NAIST-IS-MT1651108

修士論文

# 外部知識上の推論により構築した特徴量ベクトルを用いた対話状態推定

村瀬　行俊

2018 年 3 月 11 日

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士（工学）授与の要件として提出した修士論文である．

村瀬　行俊

審査委員：

| | | |
|---|---|---|
| 中村　哲 | 教授 | （主指導教員） |
| 松本　裕治 | 教授 | （副指導教官） |
| 吉野　幸一郎 | 助教 | （副指導教官） |

# 外部知識上の推論により構築した特徴量ベクトルを用いた対話状態推定*

村瀬　行俊

内容梗概

対話システムにおいて対話状態推定の精度向上が応答精度の向上に重要であることが知られている．対話状態推定は，ユーザの各発話からそれまでに行った対話の履歴を考慮してユーザの状態を推定するモジュールである．

近年では，機械学習モデルであるニューラルネットワークにより対話状態推定を行うことで精度が向上することが報告されている．機械学習モデルの入力では発話文から得られる特徴量をベクトル化して用いる．ただし，入力文で観測した特徴を用いた単純な特徴量ベクトルではベクトル空間が疎になり，機械学習モデルにより統計値を取れない問題がある．この問題はデータスパスネス問題と呼ばれる．これに対し，発話文で観測した情報を外部の知識により拡張することで多くの情報を持つ特徴量ベクトルの構築ができ，機械学習モデルの入力として用いる際に十分な統計値を取ることが可能になると考えられる．具体的には，外部の知識から発話文で観測した単語のエンティティと関連のあるエンティティを抽出することで情報の拡張を行う．

本研究では，ユーザ発話から得られた観測済み情報を元に外部知識上での推論を行う事で周辺知識からなる特徴量ベクトルを構築する．外部知識には知識ベースである Wikidata を用い，推論にはラベル伝搬法を用いる．構築された特徴量ベクトルを全結合ニューラルネットワークの入力として用いて対話状態推定を行った。また，畳み込みニューラルネットワークと組み合わせたアンサンブルモデルにより精度の向上を確認した．

*奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 修士論文, NAIST-IS-MT1651108, 2018 年 3 月 11 日.

キーワード

タスク指向型対話システム，対話状態推定，DSTC4，知識ベース，知識グラフ，ラベル伝搬法，ニューラルネットワーク

# Dialog State Tracking by Inferring on External Knowledge Feature Vector*

Yukitoshi Murase

**Abstract**

Dialog state tracker is one of the most important modules on task-oriented dialog systems to improve quality of system responses. Dialog state tracker tracks the user intention from each utterance and dialog history.

Recent years, neural network based approaches, which are one of machine learning approaches, improve dialog state trackers. Feature vectors from user utterances are used for input of these machine learning method. However, vector space constituted by simple features to be sparse, and it is difficult for machine learning model to obtain good statistics from them. This phenomenon is called data sparseness problem. Therefore, feature vector creation by information expansion with external knowledge can contain more information than the observed information. Unobserved entities are related to observed entities in external knowledge.

In this research, we propose a creation of a feature vectors by inferring on the external knowledge base. We use Wikidata as external knowledge and label propagation as the inferring method on a graph structure. We use the feature vectors for input of the fully connected neural network (FCNN) that estimates user intention. We integrated the proposed tracker with convolutional neural network (CNN) based dialog state tracker. The ensemble models of FCNN and CNN models improved dialog state tracking results.

**Keywords:**

*Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1651108, March 11, 2018.

Task-oriented dialog system, Dialog state Tracker, DSTC4, Knowledge base, Knowledge graph, Label propagation, Neural network

# 目次

# 図目次

# 第 1 章　緒言

## 1.1　背景

対話システムは非タスク指向型対話システムとタスク指向型対話システムに大別できる．非タスク指向型対話システムでは雑談等のようにドメインに依存しない対話を行う事で，ユーザがシステムとの対話を楽しむ事を目的としたシステムである [1]．一方で，タスク指向型対話システムはユーザが持つ具体的な目標を対話を通じて達成するシステムである [2]．例えば，フライトスケジュールを検索する対話システムでは旅券検索というドメインの下で，「日時」，「出発地」，「目的地」，「航空会社」をユーザとの対話により明確にしながら旅券の検索を行う．検索の結果としてユーザの目的に応じた旅券を見つけることができれば，それを提案をする事でタスクを達成する．タスク指向型対話システムではユーザの目的を達成するため自然言語であるユーザの発話から意図・目的を的確に推定してシステムに不足した情報を得るための応答をすることで，段階的にユーザの目的を明確化する必要がある．そのためユーザの意図・目的を推定するための構成要素である対話状態推定がタスク指向型対話システムに重要である [3, 4, 5]．

タスク指向型対話システムでは，タスクを達成するために必要なドメインの知識としてオントロジ（コンセプトとコンセプト間の関係が記述された集合）が与えられ，ユーザの発話から意図・目的を知識と照合することで対話状態推定をする．オントロジには「飲食店」，「レストラン」，「喫茶店」，「イタリア料理店」など抽象的なコンセプトから具体的なコンセプトを記述している [2]．また，コンセプト間の関係性は「**イタリア料理店**は**レストラン**である」や「**レストラン**は**飲食店**である」のような上位下位概念の関係性を記述している（図 1.1）．このような概念とその関係性を拡張するこで対話状態推定の精度が向上する可能性が考えられる．しかし，オントロジは手作業で構築されることが多く，コンセプトの拡張には膨大な作業量を要するため困難になるという問題がある．この問題に関しては Wikipedia から対話に必要な知識の獲得を自動的に行ったり [16]，外部の知識ベースを用いた対話システムの知識の拡張を提案している [17, 18]．

近年の対話状態推定の取り組みでは，精度の向上を目的としたシェアードタスクである Dialog State Tracking Challenges (DSTCs) が過去五年にわたり開催さ

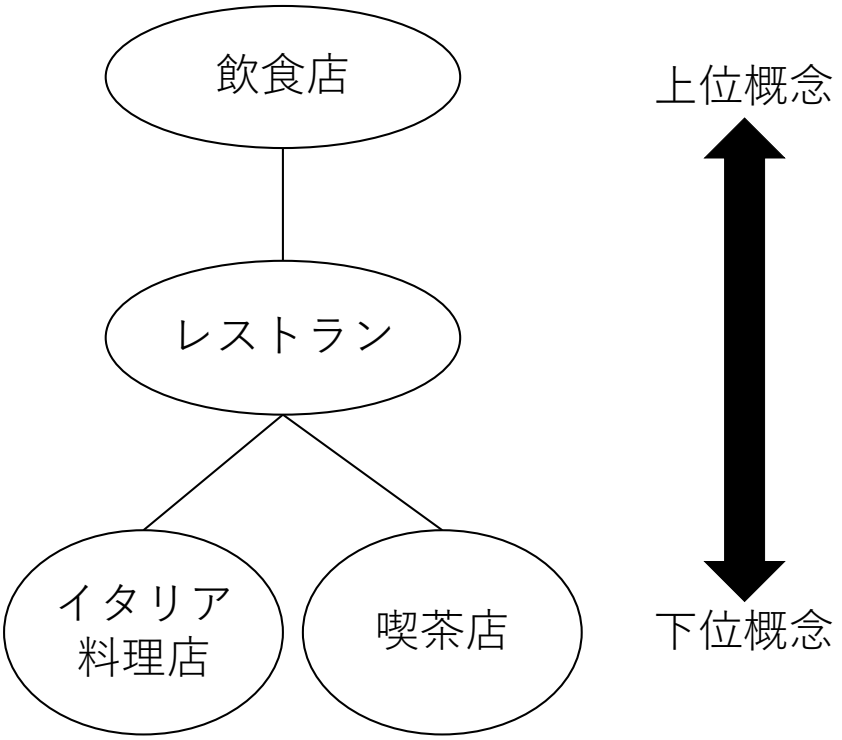


図 1.1　概念関係図

れている [6, 7, 8, 9, 10]．DSTCs を通して対話状態を機械学習による識別的アプローチで推定を行うことが，従来のルールベースや生成モデルに比べ，高い精度を達成している [11]．機械学習による識別的アプローチは，自然言語によるユーザの入力発話を機械が認識できるベクトル表現に変換して入力とする．ベクトル表現の変換方法はいくつかあるが，典型的なものとして bag-of-wards（BoW）があげられる [24]．BoW では文章で出現した単語の頻度をベクトルとして表現しており，各単語の意味を考慮していない．また，対話における発話文では出現する単語数は少なく，頻度も少ないため疎なベクトルになる．疎なベクトルを機械学習の入力として用いても十分に統計値を取ることは難しく，学習が行えない．そのため，単語埋め込み表現を用いたニューラルネットワークによる対話状態推定のアプローチが盛んである．単語埋め込み表現は各単語をベクトルで表現しており，概念の近い単語は発話文の同じ位置に出現するという仮定で構築されている．そのため，単語埋め

込みベクトルでは単語が同じ概念の下にあるなら似たベクトル表現となる．対話状態推定における単語埋め込み表現を用いた機械学習のモデルには再帰型ニューラルネットワーク（RNN）や畳み込みニューラルネットワーク（CNN）がある．再帰型ニューラルネットワークは各発話の単語系列を逐次的に確認することで，発話内における単語の履歴を考慮した特徴抽出により推定を行っている [12, 13]．畳み込みニューラルネットワークでは発話の複数単語の局所的特徴の組み合わせを用いる事で推定を行っている [14, 15]．過去の DSTCs ではこのような方法で対話状態の精度が向上したと報告されている [5]．

## 1.2 目的

本研究では，対話システムの持つ意味的情報の拡張のため外部の知識ベースを用い，知識ベース上での推論による特徴ベクトルを構築する方法を提案する．本稿ではこの特徴量ベクトルを Inference on Knowledge Graph（IKG）と呼ぶ．IKG を識別モデルである全結合ニューラルネットワーク（FCNN）の入力として対話状態推定を行う．また，Dialog State Tracking Challenge 5（DSTC5）における最も精度の高い識別モデルである CNN による対話状態推定も行う．本提案の知識ベース上の概念と概念間の関係性から特徴量の抽出お行う事に対し，単語埋め込みベクトルでは分布仮説を基に統計処理により同じ上位概念を持つ単語ベクトルを構築するため，CNN で用いる特徴量が違うものであると考えられる．このことから，CNN は本提案の対話状態推定とは異なった推定をすると考えられる．また，本提案を用いた推定器と CNN による推定器のアンサンブルにより対話状態推定の結果の相互補完が可能なモデルについても提案する．

本提案の評価には Dialog State Tracking Challenge 4（DSTC4）のデータセットと評価尺度用をいる [10]．知識ベース上の推論により構築した特徴量ベクトルを複数の特徴量ベクトルと比較することで評価を行った．また，アンサンブルモデルを含めた提案した複数のモデルの評価実験も行い，機械学習による最も精度の高い結果（MSIIP） [10, 19] と比較を行った．その結果，知識ベース上の推論により構築した特徴量ベクトルを用いた対話状態推定は MSIIP と同等の精度であることを確認し，アンサンブルモデルは MSIIP の精度を上回った事を確認した．事例分析

では，提案手法により発話文上で関連した語から未観測である対話状態を推定できている事を確認し，アンサンブルによりお互いが推定できていない状態の推定が可能であること確認した．

## 1.3　論文の構成

本論文の構成は以下のとおりである．２章では，一般的なタスク指向型対話システムの構成について述べ，対話状態推定の役割を示す．また，本研究で用いるシェアードタスク（DSTC4）の詳細について紹介する．３章では，統計的機械学習による対話状態推定の従来手法について述べ，統計的処理における従来の特徴量ベクトルの問題を提起する．４章で知識ベースを用いた推論による特徴量ベクトルの作成方法を提案し，５章では複数の特徴量ベクトルととニューラルネットワークを用いた比較実験により推定結果の精度向上の確認と分析を行った．６章では，本論文のまとめと今後の展望について述べる．

# 第 2 章　対話状態推定

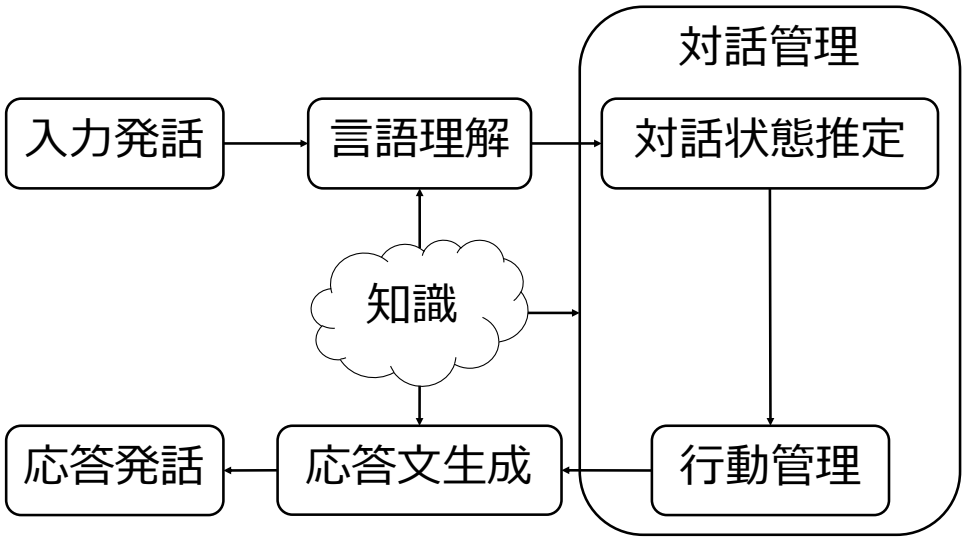


図 2.1　対話システムの概要図 1

本章では，まずタスク指向型対話システムにおける対話状態推定の位置づけを述べ（ 2.1 節），対話状態推定のシェアードタスクである Dialog State Tracking Challenge 4 について紹介する（ 2.2 節）．

## 2.1　タスク指向型対話システムの構成と対話状態推定の役割

典型的なタスク指向型対話システムは，1）言語理解部，2）対話制御部，3）言語生成部，4）オントロジのモジュールから構成される [1, 2, 20]（図 2.1）．1）言語理解部では，話し言葉による入力発話を計算機が扱える表現に変換する．2）対話制御部では，各入力発話における言語理解結果を受け取り，対話履歴として蓄積して対話状態を更新する．次に，対話状態を基に対話システムが行う行動を Policy として決定する．そのため，対話制御部の前半部として対話状態推定と後半部として Policy 選択に細分化される．3) 言語生成部では，対話制御部で選択された

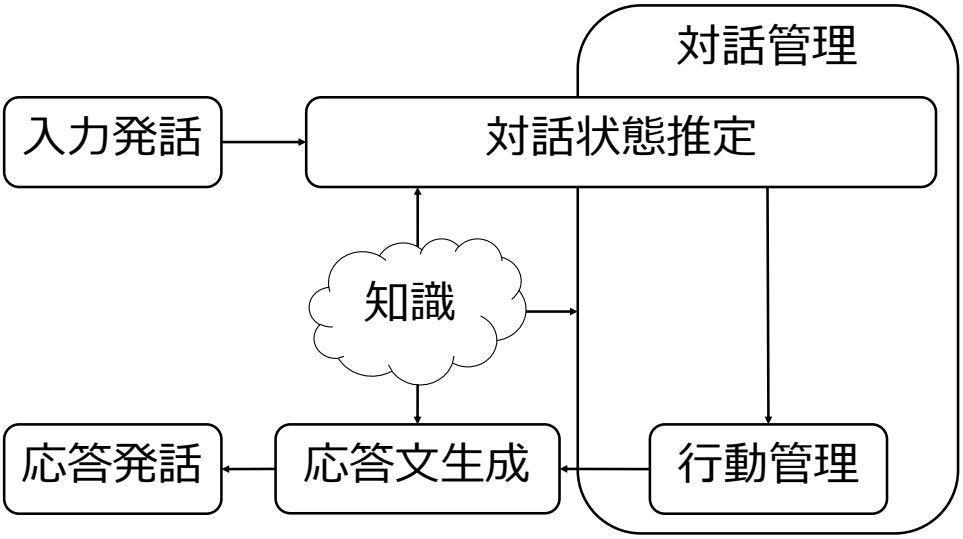


図 2.2　対話システムの概要図 2

Policy を元にどのように話し言葉で応答を生成する．その方法として，テンプレート方式 [21] や言語モデル [22] を用いた方法がある．４）オントロジは対話システムの持つ知識としてデータベース方式で構築されている．対話システムではこれまで述べた３つのモジュールで使用することが想定されている．

以前の対話状態推定では，言語理解部の結果を元に対話履歴を蓄積し対話状態を更新していた．現在では，言語理解部と対話状態推定を組み合わせて対話推定推定モジュール（DST）としている（図 2.2）．対話状態推定モジュールでは入力として発話を受け取り，データ構造化されたフレームを対話状態として計算機が扱える表現に変換する [23]．このフレームを対話状態フレーム（状態フレーム）と呼ぶ．対話状態フレームにドメイン（トピック）を定義することで，扱うことが可能な対話状態（意図・目的）の範囲を制限した物を Slot-Value の組として含有して構成される．対話状態推定は各発話における Slot-Value の組を推定して，新たに追加された Slot-Value を状態フレームを反映することで更新する．そのため，対話状態推定では最初に入力発話からトピックを推定してフレームを定義する．そして，対話を継続することでより多くの情報を入力発話から抽出して Slot-Value の推定・フレー

ムの更新をすることで現在の状態フレームを形成する．例えば旅券検索の場合，フレームには現在対話で扱われているトピックである「旅券予約」を入力発話から抽出してフレームを定義する．また，入力発話から目的地として「アメリカ」という情報を抽出すれば Slot を「目的地」とし，この Slot の Value に「アメリカ」を与える．既出の情報を保持しつつ対話を継続することで，「日時」，「出発地」，「航空会社」などの旅券予約に必要な情報を取得する．最後に，必要十分な情報が得られた時点でユーザの求めている旅券情報を推定し，提示する．

## 2.2 対話状態推定のシェアードタスク

### 2.2.1 Dialog State Tracking Challenge 4

Dialog State Tracking Challenge 4 (DSTC4) では，人間同士の対話ログに対して対話状態推定を行う [10]．これは，人間同士の話し言葉によ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が多く出現する．過去の DSTCs で用いられたタスク指向型対話システムに対する人間の対話ログの対話状態推定と比較すると DST のモデル化がより困難であり，タスクとしてより難しくなっている．

DSTC4 の対話コーパスのドメインはシンガポールにおける旅行対話であり，Skype により 3 人のガイド（*guide-1*，*guide-2*，*guide-3*）と 35 人のツーリストから計 35 対話セッションが合計 21 時間で収録されている．このコーパスはで 31,034 発話・273,580 単語含んでおり，各対話セッションにおいて任意の副対話をセグメンテーションしている．副対話とは対話の一部分を区切ったものである．各副対話で対話状態を含めた複数種類のアノテーションがされており，各副対話のアノテーションは人手により付与され，発話文も人手により書き起こしされている．また，35 対話セッションを 3 つに分割しており，トレーニングセットでは 14 対話（*guide-1* による 7 対話，*guide-2* による 7 対話），開発セットでは 6 対話（*guide-1* による 3 対話，*guide-2* による 3 対話），テストセットでは 15 対話（*guide-1* による 5 対話，*guide-2* による 5 対話，*guide-3* による 5 対話）である．

DSTC4 の対話コーパスは各副対話でアノテーションされており，「話し手」，「BIO」，「トピック」，「各話し手の対話行為」，「Slot-Value の組」がラベルとして付けられている．話し手のラベルは「Guide」と「Tourist」の 2 種類であり，各

表 2.1　副対話区間におけるアノテーション例 [10].

| **Speaker** | **Transcription** | **Annotation** |
|---|---|---|
| Guide | Let's try this one, okay? | **Topic**: Accommodation **GuidAct**: RECOMMEND **TouristAct**: ACK **INFO**: pricerange **Name**: InnCrowd Backpackers Hostel |
| Tourist | Okay. | |
| Guide | It's InnCrowd Backpackers Hostel in Singapore. If you take a dorm bed per person only twenty dollars. If you take a room, it's two single beds at fifty nine dollars. | |
| Tourist | Um. Wow, that's good. | |
| Guide | Yah, the prices are base on per person per bed or dorm. But this one is room. So it should be fifty nine for the two room. So you're actually paying at fifty nine dollars. | |
| Tourist | Oh okay. That's- the price is reasonable actually. It's good. | |

発話がどちらにより発せられたかを定義している．BIO は副対話の最初の発話に「B」，最初の発話以外に「I」，副対話に含まれない発話に「O」を，各発話に対して付与している．トピックのラベルは「Accommodation」，「Attraction」，「Food」，「Shopping」，「Transportation」の 5 種類であり，各副対話における対話がどのトピックに属しているか付与されている．話し手の対話行為は副対話において各発話者がどのような類の行為を意図しているか付与している．例えば，ガイドが何かを推薦している場合にはガイドの対話行為として「Recommend」と付与される．Slot-Value の組は話し手に扱われた重要な内容を副対話に対して記述されている（表 2.1）．

DSTC4 にはアノテーション済みの対話コーパス以外に対話状態推定に用いることができる概念知識としてオントロジがある．オントロジは人手により各トピックに属している Slot-Value が全て記述されている（表 2.2）．例えば，

表 2.2　オントロジの例.

| **Topic** | **Slot** | **Value** |
|---|---|---|
| Accommodation | Name | InnCrowd Backpackers Hostel |
| | | ・・・ |
| | Type | Hotel |
| | | Hostel |
| | | ・・・ |
| | ・・・ | ・・・ |
| Food | ・・・ | ・・・ |
| | | ・・・ |
| ・・・ | ・・・ | ・・・ |

「Accommodation」のトピックでは旅行者が希望する宿泊施設の種類を知る必要がある．そのため，Slot の種類を意味する「Type」，Value には候補となる「Hotel」や「Hostel」等の具体的な情報が記されている．

データやオントロジ以外に DSTC4 ではベースラインシステムとして曖昧性マッチング（ルールベースシステム）による対話状態推定器が提供されている．曖昧性マッチングでは入力発話文で観測された単語とオントロジ内の Value 間の Levenstein 距離を測る．その結果は 0 から 1 で得られ，任意の閾値で発話から推定できる状態フレームを決定する．

### 2.2.2　評価指標

DSTC4 はシェアードタスクとして競うため共通の評価指標がある．評価指標には「*Accuracy*」，「*F-measure*」があり，２つの評価スケジュール（*Schedule1*, *Schedule2*）の下でこれらの累積スコアを得る．*Accuracy* は対話状態フレームが完全一致率であり，Slot-Value の組が完全に一致していることを要求するため，高い精度での回答が比較的困難である（図 2.3）．一方，*F-measure* は各 Slot-Value の一致を *Precision* と *Recall* の調和平均で評価している（図 2.4）．*Precision* は対話

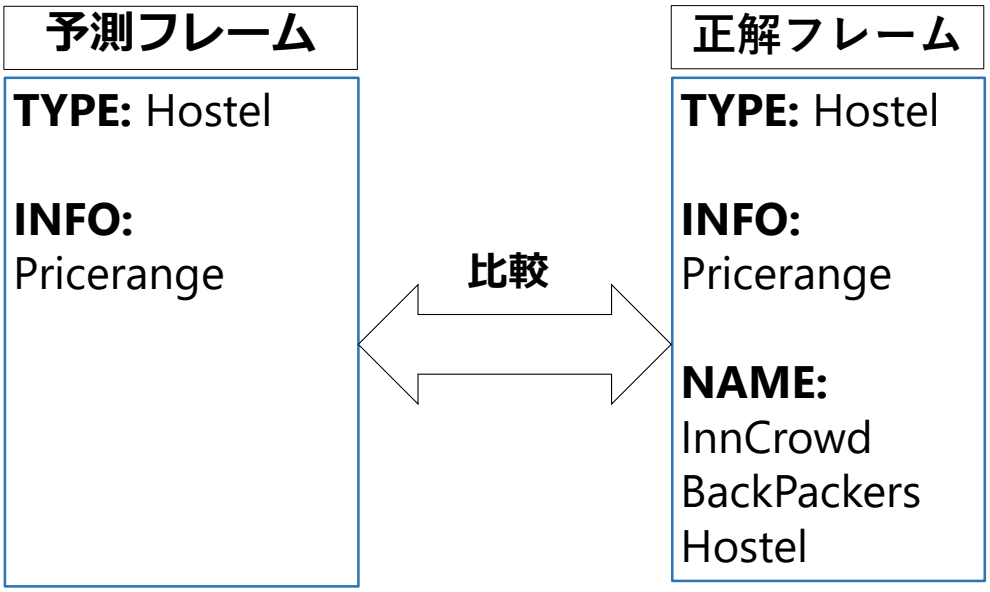


図 2.3　Accuracy の評価指標のイメージ.

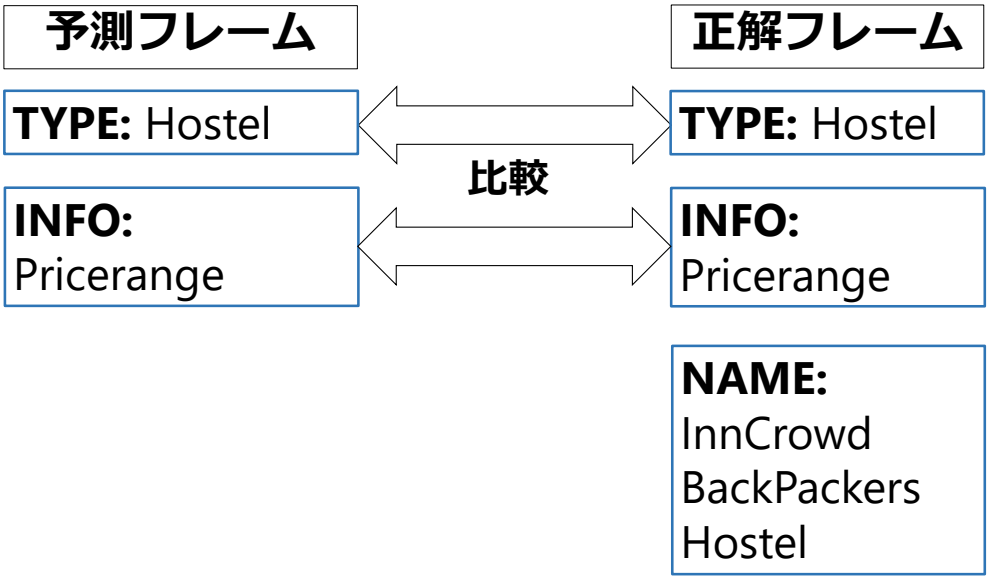


図 2.4　F-measure の評価指標のイメージ (各 Slot-Value を比較)

状態推定器が出力した Slot-Value のうちの正答率であり，余分な Slot-Value を出力すると精度が低くなるため，対話システムが混乱を起こす度合い測ることができる．*Recall* は対話状態推定器正解を当てた割合であり，正答率が低いとシステムに必要な情報が不足するため，システムがタスク達成できなかったり，タスク達成に時間がかかる．*Schedule1* では各副対話の各発話において評価を行う．これは，発話が進むにつれてどの程度理解ができているかを評価する．*Schedule2* では各副対話の最後の発話で評価を行う．これは，最終的にどの程度理解ができているかを評価する．また，副対話の各発話に同じ Slot-Value が付与されているため，前半部では情報抽出できないアノテーションが含まれるており，*Schedule2* は *Schedule1* に比べて精度が高くなる．

# 第 3 章　統計的機械学習による対話状態推定

本章では，まず自然言語処理における機械学習に用いる典型的な特徴量ベクトルの問題点を述べ（ 3.1 節），機械学習による知識ベースを用いた対話状態推定を関連研究として紹介する（ 3.2 節）．また，Dialog State Tracking Challenge 5（DSTC5）において最も精度が高かったモデルを紹介する（ 3.3 節）．

## 3.1　統計的処理における特徴量ベクトルの問題

一般的な機械学習による識別的アプローチでは大量のデータを用いて統計的に分類を行う．自然言語処理の分野では，機械学習に入力する特徴量として入力発話をコンピュータが理解できるベクトル表現に変換する．典型的なベクトル表現では発話上の単語の頻度，連接，共起を用いて構築される．例として bag-of-words（BoW）を取り上げる．BoW は観測しうる単語をベクトルの要素に対応付けをして，入力発話で観測された単語に対応するベクトルの要素として構築される．例えば，入力発話文が「I am planning to travel to Singapore or Malaysia.」であった場合，観測された単語の対応するベクトルの要素を 1 として，観測されていないされていない他の要素を 0 とする（図 3.1）．

このような特徴量ベクトルの問題としてベクトル空間が疎となるデータスパースネス問題がある [24]．これは，データの量が不十分であったり，ベクトルの要素が 0 か極端に小さい場合に表現力が乏しくなり，十分に統計値が取れないという問題である．また，BoW のように単語の頻度，単語の連接，単語の共起のような情報には発話文の表面上の特徴的情報が含まれているが，発話文の各単語における意味的特徴を十分に捉えることは難しいため，もちうる情報量が発話の情報と比べると少なくなってしまう問題がある．

## 3.2　知識グラフ上の推論による推定

Ma 等 [17] は，知識ベースをグラフィカルモデル [25] に変換し，そのモデル上で推論を行うことで対話状態推定をするモデルを提案した．ここでグラフィカルモ

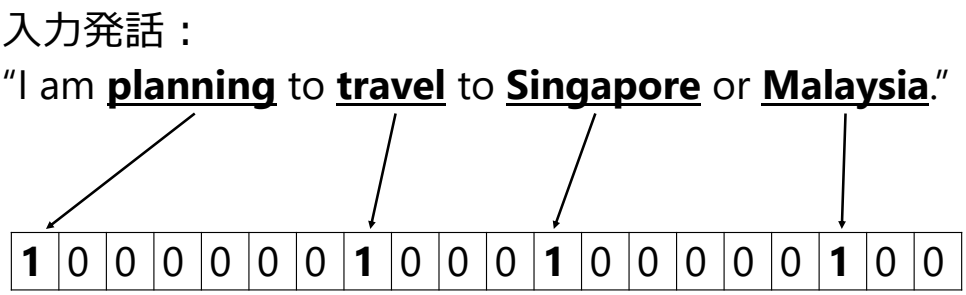


図 3.1　Bag-of-Words のベクトル表現.

デルは知識ベースのコンセプトとその関係性の表現形式をグラフに変換しており，グラフィカルモデルにはマルコフ確率場を用いている．モデル上の推論には任意の手法を用いることができる [17]．このモデルでは，入力発話文で観測した単語を因子とするため，グラフ上の対応するノードの入力値を 1 とし，未観測のノードの入力値を 0 とする．この入力状態からグラフ上で推論を行うことで，観測されたノードの確率が未観測のノードに伝搬し，全てのノードの確率を求める．その結果，因子に近いノードや複数の因子と関連（関連が強い）のあるノードの確率が相対的に高くなり，ユーザの目的に近い状態を推定することできる．例えば，入力発話文に「American」と「Expensive」が観測された場合を図 3.2 のグラフに示す．このグラフ上で推論を行った結果，因子となったノードとより関連の強いノードとて「John Howie Steak House」が得られる．このノードはグラフ上で 2 つの因子と関連を持っており，「Wiild」や「Mc Donald’s」のノードと比べて因子と強い関連を持っていることから高い確率値を得ることとなる．このようにグラフ上で推論を行うことで，観測された単語からグラフ上でより関連の強い意味的情報の抽出が行える．このモデルでは推論した局所的結果のみを用いているが，本研究では推論した結果の全てを用いることで各エンティティノードのスコアの組み合わせにより知識全体を表現した特徴量ベクトルの構築を行う．

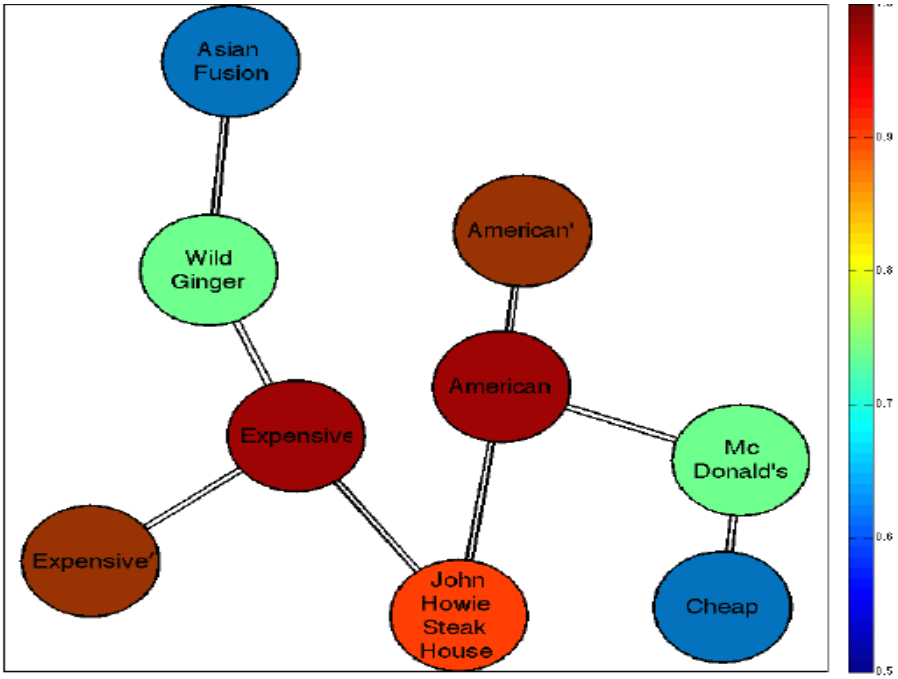


図 3.2　マルコフ確率場上の推論結果　(ヒートマップ) [17].

## 3.3　畳み込みニューラルネットワークによる対話状態推定

近年の対話状態推定のタスクである DSTC5 で最も精度が高かった推定器が畳み込みニューラルネットワーク（CNN）を用いたものである [14]．このモデルは自然言語処理における文章分類のタスクで高精度であることが示されており [26]，対話状態推定を識別モデルによる分類問題として解くことができる [27]．CNN では単語数を $n$ 個含む入力発話 ($S \in \mathbb{R}^n$) の単語ごとに学習済み Word2Vec [30] により任意の $k$ 次元の単語埋め込みベクトル ($emb_i \in \mathbb{R}^k$) を構築する（図 3.3）．この単語埋め込みベクトルを重ねることで，行で発話，列で単語を表現した行列を構築して入力とする．この入力を元に畳み込み層（Convolutional Layer）では単語埋め込み表現から局所特徴を得る．図 3.3 では各単語埋め込み表現（$emb_i$）から局所特徴を得ているが，フィルタ高を変えることで複数の単語埋め込み表現（例：$emb_1$ と $emb_2$）から局所特徴を得ることも可能である．この局所特徴値を集約するのが最大値プーリング（Max-poolong）であり，最大値のみを用いている．このように

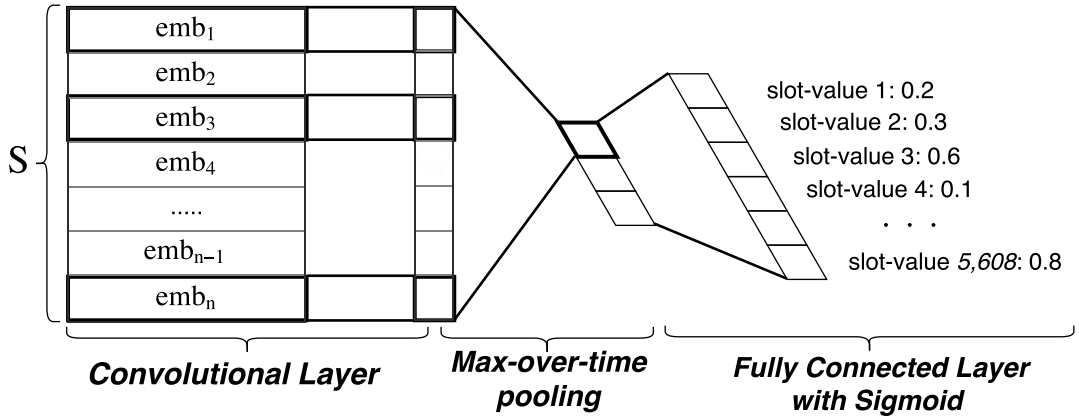


図 3.3　CNN モデルの概要図.

得られた文の特徴を用いて，CNN では対話状態推定の学習を行う.

# 第 4 章　知識グラフ上での推論による特徴量ベクトル作成の提案

本章では，まず知識ベースとグラフ構造の関係性を述べ，本研究で用いる Wikidata を紹介する（ 4.1 節）．次に，Wikidata からどのようにグラフを構築したかを述べ（ 4.2 節），特徴量ベクトルを構築するために必要なグラフ上の推論方法であるラベル伝搬法を紹介する（ 4.3 節）．また，対話の系列的情報（対話履歴）を知識グラフの推論時に考慮する方法を述べ（ 4.4 節），主成分分析（PCA）による特徴量ベクトルの次元圧縮をする理由と累積寄与率の設定を紹介する（ 4.5 節）．

## 4.1　知識ベース

### 4.1.1　知識ベースと知識グラフ

知識ベースとは，コンピュータで処理可能な形に物の概念と，これらの概念的関係性を記述したデータベースである．知識ベースは，2 つのエンティティ（物の概念）とプロパティ（関係性）の「三項関係」を一つの知識として持っており，それらの集合を構成要素として構築されている．例えば，"Singapore"と"Asia"には「Singapore *is Part of* Asia」という関係があり，"Singapore"と"City"には「Singapore *is instance of* City」という関係が記述されている (表 4.1)．

このように三項関係の集合を持つ知識ベースはグラフ構造（ネットワーク構造）に変換ができ，本研究では知識ベースを変換したグラフ構造を知識グラフと呼ぶ．知識グラフは基本的にノードとエッジにより構成されており，ノードにより概念を表現することができ，エッジにより概念間に何の関係性を表現することができる．つまり，知識ベースのエンティティをノード，プロパティをエッジとすることで知識グラフに変換できる．例えば，"Singapore"，"Asia"，"City"の関係性をグラフ構造で表現すると図 4.1 のようになる．

表 4.1　知識ベースの例.

| エンティティ 1 | プロパティ | エンティティ 2 |
|---|---|---|
| Singapore | Part of | Asia |
| | Instance of | Country |

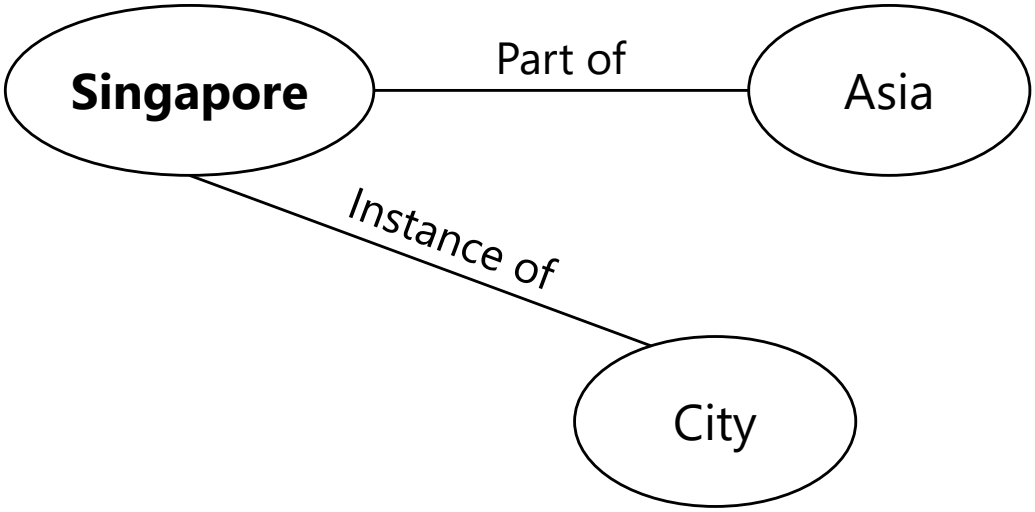


図 4.1　三項関係のグラフ化例.

### 4.1.2　Wikidata

DSTC4 で与えられている知識として Slot-Value の関係性が記述されたオントロジーを紹介したが，知識量としては制限されているため人手によりオントロジーの情報の拡張をすることは難しい．そのため，外部知識として知識ベースである Wikidata を用いることで，知識の自動拡張により意味的情報の拡張を行う．Wikidata とは，Wikipedia の関係データベースを提供するプロジェクトとして構築された知識ベースである [28]．Wikidata では，通常の知識ベース同様，エンティティとプロパティにより構成されており，エンティティに名前として「label」が与えられている．また，多数の言語に対応しているのも Wikidata の特徴である．本研究では英語による対話に焦点を当てるが，他言語へ適用する際には非常に容易で

あるため Wikidata を使用した．

## 4.2　知識ベースからの知識グラフの構築

Wikidata はエンティティをノード，プロパティをエッジとしたグラフ構造に変換可能である．しかし，Wikidata そのまま用いようとすると，エンティティの数が膨大すぎるため実時間での推論が難しい．そこで，今回は以下の手順でノードとエッジを削減して知識グラフを構築した．まず，対話状態推定タスク・ドメインにおける学習データと開発データに出現する単語に対応するノードをシードとして残す．今回は単語を抽出するために各発話文に対して単語を分割し，その単語とエンティティの名前が完全一致した場合のみ用いた．これによりドメイン内で必要最低限のエンティティが残り，ここで得られたノードをシードノードと呼ぶ．次に，A や The などのストップワードはストップワード辞書を用いて取り除いた．これらの単語が対応するエンティティの数は多く，対話状態推定では対話の内容を推定することが目的であるため，対話の内容に強いかかわりがないと考えた．また，「！」や「？」のような記号も考慮する必要はないので取り除いた．こうして構築したシードノードとそのエッジをもとに，シードノードから 1-hop のノードとそれらの間のエッジをすべてグラフ構造に残した．これらの隣接ノードは各発話文では未観測の単語だが，発話文で観測された単語との関係性があると推定され，推論時に考慮可能となる．同じ名前を持った別のエンティティについては，一意に特定できる名前を付けて別のノードとして扱った．図 4.2 に示す例は上記の手順でシードノードとして観測された"Singapore"と"Malaysia"を元に，隣接ノードと関係性のエッジを付与した結果である．なお，今回はこの関係性のラベルは使用しない．

## 4.3　知識グラフ上でのラベル伝搬法による推論

このように構築した部分知識グラフ上での推論にラベル伝搬法を用いて関連した情報の抽出を行う [29]．ラベル伝搬法ではグラフ上のいくつかのノードのラベルが与えられた際に，クラス未知のノードのクラスラベルを予測する．クラスラベルはグラフ上の隣接したノードは同じクラスに属する傾向があるという仮定のもと予

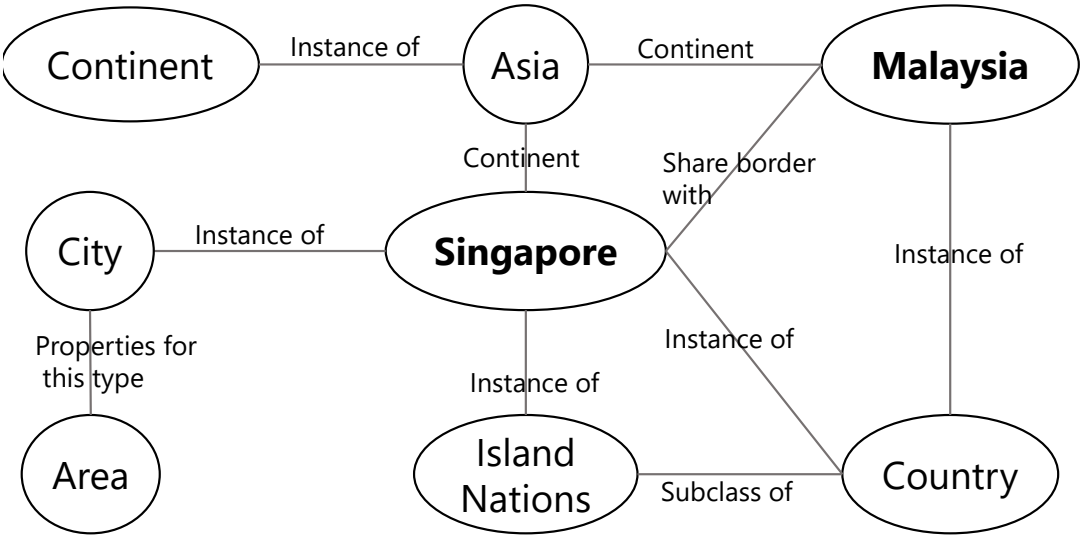


図 4.2　Wikidata から構築した部分グラフ．

測される．本研究では複数のエンティティが発話文で観測されたとき，それ以外を観測されていない状態としてラベル伝搬法で潜在的なエンティティとして予測する．まず $N \times N$ 行列の $\mathbf{W}$ はグラフ上のノード間のエッジの有無を表しているとする，$N$ はグラフ上にあるノードの数である．入力値となる $y_i$ はある単語が発話文中で観測されたかを 0 か 1 で表している．この例を図 4.3 で矢印の左の値を入力値として示す．出力値となる $f_i$ は，入力値で観測されたノードとそれに近いノードの値がそれぞれ近い値を取るように出力される．よって $f_i$ が観測されたノードに近い値をとれば潜在的に観測されたとして観測されたノードと同じクラスに属すると定義する．このラベル伝搬法の最小化問題は式（4.3.1） として定式化される．

$$J(f) = \sum_{i=1}^{n}(y_i - f_i)^2 + \lambda \sum_{i<j} w_{i,j}(f_i - f_j)^2 \tag{4.3.1}$$

式（4.3.1）の第 1 項は入力値と予測値をできるだけ近くする働きを持っており，第 2 項は隣り合ったノードの予測値を近づける働きを持っている．$\lambda$ は第 1 項と第 2 項のバランスをとるための係数である．式（4.3.1）は，

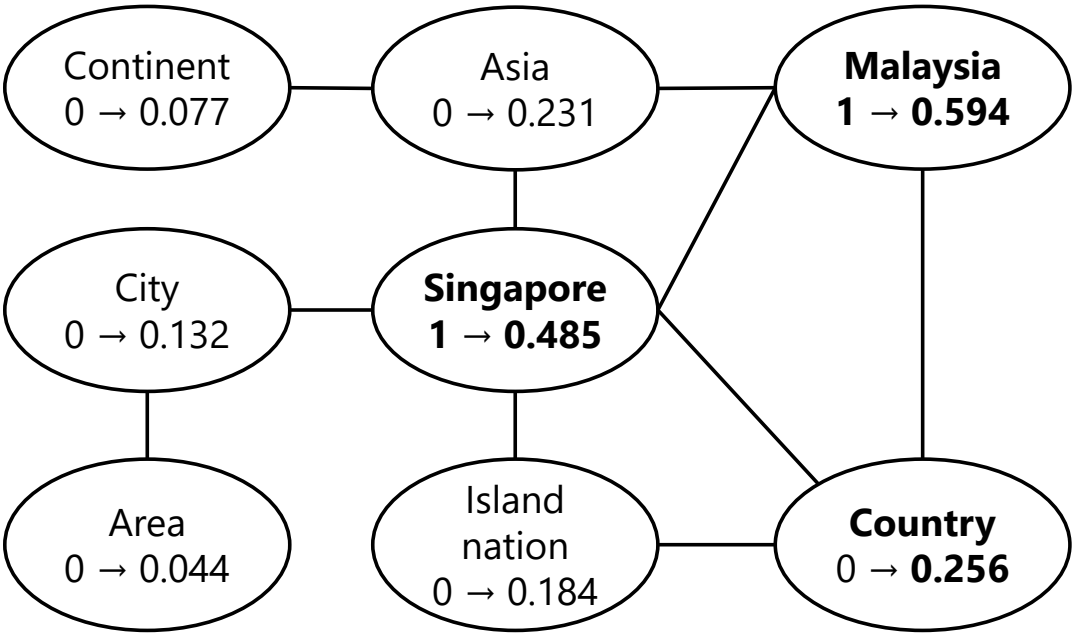


図 4.3　ラベル伝搬法によい推論したグラフ.

$$J(f) = \| \mathbf{y} - \mathbf{f} \|_{\mathbf{2}}^{\mathbf{2}} + \lambda \mathbf{f}^{\mathbf{T}} \mathbf{L} \mathbf{f} \tag{4.3.2}$$

と書き換えることができ，この最小化問題の解として解くために

$$(\mathbf{I} + \lambda \mathbf{L})\mathbf{f} = \mathbf{y} \tag{4.3.3}$$

を用いる．式（4.3.2）と式（4.3.3）の $\mathbf{L} = \mathbf{D} - \mathbf{W}$ はラプラシアン行列である．なお，$\mathbf{D}$ は対角行列の成分であり，各行の総和を要素とした行列である．本研究で必要になるのは式（4.3.3） の予測値である $\mathbf{f}$ であるため，式変形を行い以下の式で予測値を求めた.

$$\mathbf{f} = (\mathbf{I} + \lambda \mathbf{L})^{-\mathbf{1}} \mathbf{y} \tag{4.3.4}$$

これによって図 4.3 の入力値に対し，出力である $\mathbf{f}$ の値の結果を矢印の右に示す.

**Algorithm 1** ラベル伝搬法の割引率.

**Require:** $\lambda > 0$, $0 \leq \gamma \leq 1$, $i = index$ and $t = time$

```
if 最初の発話文 then
  for 単語リスト内の y_{i,t} do
    y_{i,t} = 1
  end for
else
  二発話目以降：
  for y_{i,t} do
    if 単語リスト内の y_{i,t} then
      y_{i,t} = 1 + γ y_{i,t-1}
    else
      y_{i,t} = γ y_{i,t-1}
    end if
  end for
end if
f = y(I + λL)^{-1}
return f
```

## 4.4 知識グラフ上での対話履歴の考慮

節 4.1 の手法では各発話文ごとに観測された単語のみを用いた推論を行っており，継続的に行われている対話の履歴を考慮していない．そこで，対話履歴中に含まれる単語についても同様にラベル伝搬法のシードとして扱うこととした．ただ，対話の内容は経過に伴って変化するので古い発話の影響は小さくなる．そのため入力値に割引率 $\gamma(0 \leq \gamma \leq 1)$ を取り入れた．一つ前の発話中に存在する単語を出現単語とし，そのラベルの値に割引率 $\gamma$ を掛けあわせて，現在の発話文の入力値と足しすことにより新たな入力値とする．このプロセスを Algorithm 1 に示す.

## 4.5　特徴量ベクトルの次元圧縮

本研究で提案する特徴量ベクトルでは，次元数が大きくなり，機械学習の学習時間が長くなるため，主成分分析により次元圧縮を試みた．この特徴量ベクトルは，Wikidata から知識グラフをドメインに適したエンティティとその 1-hop 先までを含み，推論（ラベル伝搬法）により全てのノードが任意の値を取る．このように推定した特徴量ベクトルは知識グラフのノードの数が特徴量ベクトルの要素数（次元数）となるため，モデルへの入力ベクトルのサイズが大きくなる．また，主成分分析を用いる際に特徴量ベクトルが持っている情報量を減らさないために，累積寄与率が 1 を保っているサイズまでの圧縮を行った．

# 第 5 章　特徴量ベクトルをニューラルネットワークに用いた実験

本章では，複数のモデルを用いて 2 通りの実験を DSTC4 のコーパスを用いて行う．1 つ目のモデルは提案法による特徴量ベクトルを全結合ニューラルネットワーク（FCNN）の入力として用いるモデルを紹介する (5.1 節)．また，畳み込みニューラルネットワーク（CNN）による対話状態推定のモデルについても紹介をする（5.2 節）．5.3 節では FCNN と CNN を用いた 2 通りのアンサンブルモデルの構築方法を紹介する．これら 4 つのモデルを用いて実験を行う．実験 1 では，本提案の特徴量ベクトルによる対話状態推定が有効であることを FCNN モデルにより検証を行う（5.4.1 節）．この実験では DSTC4 のベースラインシステムの結果と比較を行い，典型的な特徴量ベクトルである Bag-of-Words（BoW）を FCNN の入力として用いた対話状態推定の結果と単語埋め込み表現である Word2Vec（W2V）を入力とした結果との精度の比較を行う．分析と考察では意味的情報の拡張と推論により入力発話文で観測した単語から未観測の状態を推定しているか確認する．実験 2 では，前述の 4 つのモデルによる推定精度の差を検証し，アンサンブルモデルにより精度が向上するか検証する（5.4.3 節）．また，精度の比較にベースラインとして DSTC4 で CNN による推定をしたモデル [15] と機械学習を用いた最高精度のモデル [31] の結果を用いる．この実験の分析と考察では，FCNN と CNN からアンサンブルをすることで推定する対話状態に変化があるか確認する．

## 5.1　全結合ニューラルネットワークと特徴量ベクトルによる対話状態推定モデル

本研究では，提案法の特徴量ベクトルを全結合ニューラルネットワーク（FCNN）の入力として用いた実験を行った．このニューラルネットワークの構造は，全結合層の三層で構成されており，中間層は入力層と同じサイズである．活性化関数には Rectified Linear Unit を用いており，出力層は全ての Slot-Value の組の数である 5,608 個の確率を同時に推定するため，出力層にはシグモイド関数を用いる．誤差を最小化するための optimizer には Adam optimizer を使用した．また，Weight

Decay と，各層においてドロップアウトとバッチ正規化を行った．出力層の誤差の計算ではシグモイドクロスエントロピーを用いている．このモデルを用いて各入力発話から対話状態推定を行い，副対話が終了するまで状態を積み上げることで対話状態推定の更新を行う．

## 5.2 畳み込みニューラルネットワークと Word2Vec による対話状態推定モデル

畳み込みニューラルネットワーク（CNN）では提案法と違う特徴量抽出をしているため，提案法との比較やアンサンブルを目的として実験を行った．CNN の入力には Wikipedia による学習済みの 200 次元の Word2Vec による単語埋め込みベクトルを用いた．

$$S = (emb_1, emb_2, ..., emb_n) \tag{5.2.1}$$

ここで，$S$ は発話を表現しており $S \in \mathbb{R}^{n \times k}$ であり，含んでいる単語数分の単語埋め込みベクトル $emb_i \in \mathbb{R}^{k \times 1}$ を持っている．畳み込み層のフィルターは $m \in \mathbb{R}^{d \times k}$ であり，特徴マップは $c \in \mathbb{R}^{n-d+1}$ である．ここで k は単語埋め込みベクトルの次元数であり，d はフィルタの高さである．

$$c = f(m * s + b) \tag{5.2.2}$$

ここで，活性化関数 $f$ には ReLU を用いている．プーリング層では Max-over-time pooling を用いており，$\hat{c} = max\{c\}$ となるように特徴マップから最大の値をとる．プーリング後には全結合層とシグモイド関数により CNN の出力として $y_{cnn}$ を得る．

$$y_{cnn} = \sigma(\hat{c} * w + b) \tag{5.2.3}$$

その他の設定には FCNN で用いた設定を使用している．

## 5.3 フィードフォワードニューラルネットワークと畳み込みニューラルネットワークのアンサンブルによる推定

本節では，本提案の特徴量ベクトルを用いた FCNN と Wikipedia により学習済みの Word2Vec を CNN のアンサブルを 2 通りの方法で行い，これら 2 つのモデルを合わせて計 4 つのモデルにより実験を行った．

### 5.3.1 出力結果の線形補間によるアンサンブル

1 つ目のアンサンブルモデル（Ensemble-1）では，FCNN と CNN の出力結果の線形補間により対話状態推定を行う．

$$\mathbf{y}_{ensemble_1} = \mathbf{y}_{fcnn} \times w_{fcnn} + \mathbf{y}_{cnn} \times w_{cnn} \tag{5.3.1}$$

式（5.3.1）の $w_{fcnn}$ と $w_{cnn}$ は $0 \leq w_{fcnn}, w_{cnn} \leq 1$，$w_{fcnn} + w_{cnn} = 1$ を満たす変数であり，2 つのモデルの出力を調整する重み付けの働きをしている．$w_{fcnn}$ と $w_{cnn}$ のうち，どちらか一方の重みが 0.5 より大きいときは識別タスクを行う上でその重みに対応するモデルの出力がより重視されていると考えることができる．各モデルの特徴量は違うため，このように複数の重みの組み合わせで評価することで，どちらの特徴量を重視することがより精度の向上に寄与したかを確認することができる．

### 5.3.2 1 つのモデルに統一化したアンサンブル

もう一方のアンサンブルモデル（Ensemble-2）では，2 つの特徴量を 1 つのモデルの入力として学習を行う．このモデルでは，FCNN の中間層と，CNN のプーリング後に活性化関数を使用した後の層を連結，2 つの特徴量を考慮した中間層とすることでモデルの学習を行った．

$$\mathbf{h}_{ensemble_2} = \mathbf{h}_{fcnn} \otimes ReLU(\hat{c}_{cnn} + b), \tag{5.3.2}$$

また，連結した $\mathbf{h}_{ensemble_2}$ をシグモイド関数の入力とすることで出力層を構築した．

$$\mathbf{y}_{ensemble_2} = \sigma(\mathbf{h}_{ensemble_2} * w + b). \tag{5.3.3}$$

CNN 同様，その他の設定は FFNN で使用したものを用いた．

## 5.4 DSTC4 による実験

本節では，2 つの実験を DSTC4 のコーパス，タスク，ベースラインシステムを用いて行う．実験 1 では特徴量ベクトルのハイパーパラメータの組み合わせにより対話状態推定の精度にどのような差があるか，また，その他の特徴量ベクトルを同じ識別モデルに使用した場合との精度の差を検証する．提案法の特徴量ベクトルは発話内で未知の概念の推論結果を含むため，分析では発話内で観測済みの単語から未知の Value を状態として推定できているか確認する．実験 2 では FCNN と CNN のアンサンブルによる精度の変化を検証し，各モデルで推定できなかった状態を違う特徴量から推定可能であるか確認する．

### 5.4.1 実験 1 の比較

実験 1 では，提案した特徴量ベクトルのハイパーパラメータの組み合わせを比較し，より良いハイパーパラメータの組み合わせを FCNN による対話状態推定の結果から推定する．これにより最も精度の高いハイパーパラメータの組による特徴量ベクトルの結果と，その他の特徴量ベクトルと，ベースラインシステムとの比較をする．比較する特徴量ベクトルには Bag-of-Words (Bow) と，発話内の単語の埋め込みベクトルを線形補間した Word2Vec と BoW（BoW w/ W2V）を FCNN に入力として対話状態推定を行った結果を比較する．また，この実験では提案する特徴量ベクトルを次元圧縮せずに用いている．

特徴量ベクトルのハイパーパラメータでは，ラベル伝搬法の第 1 項と第 2 項のバランスをとる $\lambda$ の値を設定し，対話履歴をどの程度引き継ぐか決める割引率 $\gamma$ を設定する．これに加えて，FCNN モデルの出力の閾値（$\tau$）の設定をハイパーパラメータとして設定する．表 5.1 が本実験に用いる各ハイパーパラメータのリストで

表 5.1　各ハイパーパラメータの値（合計 432 組）

| $\lambda$ | 割引率 ($\gamma$) | 閾値 ($\tau$) |
|---|---|---|
| 0.5 | 0 | 0.1 |
| 1 | 0.125 | 0.2 |
| 1.5 | 0.25 | 0.3 |
| 2 | 0.5 | 0.4 |
| 3 | 0.7 | 0.5 |
| 8 | 0.8 | 0.6 |
| | 0.9 | 0.7 |
| | 1 | 0.8 |
| | | 0.9 |

あり，これらの全ての組み合わせにより実験を行った．

$\lambda = 1$ である時に 1 つの項に同等のバランスを取るため，第 1 項を重視するように $\lambda < 1$ となる値を設けた．また，第 2 項を重視するように $\lambda > 1$ となる値を複数設けて，極端に第 2 項を重視した場合の検証のために $\lambda = 8$ を設けた．割引率では，$\gamma = 1$ は副対話区間の全ての履歴を考慮し，$\gamma = 0$ では一切の対話履歴を考慮していないと考えられるため，これら 2 つの割引率を軸として考えた．また，対話履歴の考慮は対話を理解する上で重要と考えられるため，割引率の高い $0.7 \leq \gamma \leq 0.9$ の間では 0.1 刻みと細かく設定をして，$\gamma < 0.7$ である場合の刻みを大きくした．このように変則的な刻みを設定した理由としては，全ての発話でラベル伝搬法を用いるには時間がかかるためハイパーパラメータの組み合わせの数を減らしたかったからである．FFNN モデルの出力値に対する閾値の設定は $0.1 \leq \tau \leq 0.9$ を 0.1 刻みにすることで推定結果の変化を分析する．

実験の結果（表 5.2〜表 5.5），テストによる推定精度では *Schedule1* と *Schedule2* おいて *Accuracy* と *F-measure* の上位の組み合わせで割引率は $\gamma = 1$ となった．このことから，対話履歴を強く考慮することで DSTC4 における対話状態推定の精度がより高くなることが分かった．$\lambda$ は，*F-measure* において様々な値が上位になるため，Slot-Value の正答率に大きな影響を与えないと考えられる．*Accuracy* では

表 5.2　Schedule1 の Accuracy

| $\lambda$ | 割引率 ($\gamma$) | 閾値 ($\tau$) | **Accuracy** |
|---|---|---|---|
| 0.5 | 1.0 | 0.3 | 0.0490 |
| 0.5 | 1.0 | 0.2 | 0.0481 |
| 0.5 | 1.0 | 0.4 | 0.0456 |
| 1 | 1.0 | 0.3 | 0.0452 |
| 3 | 1.0 | 0.3 | 0.0427 |
| baseline | | | 0.0374 |

表 5.3　Schedule2 の Accuracy

| $\lambda$ | 割引率 ($\gamma$) | 閾値 ($\tau$) | **Accuracy** |
|---|---|---|---|
| 0.5 | 1.0 | 0.3 | 0.0559 |
| 0.5 | 1.0 | 0.2 | 0.0559 |
| 0.5 | 1.0 | 0.4 | 0.0549 |
| 0.5 | 1.0 | 0.5 | 0.0521 |
| 3 | 1.0 | 0.3 | 0.0502 |
| baseline | | | 0.0488 |

表 5.4　Schedule1 の F-measure

| $\lambda$ | 割引率 ($\gamma$) | 閾値 ($\tau$) | **F-measure** |
|---|---|---|---|
| 0.5 | 1.0 | 0.2 | 0.3444 |
| 3 | 1.0 | 0.2 | 0.3397 |
| 1 | 1.0 | 0.2 | 0.3391 |
| 8 | 1.0 | 0.2 | 0.3381 |
| 2 | 1.0 | 0.2 | 0.3371 |
| baseline | | | 0.2506 |

表 5.5　Schedule2 の F-measure

| $\lambda$ | 割引率 ($\gamma$) | 閾値 ($\tau$) | **F-measure** |
|---|---|---|---|
| 1 | 1.0 | 0.2 | 0.3759 |
| 3 | 1.0 | 0.2 | 0.3763 |
| 8 | 1.0 | 0.2 | 0.3754 |
| 0.5 | 1.0 | 0.2 | 0.3750 |
| 2 | 1.0 | 0.2 | 0.3727 |
| baseline | | | 0.3014 |

表 5.6　Schedule1 の F-measure による特徴量の比較表

| | **Baseline** | **BoW** | **BoW w/ W2V** | 提案法 |
|---|---|---|---|---|
| **Accuracy** | 0.0374 | 0.0036 | 0.0097 | 0.0389 |
| **Precision** | 0.3589 | 0.0099 | 0.2850 | 0.4445 |
| **Recall** | 0.1925 | 0.0440 | 0.0659 | 0.2749 |
| **F-measure** | 0.2506 | 0.0163 | 0.1070 | 0.3397 |

表 5.7　Schedule2 の F-measure による特徴量の比較表

| | **Baseline** | **BoW** | **BoW w/ W2V** | 提案法 |
|---|---|---|---|---|
| **Accuracy** | 0.0488 | 0.0066 | 0.0132 | 0.0502 |
| **Precision** | 0.3750 | 0.0093 | 0.2563 | 0.4596 |
| **Recall** | 0.2519 | 0.5335 | 0.0736 | 0.3186 |
| **F-measure** | 0.3014 | 0.0159 | 0.1143 | 0.3763 |

$\lambda = 0.5$ が *Schedule1* と *Schedule2* 共に高い精度を出している．閾値は *Accuracy* においてばらつきがあるが，低い閾値の精度が高い傾向がみられる，*F-measure* においては全てにおいて 0.2 であるため，ほとんどの場合は 0.2 程度であることが望ましいと考えられる．

ベースラインシステムとその他の特徴量ベクトルとの比較実験の結果，本提案手

法の *Schedule1* と *Schedule2* における *Accuracy* と *F-measure* 共にの精度が最も高かった（表 5.6）．表 5.6 では提案法の特徴量ベクトルのハイパーパラメータの組み合わせを $\lambda = 1$ と $\gamma = 1$ として用いている．*Accuracy* は状態フレームが完全一致しなければいけない非常に難しい指標となっているため，*F-measure* による結果の考察を行う．両方のスケジュールにおいて提案法はベースラインシステムに比べ 8% 程度の精度向上が達成されている．また，BoW では *Precision* が 0.01 以下となっており，ほとんどの Slot-Value の推定が行えていないことが解る．これは，BoW ではベクトル空間が疎であるため，統計値を得ることができず識別が行えないためである．BoW w/ W2V では BoW よりも，発話文の特徴を表現可能な特徴量ベクトルであると考えられるため，BoW と比較すると精度が高くなっている．しかし，BoW w/ W2V はベースラインシステムの精度と比べて 15% 程度下回っている．提案法は，その他の特徴量ベクトルを用いた場合と比べて精度が高く，この結果から意味的情報の拡張により発話文の特徴を広くとらえることが出来たと考えられる．

### 5.4.2 実験 1 の分析

本節では，実験 1 によるハイパーパラメータの分析と，実際に対話状態推定をして正解ラベルを推定できていた Slot-Value の組についての分析・考察をする．ハイパーパラメータの実験において割引率は $\gamma = 1$ で精度が高いことが明らかであったが，この割引率が高くなるにつれ *F-measure* においてどのような変化をするか調査した（図 5.1，図 5.2）．

図 5.1 と図 5.2 は *Schedule1* と *Schedule2* の *F-measure* と割引率の変化であり，縦軸が *F-measure* で横軸が割引率に対応している．また，各折れ線グラフは閾値に対応している．これらの図から，割引率 $\gamma = 0$ から 0.9 までは徐々に *F-measure* が上昇しているが，$\gamma = 1$ である時に全ての閾値において大きく上昇していることが分かる．このことより提案法を DSTC4 に用いる際は割引率を $\gamma = 1$ することで最も精度が高くなり，対話履歴の考慮を強くすることが重要であると考えられる．閾値は 0.2 程度が望ましいと 5.4 と 5.5 から考えられるが，各 Slot-Valu においては 2 値分類をするタスクであるため，閾値が 0.5 で精度が高くなることが理想であ

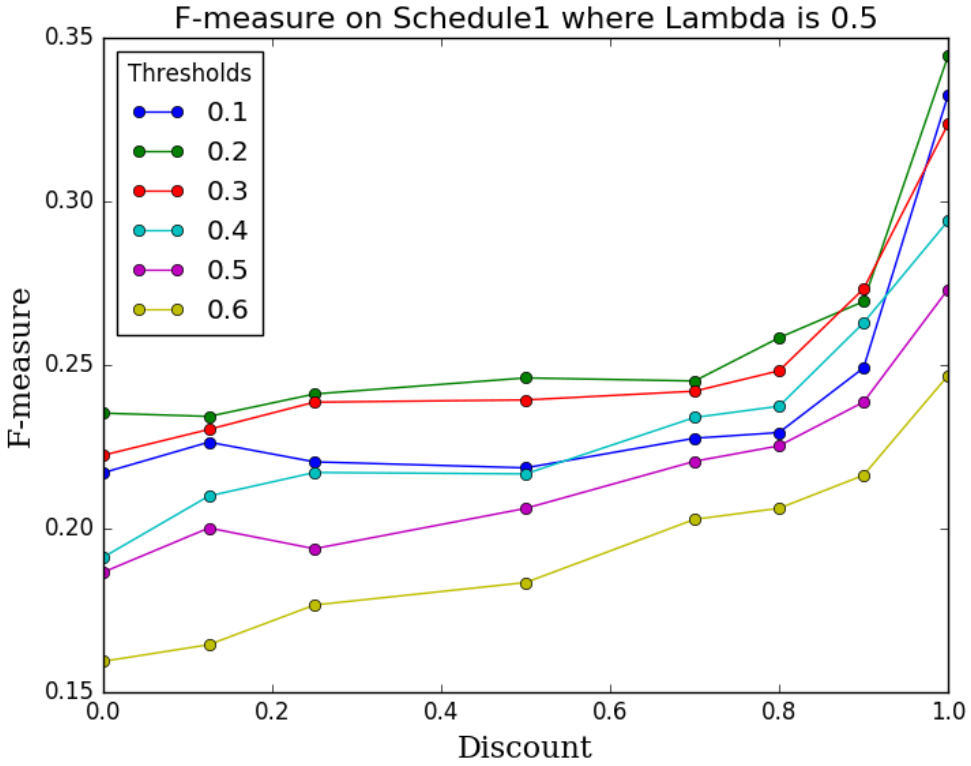


図 5.1　Schedule1 における F-measure と割引率の変化.

る．そのため，閾値により *Precision*，*Recall*，*F-measure* がどのように変化しているのかを分析する（図 5.3，図 5.4）．図 5.3 と図 5.4 は *Schedule1* と *Schedule2* の *F-measure* と閾値の変化であり，縦軸が各スコアの数値・横軸が閾値に対応している．また，各折れ線グラフは *Precision*・*Recall*・*F-measure* である．これらの図では *Precision* は高いところで推移しているが，*Recall* は低いところで推移しているため，*Recall* の影響により *F-measure* も低いところで推移をしている．また，閾値が 0.5 程度で *Precision* は 0.6 程度であるが，*Recall* は 0.2 を割っている．これらの事から，*Precision* を下げずに *Recall* をより高くすることが重要であると考えられる．

提案手法で提案手法で推定できていた Slot-Value の例を表 5.8 に挙げる．この例では，ルールベースであるベースラインシステムと実際の正解データとの

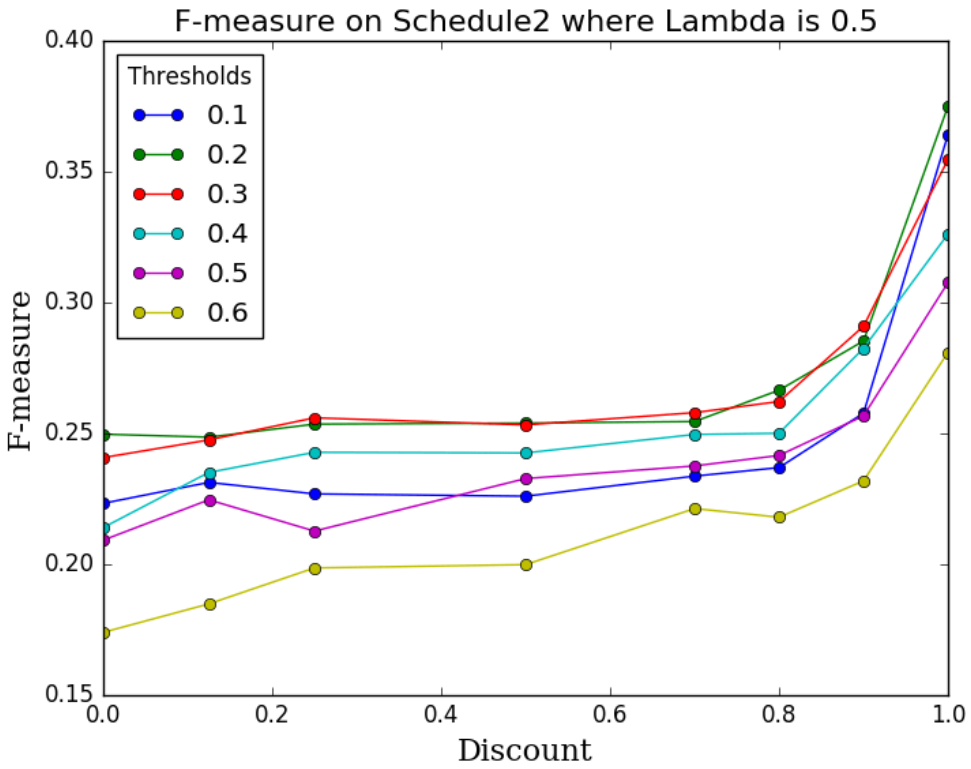


図 5.2　Schedule2 における F-measure と割引率の変化.

比較を行う．表 5.8 ではベースラインシステムにより 1 つの Slot-Value も推定されていないが，提案法では Slot=INFO に対して Value = Fee が推定されている．これは，発話文中の”free entry” から意味的情報の拡張を行ったことで推定できたのではないかと考えられる．だが，正解データにある Slot=PLACE に対する Value = National Museum of Singapore は推定できておらず，発話文中の”National Museum” から意味的情報の拡張により推定できると考えられる．これは，知識グラフの構築で”National” と”Museum” の概念から National Museum of Singapore を推定するための特徴が推論されていないからではないかと考えられる．そのため，Wikidata からの知識グラフの抽出方法の改善が必要であると考えている．

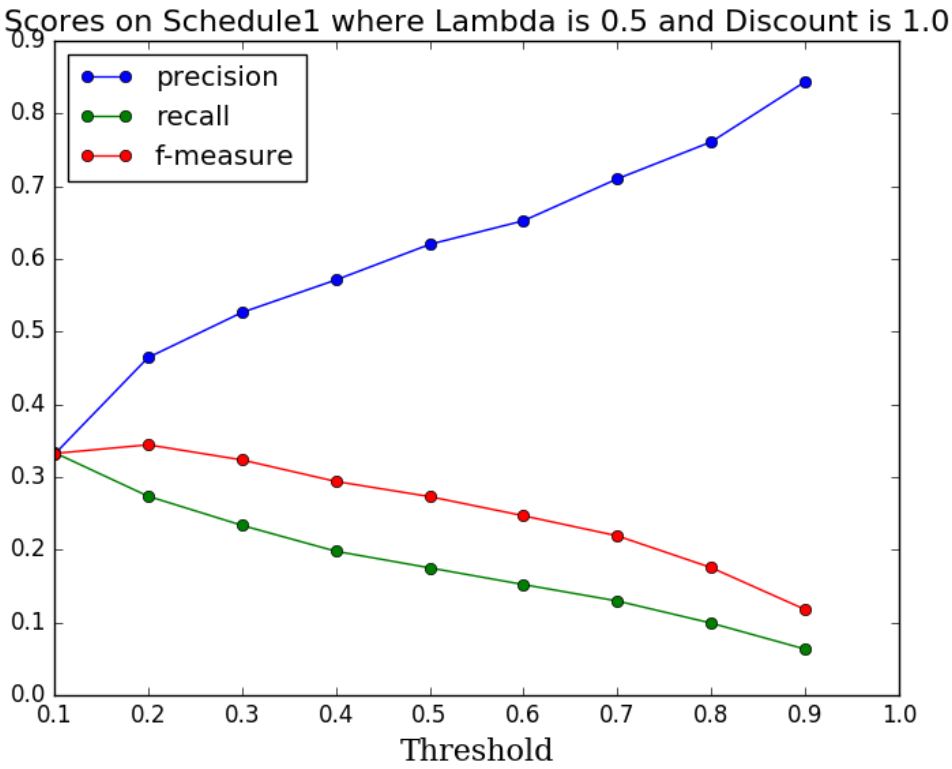


図 5.3　Schedule1 における Precision，Recall，F-measure の変化.

表 5.8　推定したフレームの比較表

| 発話文 | ベースライン | 提案法 | 正解データ |
|---|---|---|---|
| Uh National Museum, you may even get free entry because it's a-if it's a public holiday. | {} | {'INFO': ['Fee']} | {'INFO': ['Fee'], 'PLACE': ['National Museum of Singapore']} |

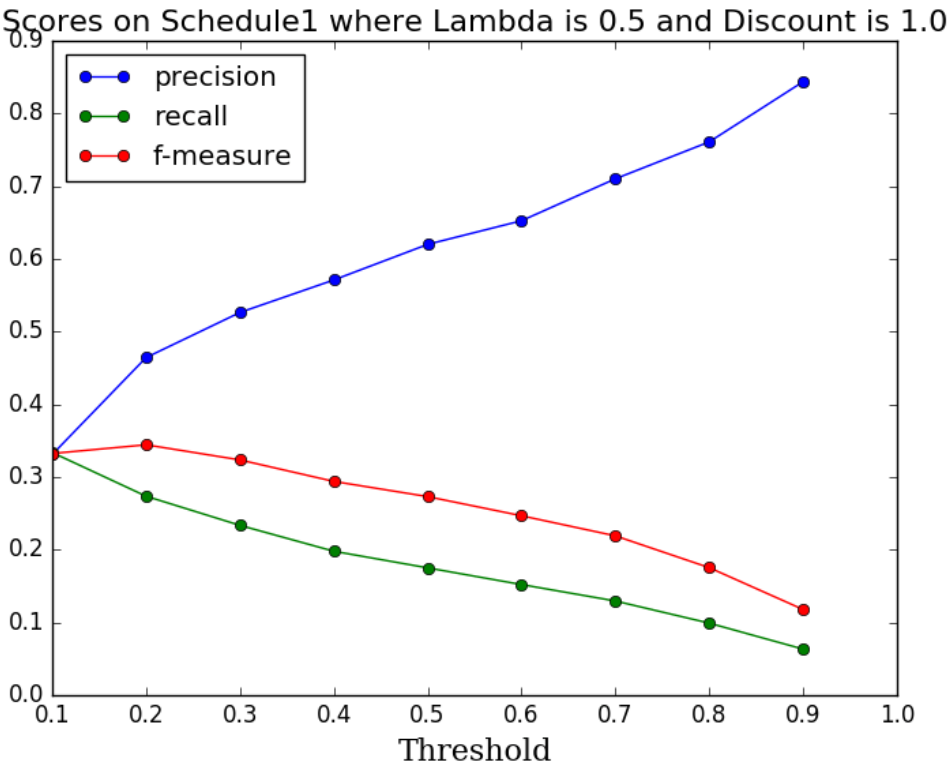


図 5.4　Schedule2 における Precision，Recall，F-measure の変化.

### 5.4.3　実験 2 の比較

実験 2 では，提案した特徴量ベクトルを用いた FCNN モデルと CNN のアンサンブルによる対話状態推定の精度の向上を確認した．まず，ニューラルネットワークのハイパーパラメータチューニングを開発セットで行い，DSTC4 のレギュレーションにより開発セットで上位 4 つのハイパーパラメータの組み合わせを探索する．テストセットでは，開発セットの上位 4 つの中からそれぞれ最も精度が高いモデルを選択し，DSTC4 で最も精度の高い機械学習によるモデル（MSIIP） [31] との比較をする．また，ベースライン（Base）として CNN をベースとしたモデル [15] との比較も行う．この CNN のモデルは Slot=INFO のみを推定しており，残りの Slot は DSTC4 のベースラインシステムである曖昧性マッチングにより行われている．

表 5.9　FCNN の開発セットにおける上位 4 つのハイパーパラメータの組 (Schedule2)

| **BS** | **DR** | **TH** | **F-measure** |
|---|---|---|---|
| 60 | 0.1 | 0.2 | 0.4445 |
| 60 | 0.3 | 0.2 | 0.4406 |
| 30 | 0.4 | 0.2 | 0.4401 |
| 40 | 0.3 | 0.2 | 0.4339 |

表 5.10　CNN の開発セットにおける上位 4 つのハイパーパラメータの組 (Schedule2)

| **FH** | **OC** | **BS** | **DR** | **TH** | **F-measure** |
|---|---|---|---|---|---|
| 1 | 1500 | 50 | 0.4 | 0.2 | 0.3953 |
| 1 | 1500 | 50 | 0.2 | 0.2 | 0.3934 |
| 1 | 1500 | 75 | 0.5 | 0.2 | 0.3922 |
| 1 | 1500 | 100 | 0.4 | 0.2 | 0.3880 |

全てのモデルにおいてドロップアウト率（DR）と閾値（TH）は 0.1 から 0.5 を 0.1 刻みで探索した．FCNN におけるハイパーパラメータとして，DR と TH に加えバッチサイズ（BS）を 20 から 110 の間を 10 刻みで探索する．CNN におけるハイパーパラメータとして，フィルターの高さ（FH）は 1 と 2 を，アウトプットチャンネル（OC）は 500 から 1500 を 500 刻みで，バッチサイズ（BS）を 50 から 100 の間を 25 刻みで探索した．その結果，FCNN と CNN では実験 1 と同様に閾値が 0.2 で精度が高く，バッチサイズは大きくないほうが良いのではないかと考えられる（表 5.9，表 5.10）．CNN では FH が 1・OC が 1500 の時が精度が高くなる傾向がわかった．また，全体の結果から下位のほとんどで FH が 2 であった．

Ensemble-1 では，トレーニングした FCNN と CNN のモデルを全て用いて，それぞれのモデル出力に対して重みの組（9 組）の探索をした（表 5.11）．上位 4 つの結果には FCNN と CNN の重みのバランスとして $w_{fcnn} = w_{cnn} = 0.5$ が 3 つ含まれており，各モデルを同等に重視することで精度が高くなると考えられる．また，各モデルのパラメータは各モデルの結果と異なり，BS と DR が低いモデルが上位に確認された．

表 5.11　Ensemble-1 の開発セットにおける上位 4 つのハイパーパラメータの組 (Schedule2)

| $w_{fcnn}$ | $\mathbf{w_{cnn}}$ | **FCNN BS** | **FCNN DR** | **FH** |
|---|---|---|---|---|
| 0.5 | 0.5 | 30 | 0.1 | 1 |
| 0.5 | 0.5 | 20 | 0.2 | 1 |
| 0.4 | 0.6 | 20 | 0.2 | 1 |
| 0.5 | 0.5 | 30 | 0.1 | 1 |
| **OC** | **CNN BS** | **CNN DR** | **TH** | **F-measure** |
| 1500 | 50 | 0.2 | 0.2 | 0.4708 |
| 1500 | 50 | 0.2 | 0.2 | 0.4704 |
| 1500 | 50 | 0.2 | 0.2 | 0.4703 |
| 1500 | 100 | 0.2 | 0.2 | 0.4688 |

表 5.12　Ensemble-2 の開発セットにおける上位 4 つのハイパーパラメータの組 (Schedule2)

| **OC** | **BS** | **DR** | **TH** | **F-measure** |
|---|---|---|---|---|
| 500 | 50 | 0.3 | 0.3 | 0.4296 |
| 500 | 50 | 0.4 | 0.3 | 0.4287 |
| 500 | 50 | 0.5 | 0.2 | 0.4277 |
| 500 | 50 | 0.3 | 0.2 | 0.4255 |

Ensemble-2 では，FH は 1 として設定している．また，BS は 25，50, 100 とした（表 5.12）．Ensemble-2 では，CNN と異なり上位の OC が全て 500 であった．これは，1 つのモデルとすることで提案法である特徴量ベクトルを畳み込みによる特徴量よりも重視したのではないかと考えられる．

テストセットでは，全てのモデルの上位 4 つから最も精度が高かった結果を DSTC4 の最も精度が高かった機械学習モデルの結果を交えて比較する（表 5.13，表 5.14）．表の太文字が *Accuracy*･*Precision*･*Recall*･*F-measure* の最も高い値を示しいる．両スケジュールにおいて *F-measure* は Ensemble-1 が最も高く，DSTC4

表 5.13　Schedule1 の結果比較

| | Base | **FCNN** | **CNN** | **Ensemble-1** | **Ensemble-2** | **MSIIP [31]** |
|---|---|---|---|---|---|---|
| *Accuracy* | 0.0371 | 0.0455 | 0.0332 | 0.0446 | 0.0435 | **0.0489** |
| *Precision* | 0.4179 | 0.4915 | 0.4146 | **0.4951** | 0.4221 | 0.4440 |
| *Recall* | 0.2804 | 0.2586 | 0.2538 | 0.2831 | **0.3457** | 0.2703 |
| *F-measure* | 0.3356 | 0.3389 | 0.3148 | **0.3602** | 0.3378 | 0.3361 |

表 5.14　Schedule2 の結果比較

| | Base | **FCNN** | **CNN** | **Ensemble-1** | **Ensemble-2** | **MSIIP [31]** |
|---|---|---|---|---|---|---|
| *Accuracy* | 0.0584 | 0.0502 | 0.0407 | 0.0559 | 0.0435 | **0.0697** |
| *Precision* | 0.4384 | 0.5082 | 0.4366 | **0.5156** | 0.4220 | 0.4634 |
| *Recall* | 0.3426 | 0.3056 | 0.2976 | 0.3335 | **0.3475** | 0.3335 |
| *F-measure* | 0.3846 | 0.3817 | 0.3540 | **0.4050** | 0.3801 | 0.3878 |

で最も *F-measure* の高かったモデルよりも *Schedule1* で約２．４％，*Schedule2* で約 1.7 %の精度が上がったことを確認した．また，Ensemble-1 の基になったモデルである FCNN と CNN と比べても Schedule2 でそれぞれ約 2 %と約 5 %精度が上がっている．*Recall* においては，FCNN と CNN と比較して値が高くなっており，*Precision* は FCNN からさほど変化がなかった．また，Ensemble-2 でも *Recall* が高くなっていたため，アンサンブルに各モデル特有の推定が行われることで，それぞれの弱点を補完しあっているのではないかと考えられる．

### 5.4.4　実験２の分析

本節では，実験 2 で得られた結果からアンサンブルによる対話状態推定をして正解ラベルを推定できていた Slot-Value の組と出来ていなかった組についての分析・考察をする（表 5.15，表 5.16）．

表 5.15　State Frames of 4 models at utterance-id 566 in dialog #21

| **Transcription** | **Gold Standard** | **FCNN** |
|---|---|---|
| also for certain rides in the Universal Studio, there's a height limit. | **PLACE**: 'Universal Studios Singapore' **ACTIVITY**: 'Amusement ride' **INFO**: 'Restriction' | **PLACE**: 'Universal Studios Singapore' |
| **CNN** | **Ensemble-1** | **Ensemble-2** |
| **PLACE**: 'Universal Studios Singapore' | **PLACE**: 'Universal Studios Singapore' **ACTIVITY**: 'Amusement ride' **INFO**: 'Restriction' | **PLACE**: 'Universal Studios Singapore' **ACTIVITY**: 'Amusement ride' **INFO**: 'Restriction' |

表 5.16　State Frames of 4 models at utterance-id 144 in dialog #23

| **Transcription** | **Gold Standard** | **FCNN** |
|---|---|---|
| So this one in M Hotel, how much it will cost? | **INFO**: 'Pricerange' **PLACE**: 'M Hotel Singapore' | **INFO**: 'Pricerange' **PLACE**: 'M Hotel Singapore' |
| **CNN** | **Ensemble-1** | **Ensemble-2** |
| **INFO**: 'Pricerange' | **INFO**: 'Pricerange' | **INFO**: 'Pricerange' **PLACE**: 'M Hotel Singapore' |

表 5.15 では独立したモデルでは推定できいなかった Slot-Value の組がアンサンブルにより推定されていた．これは，FCNN と CNN のそれぞれがこの Slot-Value を推定するための特徴量を潜在的に持っていたが，推定するための要因として弱かったためアンサンブルをすることで推定可能になったのではないかと考えられる．5.16 では FCNN で推定できていた Slot=PLACE の Value=M Hotel Singapore が Ensemble-1 で推定できていなかった．これは線形補間による単純なアンサンブルをすることで CNN の影響を強く受けすぎたためだと考えられる．そのため，CNN の影響をあまり強く受けていないと考えられる Ensemble-2 では正答できている．これらの事例より，アンサンブルのデメリットはあるものの，DCTC4 において本研究での提案である特徴量ベクトルを用いた FCNN と CNN をアンサ

ンブルすることが精度の向上に貢献したといえる.

# 第 6 章　結言

## 6.1　本論文のまとめ

本研究では，対話状態推定の精度向上のために発話文を知識グラフによる推論で意味的情報を拡張する特徴量ベクトルの作成を提案し，２つの独立したモデルと２つのアンサンブルモデルによる実験を行った．提案した特徴量ベクトルの作成では対話履歴の考慮を強くすることで，より発話の特徴が得られることがわかり，DSTC4 における対話状態推定に有効な手法であることを示した．また，FCNN と CNN のアンサンブルにより本提案の特徴量ベクトルと畳み込みによる特徴量ベクトルでは違った特徴量を抽出していると考えられるため，アンサンブルをすることで独立したモデルでは推定不可能であった Slot-Value の組が推定可能になった．

## 6.2　今後の課題

本提案の特徴量ベクトル作成では，知識グラフの構築は単純な文字列マッチングによりエンティティの抽出を行い，1-hop 先のノードまでを付け加えた．このように作成した知識グラフがどのように構成されているか確認をしておらず，おそらくスパースな部分や特徴量抽出に無用なノードが付加されていると考えている，知識グラフの分析を行う必要がある．また，ラベル伝搬法による推論ではプロパティ，グラフにおけるエッジはどのようなラベルであっても同等にあつかっている．プロパティはエンティティ間の関係性であり，任意のエンティティとその他の関連（プロパティ）のあるエンティティが全て同等の関連であるとは考え難いため，グラフ上のエッジを関連の度合いにより自動的に推定することが必要である．

本実験では FCNN と CNN を対話状態推定のモデルとしてもちいアンサンブルをしたが，本提案の特徴量ベクトルを CNN の入力に出来ないかを考えている．また，対話では対話履歴を時系列的に扱うため，再帰型ニューラルネットワーク等のような時系列を表現できるモデルの適用も考えるべきである．

# 謝辞

本学情報科学研究科の中村哲教授には，主指導教官として終始熱心なご教示と数多くの貴重なご助言をいただきました．また，研究を進めるにあたり充実した研究環境・設備を与えていただきました．心より感謝申し上げます．

本学情報科学研究科の松本裕治教授には，副指導教官として貴重なご助言をいただきました．心より感謝申し上げます．

本学情報科学研究科の吉野幸一郎助教には，副指導教官として熱心なご教示と数多くの貴重なご助言をいただきました．研究環境の設備の設定や論文執筆では大変なご支援をいただきました．また，対話システム研究に対する姿勢を近くで学べたことが一番の宝であり，この先の研究生活の基礎となると確信しております．心より感謝申し上げます．

Carnegie Mellon University の Graham Neubig 助教，本学研究科の須藤克仁准教授，Sakriani Sakti 特任准教授，鈴木優特任准教授，田中宏季助教には研究における貴重なご助言をいただきました．心より感謝申し上げます．

松田真奈美様，林美穂様，軽部瑞穂様には諸手続きでお世話になり，研究室運営でご尽力いただきました．心より感謝申し上げます．

研究室の同期の皆様には，年齢差の壁を越えて様々な議論により楽しい研究生活を与えていただきました．心より感謝申し上げます．

研究室の皆様には，研究に関するご支援と多くのご助言をいただきました．心より感謝申し上げます．

私のわがままをずっと聞いていただいた母・兄・姉には，今までのご恩を忘れることはありません．この先の皆様の健康をお祈りいたします．

最後に，修士課程の間に結婚をした村瀬弥世様には，生活面や精神面の全てにおいてご支援いただき心より感謝申し上げます．この先も末永く宜しくお願い致します．

# 参考文献


[1] 河原達也，荒木正弘.「音声対話システム」. オーム社（2006）

[2] 中野幹夫，駒谷和範，船越孝太郎，中野有紀子.「対話システム」. コロナ社（2015）

[3] Steve Young, Milica Gašić, Simon Keizer,François Mairesse, and Jost Shatzmanm. The Hidden Information State Model: A Practical Framework for POMDP-based Spoken Dialogue Management, Computer Speech and Language, Vol. 24, No.2, pp. 150-174, 2010.

[4] Blaise Thomson and Steve Young. Bayesian Update of Dialogue State: A POMDP Framework for Spoken Dialog System, Computer Speech and Language, Vol. 24, No. 4, pp. 562-588, 2010.

[5] Jason Williams, Antoine Raux and Matthew Henderson. The Dialog State Tracking Challenge Series: A Review, Dialog and Discourse, Vol. 91, No. 4, pp. 4-33, 2016.

[6] Jason Williams, Antonie Raux, Deepak Ramachandran, and Alan Black. The Dialog State Tracking Challenge, In Proceedings of SIGDIAL, pp. 404-413, 2013.

[7] Matthew Henderson, Blaise Thomson, Jason Williams. The Second Dialog State Tracking Challenge, In Proceedings of SIGDIAL, pp. 263-272, 2014.

[8] Matthew Henderson, Blaise Thomson, Jason Williams. The Third Dialog State Tracking Challenge, In Proceedings of SLT, pp. 324-329, 2014.

[9] Seokhwan Kim, Luis D'Haro, Rafael Banchs, Jason Williams, Matthew Henderson, and Koichiro Yoshino. The Fifth Dialog State Tracking Challenge, In Proceedings of SLT, pp. 511-517, 2016.

[10] Seokhwan Kim, Luis D'Haro, Rafael Banchs, Jason Williams, and Matthew Henderson. The Fourth Dialog State Tracking Challenge, Dialogues with Social Robots, pp. 435-449, 2017.

[11] Matthew Henderson. Machine Learning for Dialog State Tracking: A Review, Proceedings of MLSLP, 2015.

[12] LukasZilka and Filip Jurcicek. Incremental LSTM-based Dialog State Tracker, Proceedings of ASRU, pp. 757-762, 2015.

[13] Koichiro Yoshino, Tatsuya Hiraoka, Graham Neubig, and Satoshi Nakamura. Dialog State Tracking Using Long Short Term Memory Neural Networks, Proceedings of IWSDS, 2016.

[14] Hongjie Shi, Takashi Ushino, Mitsuruand Endo, Katsuyoshi Yamagami, and Noriaki Horii. A Multichannel Convolutional Neural Network for Cross-Language Dialog State Tracking, Proceedings of SLT, pp. 559-564, 2016.

[15] Hongjie Shi, Takashi Ushino, Mitsuruand Endo, Katsuyoshi Yamagami, and Noriaki Horii. Convolutional Neural Networks for Multi-Topic Dialog State Tracking, Dialogues with Social Robots, pp. 451-463, 2017.

[16] Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi, and Kyoko Kanzaki. Enhancing the Japanese WordNet, Proceedings of ALR, pp. 1-8, 2009.

[17] Yi Ma, Paul Crook, Rushi Sarikayu, and Eric Fosler-Lussier. Knoewledge Graph Inference for Spoken Dialog System, Proceedings of INTERSPEECH, pp. 5346-5305, 2015.

[18] 石川　葉子，平岡　拓也，水上　雅博，吉野　幸一郎，Graham　Nuebig，中村　哲．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．情報処理学会　第 109 回　音声言語情報処理研究会（SIG-SLP），2015 年 12 月，名古屋．

[19] Miao Li and Ji Wu. The MSIIP System for Dialog State Tracking Challenge 4, Dialogues with Social Robots, pp. 465-474, 2017.

[20] 河原　達也．音声対話システムの進化と淘汰: 歴史と最近の技術動向 (< 特集 > 音声対話システムの実用化に向けて)，人工知能学会誌，Vol.. 28，No. 1，pp. 45-51，2013 年．

[21] 斎藤　哲也，広田　健一，星野　准一．Web 情報を用いたキャラクタの発話・世間話モデル．情報処理学会　第 94 回　自然言語処理研究会（SIG-SLP），2007 年．

[22] Teruhisa Misu and Tatsuya Kawahara. A Bootstrapping Approach for Developing Language Model of New Spoken Dialogue Systems by Selecting Web Texts, Proceedings of INTERSPEECH, pp. 9-12, 2006.

[23] 野村　浩郷．自然言語理解の構造：理解の表現，情報処理学会誌，Vol.．30，No．10，pp．1161-1168，1989 年．

[24] 奥村　学，高村　大也．「自然言語処理のための機械学習入門」．コロナ社（2010）

[25] Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[26] Ye Zhang and Byron Wallace. A Sensitive Analysis of (and practitioners' guide to) Convolutional Neural Networks for Sentence Classification, Proceedings of IJCNLP, 2017.

[27] Yoon Kim. Convolutional Neural Networks for Sentence Classification, Proceedings of EMNLP, 2014.

[28] Deny Vrandečić and Markus Krötzsch. Wikidata: A Free Collaborative Knowledgebase, Proceedings of Communication of the ACS, pp. 78-85, 2014.

[29] Tsuyoshi Kato, Hisashi Kashima, Hisashi Sugiyama. Robust Label Propagation on Multiple Networks, IEEE Transactions on Neural Networks, Vol. 20, No. 1, pp. 35-44, 2009.

[30] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean. Distributed Represenatational of Words and Phrases and their Compositionality, Proceedings of Advances in NIPS, pp. 3111-3119, 2013.

[31] Miao Li, Ji Wu. The MSIIP system for dialog state tracking challenge 4, Dialogues with Social Robots, pp. 465-474, 2017.

# 発表リスト

### 国際会議 [査読付]

[1] Yukitoshi Murase, Koichiro Yoshino, Masahiro Mizukami, Satoshi Nakamura. "Feature Inference Based on Label Propagation on Wikidata Graph for DST." Proceedings of International Workshop on Spoken Dialog System（IWSDS），pp.1-12, June 2017, Farmington, Pennsylvania, USA.

### 全国大会 [査読無]

[1] 村瀬　行俊，吉野　幸一郎，中村　哲．知識グラフによる特徴量ベクトルを用いたニューラルネットワークによる対話状態推定．自然言語処理学会第２３回年次大会，2018 年 3 月，岡山.

### 研究会 [査読無]

[1] 村瀬　行俊，吉野　幸一郎，水上　雅博，中村　哲．Wikidata 上でのラベル伝搬法を用いた対話状態推定のための素性抽出．情報処理学会　第１１４回　音声言語情報処理研究会（SIG-SLP），2016 年 12 月，東京.
[2] 村瀬　行俊，吉野　幸一郎，中村　哲．英語・中国語の 2 言語による知識グラフの推論を用いた対話状態推定のための素性抽出.. 人工知能学会　第８１回　言語・音声理解と対話処理研究会（SIG-SLUD），2017 年 10 月，東京.