

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

村瀬 行俊^{1,a)} 吉野 幸一郎¹ 水上 雅博¹ 中村 哲¹

概要: 近年、対話システムの構築に必要なタスクとして対話状態推定 (Dialogue State Tracking; DST) の研究が盛んにおこなわれている。DST では発話文そのものと過去の発話履歴を統合して発話の意図を推定する。このとき、話している話題などの関連情報を用いなければ推定が難しい場合があり、人間が行っているように知識を用いた推論が重要となる。そこで本研究では、Web 上の関係知識データベースである Wikidata を知識として使用し、ラベル伝搬法によって話題の中心となっている単語の推論を行う。この推論結果を対話状態推定の素性として利用することで、対話状態推定の精度が向上するかを確認した。

キーワード: 知識グラフ, 対話システム, ラベル伝搬法, Wikidata, 対話状態推定

1. はじめに

対話状態推定 (Dialogue State Tracking; DST) は対話中の発話内容を理解するためのタスクであり、現在のユーザの発話と過去の対話履歴を考慮してユーザの意図・状態を推定する [5], [11]。対話中のユーザの意図・状態は多岐に渡り、それらをすべてカバーできるようにデータを集めるのは非常に困難である。仮に相当量の学習データを集められたとしても、テストデータや実際の対話ではどうしても未知の状態が出現してしまう。また、ユーザの意図はタスクやドメインにより多様であり、タスク・ドメインごとに必要な量のデータを作成することは不可能に近い [4]。そこで、本研究では発話内容に対するカバー率を上げるために外部知識のグラフ構造上での推論を行った。具体的には、外部データのグラフとして Wikidata^{*1}のグラフ構造を用い、グラフ上でラベル伝搬法により発話に関連する単語の推論を行い、抽象化する。

Wikidata は Wikimedia による Web 上のプロジェクトで複数言語を含む知識基盤である [1]。Wikidata の目的は Wikimedia の他のプロジェクトのデータを構造化することである。Wikidata はそれぞれのエンティティが様々な関係性で結ばれている。本研究ではエンティティを発話文の単語と結び付け、グラフ上の関係性をそれらの単語間の関係性として取り扱った。Wikidata はグラフ構造をなして

おり、任意の単語に関連語との繋がりを記述した知識基盤であるといえる。このような知識基盤は対話システム研究において広く利用され、手作業や情報検索のクエリから教師なしで作成したものが使われてきた [2], [3], [7], [8], [10]。これに対して近年 Wikidata のような Web 上での知識基盤が整備されつつあり、Ma らによる研究ではこうしたグラフ構造をそのまま対話状態の推定をするための推論に利用している [6]。これに対して本研究では、既存の対話状態推定にこうした知識基盤を用いた推論結果を付け加えることで、対話状態を推定する。

対話状態推定において各発話の各単語は素性として利用されるが、統計的に推定器を学習するにはスパースネスが問題となる。そこで学習データに出現する各単語をシードとして Wikidata のグラフ上でラベル伝搬法によりノード判別を行う [9], [12]。ラベル伝搬法では隣り合ったノードは同じクラスに属するという仮定のもと、観測されたというラベルを観測語に付与してグラフ上で伝搬させる。これにより、未観測だが発話文中の単語との関連が大きいノードを、対話中に出現する話題語として捉えることができる。Wikidata の利点は大規模で統制された構造化データが利用できる点で、ラベル伝搬法によって必要な範囲のグラフ上で適切な推論が可能である。これによって、対話中のユーザ発話をうまく抽象化する単語・素性を推測し、対話意図推定の精度向上に役立てることができると考えられる。

2. 知識グラフを用いた対話状態の推定

本研究は、Web 上の知識グラフをそのまま対話状態推定

¹ 奈良先端科学技術大学院大学
Nara Institute of Science Technology
^{a)} y-murase@is.naist.jp
^{*1} Wikidata: <https://www.wikidata.org/>

のための知識構造とする先行研究 [6] と強い関わりがある。この先行研究では、ウェブ上の知識グラフをマルコフ確率場における推論グラフとして用い、いくつかのノードが観測値となった時の他のノードの事後分布を推論アルゴリズムによって求めている。知識グラフはエンティティとアトリビュートを含んでおり、アトリビュートはエンティティが属するクラスとなっている。知識グラフを形成する上でエンティティとアトリビュートの両方がノードとなり、アトリビュートに属するノードとそのアトリビュートの間に関係性が存在する。ここでアトリビュートノードを因子とすることによりそれぞれのノードが条件付独立性をなす。推論の際には発話文からノードに値が与えられ、この知識グラフ上で隣接したノードへ伝搬させる。この結果、確率の高い値を持つと推論されたエンティティノードを発話文との関係性が深いとして、対話状態推定の出力値としている。これに対して本研究では、エンティティ同士の関係性であるアトリビュートのラベルは使わず、関係があるという点にだけ着目してグラフ上での推論を行う。この結果得られた関連ノードを、既存の対話状態推定の入力に対するクエリ拡張に利用する。

3. Wikidata からの推論グラフ構築

Wikidata はアイテムをノード、ノード間の関係性をエッジとしたグラフ構造を持っている。しかしこのグラフをそのまま用いようとすると、アイテムの数が膨大すぎるため実時間での推論が難しい。そこで、今回は推論のための部分グラフを構築するため、Wikidata のグラフから以下の手順によってノードおよびエッジの削減を行った。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、その単語とアイテムの名前が完全一致するもののみを用いた。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。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リンクをもとに、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た場合話題語として出力される。同じ名前を持った別のアイテムについては、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 1 に示す例は上記の手順でシードとして観測された "Singapore", "City", "Asia" に対して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。

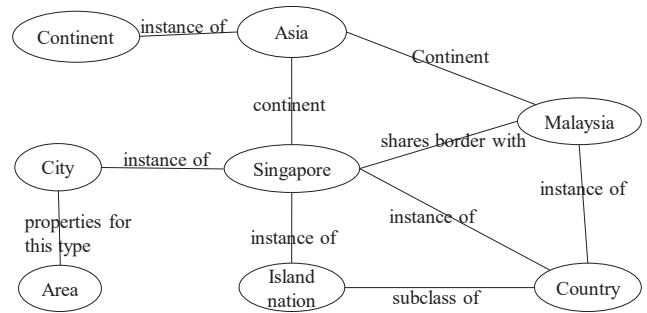


図 1 "Singapore" と "Malaysia" が観測されたときこれらのノードにより四つのノードが追加され、観測されたノードと追加されたノードの間にエッジが追加される。次に "Asia" が観測されれば追加されるノードは一つだが、既に存在するノード "Singapore" と "Malaysia" の間に関係性があるので三つのエッジが追加される。"City" も同様に隣接ノードを追加して既存のノードと関係性があればエッジを追加する。

4. ラベル伝搬法とクエリ拡張

4.1 ラベル伝搬法

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状態のとき、それ以外のもの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を予測する。

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 2 に示す。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する。このラベル伝搬法の最小化問題は式 (1) のように定式化される。

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

式 (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。 λ 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 (1) は、

$$J(f) = \| \mathbf{y} - \mathbf{f} \|_2^2 + \lambda \mathbf{f}^T \mathbf{L} \mathbf{f} \quad (2)$$

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

$$(\mathbf{I} + \lambda\mathbf{L})\mathbf{f} = \mathbf{y} \quad (3)$$

をもちいる。式 (2)(3) のラプラシアン行列は $\mathbf{L} \equiv \mathbf{D} - \mathbf{W}$ である。なお、 \mathbf{D} は対角行列の成分は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 (3) によって算出され予測値なので、以下の式で予測値を直接もとめた。

$$\mathbf{f} = (\mathbf{I} + \lambda\mathbf{L})^{-1}\mathbf{y} \quad (4)$$

これによって図 2 の例に対し \mathbf{f} の値を更新した結果を図 3 に示す。

4.2 割引率を考慮した履歴の利用

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 $d(0 < d < 1)$ 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 d 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。このアルゴリズムを **Algorithm 1** に示す。

Algorithm 1 ラベル伝搬法の割引率

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

```

if 最初の発話文 then
  for 単語のリスト内の  $y_{i,t}$  do
     $y_{i,t} = 1$ 
  end for
else
  二発話目以降:
  for  $y_{i,t}$  do
    if 単語のリスト内の  $y_{i,t}$  then
       $y_{i,t} = 1 + dy_{i,t-1}$ 
    else
       $y_{i,t} = dy_{i,t-1}$ 
    end if
  end for
end if
 $\mathbf{f} = \mathbf{y}(\mathbf{I} + \lambda\mathbf{L})^{-1}$ 
return  $\mathbf{f}$  の上位  $n$  個
  
```

4.3 対話状態推定の入力クエリ拡張

ここまで述べた手法で得た予測値の上位 n 個は実際に入力された発話文の潜在的に観測された単語として考えられる。そこで、得られた単語が入力された発話文と、この上位 n 件を発話文の末尾に付与することで対話状態推定器に対する入力クエリ拡張をおこなった。

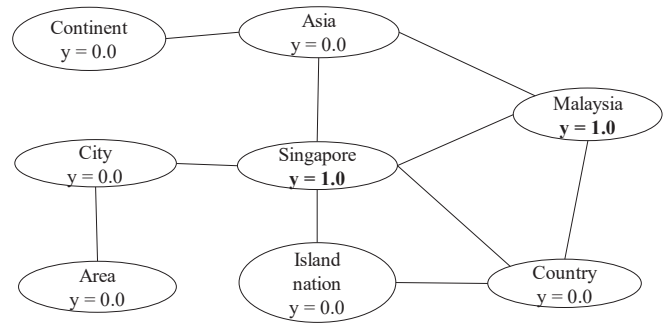


図 2 “Singapore” と “Malaysia” が発話文から観測されたときの入力値 \mathbf{y} の値の例。現在の発話文までにすべてのノードが未観測なら “Singapore” と “Malaysia” の入力値が 1 となり、その他のノードの入力値は 0 になる。

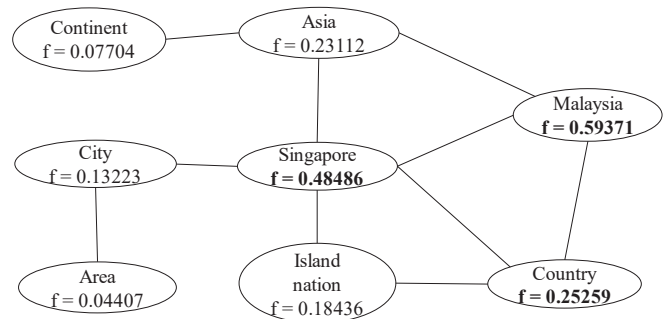


図 3 図 2 のグラフ上にラベル伝搬法を適用し、計算した結果の予測値の例。各ノードは入力値 \mathbf{y} に代わり \mathbf{f} を持っている。観測されたノードである “Singapore” と “Malaysia” は高い値をとっているが、それに加えて観測されたノード両方と関連があるのは “Country” が高い値を取っている。これは、式 (1) の第 2 項により高い予測値をとりやすい。加えて “Country” の隣接ノードに “Island nation” があるが、このノードも隣接ノードに観測ノード “Singapore” があり高い値をとる。それにより “Country” はより高い値をとる。 “Asia” も観測されたノードと関係性を持つてはいるが、隣接ノードの “Continent” は観測ノードとは隣接していないため式 (1) の第 2 項により少し小さな予測値をとる。

λ	割引率	上位 n
0.5	0	1
1	0.125	3
1.5	0.25	5
2	0.5	8
3	0.7	10
8	0.8	
	0.9	
	1	

表 1 ラベル伝搬法に与えてテストしたパラメータ: 240 通りの組み合わせがある。 λ は式 (1) における第 1 項と第 2 項のバランスをとる係数、割引率は対話履歴を考慮するための値、 n は推定されたノードのうちクエリ拡張に用いるノードの数である。

フレーム



図 4 外の枠は、ある対話系列の区間に設けられているフレームとなる。その区間の対話内容が”Accommodation”であればスロットにその値が与えられる。”Accommodation”についての詳細な情報が一番内側の枠、すなわちスロットの中の値として与えられている。この例では宿泊施設の種類と名前、情報の種類の値が与えられている。

5. 実験

本研究では Dialog State Tracking Challenge 5 (DSTC5)[5] のデータセットとベースライントラッカーを使用し、提案したラベル伝搬法によるクエリ拡張結果をトラッカーの入力としてその効果を検証する。DSTC5 のデータセットは学習セット、開発セット、テストセットに分かれている。全てのデータセットの対話数が 35、発話数は 30134、単語数は 273480 個あり、オントロジーのスロット数が 30 個、スロットが含む可能な値が 1667 個ある。正解データは各発話区間をフレームとしていくつかのスロットと、スロットに対応したラベルから構成されている。図 4 にフレーム、スロット、ラベルの例を示す。

このデータでは中国語の発話を英語に翻訳して、翻訳された発話文に対して発話文の状態をするタスクとなっている。今回は提案したラベル伝搬法を翻訳された発話文に対して実行し、その結果を素性として付け加えている。ベースライントラッカーは入力された文字列とオントロジーのエントリーに関してファジーマッチングを取り、フレームのスロットに値を与える。このベースライントラッカーの入力として、提案するラベル伝搬法によってクエリ拡張した発話文を入力した。この際に設定が必要となるラベル伝搬法のハイパーパラメータ λ 、割引率 d 、追加するクエリ数 n の三つのパラメータは、それぞれ表 1 に示す値の組み合わせを試行した。この試行は開発セットで行う。このクエリ拡張した発話を、クエリ拡張を行わない場合と比較する。

DSTC5 では 2 種類の評価単位をそれぞれ 2 種類の指標で評価している。一つ目の評価単位 (スケジュール 1) は各発話ターンごとに測っている。二つ目の評価単位 (スケジュール 2) は一定区間で発話を区切り、セグメントが終わる度に精度を評価している。また、一つ目の指標

F 値	λ	割引率	上位 n
0.18776	0.5	0.9	8
0.18294	0.5	0.9	10
0.17953	0.5	0.9	5
0.17178	0.5	0.8	5
0.17009	1.5	0.9	5
0.16990	1.5	0.9	8
0.16929	1.5	0.9	3
0.16928	0.5	0.8	8
0.13281	(ベースライン)		

表 2 スケジュール 1:F 値の高いパラメータの組み合わせ上位 8 件

(Accuracy) はフレームが完全に一致しているかを、二つ目の指標は各スロットの値の一致を Precision, Recall, F 値で評価している。

表 2, 3 に、表 1 で示したパラメータのうち、F 値が高く出る組み合わせ上位 8 件を示す。比較のため、クエリ拡張を行わないベースライントラッカーの F 値はスケジュール 1 で 0.13281、スケジュール 2 で 0.18111 となっている。表 2 の結果より、スケジュール 1 では F 値は最大で 0.05495 ポイント高くなっている。また表 3 より、スケジュール 2 では最大 0.01710 ポイントの向上が見られる。パラメータ λ は高いスコアを記録した組み合わせの多くで 0.5 をとっている。式 (1) の第 2 項よりも第 1 項の方が F 値が高くなるように影響していると推測できる。第 1 項を優先することは、隣り合った予測値の値を近づけるよりも入力値により近づけるように予測値を出した方がよいということで、あまりクエリ拡張を大きくやり過ぎない方がよいことが見てとれる。

ディスカウント値は両方のスケジュールで 0.7 以上が上位として残っており、対話履歴をより強く反映するほうがよいということがわかる。 n は予測値の上位 n 個の素性を使用するためのパラメータであり、結果からは 5 から 8 個の素性を用いることで F 値が高くなっている。

これらのパラメータの組み合わせによる結果の最上位のパラメータの組み合わせは λ が 0.5、割引率が 0.9、 n が 8 となり、二つのスケジュールともに全く同じ組み合わせになった。図 5 と 6 はそれぞれのパラメータの一つを変化させて他の二つを固定したときの F 値の変化をグラフ化した図であり、それぞれのパラメータが λ が 0.5、割引率が 0.9、 n が 8 に近づくにつれて F 値が高くなっている。

これらの結果よりパラメータを決定し、テストセットに対して同様のセッティングで実験を行う。

表 4 はテストセットに対してクエリ拡張を行わずベースライントラッカーを実行したときの結果で、表 5 ではクエリ拡張を行った結果を入力としたときの結果である。F 値は各スケジュール共に 0.04 ポイント近く上昇しており、Precision が大幅に高くなっている。しかし、Recall はあまり変化しておらず、Accuracy にいたっては 0.01 ポイン

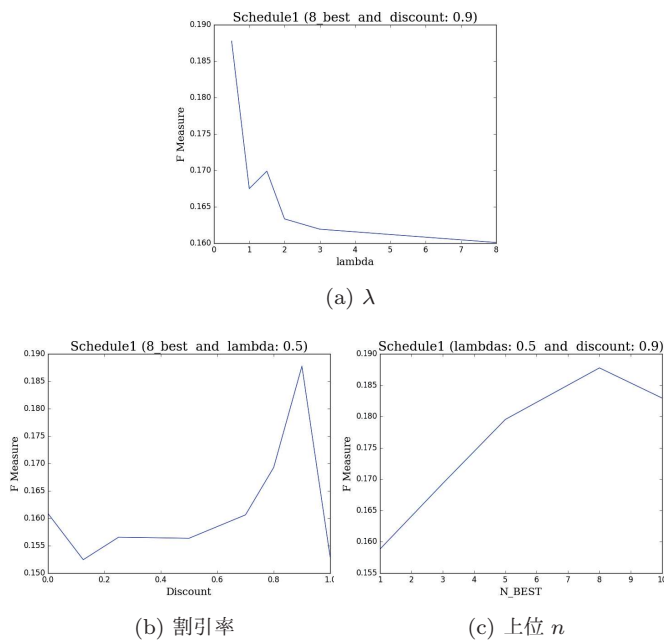


図 5 y 軸はスケジュール 1 の F 値を表している。(a): x 軸は λ が 0.5 から 8 の間の値をとる。(b): x 軸は割引率が 0 から 1 の間の値をとる。(c): x 軸は 1 から 10 個で上位 n 個の数を示す。

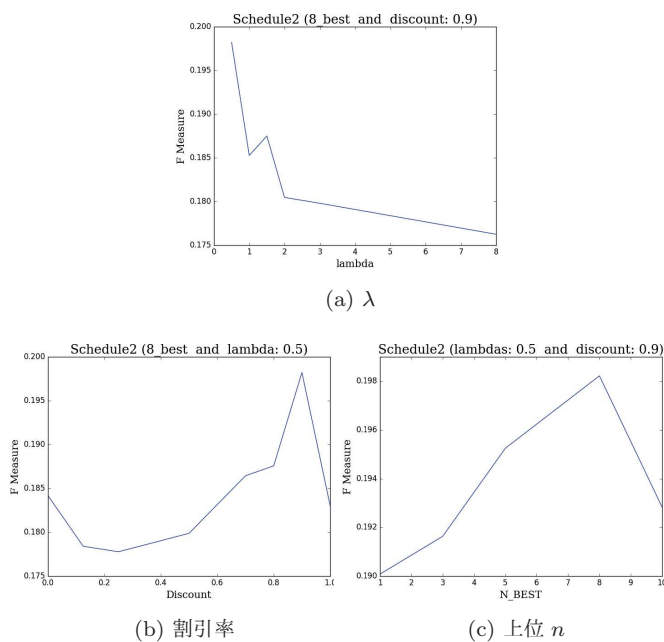


図 6 y 軸はスケジュール 2 の F 値を表している。(a): x 軸は λ が 0.5 から 8 の間の値をとる。(b): x 軸は割引率が 0 から 1 の間の値をとる。(c): x 軸は 1 から 10 個で上位 n 個の数を示す。

ト近く下がってしまっている。Accuracy の指標はフレーム全体の完全一致の評価であるため、提案手法ではスロットに余分な値を与えてしまう、あるいはこれまで推定できていたスロットを削ってしまったような場合があったのではないかと考えられる。Precision の値が上がりつつ

F 値	λ	割引率	上位 n
0.19822	0.5	0.9	8
0.19525	0.5	0.9	5
0.19364	0.5	1	1
0.19283	0.5	0.9	10
0.19267	1.5	0.7	1
0.19241	2	0.7	5
0.19241	0.5	0.7	5
0.19231	0.5	0.8	1
0.18111	(ベースライン)		

表 3 スケジュール 2: F 値の高いパラメータの組み合わせ上位 8 件

	スケジュール 1	スケジュール 2
Accuracy	0.02311	0.03137
Precision	0.11394	0.14123
Recall	0.10898	0.14868
F 値	0.11140	0.14486

表 4 ベースラインのスコアの結果

	スケジュール 1	スケジュール 2
Accuracy	0.01570	0.02296
Precision	0.25944	0.27764
Recall	0.10950	0.14601
F 値	0.15400	0.19137

表 5 ベースラインに素性を加えたスコアの結果

Recall を維持しているのは、トラッカー自体の間違いを減らしつつも、解けるものはあまり増えていないということが考えられる。

ここからは実際に得たラベルとスロットを例に考察を行う。"In this way, you can walk to the Jinsha." が入力されたとき次の発話の素性として"shape, straight-0, straight-1, straight-2, walk, bar, lake, walking" が拡張単語として抽出された。このときベースラインのシステムでは'Type' というスロットは空だったのに対して、提案手法ではそのスロットに'Walking' という正解のラベルを与えている。ただし、素性を加えることで正解ラベルで必要とされないスロットもラベルを与えられている場合があった。

6. まとめ

本研究では DSTC5 のデータにおいて、Wikidata から得られるグラフ上での推論を行うことで入力クエリを拡張し対話状態推定の精度向上を試みた。この拡張クエリをベースライントラッカーに対して用いた結果、Precision が大幅に向上するという結果が得られた。事例から、特に発話文上では未観測の単語を外部の知識基盤の利用により抽出できることがわかった。ただ、ベースライントラッカーの手法が単純なファジーマッチングであるため、今後はより一般的な機械学習手法を用いた対話状態推定器において本提案手法を試行することを検討している。

7. 謝辞

本研究開発の一部は総務省 SCOPE (受付番号 152307004) の委託を受けたものです。

参考文献

- [1] Bollacker, K., Evans, C., Paritosh, P., Sturge, T. and Taylor, J.: Freebase: a collaboratively created graph database for structuring human knowledge, *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, ACM, pp. 1247–1250 (2008).
- [2] Crestani, F. and Lee, P. L.: Searching the web by constrained spreading activation, *Information Processing & Management*, Vol. 36, No. 4, pp. 585–605 (2000).
- [3] El-Kahky, A., Liu, X., Sarikaya, R., Tur, G., Hakkani-Tur, D. and Heck, L.: Extending domain coverage of language understanding systems via intent transfer between domains using knowledge graphs and search query click logs, *2014 IEEE ICASSP*, IEEE, pp. 4067–4071 (2014).
- [4] Henderson, M., Thomson, B. and Williams, J.: Dialog state tracking challenge 2 & 3 (2014).
- [5] Kim, S., D’Haro, L. F., Banchs, R. E., Williams, J., Henderson, M. and Yoshino, K.: The Fifth Dialog State Tracking Challenge, *Proceedings of the 2016 IEEE Workshop on SLT* (2016).
- [6] Ma, Y., Crook, P. A., Sarikaya, R. and Fosler-Lussier, E.: Knowledge graph inference for spoken dialog systems, *2015 IEEE ICASSP*, IEEE, pp. 5346–5350 (2015).
- [7] Pearl, J.: *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann (2014).
- [8] Raux, A. and Ma, Y.: Efficient Probabilistic Tracking of User Goal and Dialog History for Spoken Dialog Systems., *INTERSPEECH*, pp. 801–804 (2011).
- [9] Scholkopf, O. C. B. and eds, A. Z.: *Semi-supervised Learning*, MIT Press (2006).
- [10] Wang, L., Heck, L. and Hakkani-Tür, D.: Leveraging semantic web search and browse sessions for multi-turn spoken dialog systems, *2014 IEEE ICASSP*, IEEE, pp. 4082–4086 (2014).
- [11] Williams, J., Raux, A., Ramachandran, D. and Black, A.: The dialog state tracking challenge, *Proceedings of the SIGDIAL 2013 Conference*, pp. 404–413 (2013).
- [12] 鹿島久嗣: グラフとネットワークの構造データマイニング, *電子情報通信学会誌*, Vol. 93, No. 9, pp. 797–802 (2010).