した場合は平均値では完全に優越し，また最も単純な手法でも $ACC_{co} = 1.0$ 以外は性能が大きく優越する結果となった．

今後の課題としては，ACC_{co} がほぼ 1.0 となることを求められる問題において，高い EFF_{co} の性能を得ることである．単純な乱数生成だけでは元々の学習データと重なりが生まれてしまうため，その部分を除去する手法が求められる．

参考文献

[Kawaguchi 18-1] 川口英俊, 中谷裕一： ミッションクリティカルな多クラス分類を支援するエージェントとその性能評価トレードオフ曲線，合同エージェントワークショップ &シンポジウム 2018(JAWS-2018),

[Yamasaki 15] 山崎俊彦, 大島辰之輔, 相澤清晴： 各クラスへの中間出力値を用いた多クラス認識のための確信度処理，映像情報メディア学会誌，Vo.69， No.8， pp.J257–J260(2015).

[Yamasaki 12] Toshihiko, Y., Tsuhan,C.: Confidence-assisted classification result refinement for object recognition featuring TopN-Exemplar-SVM, Proc. *International Conference on Pattern Recognition*(*ICPR*), pp.1783-1786(2012).

[Maeda 14] Takaki, M., Toshihiko, Y. and Kiyoharu, A.: Multi-stage object classification featuring confidence analysis of classifier and inclined local naive bayes nearest neighbor, Proc. *IEEE International Conference on Image Processing*(*ICIP*)(2014).

[Kawaguchi 18-2] 川口英俊，石原裕一：脅威情報の評価を支援する知的エージェント，人工知能学会全国大会 (第 32 回) 論文集,2P1-05(2018)