Vol.2015-SLP-109 No.13
2015/12/2

対話状態推定のための外部知識ベースを利用した意味的素性の提案

石川 葉子[1]　平岡 拓也[1]　水上 雅博[1]　吉野 幸一郎[1]　Graham Neubig[1]　中村 哲[1]

概要：対話を通して話者の意図を推定することを対話状態推定と呼び，対話システムにとって非常に重要な課題のひとつである．人同士の対話において対話状態推定を行い，その推定精度を比較する Shared Task として Dialog State Tracking Challenge 4 (DSTC4) がある．本研究では，DSTC4 において Long-Short Term Memory (LSTM) を用いた推定器を構築した．また，発話文の表層から得ることが難しい単語の意味的情報を考慮するため，大規模関係データベースである Wikidata を用いて拡張を行った．Wikidata には，単語の同義語や説明文，単語間の関係などの情報が含まれている．これらの情報を用いた結果，対話状態の推定精度が向上した．

1. はじめに

対話システムにとって，対話相手の意図を理解することは非常に重要な課題のひとつである．相手の意図や目的は，対話が進むにつれて徐々に変わる．このように，対話を通して変化する相手の意図を推定することを，対話状態推定と呼ぶ．対話システムが，逐次対話状態を正しく推定することで，対話相手にあわせた適切な行動選択が可能となる．特に相手の発話に対し，複数の返答候補がある場合，相手の意図や目的を正しく把握していれば，対話システムはより正しい返答を行うことが期待できる．

Dialog State Tracking Challenge (DSTC) とは，対話における対話状態を推定し，その推定精度を比較する Shared Task である．DSTC3 までの DSTC では，タスク指向型対話システムと人間の対話ログに対して，人間の対話状態を推定していた．これに対し，今回我々が取り組んだ DSTC4 では，人同士の対話ログに対して対話状態推定を行う．これは，これまでのシステムに対する人間の発話と異なり，人同士の対話における発話を対象とするため，より幅広い言葉遣いやくだけた文法を多用した発話が出現する．そのため，対話状態推定のためのモデル化がより困難になり，これまでの DSTC よりも難しいタスクであるといえる．

本研究では，DSTC4 に取り組むにあたり，(Long-Short Term Memory) LSTM を用いた対話状態推定器を構築し，単語の意味的情報を表す素性の拡張を行った．LSTM を用いた対話状態推定器の構築では，対話中の発話文の分散表現を入力として受け取るベーストライントラッカーを構築する（２節）．そして，発話中にある単語の意味的情報を表す素性の拡張を行う（３節）．ユーザ状態の推定に，外部にある知識表現を利用することが有効であることが示されている [1]．そこで，単語の意味的情報を扱うため，大規模関係データベースである Wikidata を用いた．一般的に，発話文の表層のみから各単語が持つすべての概念的・意味的情報を扱うことは難しい．Wikidata のような関係データベースには，単語の同義語や説明，また単語間の関係などの情報が記述されている．そのため，発話文に含まれている単語のエントリを参照することで，単語の意味的情報を引き出すことができると期待される．外部知識ベースの拡張を行い，意味的情報を用いた結果，対話状態の推定精度が向上したことを実験によって確認した（４節）．

[1] 奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology

2. LSTM を用いた対話状態推定器の構築

過去の DSTC において，Recurrent Neural Networks (RNN) が対話状態推定に有効であることが示されている [2][3]．しかし，RNN は長い系列データを扱う場合，系列が進むにつれて隠れ層に保持されている古い情報が減衰し，長距離の情報が保持できないという問題があった．LSTM は，この減衰の問題を軽減するように隠れ層の入力，忘却，出力を制御したものである [4]．そのため，RNN と比較して，長い系列データの状態推定に対してより効果的であることが期待される．

DSTC4 では，あらかじめ用意されたベースラインとなる推定器がある．このベースラインとなる推定器と，LSTM

による対話状態推定器をあわせ，対話状態推定器を構築した．これを本研究におけるベースライントラッカーとする．

ベースライントラッカーでは，単語と発話内の語順を用いる．一般に対話状態推定では，推定するべき状態に対して訓練データの量が少ないため，単に単語ベクトルを用いるとスパースになりやすい．このスパース性を解決するため，我々は doc2vec[5][6] を用いて単語ベクトルを 300 次元のベクトルに圧縮した．これによって，出力されるベクトルは，単語とその語順の情報を持つ意味ベクトルとなる．これを LSTM の入力とし，出力を対話状態（フレームとそのスロット）とした．

3. 外部知識ベースを用いた拡張

発話文の表層から得ることが難しい単語の意味的情報を考慮するため，大規模関係データベースである Wikidata を用いる．Wikidata とは，Wikipedia の関係データベースを提供するプロジェクトで作成されているデータベースである [7]．Wikidata には，それぞれの単語に対して labels（単語名），aliases（同義語），descriptions（説明文），claims（関連語）などの情報が付与されている．これらの情報を用いることで，単語表層には現れない意味的な素性の拡張が可能である．また，関連語に関しては単語間の関係が複数定義されており，その関係を辿ることで出現した語の抽象化が可能になると期待される．そこで，本研究では，これらの意味的情報を利用した素性を提案する．

素性のラベル名としては，labels（単語名）と aliases（同義語）として定義されている単語を用いる．ラベル名が発話中の単語と一致した場合，そのエントリから生成される素性を発話に対する素性として追加する．なお，ラベル名が発話中の複数の単語と一致した場合は，それぞれの素性の論理和をとった結果を発話に対する素性として追加した．拡張としては，意味タグを参照した拡張と，関連語による拡張を提案する．

3.1 意味タグを参照した拡張

DSTC4 では，発話内の単語に対し，いくつかの意味的情報が意味タグとして付与されている [8]．表 1 にこの意味タグの定義を示す．これらの単語はユーザ発話の意図に関わる重要な単語であり，これらの単語に関する素性を拡張することによって，より効果的に推定器を拡張することができると考えられる．そこで，これらの意味タグを持つ単語が，Wikidata の当該単語エントリにある claims（関連語）と descriptions（説明文）の中に含まれるか調べ，その頻度をベクトル化した．この拡張素性のスロットは，表 1 に掲載する MAIN と SUBCATEGORY の全 44 個である．これらの単語が Wikidata 内の claims（関連語）または descriptions（説明文）に含まれた場合，その単語のスロットを 1 とする．この際，SUBCATEGORY の単語が一致した場合は，その SUBCATEGORY の上位クラスとして定義されている MAIN の単語スロットも 1 とする．これにより抽出される素性の具体例を図 1 に示す．ここでは，Belgium の descriptions(説明文) に country が一致したため，country スロットを 1 とし，その上位クラスの area スロットも 1 としている．

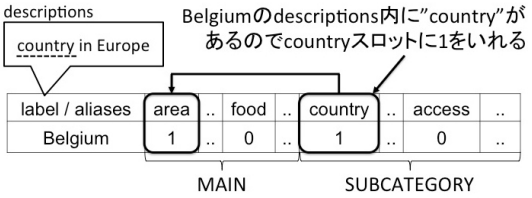

図 1 意味タグを参照した素性

表 1 DSTC4 で定義されている意味タグ

MAIN	SUBCATEGORY
AREA	COUNTRY, CITY, DISTRICT, NEIGHBORHOOD
DET	ACCESS, BELIEF, BUILDING, EVENT, PRICE, NATURE, HISTORY, MEAL, MONUMENT, STROLL, VIEW
FEE	ATTRACTION, SERVICES, PRODUCTS
FOOD	—
LOC	TEMPLE, RESTAURANT, SHOP, CULTURAL, GARDEN, ATTRACTION, HOTEL, WATERSIDE, EDUCATION, ROAD, AIRPORT
TIME	DATE, INTERVAL, START, END, OPEN, CLOSE
TRSP	STATION, TYPE
WEATHER	—

3.2 関連語による拡張

Wikidata 内の claims（関連語）には，各 labels（単語名）に対し，いくつかの関連語が含まれている．単語とその関連語間の関係も併せて定義されており，今回はその関係の種類に注目して素性を拡張した．単語間の関係によって紐づけられた関連語の例を図 2 に示す．例えば，Ramen の claims（関連語）内では，"instance of（部分関係にあるもの）"として Japanese cuisine がある．単語ベクトルを作る際，doc2vec によって意味ベクトルに変換される前と後のそれぞれに追加する 2 通りの素性を試す．

3.2.1 関連語を単語ベクトルに追加

まず，出現語に対して以下の 3 つの関係を持つ関連語の頻度をベクトル化した．Wikidata にある全エントリの claims（関連語）で，以下の関係によって参照された単語を全て書き出し，参照された回数が 1500 回以上の計 287 個の単語名を素性のスロットとして与える．

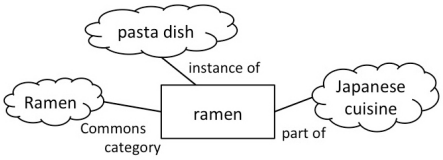

図 2　単語間の関係と関連語の具体例

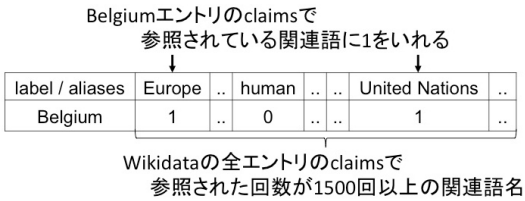

図 3　関連語を単語ベクトルに追加する素性の例

- Commons category（共通カテゴリ）
- part of（部分関係にあるもの）
- instance of（実例）

Wikidata の各エントリの claims（関連語）内に上記で定めた頻出関連語が含まれている場合，その単語のスロットを 1 とする．この素性は，直接素性ベクトルに対して追加する．素性の具体例を図 3 に示す．ここでは，Belgium の claims（関連語）内に頻出関連語の Europe と United Nations が含まれていたため，そのスロットを 1 としている．

3.2.2　関連語を意味ベクトルに追加

1 発話の出現語に対して，Wikidata 内で定義されている関連語を，そのまま発話内容として直接追加する．こちらでは，以下の 7 つの関係に注目し，それぞれの関係で結ばれた関連語を追加する．

- subclass of（サブクラス）
- Commons category（共通カテゴリ）
- capital（首都名）
- continent（所属する地域）
- part of（部分関係にあるもの）
- member of（一員関係）
- topic's main category（トピックのメインカテゴリ）

この素性は発話に直接追加しているため，doc2vec の入力として追加され，意味ベクトルに変換される．

4.　評価実験

提案した素性の評価を行うため，以下の実験を行った．ベースライントラッカーとして，2 節で構築した単語とその意味ベクトルのみから学習した推定器を用いる．ベースライントラッカーに，提案した素性を追加することで推定精度がどの程度向上するかを評価する．

4.1　実験条件・評価指標

今回は各ターンの対話状態に注目し，以下の評価指標を用いて推定精度を評価する．

表 2　ベースライン，Sem. Tags，Sem. Tags + Rel. Words の結果

手法	Accuracy	Precision	Recall	F1
ベースライン	0.0239	0.300	0.287	0.293
+ Sem. Tags	0.0239	0.303	0.292	0.297
+ Sem. Tags + Rel. Words	0.0244	0.301	0.288	0.295

表 3　意味タグを参照した拡張による TRANSPORTATION カテゴリの精度向上

	N	F1
ベースライン	471	0.425
+ Sem. Tags	513	0.524
差分	42	0.0990

- Accuracy：スロット，フレームを含む 1 発話に対する推定器の出力結果の正答率
- Precision：推定器の出力結果のうち正解したスロットの割合
- Recall：正解のうち正しく推定されたスロットの割合
- F-measure：Precision と Recall の調和平均

実験として，以下の拡張された 3 つの推定器を比較した．

(1) 意味タグを参照した素性を拡張 (Sem. Tags)
(2) 意味タグを参照した素性と関連語を単語ベクトルに追加した素性を拡張 (Sem. Tags + Rel. Words)
(3) 関連語を意味ベクトルに追加した素性を拡張

それぞれの実験結果について，次節以降で述べる．

4.2　実験結果

ベースライントラッカーの結果と，意味タグを参照した素性を用いた結果 (Sem. Tags)，意味タグを参照した素性と関連語を意味ベクトルに追加した素性を拡張した結果 (Sem. Tags + Rel. Words) を表 2 に示す．

4.2.1　意味タグを参照した拡張

意味タグを参照した素性を用いた結果 (Sem. Tags)，F 値は約 0.4% の向上が見られた．特に，TRANSPORTATION トピックで精度向上が見られ，そのなかでも今回加えた素性によって TYPE スロットの推定精度が向上した（表 3）．

4.2.2　単語ベクトルとして関連語を用いた素性

4.2.1 に加え，Wikidata における関連語の意味的情報を単語ベクトルとして用いた素性を加えた結果 (Sem. tags + Rel. Words)，F 値は約 0.1% の精度向上に留まった．4.2.1 のときと同様に，TRANSPORTATION トピックの TYPE スロットに効果がみられたが，意味タグのみを用いた場合より精度の向上が小さく，結果として全体の精度向上は 4.2.1 に及ばなかった．

4.2.3　意味ベクトルとして関連語を用いた素性

表 4 に関連語を意味ベクトルに追加した素性を拡張した結果を示す．表 4 から，最も精度向上が見られたのは"topic's main category"であった．以降では，それぞれの素性を加

表 4　関連語を意味ベクトルに追加した拡張の実験結果

	Accuracy	Precision	Recall	F1
subclass of	0.0246	0.300	0.285	0.292
Commons category	0.0179	0.281	0.309	0.294
capital	0.0230	0.303	0.288	0.295
continent	0.0234	0.301	0.291	0.296
part of	0.0249	0.312	0.287	0.299
member of	0.0237	0.314	0.292	0.302
topic's main category	0.0254	0.317	0.292	0.304

えた結果，具体的にどのスロットが変化したか考察する．

SHOPPING トピックに関し，”Commons category”の関係を追加した時のみ約 0.2% 程度のわずかな精度向上がみられた．具体的には，SHOPPING トピックの TYPE_OF_PLACE に関する状態推定の精度が約 3.9% 程度向上した一方で，NEIGHBOURHOOD スロットに関する推定精度は約 2.9% 程度下がっていた．”Commons category”以外の関係を追加した場合，共通して NEIGHBOURHOOD スロットに関する推定精度が約 0.6% 程度下がり，それ以外のスロットは変化していなかった．

FOOD トピックに関し，”continent”と”capital”の関係を追加しても精度に変化は見られず，それ以外の関係については僅かな低下が見られた．特に，”part of”, ”topic's main category”, ”subclass of”の 3 種類の関係を追加した場合，FOOD トピックの推定精度がそれぞれ約 1% 程度下がった．”continent”と”capital”を除く他の関係を追加した場合，FOOD トピックの CUISINE, TYPE_OF_PLACE, PLACE 等のスロットに対し，共通して悪影響があった．一方で，”Common category”の関係を追加した時のみ，FOOD スロットの MEAL_TIME，DISH スロットに対してはそれぞれ 2% と 1% 程度の精度向上が見られた．

ATTRACTION トピックに関しては，”part of”, ”subclass of”, ”topic's main category”の関連語を追加することで約 0.2% 程度の向上が見られた．一方で，”continent”, ”capital”を追加することで 0.1% 程度の悪影響があった．”Commons category”の関連語を追加することで，ATTRACTION トピックの TYPE_OF_PLACE スロットと TIME スロットに対し，それぞれ 9%，7.6% 程度の精度向上が見られたが，一方で PLACE スロットに対しては 3.7% 程度の精度の低下が見られた．全ての関係を追加した結果，共通して ATTRACTION トピックの NEIGHBOUR スロットに悪影響を与えていた．

TRANSPORTATION トピックに関しては，いずれの関係も推定に対し良い影響を与えている．特に，”Commons category”, ”topic's main category”の関連語を追加したことで 4% 程度の向上が，”member of”の関連語を追加したことで 3.2% 程度の向上が見られた．上記の 3 種類の関係を追加することで，TRANSPORTATION トピックの TYPE スロットの推定精度が 10% 以上向上していた．ただし，いくつかのスロットで精度が下がった例もあった．特に，FROM スロットや TO スロットの精度は低下傾向にあった．また，”Commons category”の関連語を追加することで，TICKET スロットの推定精度が 5% 弱下がっている．

ACCOMMODATION トピックでは，”part of”, ”member of”, ”topic's main category”を追加した場合，約 3% 程度の精度向上が見られた．いずれの関係を追加した場合も，ACCOMMODATION トピックの PLACE スロットに関しては精度向上が見られ，特に，上記の 3 種類の関係を追加した場合はいずれも 4% 前後の精度向上があった．一方で，”continent”, ”subclass of”, ”Commons category”の関係を持つ関連語を追加した場合は ACCOMMODATION トピックの推定精度が下がっていた．ACCOMMODATION トピックの NEIGHBOURHOOD スロットに関しては，いずれの関係を追加しても精度の低下が見られた．

5. まとめ

LSTM に基づいた対話状態推定器に外部知識ベースを拡張する手法を提案した．この結果，外部知識ベースを用いた意味的素性を加えることで，いくつかの対話状態の推定精度が上がることがわかった．Wikidata には国名や地名の情報が多く含まれており，特に交通機関に関連した対話状態の推定精度が上がったのではないかと考えられる．

今後の課題としては，複数の関係を組み合わせた結果，対話状態推定の精度にどのような変化がみられるか確認する必要があると考えている．

参考文献

[1] Yi Ma, Paul A. Crook, Ruhi Sarikaya, Eric Fosler-Lussier: Knowledge graph inference for spoken dialog systems. In *Proc. IEEE-ICASSP*, 2015.
[2] Matthew Henderson, Blaise Thomson, and Steve Young: Robust dialog state tracking using delexicalised recurrent neural networks and unsupervised adaptation. In *Proc. IEEE-SLT*, pages 360–365. IEEE, 2014.
[3] Matthew Henderson, Blaise Thomson, and Steve Young: Word-based dialog state tracking with recurrent neural networks. In *Proc. SIGDIAL*, page 292, 2014.
[4] Sepp Hochreiter and Jürgen Schmidhuber: Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
[5] Quoc V Le and Tomas Mikolov: Distributed representations of sentences and documents. In *Proc. ICML*, 2014.
[6] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean: Distributed representations of words and phrases and their compositionality. In *Proc. NIPS*, pages 3111–3119, 2013.
[7] D. Vrandečić and M. Krötzsch: Wikidata: a free collaborative knowledgebase, *Communications of the ACM*, vol. 57, no. 10, pp. 7885, 2014.
[8] Seokwan Kim, Luis Fernando D'Haro, Rafael E. Banchs, Jason Williams, and Matthew Henderson: The Fourth Dialog State Tracking Challenge. In *Proc. IWSDS*, 2016.