

# 暗黙的確認による未知語獲得のための ユーザ応答に基づくクラス推定結果の正誤判別

大野 航平<sup>†</sup> 武田 龍<sup>†</sup> Eric Nichols<sup>‡</sup> 中野 幹生<sup>‡</sup> 駒谷 和範<sup>†</sup>

<sup>†</sup>大阪大学 産業科学研究所

<sup>‡</sup>(株)ホンダ・リサーチ・インスティテュート・ジャパン

ohno@ei.sanken.osaka-u.ac.jp

## 1 はじめに

我々は、特定ドメインの非タスク指向型対話システムの構築を目指しており、特に食べ物や飲み物に関する対話を行うものを開発している。ここで課題となるのが、システムの知識にない単語、すなわち未知語への対応である。

この課題に対し、我々は暗黙的確認により対話を通じて未知語を獲得する手法の開発に取り組んでいる。ここで未知語の獲得とは、システムの知識に登録されていない飲食物の、料理ジャンルに関するクラスの獲得であるとする。「プッタネスカ」が未知語の場合、(1) 所属するクラスが「イタリアン」であることを推定し、(2)「イタリアンは洒落た料理が多いですね」といった暗黙的確認要求を出力し、(3) 確認要求に含めたクラス推定結果の正誤をユーザの応答から判別する [5]。暗黙的確認により、明示的な質問を用いる手法 [3] よりも自然に対話を継続できると考える。

上記の手順 (3) を実現するために、2 つの課題に取り組む。まず、単一ユーザ応答の表現に基づいて、確認要求に含めたクラスの正誤判別を行う。明示的な質問の場合とは異なり、システムの暗黙的確認要求に対して、ユーザが正しい/正しくないことを明示的な同意や反対の表現で示すとは限らない。そのため、同意や反対を明示的に示す表現だけでなく、それ以外の表現も考慮して正誤判別を行う。このとき、確認要求後のユーザ発話だけでなく、確認要求前のユーザ発話や、両者の関係も考慮する。次に、同じ確認要求に対する複数の判別結果によってさらに推定を行う。未知語についての誤ったクラス推定結果を含む確認要求に対して、ユーザが投げやりになって同意を示す表現を用いて応答し、システムが誤った知識を獲得してしまう場合がある。異なるユーザに対する複数回の暗黙的確認の結果を用いて、正しい知識のみを獲得することを

狙う。

ルールや機械学習を用いて文が肯定的なものか否定的なものかを分類する研究が多く行われている。Marnette らは、Yes/No 質問に対する応答が、明示的にどちらかを示すものではなかった場合に、それがどちらを意図したものか判定するためのルールを作成した [4]。Gokcen らは、Web 上での議論のコーパスから、反対の立場をとる投稿を識別するための素性を設計した [1]。これに対して本研究では、応答そのものを分類するためではなく、確認要求に含めたクラスの正誤判別を行う。さらに、複数の判別結果を利用して推定を行う。

## 2 ユーザ応答に基づくクラス推定結果の正誤判別

ユーザ応答に基づくクラス推定結果の正誤判別について、全体像を図 1 に示す。本手法は以下の 4 ステップからなる。ここで、 $i$  の初期値は  $i = 1$  である。

1. 未知語  $w$  を含む発話を行った  $i$  番目のユーザに対し、クラス推定結果  $c$  を含む暗黙的確認要求を出力する。
2. 確認要求前後のユーザ応答から判別の確信度  $p_i(w, c)$  を算出する (提案手法 1)。
3.  $p_1(w, c), \dots, p_i(w, c)$  から素性を抽出し、確信度  $Conf(w, c)$  を算出する (提案手法 2)。
4.  $Conf(w, c)$  がしきい値を超えていれば、クラス推定結果  $c$  は正しいとみなしてシステムの知識に獲得する。そうでなければ、 $i = i + 1$  としてステップ 1 から再度処理を行う。

ステップ 2 では、単一のユーザ発話に基づいてクラス推定結果の正誤判別を行う。正しいクラスあるいは

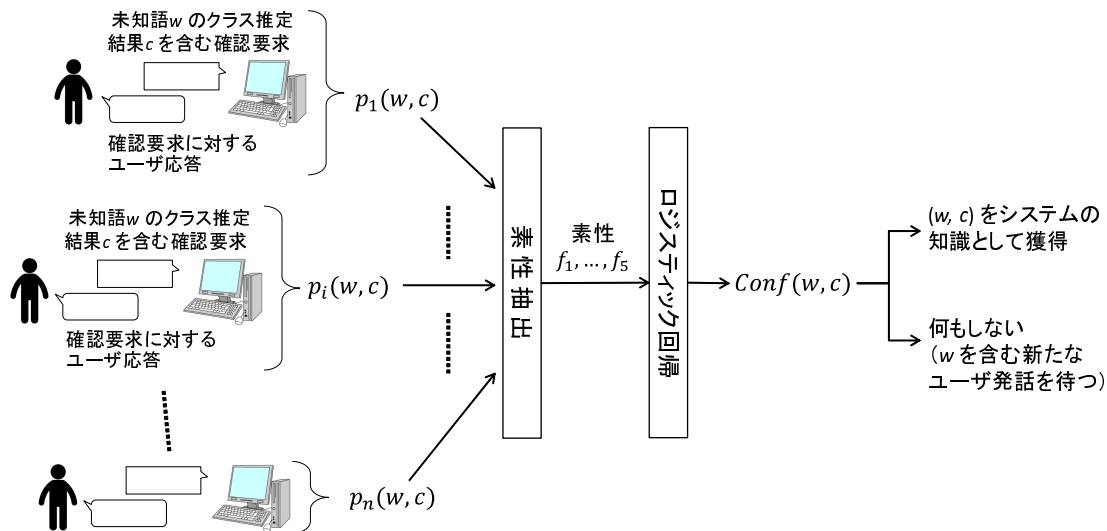


図 1: ユーザ応答に基づくクラス推定結果の正誤判別の全体像

誤ったクラスを含む確認要求に対するユーザ応答の傾向の違いを分析し、確認要求に含めたクラスの正誤判別に有効と考えられるユーザ発話中の表現を素性として設計する。このとき、確認要求前のユーザ発話も含めて考慮する。

ステップ 3 では、複数のユーザ発話に基づいてクラス推定結果の正誤判別を行う。ステップ 2 で得られた複数の推定結果から素性を抽出し、それを入力としてさらに推定を行う。推定の確信度  $Conf(w, c)$  がしきい値を超えていれば、未知語とそのクラスをシステムの知識として獲得する。

## 2.1 単一のユーザ応答に基づいた確信度 $p_i(w, c)$ の算出

単一ユーザ応答の表現から確認要求に含めたクラスの正誤判別を行うために、11 種類の素性を設計した。その一覧を表 1 に示す。U1, S1, U2 はそれぞれ、確認要求前のユーザ発話、確認要求、確認要求に対するユーザ発話、を表す。以下に例を示す。

- U1: 鍋料理と言えば常夜鍋だね。
- S1: 冬になるとお鍋が食べたくくなりますね。
- U2: そうですね。

また、素性の名称の後の丸括弧内は、着目する発話を示す。これらを入力とし、確信度  $p_i(w, c)$  を算出する。

ここでは、システムがある単語についての正しいクラスあるいは誤ったクラスを用いた確認要求を行った場合、確認要求前後のユーザ発話間にはそれぞれ異なっ

表 1: 単一のユーザ発話から  $p_i(w, c)$  を算出するための素性

g1	U2 が S1 への同意表現を含む (U2)
g2	U2 が S1 への反対表現を含む (U2)
g3	U2 が S1 への訂正表現を含む (U2)
g4	U1 と U2 とで共通した単語が出現 (U1, U2)
g5	U2 が S1 で用いられたクラスを含む (U2)
g6	U2 が S1 で用いられていないクラスを含み、かつ S1 への訂正表現を含まない (U2)
g7	U2 が S1 による話題転換を防止する語を含む (U2)
g8	U1 が S1 で用いられたクラスを含む (U1)
g9	U1 が S1 で用いられていないクラスを含む (U1)
g10	U1 が疑問詞を含む (U1)
g11	U1 の表現と S1 で用いられたクラスとの不一致 (U1)

た傾向が現れることを仮定している。収集した応答のデータからそれぞれの傾向を分析し、それらを素性で表現する。設定した表現やクラスについては文献 [5] を参照されたい。

## 2.2 複数のユーザ応答に基づいた確信度 $Conf(w, c)$ の算出

本手法の目標は、誤った知識をシステムが獲得するのを防ぐために、異なるユーザに対する複数回の暗黙の確認の結果を用いて、正しい知識のみを獲得することである。複数のユーザ応答を考慮することで、確認要求に含めたクラスの正誤判別をより正確に行うことを狙う。

推定の確信度  $Conf(w, c)$  の算出は、回帰問題として定式化する。異なる  $n$  人のユーザからそれぞれ得ら

表 2:  $n$  人分のユーザ応答から確信度  $Conf(w, c)$  を算出するための素性 ( $1 \leq i \leq n$ )

f1	$p_i(w, c)$ の平均値
f2	$n$
f3	$p_i(w, c)$ の最大値
f4	$p_i(w, c)$ の最小値
f5	$\{p_i(w, c) \geq 0.5 \text{ となる応答数}\} / n$

れた  $p_1(w, c), \dots, p_n(w, c)$  を入力とし, 機械学習の手法により  $Conf(w, c)$  を算出する. ここで入力とする素性を表 2 に示す. これらは,  $p_i$  が  $n$  個得られたときの素性を表している. 機械学習に基づく回帰問題として定式化するには,  $n$  が変わった場合でも, 素性の数が一定である必要がある. このことを考慮して,  $n$  回の暗黙的確認から得られる情報を表現できるようにした.

### 3 評価実験

#### 3.1 単一のユーザ応答に基づくクラス推定結果の正誤判別

本節では, 表 1 に示す素性を利用した正誤判別の性能を検証するために行った, 評価実験について述べる. 対象とするデータは, クラウドソーシングにより収集した, 1956 のやりとりを対象とした [5]. これらはそれぞれ U1, S1, U2 の 3 発話からなる. 正しいクラスあるいは誤ったクラスを含む確認要求 10 文ずつに対し, それぞれ 98 人 (一部 97 人) のワーカから応答を収集した. まず, 収集した 1956 のやりとりから, 表 1 に示す素性として定義した表現やクラスの有無をそれぞれ検出した. 各々の素性値は, 設定した表現やクラスなどが含まれていれば 1, そうでなければ 0 とした. 次に, ロジスティック回帰を用いて 10 分割交差検証により分類を行った. ロジスティック回帰は Weka [2] (Version 3.8.1) のものを使用し, パラメータはデフォルト値を用いた.

確認要求に対する同意表現, 反対表現のみを考慮して設計した素性 f1, f2 を用いる手法をベースラインとし, 11 種類の素性全てを用いた場合の結果との比較を行った. それぞれの分類結果を表 3 に示す. ベースライン手法に比べ, 11 種類の素性を用いることで, 正解率は 0.551(1078/1956) から 0.719(1407/1956) に増加した. これにより, 確認要求に対する同意表現や反対表現以外の表現を考慮することで, 確認要求に含まれたクラスの正誤判別の性能向上が示された.

表 3: 11 種類の素性全てを用いた場合と素性 f1, f2 のみを用いた場合との分類結果

用いた素性	出力	正解	
		正しい	誤り
11 種全て	正しい	724	298
	誤り	254	680
f1, f2 のみ	正しい	391	297
	誤り	587	681

未知語の正しいクラスのみを獲得するには, より精度の高い推定を行う必要がある. そこで本研究では, 2.2 節で提案した複数ユーザ応答に基づくクラスの正誤判別を行う. その有効性を検証するために行った実験について次節で述べる.

#### 3.2 複数のユーザ応答に基づくクラス推定結果の正誤判別

##### 3.2.1 学習データの作成

実験における学習および評価は, ワーカ open, 確認要求 open で実施した. 3.1 節で述べた通り, 我々は正しいクラスあるいは誤ったクラスを含む確認要求 10 文ずつに対し, それぞれ 98 人 (一部 97 人) のワーカから応答を収集した [5]. 正しいクラスあるいは誤ったクラスを含む確認要求 5 文ずつに対する 49 人 (一部 48 人) 分の応答を 1 グループとし, 収集した応答を 4 つのグループに分割した.

学習データには, 複数のユーザから得た応答を用いる. 1 つの暗黙的確認に対して  $N$  人から応答を得た場合,  $n$  人から応答を得た時点での学習データは,  ${}_N C_n$  通りを得ることができる. ここで,  $n$  が大きくなると  ${}_N C_n$  の値が非常に大きくなってしまいうため, 組み合わせの上限値を 1000 と定め,  ${}_N C_n$  が 1000 を超える場合にはランダムで組み合わせを 1000 通り選ぶ操作を行った. 組み合わせのそれぞれから表 2 に示す素性の値を計算し, 学習に用いる. 正解ラベルは, 確認要求に含まれるクラスが正しい場合には 1, そうでない場合は 0 とした.

学習および評価に用いるグループは, ワーカと確認要求との重なりがないよう選択した. つまり, 全 4 通りの学習および評価を行った. 確信度  $Conf(w, c)$  の算出には Weka [2] (Version 3.8.1) のロジスティック回帰を使用し, パラメータはデフォルト値を用いた.

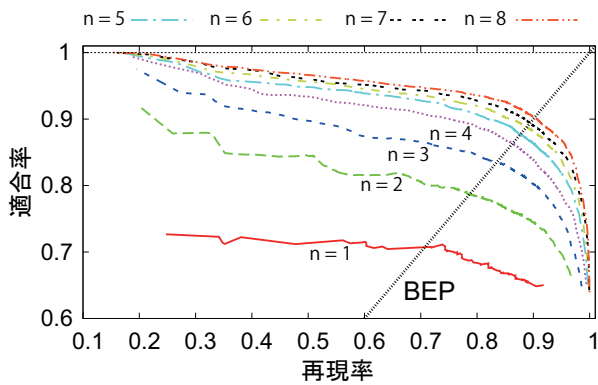


図 2: precision-recall 曲線と BEP の直線 ( $n \leq 8$ )

### 3.2.2 実験結果

まず、複数のユーザ応答を考慮することで、クラス推定結果の正誤判別の性能が向上するかを調査した。確信度  $Conf(w, c)$  のしきい値を種々の値に変化させ、それぞれの場合について正しいクラスを抽出する際の適合率および再現率を算出した。

図 2 に、 $n \leq 8$  における適合率と再現率との変化をそれぞれ示す。図中の直線は、breakeven point (BEP) を示す。これは、適合率 = 再現率となる値である。これは  $n$  の値が増加するにつれて、BEP も増加していることを示している。よって、単独の応答を用いた分類 ( $n = 1$ ) よりも複数の応答を用いた分類 ( $n \leq 2$ ) の方が性能が良いと言える。

次に、複数の応答を用いて正誤判別を行う際に応答を得る人数について考察する。図 2 より、 $n \geq 5$  の場合は BEP の増加値が比較的大きく、それ以降は増加値が小さくなっていくことがわかる。ここから、複数の応答を用いて分類を行う際には少なくとも 5 人から応答を得る必要があると判断した。

本手法によるクラス推定結果の正誤判別のためには、適切な  $Conf(w, c)$  のしきい値について調査を行う必要がある。我々は、獲得すべき未知語について、開発者に提案をする機能を実装予定である。自動で未知語の獲得を行うのではなく、対話のログに基づいて事後的に推定を行い、追加すべき未知語の候補を開発者に示す。自動で未知語の獲得を行うのであれば、誤った知識の獲得を防ぐため、適合率を重視する必要がある。一方、開発者に提案を行う場合、未知語を追加するかどうかの最終的な判断は人が行う。そのため、多くの候補を見せられるように再現率にも重きを置くことが求められる。

## 4 おわりに

本稿では、ユーザ応答に基づいて、暗黙的確認要求に含めたクラス推定結果の正誤判別を行う手法を提案した。実験により、(1) 単一のユーザ応答に基づくクラスの正誤判別には、確認要求前のユーザ発話も考慮することが有効であること、(2) 複数のユーザ応答を利用することで正誤判別性能が向上すること、を示した。

今後は、実際の対話システムに暗黙的確認の機能を組み込み、提案手法による未知語獲得手法の有効性を検証する。また、暗黙的確認に基づいた未知語獲得手法が、明示的確認に基づいたものに比べてユーザに煩わしさを感じさせないという仮説の検証も検討している。

## 謝辞

本研究の一部は JSPS 科研費 JP16H02869 の助成を受けた。

## 参考文献

- [1] Gokcen, A. and Marneffe, de M.-C.: I do not disagree: Leveraging monolingual alignment to detect disagreement in dialogue. *Proc. of ACL*, pp. 94-99, 2015.
- [2] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H.: The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1), pp. 10-18, 2009.
- [3] Komatani, K., Otsuka, T., Sato, S., Nakano, M.: Question Selection based on Expected Utility to Acquire Information through Dialogue, *Proc. of IWSIDS*, 2016.
- [4] Marneffe, de M.-C., Grimm, S., and Potts, C.: Not a simple yes or no: Uncertainty in indirect answers. *Proc. of SIGDIAL*, pp. 136-143, 2009.
- [5] 大野, 武田, ニコルズ, 中野, 駒谷.: 対話を通じた未知語のクラス獲得に向けた暗黙的確認の提案. *人工知能学会論文誌*, Vol. 33, No. 1, pp. DSH-E.1-10, 2018.