

# 識別に悪影響を及ぼす変数の推定・影響排除が可能な Autoencoder ベース識別器の提案

石井良尚<sup>†</sup>

(株)豊田中央研究所<sup>†</sup>

## 1. はじめに

近年, Neural Network (NN) ベースの識別器が, その識別性能の高さから, センサデータを用いた異常検知, 状態推定のタスクで利用され始めている. しかし, 異常時のデータや稀な状態のデータを網羅的に収集することは通常困難であり, NN 識別器に対する学習データは不足しやすい. 学習データが十分でない場合, 図 1 の例のように, NN 識別器は本来識別に無関係な変数の悪影響により誤識別する可能性がある. そこで今回, 誤識別の際に悪影響を与えた変数を推定でき, また, それら変数の悪影響を排除して学習・識別が可能な Autoencoder (AE) [1] ベースの識別器を提案する.

ラベル1のデータ(変数1 ≤ 0.2)					ラベル2のデータ(変数1 > 0.2)				
sample	変数1	変数2	変数3	変数4	sample	変数1	変数2	変数3	変数4
1	0.1	1.5	32	997	1	0.21	2.1	15	990
2	0.12	1.4	32	997	2	0.23	2.1	15	990
3	0.13	1.3	32	997	3	0.25	2.2	15	990
...	...	...	...	...	...	...	...	...	...
999	0.2	2	33	998	999	0.39	3.7	14	990
1000	0.19	1.6	33	998	1000	0.4	3.8	14	990

上記データ(変数3, 4はラベルに無関係な変数)でNN識別器を学習

ラベル2のテストデータ				
sample	変数1	変数2	変数3	変数4
1	0.22	2.2	33	999

図 1: 一般的な NN 識別器の課題

## 2. Autoencoder

AE は NN モデルの一種で, 入力と出力のニューロン数が同一なネットワーク構造を持ち,  $\{x_i\}$  ( $i = 1, \dots, N$ ) を学習データ,  $\{\bar{x}_i\}$  ( $i = 1, \dots, N$ ) を各学習データに対応する AE からの出力とすると, 一般的に次式で定義されるデータごとの再構成誤差  $e_i$  を最小化するように学習することを特徴とする ( $N$  はサンプル数).

$$e_i = \|x_i - \bar{x}_i\|_2 \quad (1)$$

ただし,  $\|\cdot\|_2$  はユークリッドノルムである. これ以降,  $x_i, \bar{x}_i$  の各要素の値  $x_{i,j}, \bar{x}_{i,j}$  ( $j = 1, \dots, D$ ) ごととの二乗誤差  $(x_{i,j} - \bar{x}_{i,j})^2$  を, 変数ごとの再構成誤差と呼ぶ ( $D$  は次元数).

学習済みの AE に対してテストデータを入力し, その再構成誤差があらかじめ設定した閾値を超えた場合, そのテストデータは学習データ中に存在しないラベルを持つと判定できる [2]. ここで, AE が誤判定した場合, すなわち, あるテストデータが閾値を超えたにも関わらず, 実際にはそのテストデータのラベルが学習

### AE-based classifier estimating and eliminating variables that deteriorate the classification

<sup>†</sup>Yoshinao Ishii

<sup>†</sup>Toyota Central R&D Labs., Inc.

480-1118, Aichi, Japan

e1597@mosk.tytlabs.co.jp

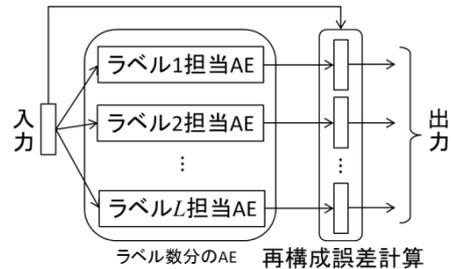


図 2: AE ベース識別器のネットワーク構造

データ中に含まれていた場合を考える. このとき, 誤判定に大きく寄与した変数は, 変数ごとの再構成誤差の値が大きい変数であると推定できる [3]. また, 誤判定に大きく寄与したと推定された変数が, 本来判定に必要な変数である場合, その変数を利用せずに再学習・判定することで, 未知のテストデータに対する判定精度が向上すると考えられる. この特性を利用し, 1 章で述べた特徴を持つ AE ベースの識別器を提案する.

## 3. AE ベース識別器

### 3.1. ネットワーク構造

提案する AE ベース識別器のネットワーク構造を図 2 に示す. AE ベース識別器は学習データのラベルと同数の AE を持ち, 各 AE は担当するラベルのデータのみを再構成することを目的としている. 以降, ラベル  $l \in \{1, \dots, L\}$  のデータの再構成を担当する AE を,  $AE(l)$  と表記する ( $L$  はラベル数). また, 各 AE の出力は入力と比較され, 式 (1) より得られる再構成誤差を最終的に出力する. テストデータの識別は, 最小の再構成誤差を出力した AE が担当するラベルを参照することで行う. つまり, AE ベース識別器は, 各 AE の再構成誤差を元に識別を行う.

### 3.2. 学習方法

$x_i$  のラベルを  $y_i \in \{1, \dots, L\}$ ,  $AE(l)$  からの出力を  $\bar{x}_i^l$  とすると, AE ベース識別器は以下の誤差関数を最小化するように学習する.

$$Loss = \|x_i - \bar{x}_i^{y_i}\|_2 + \alpha \sum_{l=1, l \neq y_i}^L \max(0, \beta \bar{x}_i^l - \bar{x}_i^l) \quad (2)$$

ただし,  $\alpha (\geq 0), \beta (\geq 1)$  はパラメータである. 上式右辺の第 1 項は, 入力データのラベルを担当する AE の再構成誤差を最小化するための項である. 第 2 項はヒンジ関数をベースとするペナルティ項であり, 入力データのラベルを担当する AE の再構成誤差が, その他すべての AE の再構成誤差よりも小さい場合のみ 0 となる. すなわち, AE ベース識別器は, 入力データのラベルを担当する AE の再構成誤差が最も小さくなる

ように、すべての AE を同時に学習する。

また、ラベル  $l$  の識別に悪影響を及ぼす変数の集合  $X_l \in \{1, \dots, D\}$  が与えられている場合、以下の制約をモデルに課しながら学習及び識別することで、 $X_l$  が AE( $l$ )に与える悪影響を排除することができる。なお、各集合  $X_l$  はラベルごとに独立して設定可能である。

- ① AE( $l$ )の再構成誤差を計算する際、 $X_l$ に含まれる変数における変数ごとの再構成誤差を 0 として計算する。
- ② AE( $l$ )の入力層において、 $X_l$ に関係する重みパラメータをすべて 0 に固定する。

### 3.3. 識別に悪影響を与える変数の推定方法

学習後の AE ベース識別器に対してラベル  $l$  のデータを入力した際に、誤識別した場合を考える。誤識別は、AE( $l$ )の再構成誤差が最小とならず、相対的に大きな値となったために起きている。すなわち、一般的な AE の場合と同様に、AE( $l$ )の変数ごとの再構成誤差の値が大きい変数が識別に悪影響を与えた変数であると推定できる。

## 4. 評価実験

### 4.1. 実験方法

本実験は、以下の 3 つの観点から AE ベース識別器を評価するために行う。

1. 一般的な識別問題にて、一般的な NN 識別器と同等の識別性能となるか
2. 識別に悪影響を及ぼす変数が存在するデータにて、その変数を特定できるか
3. 識別に悪影響を及ぼす変数が存在するデータにて、その変数の影響を排除することで、高い識別性能を示すか

評価に用いるデータとして、スマートフォンのセンサデータから所持者の動作（歩き、座り等）を予測する HAPT データ[4]（以降、元データと表記）、及び元データにノイズを加えたノイズデータを用いる。ノイズデータは、ラベルごとに独立に変数を 1 つ乱択し、テストデータ中のその変数の値をすべてランダムな値に置き換えることで作成する。以降、ランダム値に置き換えられた変数をノイズ変数と表記する。

以下に各観点に対する実験・評価方法を述べる。観点 1 については、元データに対して AE ベース及び NN 識別器を適用し、識別精度を比較することで評価する。観点 2 については、まず、観点 1 で学習済の AE ベース識別器に対して元データとノイズデータのテストデータを適用し、元データでは識別に成功していたがノイズデータでは誤識別したノイズテストデータ集合を導出する。次に、集合内のテストデータをラベルごとに分割し、分割ごとに学習済 AE ベース識別器に入力する。このとき、各分割のラベルを担当する AE における、変数ごとの再構成誤差をサンプル毎に計算し、変数ごとに誤差の中央値を算出する。ここで、中央値が最大となった変数が、各分割のラベルに対応するノイズ変数と一致した場合、そのラベルの悪影響を及ぼした変数の特定に成功したと判定する。この実験を、ノイズを加える変数の組み合わせを 100 通りに変化さ

表 1: 各手法の平均識別精度 [%]

手法	NN識別器	AEベース (再学習前)	AEベース (再学習後)
平均精度[%]	63.2	61.4	80.6

せて試行し、特定に成功した確率を評価する。観点 3 については、観点 2 に対する実験の際、各試行において特定された変数の悪影響を排除するように AE ベース識別器を再学習し、再学習後モデルのノイズテストデータに対する識別精度により評価する。また、比較のため、再学習前の AE ベース識別器、及び NN 識別器のノイズデータに対する識別精度も評価する。

なお、HAPT データには多くの冗長な変数が含まれるため、今回は基本となる 7 変数(変数リストの後半 7 変数)のみを用い、ウィンドウ幅を 5 とした 35 次元のデータを元データとして利用する。また、ラベルは移行動作を除く 6 ラベルを用いる。また、AE ベース識別器の各 AE 及び比較対象の NN 識別器は、共に中間層が 1 層（ニューロン数:1000, AE にはスパース制約を付与）の全結合ネットワークを用い、その実装には Chainer を用いる。また、 $\alpha = 100$ ,  $\beta = 1.1$ とする。

### 4.2. 実験結果

観点 1 に対する実験結果としては、AE ベース識別器の精度が 86.0%, NN 識別器の精度が 86.1%となり、両者の一般的な識別問題に対する識別性能は同等であることが確認できた。観点 2 に関しては、ノイズ変数の特定に成功した確率は 76.7%となり、AE ベース識別器が悪影響を与えたノイズ変数を高い精度で推定できることが示された。また、観点 3 に関して、100 回の試行における各手法の平均識別精度を表 1 に示す。また、AE ベース識別器においては、全試行において再学習後に識別精度が向上した。この結果から、AE ベース識別器は実際に悪影響を与える変数を高精度で推定でき、また、悪影響を及ぼす変数の影響を排除して学習・識別できることが検証できた。

## 5. おわりに

本稿では、識別に悪影響を与えた変数を推定でき、それら変数の悪影響を排除して学習・識別が可能な AE ベース識別器を提案した。また、実データを用いた実験により、悪影響を与える変数を高精度に推定でき、その影響を排除可能なことを確認した。今後は、学習データのみで悪影響を及ぼす変数を特定し、テスト時の識別精度が向上するか評価を進める。

### 参考文献

- [1] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks”, *Science*, Vol. 313, pp. 504-507 (2006).
- [2] 櫻田麻由, 矢入健久, “オートエンコーダを用いた次元削減による宇宙機の異常検知”, *人工知能学会全国大会論文集* (2014).
- [3] T. Tagawa, Y. Tadokoro, and T. Yairi, “Structured Denoising Autoencoder for Fault Detection and Analysis”, *In ACML* (2014).
- [4] J. L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra and D. Anguita, “Transition-Aware Human Activity Recognition Using Smartphones”, *Neurocomputing* (2015).