4R-04

誤りを含む音節認識結果に対応する知識グラフ内エンティティの同定

平川 巧人 † 大塩 幹 † 近辻 脩壱 † 武田 龍 † 駒谷 和範 †

† 大阪大学 産業科学研究所

1. はじめに

音声対話システムの知識ベースの更新は，しばしば人手により行われる．例えば，新しい商品として「地球グミ」という単語が生まれたとする．システムに対してユーザがその単語を含む発話をした場合，システムの知識ベースに「地球グミ」に関する情報はまだないため，その単語に関する話題を継続することは難しい．また，人手による知識ベースの更新では，その更新タイミングによって別の対話でも「地球グミ」は知らないものとして扱われてしまう可能性がある．

本研究では，ユーザ発話内の未知語に対するシステム内知識ベースの動的な更新を考える．具体的には，ユーザとの対話と同時に知識ベースの更新を行うことを目指す．未知語は知識ベース内に存在しない語とする．動的な更新ができれば，先ほどの例の「地球グミ」のような未知語に関する情報を，対話をもとに得ることができる．この場合，システムとしてはまず未知語を音声認識できる必要があるが，本稿では音素や音節認識に基づく方式を想定する [1]．この方式では，認識された音節列（発音記号列）を単語へ分割するため，ユーザ発話に未知語が含まれていても対応する音節列を推定できる．

しかし，音節認識誤りを含む既知語が誤って未知語と判断される場合がある．例えば，図 1 の例では，既知語である「ハンバーグ」が誤って「ハンバーブ」という未知語と認識されている．ユーザは「ハンバーグ」について話しているため，知識ベース内で「ハンバーブ」と認識して対話を行うのは誤っている．

本稿では，知識ベースとして知識グラフを扱い，音節認識誤りが生じた既知のエンティティ名に対する元のエンティティの同定手法について述べる [2, 3]．提案手法ではまず，音節列と各エンティティ名の編集距離を用いて，候補集合の順位づけを行う．この集合内のエンティティが持つリンク構造を質問に用い，回答に合致しないエンティティを除外した後に最も類似度の高いもので同定する．質問回数の上限を設定したシミュレーション実験により，質問の有無による同定性能差を明らかにする．

2. エンティティ同定問題と関連研究

2.1 問題設定

本稿では，対話中に生じた音節認識誤りを含むエンティティ e' を想定し，それが本来指していたエンティティ $e^* \in \mathscr{E}$ を，ユーザへの質問を通して $\mathscr{E}$ から同定する．ここで，知識グラフは subject entity, relation, object entity に対応する三つ組み (triplet) (s,r,o) の集合で構成され，エンティティの集合を $e \in \mathscr{E}$ で，各エンティティを $e \in \mathscr{E}$ で表すものとする．また，e' は音節列（カタカナ表記）で得られることを想定しているため，e^*, e も同様に音節列で表現されていると仮定する．e' は認識誤りを含むため，例えば，音節列上で距離が小さいエンティティが複数候補存在することもある．候補を絞り込むためにユーザに質問を繰り返すことになるが，同定までに行う最大の質問回数 N が少ないことが望ましい．

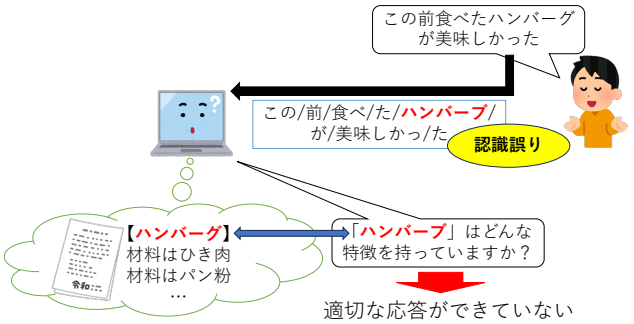

図 1: 音節認識誤りにより既知語が誤って未知語と判断される例

ユーザへの質問は知識グラフに含まれるリンク構造に基づき行う．同定先候補を絞るのに用いるリンク構造は triplet (s,r,o) を指し，特に各エンティティ s が持つ (r,o) の有無をもとに候補を絞る．(r,o) の有無はユーザへの Yes/No 質問の回答として正しい情報が得られると想定する．

2.2 関連研究

音声認識に知識グラフを用いることで，音声認識用モデルの語彙外エンティティに対する認識率の改善が報告されている [4]．知識グラフを用いることで，認識誤りのある語彙外エンティティを知識グラフ内のエンティティへ同定することを可能にしている．一方で，この研究において未知語部の再現率は 7 割程度にとどまっており，対話を行わず一発話のみで認識誤りのあるエンティティを同定するのは困難であることがうかがえる．

対話ベースの意味導出サービスに関するアプローチを提案している研究もある [5]．ユーザ発話に未知語が検出された際，様々な単語ネットワークから未知の単語の候補として関連する同義語を見つけ，これらの候補を知識グラフに対して検証し，ユーザが正誤確認をすることで未知語を習得するというものである．この研究と本研究は知識グラフを用いたアプローチをする点では同じである．しかし，この研究では「地球グミ」のような真に未知語の単語のみに対応している一方で，本稿で扱う「ハンバーブ」のような既知語が未知語化した単語の扱いには触れられていない．

3. リンク構造に基づく質問による同定手法

3.1 編集距離による候補エンティティの順位づけ

知識グラフにはエンティティ間に順位の概念が存在していないため，同定の際に何かしらの尺度で順位づけを行う必要がある．本稿では，エンティティ名の編集距離を尺度として用い，その値の小さい順に順位付けを行う．

Identifying Knowledge Graph Entity Corresponding to Erroneous Syllable Recognition Result: *Takuto Hirakawa, Miki Osio, Shuichi Chikatsuji, Ryu Takeda, and Kazunori Komatani (Osaka Univ.)*

知識グラフ内の各エンティティ$e \in \mathscr{E}$と対話中に生じた音節認識誤りを含むエンティティe'との編集距離を測り，全エンティティ$e \in \mathscr{E}$を編集距離が小さい順にソートすることで同定先エンティティ候補の順位づけを行う．この順位をもとに，リンク構造に基づく質問選択を行う．

3.2 リンク構造に基づく質問選択

質問に用いるリンク構造として，順位が 1 位と 2 位 (同率を含む) それぞれのエンティティが持つ triplet 情報で互いに異なる部分を扱う．上位 2 位から選択する理由は，認識誤りが少ない場合に確実に同定したいからである．例えば，未知語化した文字列である「ハンバーブ」の同定先エンティティ候補の 1 位が「ハンバーグ」，2 位が「ハンバーガー」である場合，それぞれが持つ triplet を探索し，どちらか片方しかもっていない triplet を 1 つ選択する．このとき，異なる triplet が複数ある場合はその中からランダムに 1 つ選択する．正解エンティティ$e^* \in \mathscr{E}$がここで選択した triplet 情報を持つかどうかを質問することに相当する．ユーザから得た回答と合致しない triplet 情報を持つエンティティを除外する．この操作を複数回行って候補を減らすことで，同定精度の向上を図る．質問をあらかじめ決めた回数行った後，残った候補の中で編集距離が最も小さいものを同定予測先のエンティティとする．

4. 評価実験

4.1 実験条件

知識グラフとして，Wikidata から抽出した料理部分グラフに対し，楽天レシピを用いて拡充したものを使用した [6]．表 1 に triplet 数および，エンティティ数とリレーション数を示す．

知識グラフのエンティティ集合$\mathscr{E}$をもとに入力データを作成した．エンティティ$e \in \mathscr{E}$すべてに対し，対応する音節列の 20〜40%の数の文字をランダムに変換したものをe'と見なした．変換は，事前実験において求めた音節ごとの誤り率（各音節が挿入・削除・置換される確率）をもとに行った．この際，各エンティティについて，知識グラフ内に持つ triplet 数が 2 以下のものは入力データから除外した．また，エンティティ名の文字数が 3 文字以下のものも除外した．

e'を入力とし，それに対する予測$e \in \mathscr{E}$を 1 つ決定した．全ての予測の正誤を判定した結果をもとに，エンティティ同定精度を Hits@1 で評価した．

編集距離に基づく順位の上位 2 つに対して正解か不正解かの 2 分化を行った場合を想定し，ベースラインを設定した．この手法では，1 度の質問で候補数を 1 つ減らすことから，編集距離に基づく順位の上位$(1+N)$位までの中に正解エンティティが含まれる確率を評価した．

また，エンティティ候補の順位において，正解エンティティe^*と同率で 1 位となるエンティティ$e \in \mathscr{E}$が他に存在した場合，正解数を数える際に，1 ではなく期待値として 1/(e^*の同率順位のエンティティ数) とカウントして評価を行った．

4.2 実験結果と考察

各手法における同定精度（%）を表 2 に示す．編集距離のみの同定精度は 84.9%であったが，提案手法は 96.5%となり，約 12 ポイント精度が向上した．編集距離と提案手法の精度を比較すると，質問自体は有効であるが，質問回数$N = 3$では同定しきれていないことがわかる．

表 1: 知識グラフの統計情報

Triplet 数	52,730
エンティティ (subject) 数	6,945
リレーション数	92
エンティティ (object) 数	1,977

表 2: 各手法の同定精度（%）

編集距離（質問なし）	84.9 (1620/1909)
ベースライン（4 位以内）	96.0 (1832/1909)
提案手法（質問 3 回）	96.5 (1842/1909)

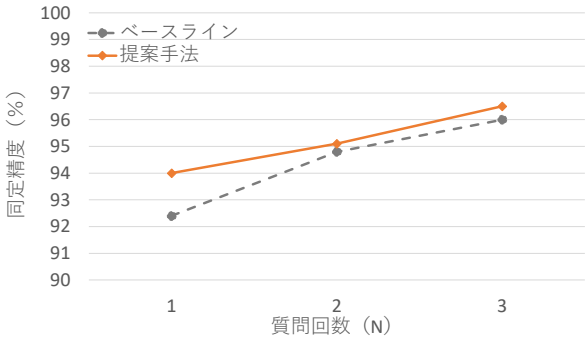

図 2: 質問回数に対する各手法の同定精度の変化

これは，文字列類似度の低いものが絞り込めていないことが原因だと考えられる．また，ベースラインと提案手法の結果の比較から，上位 2 位の単純な 2 分化より提案手法の方が精度が高いことが分かる．

次に，図 2 にベースラインと提案手法の質問回数Nと同定精度の関係を示す．各質問回数$N = 1, 2, 3$で，提案手法がベースラインを上回っていることが分かる．これは，ベースラインが上位 2 つの単純な 2 分化を行っているのに対し，提案手法は 3 位以下の候補も同時に絞ることができる点が影響していると考えられる．

5. おわりに

本稿では，認識誤りを含む音節列から，対応する知識グラフ内のエンティティを同定する手法を提案した．対話中の質問応答から得られる情報を用いることで，単純な編集距離を用いた手法と比べて約 12 ポイント同定精度が向上した．今後は未知語か認識誤りを含む既知語かの判定などに取り組む．

参考文献

[1] Miki Oshio, et al. Out-of-vocabulary word detection in spoken dialogues based on joint decoding with user response patterns. In *Proc. of APSIPA ASC*, pp. 1753–1759, 2023.

[2] Xuchen Yao and Benjamin Van Durme. Information extraction over structured data: Question answering with freebase. In *Proc. of ACL*, pp. 956–966, 2014.

[3] Hao Zhou, et al. Commonsense knowledge aware conversation generation with graph attention. In *Proc. of IJCAI*, pp. 4623–4629, 2018.

[4] Nilaksh Das, et al. Listen, know and spell: Knowledge-infused subword modeling for improving ASR performance of OOV named entities. In *Proc. of ICASSP*, pp. 7887–7891, 2022.

[5] Alexander Wachtel, et al. Dialog-based meaning derivation service for technical language domains. In *Proc. of ICSC*, pp. 375–380, 2019.

[6] Shuichi Chikatsuji, et al. Knowledge graph augmentation with entity identification for improving knowledge graph completion performance. In *Proc. of PRICAI*, pp. 480–487, 2023.