1K3-OS-17a-2

データベース検索音声対話システムにおける店舗属性値取得のための質問生成

Generating Questions for Acquiring Restaurant Attribute in Spoken Dialogue System

大塚　嗣巳[*1]　駒谷　和範[*1]　佐藤　理史[*1]　中野　幹生[*2]
Tsugumi Otsuka　Kazunori Komatani　Satoshi Sato　Mikio Nakano

[*1]名古屋大学大学院　工学研究科　電子情報システム専攻
Graduate School of Engineering, Nagoya University

[*2]ホンダ・リサーチ・インスティチュート・ジャパン
Honda Research Institute Japan Co., Ltd.

We are trying to acquire attributes of unknown words from users by generating questions more concrete than wh-questions. Such questions are desirable in spoken dialogue systems because they can narrow down the variation of the following user response and accordingly avoid possible speech recognition errors. We propose a method for calculating a well-distributed confidence measure (CM) to generate more concrete questions. The CM is obtained by integrating two basic CMs using (1) word distribution in the database and (2) frequency of occurrence of the restaurant attributes in the Web. As a criterion in experimental evaluation, we use the average error from the logistic regression function estimated on the test data. As a result, the integrated CM outperformed the two basic CMs; the average error became smaller 10% and 7% for each case.

1. はじめに

音声対話システムにおいて，未知語への対応は重要な課題である．ここでの未知語は，システム開発者が，応答に必要な属性を定義していない単語，つまり応答生成部にとっての未知語とする．すなわち，正しい音声認識結果が得られたとしても，システムが応答できない単語である．対話を通じて応答に必要な属性を取得可能とすることにより，対話をするにつれて賢くなるシステムの構築を目指している．

本稿では，未知語が現れた際に，その語に関する属性値を取得するために，適切な質問を生成する手法について述べる．ここでは具体的なタスクとして，レストランデータベース（DB）の検索を扱う．システムに入力され得る未知語は，この DB の主キーである，店舗名，つまりレストラン名とする．質問により取得する属性値として，対象 DB 内に含まれるレストランの属性（ジャンルや最寄駅など）の中から，今回はジャンルを例として考える．

ユーザに対して行う質問は，より具体的である方がよい．例を図 1 を用いて説明する．この図の上部の例では，単純にジャンルを尋ねているのに対して，下部の例では，具体的にイタリアンという候補を提示し，Yes/No 質問を行っている．音声対話システムにおいては，後者の質問の方が，次に続く応答の候補を絞り込めるため，音声認識が容易であるうえ，ユーザがジャンルに関して新たな未知語を回答するリスクを低減できる．

このような具体的な質問を生成するために，未知語として入力されるレストランのジャンルを推定し，推定したジャンルに対して適切な信頼度を付与する．これまでにも数多く，単語の属性を推定する研究は行われている [山本 00, 佐藤 03, 吉永 09]．本研究では，適切な質問の生成が主眼であるため，属性の推定自体には比較的単純な手法を用い，推定結果に対して適切な信頼度を付与することに焦点を当てる．つまり，正しく推定できている場合には高い信頼度を付与し，そもそも推定が困難な場合には低い信頼度を付与することを目指す．

連絡先: 大塚 嗣巳，464-8603 名古屋市千種区不老町 C3-1(631)，
052-789-4435，t_ootuka@nuee.nagoya-u.ac.jp

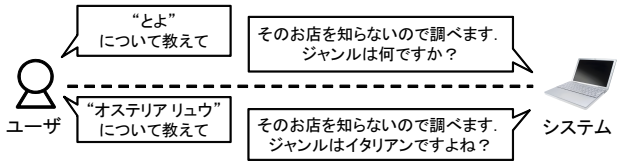

図 1: 単純な質問と具体的な質問の例

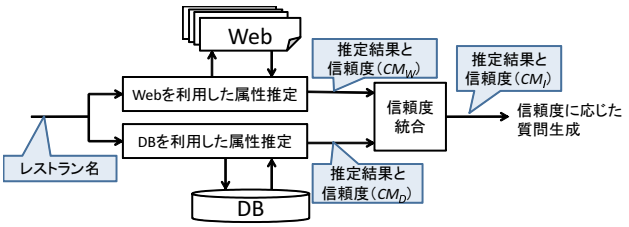

図 2: 提案手法の処理の全体像

2. ジャンルの推定と信頼度の算出

本稿では 2 種類の推定を行い，それらから得られる信頼度を統合して，質問生成に用いる．推定方法は以下の 2 つである．

1. 検索対象 DB 内の文字の分布に基づく推定
2. Web ページ上のジャンルの出現頻度に基づく推定

提案手法における処理の全体像を図 2 に示す．入力は，未知語であるレストラン名である．出力は，推定の信頼度であり，それに応じて質問を生成する．信頼度は，各ジャンル $g_j \in G$ に対して推定する．ここで G は，検索対象のデータベースに存在するジャンルの集合である．

入力にはレストラン名を仮定している．これは，レストラン名が応答生成部にとって未知語である一方で，これに対する

正しい音声認識結果が得られ，かつ，レストラン名であると言語理解部によって正しく認定されることを意味する．これは，超大語彙の音声認識エンジンや，統計的 NE (Name Entity) Tagger を利用すれば，一定の精度で実現可能な問題設定であると考える．

2.1 DB 内の文字の分布に基づくジャンル推定

検索対象 DB 内の文字の分布に基づいて，入力されたレストランのジャンルを推定する．検索対象 DB には，レストラン名とそのジャンルの組が多数収録されていることから，これらを用いて教師付き学習を行い，ジャンルとその信頼度を出力する．これは，レストラン名の文字種や部分文字列から，一部のジャンルが推定可能であるという直感に基づいている．例えば，レストラン名が漢字である場合，和食や中華の店である可能性が高いと推定できることに対応する．

識別器には Maximum Entropy (ME) model [Berger 96] を用いる．これから出力される事後確率 $p(g_j|s_i)$ を，DB を用いた推定における信頼度 $CM_D(s_i, g_j)$ とする．$CM_D(s_i, g_j)$ は以下の式により得られる．

$$CM_D(s_i, g_j) = p(g_j|s_i) = \frac{1}{Z} \exp\left[\vec{\lambda}(g_j) \cdot \vec{\phi}(s_i)\right] \quad (1)$$

ここで s_i はレストラン名，$g_j (\in G)$ は DB 内の各ジャンルを表す．$\vec{\phi}(s_i)$ は素性ベクトル，$\vec{\lambda}(g_j)$ は素性ベクトルに対する重みベクトル，Z は $\sum_{g_j} CM_D(s_i, g_j) = 1$ を保証する正規化係数である．

素性には，個々のレストラン名から，以下を生成して用いる．

- 文字 n-gram $(n = 1, 2, 3)$
- 形態素
- 文字種

文字 n-gram や形態素は，レストラン名にそれらが存在する場合，その素性値を 1 とする．形態素解析には，IPADIC から学習した辞書に基づく Mecab を用いた．文字種は，ひらがな，カタカナ，漢字，アルファベットの有無を，4 個の素性で表現する．例えば「まる寿司」というレストラン名の場合，ひらがなと漢字の素性値を 1 とし，カタカナとアルファベットの素性値を 0 とする．

さらに，この素性集合に対して素性選択を行う．入力されるレストラン名は，DB に存在しないことが前提であるため，過学習を避け，DB に依存しない識別器を構築することが必須である．このために，各素性とジャンルの集合との相互情報量を計算する．これにより，例えば,「まる寿司」というレストラン名から生成される素性のうち，ジャンル「和食」と共起することの多い 2-gram「寿司」はジャンル推定に有効だが，2-gram「まる」は有効ではない，などと判断できる．相互情報量の計算式を，式 (2) に示す．

$$I(f_k; G) = \sum_{g_j \in G} p(f_k, g_j) \log \frac{p(f_k, g_j)}{p(f_k)p(g_j)} \quad (2)$$

ここで $p(f_k)$，$p(g_j)$ は，学習データから生成される素性 f_k とジャンル $g_j (\in G)$ それぞれの生起確率，$p(f_k, g_j)$ はそれらの同時確率を表す．

各素性 f_k に対する相互情報量に基づき，この値の低いものから順に，素性を除外していく．これを，過学習が回避できていると判断できるまで行う．具体的には，closed テストと open テスト（実際には交差検定）の正解率の差が，ほぼなくなった場合に，過学習が回避できているとみなせる．

2.2 Web 上の出現頻度を利用したジャンル推定

Web 上の出現頻度を利用して，レストランのジャンルを推定し，その信頼度を得る．これは，Web ページにおいて，レストラン名がそのジャンルと共起することを仮定し，これを利用している．具体的には，以下の手順で，信頼度 $CM_W(g_j)$ を得る．

1. Web 検索 API を通じて，入力されたレストラン名に関する Web ページを取得する．本研究のドメインは，愛知県内のレストランであるため，検索クエリを「`<restaurant>` 愛知県　レストラン」とした．`<restaurant>`には，入力であるレストラン名が入る．
2. 取得した i 位のページにおける，各ジャンル名 g_j の出現頻度 $freq_i(g_j)$ を求める．i は Web 検索 API が出力するページの順位である．
3. 出現頻度を，得られたページに関して総和を求めた後に事後確率化し，信頼度 $CM_W(g_j)$ とする．この際に，各ページ内の出現頻度情報に応じた重み w_i を与える．

$$CM_W(g_j) = \frac{\sum_i w_i \cdot freq_i(g_j)}{\sum_{g_j} \sum_i w_i \cdot freq_i(g_j)} \quad (3)$$

重み w_i は，以下の 2 要素を用いて決定する．

1. 検索 API が出力するページ順位 i
   検索 API から上位として出力されるページほど，当該レストランへの関連が深い．
2. i 番目のページに出現するジャンルの異なり数 $genre(i)$
   ポータルのページなど，多くのジャンルが出現するページは，特定のジャンルを示す度合が低いと考える．例えば,「中華」のみ出現するページと,「中華」「和食」「居酒屋」「洋食」の 4 つが出現するページを比較した場合，前者の方が「中華」のレストランに関するページである可能性が高い．

上記を考慮して，以下の式により w_i を計算する．

$$w_i = \frac{1}{i \cdot genre(i)} \quad (4)$$

本稿における実装では，Web 検索 API として，Yahoo!株式会社のもの[*1]を使用した．取得するページ数は，検索クエリ 1 件に対し，20 ページとした．これは API から一度に取得できるページ数の，実験当時の上限である．

2.3 2 つの信頼度の統合

2 つの推定結果に基づく信頼度 CM_D, CM_W を統合し，最終的な信頼度 CM_I を得る．本稿では，統合方法として，2 つの信頼度の重み付き和を採用した．

重みは，データに基づき，最適な値を計算して設定する．つまり，重みの最適値を，推定結果の正解を教師信号として与えたデータ集合を用いて定める．具体的には，推定が正解であるジャンルに 1，誤りであるジャンルに 0 を与え，その場合の CM_D と CM_W を入力として，ロジスティック回帰を行う．計算式を式 (5) に示す．これにより，2 つの信頼度に対する最適な重み w_D，w_W，および定数項 w_0 を求める．

$$CM_I(g_i) = \frac{1}{1 + exp\{-(w_D CM_D + w_W CM_W + w_0)\}} \quad (5)$$

得られた最適な重み w_D，w_W，w_0 を用いて，式 (5) に基づいて 2 つの信頼度を統合し，最終的な信頼度 CM_I を得る．

*1 http://developer.yahoo.co.jp/webapi/search/websearch/v2/websearch.html

表 1: 質問に含めるジャンル数 num による質問形式とその例

num	質問形式	応答例
1	Yes/No 質問	ジャンルは g_1 ですよね？
2	2 者択一	g_1，g_2 のどちらですか？
3	3 者択一	g_1，g_2，g_3 のどれですか？
4 以上	Wh 質問	ジャンルは何ですか？

2.4 推定の信頼度に基づく質問生成

統合した信頼度 $CM_I(g_j)$ に基づき，質問形式を選択する．信頼度が高いジャンルが得られた場合には，質問文に含めるジャンルの数を絞り，より具体的な質問を生成する．一方，信頼度の低いジャンルしか得られなかった場合には，複数のジャンルを含んだ質問を生成する．質問に含めるジャンルの数 num に基づく質問文の例を，表 1 に示す．

質問に含めるジャンルの数 num を，信頼度 $CM_I(g_j)$ に基づき，式 (6) により決定する．なおここでは，g_j は $CM_I(g_j)$ の降順に並べられているとする．

$$num = min(n) \text{ s.t. } \sum_{j=1}^{n} CM_I(g_j) > \theta \tag{6}$$

θ は定数であり，信頼度 $CM_I(g_j)$ の分布を考慮して，人手で決定するものとする．式 (6) が意味するのは，例えば，$n = 1$，つまり 1 位のジャンルの信頼度 $CM_I(g_1)$ だけで θ を超えているとき，そのジャンルのみを含む具体的な質問を行う．一方，$n = 4$，つまり 4 位までのジャンルの信頼度 $CM_I(g_j)$ を加算しても θ を超えない場合は，推定したジャンルを用いず，Wh 質問を行うことになる．

3. 評価実験

具体的な質問の生成に用いる信頼度の評価実験を行った．まず，検索対象 DB を用いたジャンル推定と信頼度の算出において，相互情報量を用いた素性選択による過学習の回避を検証する．次に，得られた信頼度が，質問の生成において，適切な尺度となっているかどうかを評価する．最後に，2 つの信頼度を統合することの有効性を示す．

本稿では，検索対象 DB として，愛知県のレストラン DB [西村 12] を用いた．この DB には，2398 件のレストランが掲載されている．DB 内のジャンルは 16 種類，つまり $|G| = 16$ である．

3.1 相互情報量を用いた素性選択による過学習の回避

検索対象 DB を用いたジャンル推定において，相互情報量を用いた素性選択により，過学習が回避できるかどうかを検証する．closed テストと open テストの正解率がほぼ同等の場合には，過学習が回避できているとみなせる．ここでは，closed テストとして，検索対象 DB 内の全エントリ 2398 件を用いて識別器を学習し，同じデータに対して推定精度を算出した．一方 open テストとしては，同じ 2398 件のデータに対して，10 分割交差検定を行い，推定精度を算出した．

特徴選択を行った場合の，closed テストと 10 分割交差検定の正解率を，図 3 に示す．横軸は，全 20679 種類の素性のうち，相互情報量の上位 x%を使用したことを示し，縦軸は，それぞれの場合のジャンルの推定精度を示す．まず，素性を全て用いた場合，つまり横軸が 100%の時，closed テストと 10 分割交差検定の精度の差が 28.1%であることから，過学習が起こっていたことがわかる．次に，用いる素性を，相互情報量が

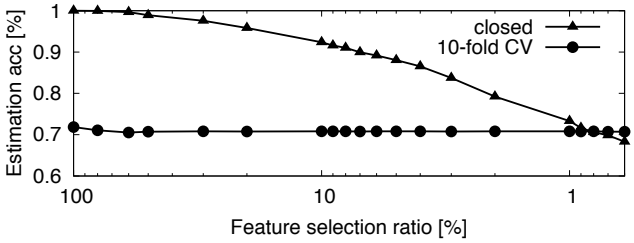

図 3: 素性選択時の closed テストと 10 分割交差検定の正解率

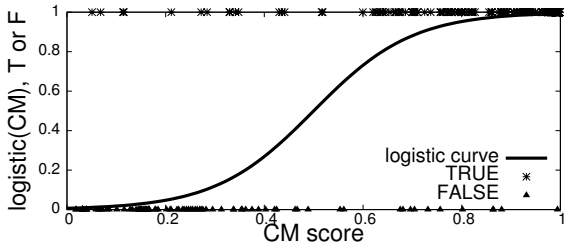

図 4: 評価対象データと最尤のロジスティック曲線の例

小さいものから順に除外していくと，$x = 0.8$ の時に，closed テストと 10 分割交差検定の間の推定精度の差がほぼなくなった．この結果得られた 165 種類の素性を，過学習が生じていない素性集合として，以降で用いる．この素性の中には，例えば，餃子という 2-gram など，直感的にもジャンル推定に有効である素性が含まれていた．

3.2 信頼度の分布の評価

推定結果とともに得られる信頼度の分布を評価する．具体的には，DB，Web のいずれかを用いた推定により得た信頼度 CM_D，CM_W と，統合した信頼度 CM_I の，3 種類の分布を評価した．評価対象データには，検索対象 DB のエントリのうち，400 件のエントリを抽出して用いた．残りの 1998 件は，CM_D を計算する際の学習データとして用いた．

評価対象データをプロットしたものの例を，図 4 に示す．図の横軸は信頼度 $CM(g_j)$ の値，縦軸は推定ジャンル g_j の正誤を表す．つまり，正解の場合は 1，誤りであった場合は 0 としている．さらに，このデータに対して最尤となるロジスティック曲線を，同じ図 4 に示している．このロジスティック曲線の式を，式 (7) に示す．最尤となる重み計数 w_0, w_1 は，Weka のロジスティック回帰を学習するモジュールを用いて求めた．

$$logistic(CM) = \frac{1}{1 + exp\{-(w_0 + w_1 CM)\}} \tag{7}$$

本稿では，信頼度の分布の評価尺度を，各信頼度 CM の値と，式 (7) で表される最尤のロジスティック曲線との，誤差の絶対値平均とする．ここでは，信頼度の理想的な分布は，評価データに対して最尤のロジスティック曲線となると仮定している．これにより，最尤のロジスティック曲線と各プロットとの距離を，信頼度の分布の評価尺度 $eval(CM)$ とする．

$$eval(CM) = \frac{\sum_i^N |logistic(CM^i) - \phi(i)|}{N} \tag{8}$$

N はテストデータの総数，つまりここでは $N = 400$ である．添字 i は，N 個のテストデータのうちの 1 つを表し，CM^i は

表 2: 各信頼度に対するの評価値

	$eval(CM_x)$
CM_D	0.30
CM_W	0.29
CM_I	0.27

表 3: 各手法における信頼度ごとの正解数

	CM_D		CM_W		CM_I	
各 CM の値	正解	誤り	正解	誤り	正解	誤り
0.0 – 0.1	0	0	0	32	2	10
0.1 – 0.2	0	0	0	11	9	15
0.2 – 0.3	1	16	14	22	15	18
0.3 – 0.4	6	19	28	19	10	8
0.4 – 0.5	11	25	29	21	13	12
0.5 – 0.6	21	29	56	9	13	12
0.6 – 0.7	22	28	85	7	15	7
0.7 – 0.8	41	16	42	3	17	6
0.8 – 0.9	21	9	19	1	19	9
0.9 – 1.0	131	4	1	1	184	10
合計	254	146	274	124	297	103

その信頼度を表す．$\phi(i)$ は以下のように定義する．

$$\phi(i) = \begin{cases} 1 & \text{推定結果 } i \text{ が正解であった場合} \\ 0 & \text{推定結果 } i \text{ が誤りであった場合} \end{cases} \quad (9)$$

具体的に説明すると，信頼度が高いにも関わらず推定が誤りである場合や，信頼度が低い際に推定が正解である場合には，式 (8) 中の $|logistic(CM_i) - \phi(i)|$ の値が大きくなる．このように不適切な信頼度が出力されているほど，$eval(CM)$ の値は大きくなる．信頼度の分布が適切である場合には，$eval(CM)$ の値は小さくなる．

表 2 に各信頼度の分布の評価結果を示す．ここでは，CM_D，CM_W，CM_I それぞれに対する，式 (8) の $eval(CM)$ の値を示している．この結果，本稿での評価尺度では，統合した信頼度 CM_I が最も良いことを確認した．統合した信頼度 CM_I の，CM_D，CM_W に対する誤り削減率は，それぞれ 10%，7%であった．

3.3 信頼度の統合の有効性

2 つの信頼度の統合の有効性を，別の視点から検証する．具体的には，統合により正解数が増えているかどうかを確かめる．

まず表 3 に，CM_D，CM_W，CM_I それぞれの，信頼度の分布と推定の正解数を示す．分布は，$CM(g_j)$ が最大であったジャンル g_j について，その $CM(g_j)$ の値と推定結果であるジャンル g_j の正誤を，集計したものである．最下欄の合計値を見ると，CM_I の正解数が 297 と最も高いことがわかる．これは統合により，正解ジャンル $\overline{g}$ の信頼度 $CM(\overline{g})$ の順位が，統合前と比べて相対的に上昇したことを示しており，統合の有効性を示している．なお，出力された各信頼度の分布を見ると，CM_W の値の分布（例えば平均値）は，他の 2 つに比べて小さいことが見て取れる．このことから，2.3 節で示したように，統合する際に，単純に 2 つの信頼度を加算するのではなく，重み付き和とすることの必要性が示されている．

さらに具体的に，統合の前後における，推定の正誤の変化を調べた．結果を表 4 に示す．表中の○は，信頼度が最大であったジャンルが正解であった場合，×は誤りであった場合を表す．列×／○と列○／×は，DB と Web のどちらか一方でのみ，レストランのジャンルを正しく推定できたことを示す．

表 4: 各推定結果による正誤の変化

		CM_D による正誤／ CM_W による正誤			
		×／×	×／○	○／×	○／○
CM_I による正誤	○	0	51	33	213
	×	85	10	8	0

このようなレストランが，400 件のうち 102 件存在した．このうち，84 件（82%）に対しては，統合によって推定結果が正解へと変化している．このことから，CM_D と CM_W は相補的であり，これらを統合することの効果が確認できている．

次の 2 例は，いずれも統合により推定結果が正解となったものである．まず,「加屋」というレストランの正解ジャンルは「お好み焼き もんじゃ焼き 鉄板焼き」である．これに対し CM_D を用いた場合では，誤って「居酒屋」と推定された．一方 CM_W を用いた場合では，正しく推定できていた．また,「玉寿司 今池」というレストランの正解ジャンルは「和食」である．これに対して，CM_D では正しく推定できていたものの，CM_W を用いると「居酒屋」であると誤って推定した．このように，CM_D と CM_W を統合した信頼度により，両推定方法の長所を取り入れていることがわかる．

4. まとめ

システムにとって未知のレストランに関して，ユーザとの対話を通じて，その属性の獲得を試みた．具体的には，ジャンルを対象とし，その推定の信頼度に応じた具体的な質問を行う方法を述べた．DB，Web に基づく 2 種類の推定を行い，その結果を統合して，信頼度を算出した．評価の結果，統合により，分布と正解率の両面で，より良い信頼度が算出できることを確認した．今後は，ジャンル以外の属性値の推定方法を考案する．

参考文献

[Berger 96] Berger, A. L., Pietra, V. J. D., and Pietra, S. A. D.: A maximum entropy approach to natural language processing, *Computational Linguistics*, Vol. 22, No. 1, pp. 39–71 (1996)

[吉永 09] 吉永 直樹, 鳥澤 健太郎, Web からの具体物の属性・属性値情報の自動獲得, 言語処理学会第 13 回年次大会発表論文集, pp. 887–890 (2009)

[佐藤 03] 佐藤 理史, 佐々木 靖弘, ウェブを利用した関連用語の自動収集, 情報処理学会研究報告, 2003-NL-153-8, pp. 57–64 (2003)

[山本 00] 山本 あゆみ, 佐藤 理史, ワールドワイドウェブからの人物情報の自動収集, 情報処理学会研究報告, 2000-ICS-119-23 (2000)

[西村 12] 西村 良太, 駒谷 和範, データベース検索音声対話システムにおける対話状態の推定, 情報処理学会研究報告, 2012-SLP-90-20 (2012)