Vol.2014-NL-216 No.17
Vol.2014-SLP-101 No.17
2014/5/23

音声対話での利用を目的とした Deep Neural Network による ユーザ発話のトピック分類方法の検討

本間 健[†1, a)]　神田 直之[†1]

ユーザの自由な言い回しを許容する音声対話システムにおいて，ユーザ発話が属する話題を分類する技術（トピック分類技術）は，基本技術の 1 つである．本研究では，近年注目されている Deep Neural Network(DNN)をトピック分類に適用し，有効性を評価した．NIST の TREC 質問応答トラックで使われた英語質問文データ（約 5,500 文）を学習データとし，入力質問文が 50 種類のトピックのうちいずれに属するかを分類するトピック分類器を構築した．分類精度の評価の結果，Dropout 法を用いて学習した DNN による分類正解率が 84.1 %となり，従来手法で最高性能であった最大エントロピー法の 83.4 %を上回った．

Topic Classification of Spoken Sentence Using Deep Neural Network for Spoken Dialogue System

TAKESHI HOMMA[†1,a)]　NAOYUKI KANDA[†1]

In spoken dialogue system, the topic classifier for a user's question is an important component because the topic classifier should be able to estimate the correct topic from user's questions with various expressions. In this research, we applied Deep Neural Network (DNN) to classify question sentences. We developed a DNN-based topic classifier which estimates one of topics to which the inputted question belongs. English questions (about 5,500 sentences) in the NIST TREC QA track were used to train the DNN-based topic classifier. Evaluation results showed that the DNN-based classifier trained with the dropout technique gave the correct classification at a rate of 84.1 %. This correct classification rate was higher than the rate given by a classifier based on the Maximum Entropy principle which had the best accuracy obtained from previous topic classifiers at 83.4 %.

1. はじめに

発話トピック分類技術とは，音声対話システムへユーザが発話した内容が，あらかじめ設定したトピックのうち，いずれに属するかを自動分類する技術である．

音声対話システムに関する知識を持たないユーザが，システムに話しかけるとき，ある 1 個の発話行為を表現する発話であっても，さまざまな言い回しや多様な語彙が発話される．そのため，ユーザにとって自然な音声対話を実現するためには，ユーザの多様な発話から，正確にユーザの発話行為に該当するトピックを推定できる発話トピック分類技術が求められる．

発話トピック分類技術は，(a) 人手で構築したルールに基づく方法[1]，(b) 機械学習を用いた統計的手法[2][3][4][5]，に分けることができる．機械学習を用いた手法では, コサイン類似度[2], Support Vector Machine(SVM)[3], 最大エントロピー法[4], AdaBoost[5]などを用いた発話トピック分類手法が提案されてきた．機械学習を用いた手法では，あらかじめ，それぞれのトピックに該当する発話文を多数収集し，学習データとして利用できれば，高い分類精度を実現できる．たとえば，飛行機フライト案内(ATIS)におけるトピック分類の実験では，トピック数 14 個，学習文数 5,822 個の条件にて，分類誤り率 4.8 %が得られている[6][7].

しかし，トピックの種類が多い場合や，学習データが少ない場合へ対応する手法は，十分に検討されていない．本研究では，学習文数に対してトピックの種類数が多い場合に着目する．具体的には，学習文数が約 5,500 個，トピック数が 50 個の場合において, トピック分類方法の検討を行う．

また，近年，認識タスクにおいて Deep Neural Network(DNN)を用いた手法が注目されており[8], 画像認識[9]や音声認識[10][11]において, 従来手法より高い認識精度が報告されている. 発話トピック分類への適用例としては，Sarikaya ら[12]が，単語素性を入力とした Deep Belief Network によるトピック分類の精度は，SVM と同程度であったと報告している．しかし，発話トピック分類に適用した研究は少なく, DNN で提案されているさまざまな学習手法の効果の検討や，公開データセットによる評価は，行われていない．

本研究では, DNN に基づいた発話トピック分類器を構築し，従来手法との精度を比較することで，DNN の有効性を検討する．DNN の学習手法として，音声認識において高い性能を示している Dropout 法[13]および識別的 pre-training[14]を適用する．また，データセットとして，Li ら[15]が公開している NIST の TREC 質問応答トラックの質問文をもとに作成された質問分類データセットを使用する．

†1 (株) 日立製作所中央研究所
Central Research Laboratory, Hitachi, Ltd., Tokyo 185-8601, Japan
a) takeshi.homma.ps@hitachi.com

Vol.2014-NL-216 No.17
Vol.2014-SLP-101 No.17
2014/5/23

2. Deep Neural Network に基づく発話トピック分類器の構築

2.1 Deep Neural Network (DNN)

DNN とは, 機械学習に使用されるニューラルネットワークのうち，とくに層の数が多いものを指す．ニューラルネットワークは，層の数を多くすることにより，複雑な入出力関係をモデル化できる能力を持つ．しかし，2000 年代前半までは，層の数が多くなった場合に，効率的にネットワークのパラメータを学習する手法が存在しなかった．そのため，実応用されるニューラルネットワークは，層の数が少ないものに限定されていた．ところが，2000 年代後半より, それぞれの層ごとのパラメータを事前学習(pre-training)する手法が提案され[8]，さまざまな用途への応用が拡大している．

DNN を用いた発話トピック分類器の構成を図 1 に示す．最初に，発話文から，DNN に入力する素性ベクトル$\boldsymbol{x}$を作成する．本研究では，素性として，発話文の単語の N-gram を使用することとする．すなわち，素性ベクトル$\boldsymbol{x}$のそれぞれの要素は，異なる N-gram 要素に対応しており，発話文のなかに素性が存在すれば 1，存在しなければ 0 に設定される．

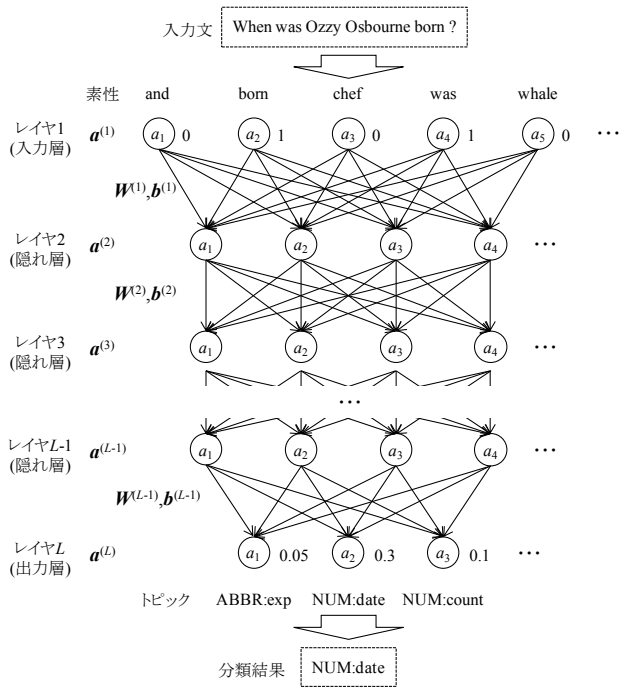


図 1　DNN を使用した発話トピック分類器
Figure 1　DNN-based topic classifier.

つぎに，素性ベクトル$\boldsymbol{x}$を，DNN の入力層のノードの値$\boldsymbol{a}^{(1)}$に設定する（$\boldsymbol{a}^{(1)} \leftarrow \boldsymbol{x}$）．入力層$\boldsymbol{a}^{(1)}$の値は，隠れ層$\boldsymbol{a}^{(2)}, \boldsymbol{a}^{(3)}, \cdots, \boldsymbol{a}^{(L-1)}$を経由して，出力層$\boldsymbol{a}^{(L)}$まで伝播する．$l+1$番目の層のノードの値$\boldsymbol{a}^{(l+1)}$は，$l$番目の層のノードの値$\boldsymbol{a}^{(l)}$を使い，以下の式により計算される．

$$\boldsymbol{a}^{(l+1)} = \begin{pmatrix} a_1^{(l+1)} \\ a_2^{(l+1)} \\ \vdots \\ a_{n_{l+1}}^{(l+1)} \end{pmatrix} = \begin{pmatrix} f\left(\sum_{j=1}^{n_l}\left(W_{1j}^{(l)} a_j^{(l)}\right) + b_1^{(l)}\right) \\ f\left(\sum_{j=1}^{n_l}\left(W_{2j}^{(l)} a_j^{(l)}\right) + b_2^{(l)}\right) \\ \vdots \\ f\left(\sum_{j=1}^{n_l}\left(W_{n_{l+1}j}^{(l)} a_j^{(l)}\right) + b_{n_{l+1}}^{(l)}\right) \end{pmatrix}$$

$$= \begin{pmatrix} f\left(z_1^{(l)}\right) \\ f\left(z_2^{(l)}\right) \\ \vdots \\ f\left(z_{n_{l+1}}^{(l)}\right) \end{pmatrix} \quad (1)$$

ただし，

$$z_i^{(l)} = \sum_{j=1}^{n_l}\left(W_{ij}^{(l)} a_j^{(l)}\right) + b_i^{(l)} \quad (2)$$

である．また，$W_{ij}^{(l)}$および$b_i^{(l)}$を行列で表現することで，式(1)は，以下のように表現される．

$$\boldsymbol{W}^{(l)} = \begin{pmatrix} W_{11}^{(l)} & \cdots & W_{1n_l}^{(l)} \\ \vdots & \ddots & \vdots \\ W_{n_{l+1}1}^{(l)} & \cdots & W_{n_{l+1}n_l}^{(l)} \end{pmatrix}, \boldsymbol{b}^{(l)} = \begin{pmatrix} b_1^{(l)} \\ \vdots \\ b_{n_{l+1}}^{(l)} \end{pmatrix}$$

$$\boldsymbol{z}^{(l)} = \begin{pmatrix} z_1^{(l)} \\ \vdots \\ z_{n_{l+1}}^{(l)} \end{pmatrix}, F\left(\boldsymbol{z}^{(l)}\right) = \begin{pmatrix} f\left(z_1^{(l)}\right) \\ \vdots \\ f\left(z_{n_{l+1}}^{(l)}\right) \end{pmatrix} \quad (3)$$

$$\boldsymbol{a}^{(l+1)} = F\left(\boldsymbol{z}^{(l)}\right) = F\left(\boldsymbol{W}^{(l)}\boldsymbol{a}^{(l)} + \boldsymbol{b}^{(l)}\right)$$

$\boldsymbol{W}^{(l)}$と$\boldsymbol{b}^{(l)}$は，DNN のパラメータであり，それぞれ重み，バイアス項と呼ばれる．これらのパラメータは，後ほど学習によって最適化される．$f()$は，活性化関数と呼ばれ，各ノードへ入力された信号を出力に変換する．$F()$は，ベクトルの各要素に$f()$をかける関数である．n_lは，l番目の層のノードの数である．n_Lは，出力層のノード数であり，トピックの種類数と一致する．

出力層のノード値を示すベクトル$\boldsymbol{a}^{(L)}$のそれぞれの要素は，それぞれ異なるトピックに対応する．ベクトル$\boldsymbol{a}^{(L)}$の値を参照し，もっとも大きかった要素に対応するトピックを，分類結果として出力する．

活性化関数$f()$には，隠れ層と出力層で異なるものを使用した．隠れ層の活性化関数には，シグモイド関数を使用した．

$$f\left(z_i^{(l)}\right) = \frac{1}{1+\exp\left(-z_i^{(l)}\right)} \quad (4)$$

発話文は，トピックのうちいずれかの 1 個にのみ属する．出力層のノードの値には,「入力文が存在する条件でのトピックの事後確率」という意味合いを持たせるため，値の総和が 1 になる制約を設けた softmax 関数を使用した．

$$f\left(z_i^{(n_L)}\right) = \frac{\exp\left(z_i^{(n_L)}\right)}{\sum_{j=1}^{n_L}\exp\left(z_j^{(n_L)}\right)} \quad (5)$$

2.2 学習アルゴリズム

DNN の学習アルゴリズムを説明する. 1 個の学習データ

は，素性ベクトル$\boldsymbol{x}$と，正解のトピックの要素だけが 1 であるトピックベクトル$\boldsymbol{y}$の組$(\boldsymbol{x},\boldsymbol{y})$で与えられる．そして，入力層$\boldsymbol{a}^{(1)}$に素性ベクトル$\boldsymbol{x}$を代入して求めた出力層のノード値$\boldsymbol{a}^{(L)}=(a_1^{(L)},a_2^{(L)},\cdots,a_{n_L}^{(L)})^T$が，トピックベクトル$\boldsymbol{y}=(y_1,y_2,\cdots,y_{n_L})^T$に近づくように, DNN のパラメータ（重み，バイアス項）を更新する．

DNN の出力層の値$\boldsymbol{a}^{(L)}$と，トピックベクトル$\boldsymbol{y}$の近さを評価する指標として, 交差エントロピーに基づく評価関数Jを使用する．

$$J=\sum_{i=1}^{n_L}\left(-y_i\log a_i^{(L)}-(1-y_i)\log\left(1-a_i^{(L)}\right)\right) \quad (6)$$

つぎに，評価関数Jに対して，重み，バイアス項の各要素に対する勾配を計算する．この勾配は，back propagation 法により計算することができる．勾配が計算されれば，勾配の符号と反対の方向にパラメータを更新していくことで，評価関数Jを最小化することができる．実際には，いくつかの学習データをグループ化し，勾配をグループ内で平均した値を計算し，その平均値に従って，パラメータを更新する．グループに属するm個の学習データを$\{(\boldsymbol{x}_1,\boldsymbol{y}_1),(\boldsymbol{x}_2,\boldsymbol{y}_2),\cdots,(\boldsymbol{x}_m,\boldsymbol{y}_m)\}$と表すと，パラメータの更新は以下の式で表現される．

$$W_{ij}^{(l)}\leftarrow W_{ij}^{(l)}-\alpha\left(\frac{1}{m}\sum_{k=1}^{m}\frac{\partial}{\partial W_{ij}^{(l)}}J(\boldsymbol{x}_k,\boldsymbol{y}_k)\right) \quad (7)$$

$$b_i^{(l)}\leftarrow b_i^{(l)}-\alpha\left(\frac{1}{m}\sum_{k=1}^{m}\frac{\partial}{\partial b_i^{(l)}}J(\boldsymbol{x}_k,\boldsymbol{y}_k)\right) \quad (8)$$

グループに属する学習データの個数mは，ミニバッチ数と呼ばれる．αは，学習率と呼ばれる値であり，1 回の更新におけるパラメータの更新度合いを示す．本研究では，学習が進むたびに，学習率を過去の勾配に基づいて徐々に減らしていく AdaGrad 法[16]を使用した．

2.3 識別的 pre-training

DNN を学習するために行う手法として，第 1 に，多層のネットワークの入力層と出力層に学習データを与えて，back propagation 法により学習する手法が考えられる．しかし，この手法の場合，層の数を増やすほどパラメータの最適化が難しくなり，適用できる層の数が限定されていた．しかし, 近年， 隠れ層のパラメータを入力層から順に更新していく事前学習(pre-training)の方法が提案され，DNN により従来手法より高い識別性能が報告された．事前学習の手法には，Restricted Boltzmann Machine による方法[8]，Stacked Auto-encoder による方法[17]，識別的 pre-training による方法[14]などが提案されている．本研究では，英語の音声認識実験により，高い性能が報告されている識別的 pre-training を使用する．

識別的 pre-training の手順を以下に説明する．

(1) 入力層-隠れ層-出力層からなるニューラルネットワークを構成し，パラメータを小さい乱数により初期化する．
(2) (1)で構築したニューラルネットを学習する．
(3) 出力層を取り除き，新たに隠れ層-出力層を接続する．新たに接続した隠れ層と出力層の間のパラメータを小さい乱数により初期化する．
(4) (3)で構築したニューラルネットワークを学習する．
(5) 所定の隠れ層の数になるまで，(3),(4)を繰り返す．

2.4 Dropout 法

DNN は，複雑な非線形関係を持つ入力と出力であっても，その入出力関係をモデル化できる高い表現力をもつ．一方, 高い表現力があることにより, 学習データに対して，過学習を引き起こすことがある．本研究では，過学習を防止するため，Dropout 法[13]を使用する．

Dropout 法では, 学習時において, 隠れ層のノードの 50 %をランダムに選出し，その出力値を 0 にする．この操作により，学習されるニューラルネットワークは，それぞれの学習データにより，異なる構成となる．そのため，最終的に学習されたニューラルネットワークには，複数のニューラルネットワークによるトピック分類結果を平均化するような作用が働く．この効果により，学習データ以外の未知のデータに対しても，分類性能を高く維持することができる．

3. 評価実験

3.1 方法

3.1.1 データセット

評価実験に使用するデータには，Li らが公開している英語の質問分類タスクのデータセットを使用した[15]．このデータセットでは，NIST が主催する TREC の質問応答トラックで使われた質問文に対して，質問分類クラスが付与されている．このデータセットに収録される質問文は，約 6,000 文である. 質問分類クラスの種類は, 50 種類である．本実験では，この質問分類クラスをトピックとみなして，質問文からの自動分類を試みる．

実際の質問文の例を表 1 に示す．また，トピックとして使用する質問分類クラスの内訳を表 2 に示す．学習文数は 5,452 個，評価文数は 500 個である．

同一のデータセットを使用した過去の研究として，Blunstom ら[18]の研究がある．Blunstom らは，最大エントロピー法を使用したトピック分類手法を提案し，単語 N-gram 素性に基づく分類正解率が，83.4 %であったと報告している．

3.1.2 素性ベクトル

発話文の素性ベクトルとして，入力される質問文の Bag-of-Words 表現を用いた．素性ベクトルのそれぞれの要

素は，それぞれ異なる単語 N-gram を表しており，該当する単語 N-gram が発話文に存在すれば 1，存在しなければ 0 となる．

単語 N-gram の次数には，1-gram，2-gram，3-gram を用いた．また，質問文のそれぞれの単語には，あらかじめ Porter のアルゴリズムによるステミング処理[19]を行い，語形変化を取り除いた状態にしたうえで，N-gram を作成した．ただし，3-gram のうち，学習データに現れた頻度が 1 回であった 3-gram は，素性から除外した．素性の種類数は，40,426 個であった．

表 1　データセットの質問文例

Table 1　Examples of question sentences in the dataset.

質問文	質問分類(トピック)
How did serfdom develop in and then leave Russia ?	DESC:manner
What films featured the character Popeye Doyle ?	ENTY:cremat
Who was the inventor of silly putty ?	HUM:ind
What fowl grabs the spotlight after the Chinese Year of the Monkey ?	ENTY:animal
What is the full form of .com ?	ABBR:exp
What sprawling U.S. state boasts the most airports ?	LOC:state

表 2　各トピックの質問文数

Table 2　Numbers of question sentences for each topic.

質問分類(トピック)	学習データ	評価データ	質問分類(トピック)	学習データ	評価データ
ABBR:abb	16	1	HUM:desc	47	3
ABBR:exp	70	8	HUM:gr	189	6
DESC:def	421	123	HUM:ind	962	55
DESC:desc	274	7	HUM:title	25	1
DESC:manner	276	2	LOC:city	129	18
DESC:reason	191	6	LOC:country	155	3
ENTY:animal	112	16	LOC:mount	21	3
ENTY:body	16	2	LOC:other	464	50
ENTY:color	40	10	LOC:state	66	7
ENTY:cremat	207	0	NUM:code	9	0
ENTY:currency	4	6	NUM:count	363	9
ENTY:dismed	103	2	NUM:date	218	47
ENTY:event	56	2	NUM:dist	34	16
ENTY:food	103	4	NUM:money	71	3
ENTY:instru	10	1	NUM:ord	6	0
ENTY:lang	16	2	NUM:other	52	12
ENTY:letter	9	0	NUM:perc	27	3
ENTY:other	217	12	NUM:period	75	8
ENTY:plant	13	5	NUM:speed	9	6
ENTY:product	42	4	NUM:temp	8	5
ENTY:religion	4	0	NUM:volsize	13	0
ENTY:sport	62	1	NUM:weight	11	4
ENTY:substance	41	15			
ENTY:symbol	11	0			
ENTY:techmeth	38	1			
ENTY:termeq	93	7			
ENTY:veh	27	4			
ENTY:word	26	0			

3.1.3 ベースライン手法

ベースラインの手法として，音声対話の発話トピック分類によく使用される，SVM および最大エントロピー法を設定し，DNN とのトピック分類精度の比較を行った．いずれの手法においても，3.1.2 節で説明した素性ベクトルを入力とし，トピックを出力とする分類器を構築した．

SVM によるトピック分類では，予備実験の結果，非線形カーネルである RBF カーネルおよび多項式カーネルを使用した場合よりも，線形カーネルを使用した場合に，もっともよい分類精度が得られた．そのため，SVM を用いた実験では，線形カーネルを使用した．SVM の実装には，オープンソースである LIBLINEAR を使用した[20]．また，SVM のパラメータは，評価データでの分類正解率がもっとも高くなるように，あらかじめ調整した．

最大エントロピー法のトピック分類では，オープンソースである Classias を使用した[21]．重みの正則化は L2 ノルムにて行い，正則化係数は，評価データの分類正解率がもっともよくなるように，あらかじめ調整した．

3.1.4 DNN による手法

DNN の実験条件を表 3 に示す．

隠れ層の数による分類精度の影響を調べるため，隠れ層の数は，1 層から 4 層までの 4 条件とした．隠れ層の 1 層目は，ノード数を 100 とし，他の隠れ層よりもノード数を小さくする構成とした．本実験の入力層のノード数は，40,000 以上と多数であるうえ，ほとんどのノード値が 0 であるスパースな層になる．そのため，パラメータの学習を効率的にするには，一度，入力情報の次元を圧縮する必要があると考えた．そのため，第 1 の隠れ層のノード数を少なくする構成とした．

また，Dropout 法の有無の両条件を評価し，Dropout 法の効果を検証した．

表 3　DNN の実験条件

Table 3　Evaluation conditions of DNN-based topic classifier.

固定条件	
入力層のノード数（素性数）	40,426
1-gram素性	8,039
2-gram素性	28,223
3-gram素性	4,164
出力層のノード数（トピック数）	50

変化条件	
隠れ層の数	1層（ノード数: 100）
	2層（ノード数: 100,500）
	3層（ノード数: 100,500,500）
	4層（ノード数: 100,500,500,500）
Dropout法	なし
	あり

パラメータの学習回数を説明する．2.3 節で説明したとおり，識別的 pre-training では，最初に 1 層目の隠れ層だけを追加したネットワークでパラメータ学習を行う．この学

Vol.2014-NL-216 No.17
Vol.2014-SLP-101 No.17
2014/5/23

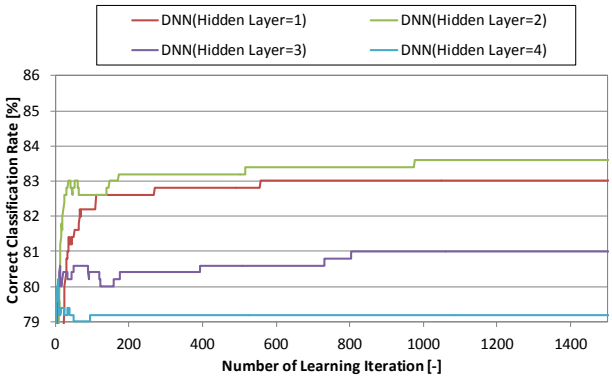
習が終了したのち，2 層目の隠れ層を追加したネットワークでパラメータ学習を行う．このプロセスを，最終の隠れ層を追加するまで繰り返す．本研究では，最終ではない隠れ層の学習では, 全データの学習を 10 回繰り返すこととした．そして，最終の隠れ層を追加したネットワークの学習では，全データの学習を 100 回以上繰り返し，分類精度の変化を観察した．

学習率の初期値は，Dropout 法なしの条件では全層にて 0.05 とし，Dropout 法ありの条件では 1 層目の隠れ層のみ 1.0，以降の隠れ層では 0.2 とした．ミニバッチ数は，50 とした．

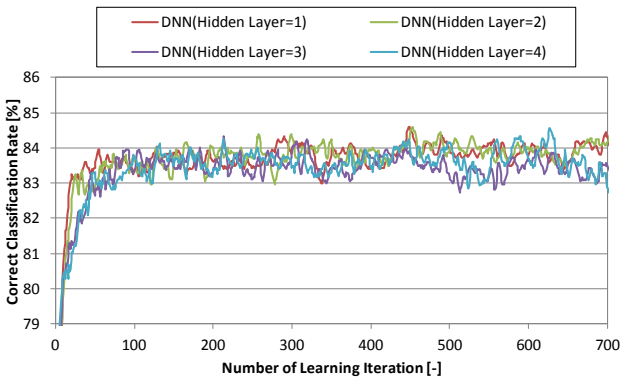
3.2 結果

DNN によるトピックの分類正解率を図 2 に示す．横軸は，最終の隠れ層の学習回数であり，縦軸が，トピックの分類正解率である．

Dropout 法の有無を比較すると，Dropout 法ありの条件のほうが，Dropout 法なしの条件より，高い分類正解率が得られた．隠れ層の層数が同一の条件で比較すると，Dropout 法ありの条件の分類正解率は，Dropout 法なしの条件に比べて 0.5～4.6 ポイント高くなった．



(a) Dropout 法なし



(b) Dropout 法あり　(学習回数 5 回の移動平均)

図 2　DNN を使用した発話トピック分類器の分類正解率

Figure 2　Correct classification rate of DNN-based topic classifier.

隠れ層の層数による分類正解率を比較する．Dropout 法なしの条件では，隠れ層が 2 層のときに分類正解率が最大となり，それ以上の隠れ層を追加した場合には，分類正解率が低下した．一方，Dropout 法ありの条件では，隠れ層の層数が 1, 2 の条件で，ほぼ同等の分類正解率となり，より層数を増やした条件でも，明確な分類正解率の向上は見られなかった．

DNN とベースライン手法との分類正解率を比較した結果を表 4 に示す．ベースライン手法では，最大エントロピー法がもっとも高い分類正解率を示し，83.4 %であった．この正解率は，Blunstom ら[18]の報告値と一致する．最大エントロピー法の分類正解率と，DNN による分類正解率を比較する．Dropout 法を使用した条件では，層数の条件によらず，最大エントロピー法の分類正解率を上回った．DNN によるもっとも高い分類正解率は, 隠れ層が 1 個および 2 個の場合に得られ，84.1 %であった．

表 4　各手法の分類正解率の比較

Table 4　Comparison of correct classification rates given by each topic classification method.

手法		分類正解率 [%]
SVM		80.0
最大エントロピー法		83.4
DNN (Dropoutなし) ※1	隠れ層1	83.0
	隠れ層2	83.6
	隠れ層3	81.0
	隠れ層4	79.2
DNN (Dropoutあり) ※2	隠れ層1	84.1
	隠れ層2	84.1
	隠れ層3	83.5
	隠れ層4	83.8

※1：学習回数1450-1500回の平均値
※2：学習回数650-700回の平均値

4. 考察

本研究の結果, DNN を用いた発話トピック分類器により，従来手法より高い分類正解率を得た．本研究で得られた結果を総括すると，とくに Dropout 法を使用する条件において，分類正解率の向上が見られた．しかし，隠れ層の層数を増やすことによる効果は，小さかった．

ネットワークの多層化による性能向上の程度が低かった原因を考察する．DNN には，入力と出力の関係に強い非線形性があっても，適切に学習できるメリットがある．しかし，今回使用した質問分類のデータセットでは，入力となる素性数が 40,000 以上であったのに対して，学習データ数が約 5,500 と少なかった．このことから，表現力の高い多層の DNN において，過学習が起こったと考えられる．実際に, ベースライン手法として実験した SVM においても，非線形カーネルではなく，より過学習の可能性が低い線形カーネルを使用したときに，もっともよい分類性能が得ら

れた．このことからも，DNN の多層化による性能向上よりも，過学習による性能低下が発生し，明確な性能向上が見られなかったと考えられる．

ただし，DNN による発話トピック分類器では，Dropout 法を適用することにより，従来手法の最大エントロピー法より高い分類正解率が見られた．この原因としては，Dropout 法の過学習を防止する作用が効果的に働いたためだと考えられる．すなわち，通常の学習では，過学習による分類精度の低下が発生するが，Dropout 法を適用することにより，過学習の程度が抑制され，DNN の本来の表現力の高さが生かされることで，分類正解率が向上したと考えられる．

5. まとめ

本研究では，近年注目されている DNN を発話トピック分類技術に適用し，その有効性を検討した．本研究の結論を以下に列挙する．

(1) DNN の学習手法のうち, Dropout 法と識別的 pre-training を適用した発話トピック分類器を開発し，公開されている英語質問文データセットにおいて，トピック分類の精度評価を行った．その結果，DNN による分類正解率は，84.1 %となった．

(2) 従来手法による最大の分類正解率は，最大エントロピー法を使用したときの 83.4 %であった．このことから，DNN による分類正解率は，従来手法を上回った．

(3) DNN の学習手法のうち，Dropout 法を使用することによる効果が高く，Dropout 法を使用しない条件と比べて 0.5～4.6 ポイント高い分類正解率が得られた．

(4) 隠れ層の多層化による分類正解率に対する効果は少なく，隠れ層が 1 層または 2 層の条件において，最大の分類正解率が見られた．

今後は，より効果的な DNN の学習方法の検討や，少ない学習データにおける精度評価を進める予定である．

参考文献

[1] Seneff, S.: Robust parsing for spoken language systems, Proc. of ICASSP, Vol.1, pp.189-192 (1992).
[2] Chu-Carroll, J. and Carpenter, B.: Vector-based natural language call routing, Computational Linguistics, Vol.25, No.3, pp.361-388 (1999).
[3] Wang, Y.-Y., Acero, A., Chelba, C., Frey, B. and Wong, L.: Combination of statistical and rule-based approaches for spoken language understanding, Proc. of ICSLP, pp.609-612 (2002).
[4] Wang, Y.-Y., Lee, J. and A. Acero: Speech utterance classification model training without manual transcriptions, Proc. of ICASSP, pp.553-556 (2006).
[5] Schapire, R.E. and Singer, Y.: BoosTexter: A boosting-based system for text categorization, Machine Learning, Vol.39, No.2/3, pp.135-168 (2000).
[6] Chelba, C., Mahajan, M. and Acero, A.: Speech utterance classification, Proc. of ICASSP, pp.280-283 (2003).
[7] Tur, G., Hakkani-Tür, D., Heck, L.: What is left to be understood in ATIS?, Proc. of IEEE Spoken Language Technology Workshop (SLT), pp.19-24 (2010).
[8] Hinton, G.E., Osindero, S. and Teh, Y.-W.: A fast learning algorithm for deep belief nets, Neural Computation, Vol.18, pp.1527-1554 (2006).
[9] Krizhevsky, A., Sutskever, I. and Hinton, G.E.: ImageNet classification with deep convolutional neural networks, Advances in Neural Information Processing, Vol.25, pp.1097-1105 (2012).
[10] Hinton, G.E., Deng, L., Yu, D., Dahl, G., Mohamed, A.-R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N. and Kingsbury, B.: Deep neural networks for acoustic modeling in speech recognition, IEEE Signal Processing Magazine, Vol.29, No.6, pp.82-97 (2012).
[11] Kanda, N., Takeda, R. and Obuchi, Y.: Elastic spectral distortion for low resource speech recognition with deep neural networks, Proc. of ASRU, pp.309-314 (2013).
[12] Sarikaya, R., Hinton, G.E. and Ramabhadran, B.: Deep belief nets for natural language call-routing, Proc. of ICASSP, pp.5680-5683 (2011).
[13] Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Improving neural networks by preventing co-adaptation of feature detectors, arXiv:1207.0580 (2012).
[14] Seide, F., Li, G., Chen, X. and Yu, D.: Feature engineering in context-dependent deep neural networks for conversational speech transcription, Proc. of ASRU, pp.24-29 (2011).
[15] Li, X. and Roth, D. Learning Question Classifiers, Proc. of COLING (2002).
[16] Senior, A., Heigold, G., Ranzato, M. and Yang, K.: An empirical study of learning rates in deep neural networks for speech recognition, Proc. of ICASSP, pp.6724-6728 (2013).
[17] Bengio, Y., Lamblin, P., Popovici, D. and Larochelle, H.: Greedy layer-wise training of deep networks, Advances in Neural Information Processing Systems, pp.153-160 (2007).
[18] Blunsom, P., Kocik, K. and Curran, J.R.: Question classification with log-linear models, Proc. of ACM SIGIR, pp.615-616 (2006).
[19] Porter, M.F.: An algorithm for suffix stripping, Program, Vol.14, No.3, pp.130-137 (1980).
[20] Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R. and Lin, C.-J.: LIBLINEAR: A library for large linear classification, Journal of Machine Learning Research, Vol.9, pp.1871-1874 (2008).
[21] Okazaki, N.: Classias: a collection of machine-learning algorithms for classification, http://www.chokkan.org/software/classias/ (2009).