2M2-OS-34a-2

ディープラーニングにおける分類パターンの意味付け支援

Labeling Support System for Classification Patterns in Deep learning

安藤 雅行 *1 Masayuki Ando
河原 吉伸 *2*4 Yoshinobu Kawahara
砂山 渡 *3 Wataru Sunayama
畑中 裕司 *3 Yuji Hatanaka
小郷原 一智 *3 Kazunori Ogohara

*1滋賀県立大学大学院工学研究科
Graduate School of Engineering, The University of Shiga Prefecture

*2大阪大学産業科学研究所
The Institute of Scientific and Industrial Research, Osaka University

*3滋賀県立大学工学部
School of Engineering, The University of Shiga Prefecture

*4理化学研究所 革新知能統合研究センター
RIKEN Center for Advanced Intelligence Project

This paper reported the classification pattern labeling support system using the contents of learning result of deep learning, and verified its effectiveness. In deep learning, there is a problem that concrete classification patterns for deriving reasons for classification are often unknown. The reported system takes out the contents of learning from the learning result of deep learning and makes meaning to that information. Then, the system displays the classification network so that anyone can easily understand the learning result. In verification experiments confirming the effectiveness of the system, based on the learning result of deep learning classifying sentences, multiple subjects made meaning to rules common to the output (classification destination) and classification patterns peculiar to each output. The results show that anyone can easily understand the meaning of the classification pattern of deep learning using the reported system.

1. はじめに

インターネットの普及に伴い，また，SNS（Social Networking Service）の出現により，世の中の文章データが大規模になり，自分が欲しい文章や要らない文章を見分けるのが難しくなってきている．そこで注目されているのが，深層学習を用いたテキストマイニングシステムである [1]．深層学習は近年流行りだした機械学習であり，従来の機械学習と比べて学習の精度が高いという利点がある [2]．例えば，株に関するニュース記事とその時の株価の関係を深層学習で学習させ，株価を予想する研究では，従来の機会学習より高い精度で予想ができることが確認されている [3]．

その一方で，深層学習は，その出力を導いた具体的な分類パターンが不明なことが多い．よって，もし文章分類などで深層学習の学習結果の意味を理解できれば，それはすなわち，人が文章の中身を容易に把握し，自分が求める内容を持った文章の判別ができることを示す．すでに画像認識の分野では，深層学習の学習済みのネットワークの中間層のノード情報から，ノードを画素に直すとこで，画像の分類に重要な部位を理解する研究が行われている [4]．また，ニューラルネットワークにおいて中間層の持つ情報に注目した研究はいくつかある [5]．しかし，文章を取り扱った場合には，中間層ノードの持つ情報の解釈が困難であるという問題がある．

そこで本研究では，文章の分類問題を例として，：深層学習の 1 つである DNN（Deep Neural Network）の学習によってネットワークの層に付けられた重みの値を取り出し，各中間層が学習した情報を，単語の集合として表現する．そして，単語の組み合わせによって中間層ごとの学習された情報を解釈し，そこから分類のパターン，つまり出力を導くルールの意味付けをする手助けを行うシステムの開発を目指す．

連絡先: 安藤 雅行，滋賀県立大学大学院工学研究科電子システム工学専攻，〒 522-8533 滋賀県彦根市八坂町 2500，oh23mandou@ec.usp.ac.jp

2. 深層学習の重みを用いた分類パターン意味付け支援システム

この章では，開発したシステムについて，システムの構成とその処理過程やシステムの機能について述べる．

2.1 システムの構成

システムの構成は図 1 に示すように，入力データとなる文章集合を DNN で学習し，分類ネットワークの重み付けを行う．そして分類ネットワーク上の重みから，提案システムの可視化処理部によって各出力（分類先）を導くネットワーク上のパスと，パス上の各ノードが持つ学習情報の決定，表示を行う．この重要パスとノード情報が表示された分類ネットワークを，意味付け支援ネットワークと呼ぶ．最後に，システム利用者が意味付け支援ネットワークの表示内容を，意味付け支援機能を用いて調整することで，表示内容から分類パターンの意味付けを行う．

2.2 文章集合

システムの入力データには，分類パターンを解釈したい文書を集めた文章集合を用いる．また，この文章集合は，深層学習にかける前に，単語抽出を行い，文章中の単語を要素とした単語集合ベクトルへ変換される．抽出される単語は，名詞，動詞，形容詞とする．これは，文章の特徴をより学習しやすくするためである．

2.3 Deep Neural Network による学習

DNN はニューラルネットワークの中間層を多段化（2 層以上）したものである．DNN の各中間層では入力信号を段階的に学習していき，学習された情報は圧縮され，出力層に送られ，出力となる．各層はノードによって構成され，ノード同士はエッジと呼ばれる情報を伝える線で連結されている．この

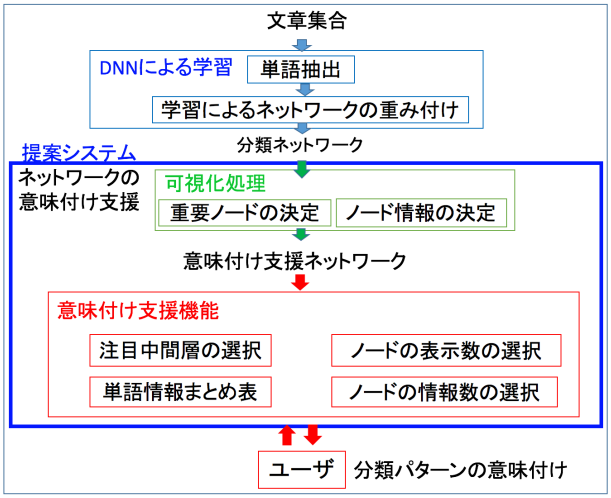

図 1: システムの構成

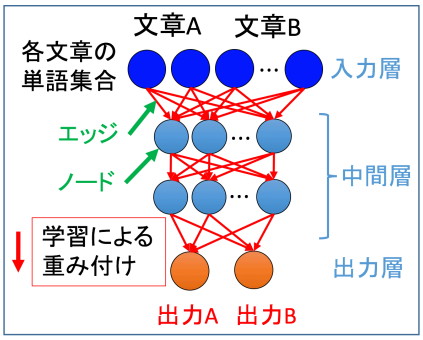

図 2: 分類ネットワークの形成

エッジには，伝わる情報が重要かどうかを判断する数値である重みが付き，重みの値は学習によって決まっていく [6]．例えば，教師付きデータ（データごとに分類先が決められているもの）を入力とした場合，入力データの特徴を学習し，その特徴が決められた分類先（出力層のノード）になるよう，エッジに重み付けがされていく．

システムでは，文章集合を変換した単語集合ベクトルを，教師付きデータとして入力し，DNN で分類を行う．この処理では，各出力を導くような重み付けがされた分類ネットワークを得る．図 2 はそのイメージ図である．

2.4 意味付け支援のための可視化処理

本提案システムの可視化処理部では，DNN によって得られた分類ネットワークから，各出力を導く最も関係が強いパス（ネットワーク上のエッジの繋がりの線）を決定し，可視化する処理を行う．この出力と繋がりが強いパス（重要パスと呼ぶ）の決定について，図 3 に示した 4 層（入力層，中間層 1，中間層 2，出力層）全結合型ネットワークモデルを例として，重要パスを決定する具体的な手順を述べる．

まず，ある分類先（図 3 の例では出力層のノード s）に到達するパスについて，パス上のエッジについた重みの積で定義される，重要度と呼ばれる値を算出する．図 3 の入力層のノード i からの赤いパス上のエッジに付いた重みを $W_{i,j}$，$W_{j,k}$，$W_{k,s}$ とすると，赤いパスの重要度 $W_{i,j,k,s}$ は以下の式で導かれる．なお，入力層の i，中間層 1 の j，中間層 2 の k，出力層の s は，各層のあるノードの番号（左から 0，1，2 $\cdots$ とし，1 ～

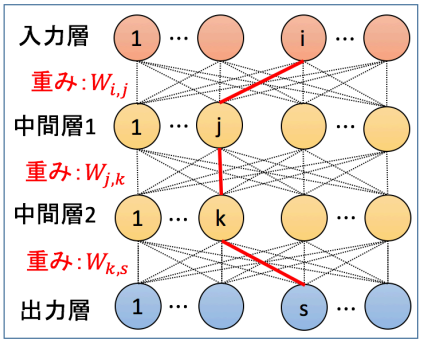

図 3: 4 層全結合型ネットワーク

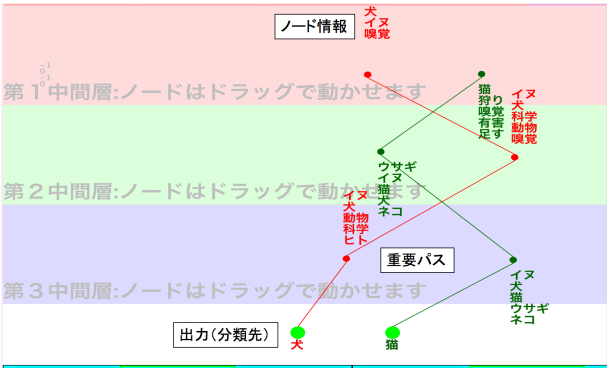

図 4: 意味付け支援ネットワークの一例

各層のノード数の間で変化する）とする．

$$W_{i,j,k,s} = W_{i,j} \times W_{j,k} \times W_{k,s} \tag{1}$$

そして，式 1 で出力 s に到達する全てのパスの重要度を計算して比較し，最も値の大きいパスを，重要パスと決定する．出力 s の重要パスの重要度を $MaxW_s$ として，その計算式を以下に示す．

$$MaxW_s = \max_{i,j,k} W_{i,j,k,s} \tag{2}$$

次に，重要パス上のノードについて，ノードの情報を決定の処理を行う．ノード情報の決定方法は，出力ごとの重要パスを決定した時と同様に，入力層から，ノード情報を決定したいノードへ到達するパスの重要度を算出し，重要パスを決める．ここで，DNN の入力には単語集合ベクトルを用いているため，入力層ノードにはノードごとに 1 種類の単語が入力されていることになる．そこで，重要パス上の入力層ノードの単語を参照し，その単語をノードの情報と決める．

こうして各出力を導く重要パスとパス上のノード情報を表示したものを，意味付け支援ネットワークと呼ぶ．図 4 に，システム上で表示される意味付け支援ネットワークの一例を示す．

図 4 は「犬」「猫」「うさぎ」「ハムスター」「オウム」の生態についての文章をインターネット上のサイト *1 で集めて文章集合としてシステムに入力し，各動物名を出力とした時の，「犬」と「猫」の意味付け支援ネットワークを表示したものである．また，使用した学習ネットワークの中間層数は 3 層であ

*1 「Wikipedia」（https://ja.wikipedia.org/）

表 1: 意味付け支援機能

機能名	効果
注目中間層の変更	注目してる中間層のノード情報を表示する.
重要パス数の変更	各出力ごとの重要パスの本数を増やす. 2 本目以降の重要パスは, 重要度が高い順に選ばれる.
出力の選択	全ての出力の重要パスの表示から, 任意の 1 出力, 2 出力のみの重要パスを表示させる. 特に任意の 2 出力の重要パスを表示させている時は,「ノード情報数の変更」,「単語情報まとめの表示」機能が使えるようになる.
ノード情報数の変更	重要パス上のノードに表示される単語の個数を増やす. 2 つ目以降の単語は, そのノードへのパスの中で, 重要度が高い順に, かつ異なる単語が選択されていく.
単語情報まとめの表示	2 つの出力の重要パス上のノード情報の単語をまとめて表示する. また, それぞれの出力の重要パス上に共通に表示されている単語や, 特有の単語も表示される.

り, システムの画面上では入力層は表示されない. なお, 図 4 ではノード情報を 1 つのノードにつき複数表示させている.

2.5 意味付け支援のための機能

システムには, 利用者が意味付け支援ネットワークの意味付けを行いやすいように, その表示内容を変更できる機能がある. その主なものを表 1 に示す.

意味付け支援機能の具体例として, 図 5 に注目中間層を変更した時の例を示す. 図 5 のネットワークでは, 図 4 で用いた文章集合を入力としており, 全ての出力の重要パスが表示されている. そして, 注目中間層の変更機能により, 3 つの中間層の内 2 つ目の中間層（図 5 中では第 2 中間層と表記）が注目中間層となっている. そのため, 第 2 中間層のノード上にノード情報が表示されるようになっている. また, 重要パス数も, 重要パスの変更機能により各出力ごとに 5 本に増加している.

図 6 は, 図 4 に示したような, 出力「犬」と「猫」を選択している状態（ただし, 重要パス数を各出力 20 本, ノード情報数を 10 個にしている）で, 単語情報まとめ機能を用いてを単語情報まとめを表示された時の具体例である. 単語情報まとめ表は, 画面に表示されている単語を, 2 つの出力ごと（「選択出力 1」と「選択出力 2」の項目）に, 種類別に並べて表示し, 単語ごとの表示数もわかるようになっている. また, 表示された単語の中で, 2 つの出力ごとに共通の単語は, 図 6 の「共通部分」の項目に別枠として表示される.

3. 分類パターンの意味付け支援システムの有効性の検証実験

検証実験では, 深層学習の重みを用いた文章の分類パターンの意味付け支援システムにより, 文章集合の分類ネットワークから, 文章の内容について, 他の文章との共通点, 相違点の意味付けができるかを判断基準とし, システムの有効性を確認する.

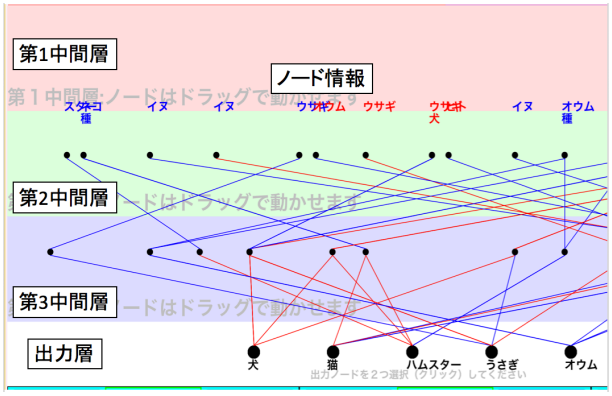

図 5: 注目中間層の変更

図 6: 単語情報まとめの表示

3.1 使用文章集合

実験に使用した文章集合は「アニメ」と「ラノベ」の 2 種類で, それぞれ分類先は 9 つとなっている（つまり出力は 9）. それぞれの文章集合の内容について, 表 2 に示す.

3.2 使用 DNN のネットワークモデル

文章集合ごとの, 実験に用いた DNN の学習ネットワーク（全結合型）の詳細を表 3 に示す. なお, 入力層のノード数は, 単語抽出で文章集合から抽出された単語の種類数と等しい.

3.3 実験内容

システムの有効性を確認するため, 大学生, 大学院生の計 10 人を対象として, 文章集合「アニメ」と「ラノベ」のそれぞれの意味付け支援ネットワークから, 指定した出力と任意で選択してもらった出力の 2 つを選び,「2 つの出力で共通の分類パターン」と「2 つの出力で特有の分類パターン」を考察してもらい, 分類パターンを説明する文を解凍してもらった. そして, 被験者の回答の内容について, 元となる文章の内容と合っているかどうかで, 回答文ごとに正誤を判定した.

3.4 実験結果

図 7 は文章集合「アニメ」と「ラノベ」について,「2 つの出力で共通の分類パターン」と「2 つの出力で特有の分類パターン」の項目で被験者 10 人中の正解の回答ができた人数を示している. ただし,「2 つの出力で特有の分類パターン」の項目はさらに「指定した出力」と「任意で選択した出力」で分けてい

*2 Wikipedia：https://ja.wikipedia.org/, ピクシブ百科事典：http://dic.pixiv.net, ニコニコ大百科：http://dic.nicovideo.jp
*3 上記+アニヲタ Wiki(仮):https://www49.atwiki.jp/aniwotawiki/

表 2: 使用文章データの詳細

データ名	内容
アニメ	アニメ「ラブライブ!サンシャイン!!」の主要人物 9 人の，それぞれの特徴や生い立ちなどについて説明された文章をインターネット上の複数のサイト *2 から 1 人あたり 3 つずつ用意した．また，各キャラには「キャラ 1」のように番号付きのラベルを付与した．
ラノベ	人気ラノベ 9 冊の，それぞれの概要や世界観，用語などについて書かれた説明文をインターネット上の複数のサイト *3 から 1 冊あたり 4 つずつ用意した．また，各ラノベには「ラノベ 1」のように番号付きのラベルを付与した．

表 3: 使用学習モデルの詳細

データ名	入力層ノード数	中間層数	中間層ノード数	出力層ノード数
アニメ	1124	3	50	9
ラノベ	5036	3	50	9

る．図 7 を見ると，文章「アニメ」の「選択した出力に特有の分類パターン」の項目を除いて正解者の人数が被験者全員の 80 %以上であり，本システムには有効性があると言える．一方，正解率が 50 %になってしまった「アニメ」の「選択した出力に特有」の項目は，特有の特徴がわかりにくい出力を選択してしまった人が多かったのではないかと考えられる．そこで，回答の中から正解文と不正解文を比較し，どのような文が不正解になってしまったのかを見ていく．

表 4 は文章集合「アニメ」と「ラノベ」のそれぞれで，「指定した出力に特有の分類パターン」についての回答の一例を示している．表 4 の正解，不正解の回答文を見てみると，まず正解文では「アイドル」「リーダー」や「実家」「旅館」,「高校」「通う」などの無理のない単語の組み合わせをしたことで，正解に導く文を考察することができたと考えられる．そして，不正解文では単に並んでいただけで，その単語間の関係がわからなかったため，「旅館」「好き」や「歳」「魔法」など無理のある組み合わせをしてしまったと考えられる．よって，より正解文を利用者が導きやすくするためには，実際の文章中で使用された単語の，組み合わせを考慮した出力が望まれると言える．

4. おわりに

本研究では，深層学習の文章の分類を行ない，学習によって重み付けがされた分類ネットワークを用いた，分類パターンの意味付け支援システムを開発した．また，その有効性の検証実験として，10 人の被験者にシステムを用いて複数の文書に共通の分類パターンや，特有の分類パターンの意味付けをしてもらい，80 %以上の人が正解文を導き出せたことから，本システムの有効性が確認された．今後の課題としては，さらに正解率を高めるため，実際の文章中の単語同士の関係がわかるような入力方法を検討する必要がある．

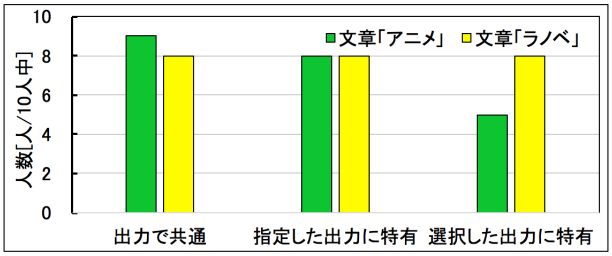

図 7: 回答項目ごとの正解者数

表 4: 回答された指定した出力に特有の分類パターンについての正解文と不正解文の一例

正誤	「アニメ」の回答内容	「ラノベ」の回答内容
正解	アイドルグループのリーダー	魔法の世界である
	アイドルが好き	高校が舞台である
	実家が旅館	目に特殊な能力がある
	姉がいる	魔法能力がある世界
	学生でアイドルをやっている	高校に通っている
不正解	姉が好き	歳をとる魔法がある
	旅館が好き	年齢を重ねると魔法使いなる
	自分が姉である	魔法の能力が年齢と関係
	お姉さんタイプのキャラ	目が魔法のキーポイント
	妹か弟がいる	目から魔法が出る

参考文献

[1] ボレガラ ダヌシカ：自然言語処理のための深層学習，人工知能学会誌，Vol.29，No.2，pp.195-201，2014

[2] Ebru Arisoy, Tare N. Sainath, Brian Kingsbury, Bhuvaba Ramabhadran：Deep Neural Network Language Models, in Proc. of the NAACLHLT Workshop, Will We Ever Really Replace the N-gram Model?, pp.20-28，2012

[3] 藤川 和樹，関 和広，上原 邦昭：言語情報を用いた経済指標の予測と分析，第 28 回人工知能学会全国大会，3L4-OS-26b-3，2014

[4] 西銘 大喜：ディープニューラルネットワークによる画像からの表情表現の学習，第 29 回人工知能学会全国大会，3L4-3，2015

[5] 松倉 健太郎，村田 昇：ニューラルネットワークの中間層における独立な特徴量の抽出，電子情報通信学会技術研究報告，NC，Vol.106，No.102，pp.63-67，2006

[6] 巣篭 悠輔：Deep Learning Java プログラミング 深層学習の理論と実装，株式会社インプレス，2016