

リアルタイムな質問に回答するための日替わり語の獲得

貞光 九月 東中 竜一郎 松尾 義博

NTT メディアインテリジェンス研究所

{sadamitsu.kugatsu,higashinaka.ryuichiro,matsuo.yoshihiro}
@lab.ntt.co.jp

1 はじめに

近年音声対話エージェントへ注目が集まっており、様々な場面での実用化が進んでいる。スマートフォン上で動作するバーチャルエージェントである、Apple 社の Siri、Microsoft 社の Cortana、NTT ドコモ社のしゃべってコンシェルや、ハードウェアを伴った音声対話エージェントである、ソフトバンク社の Pepper、Amazon 社の Echo、Jibo 社の Jibo 等も多く実用化されている。音声対話エージェントにとって基本的かつ難しいタスクの一つに質問応答が挙げられる。質問が明確な場合、現在の質問応答システムでも回答可能であるものの、ユーザがより高い満足度を得るためには、質問自体には含まれない非明示的な意図を推測することが重要となる。例えば「巨人の先発を教えてください」という質問文に対し、論理的には当該チームに在籍した過去の先発投手全てが回答となり得るが、10年前のある日の先発投手の名前を提示したとしても、ユーザの満足度は決して高くないだろう。一方、人間同士でこのような質問が投げかけられた場合、その背景には「今日の先発」というように、回答をリアルタイムな情報に制約する意図、文脈が作用することが多い。機械でもこのような非明示的な意図を読み取ることで、ユーザの満足度を向上させられると考えられる。

本稿では非明示的な質問意図の一つとして、上記例のように、ユーザがリアルタイムな情報を尋ねているか否かを判別し、その判別結果に応じて回答の探索範囲を切り替えることで、よりの確に回答を返すことが可能なリアルタイム質問応答システムを提案する。リアルタイム質問応答において達成すべき課題は2つある。1つ目の課題はリアルタイム性を有する質問を自動的に判別する質問文のリアルタイム性判定の実現である。我々は、リアルタイムな情報を尋ねる質問文を収集・分析を行った結果、それらの文には「ゲスト、先発、試合、気温」のように、語の指す対象が日によって変化する語が含まれている傾向が見られた。我々はこれを日替わり語 (daily changing words) と呼び、以下でもその呼称を用いる。日替わり語は、質問文の中に含ま

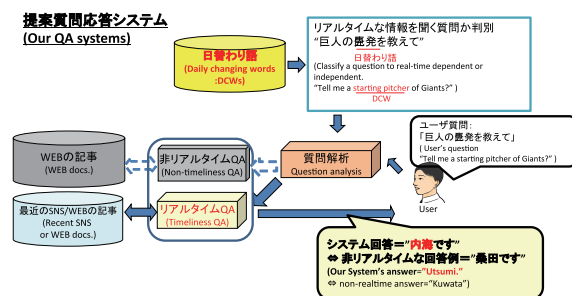


図 1: 提案質問応答システムの全体像 (Illustration of our system.)

れるリアルタイム性を暗に示す情報と考えられるため、本稿では日替わり語を自動的に獲得・利用することで、精度良く分類を行う手法を提案する。2つ目の課題は質問文のリアルタイム性が判別できたとして、判別結果に応じた質問応答を行うシステムの構築である。本稿では、検索システムをベースとする質問応答システムを用いて、その検索対象リソースを切り替えることで全体システムを構築することができることを示す。本提案システムは既に、NTT ドコモ社のしゃべってコンシェルに導入されており [Higashinaka et al.2013]。実用の中から課題を抽出し、その解決策を示した研究の一例と言えるだろう。

以下2節ではリアルタイム質問応答システムの全体像とそこで必要となる日替わり語について述べ、従来の語クラスとの類似点、相違点について述べる。3節で日替わり語の具体的な獲得手法を述べた後、4節にて実験を行い、日替わり語のリアルタイム性判定における有効性を示す。

2 リアルタイム質問応答システム

本節では、図1を用いて、リアルタイム質問応答を実現するためのシステムの全体像を説明する。まず図中右側に位置するユーザが発話入力あるいはテキスト入力をを行い、質問をシステムに投げかける(例:「巨

人の先発を教えて」). 質問解析部では入力された質問文がリアルタイム性を有するか否かを判定し, 後段のリアルタイム QA あるいは非リアルタイム QA に処理を渡す. リアルタイム QA・非リアルタイム QA のいずれも検索システムベースで質問応答機能を実現している点でほぼ同じであるが, リアルタイム性有と判別された質問文に対しては, 同じくリアルタイム性の高い情報が多く含まれる SNS データや, タイムスタンプを見て最近記述された記事のみに対象を絞り込む. 一方, リアルタイム性が無いと判別された質問文に対しては, 特に制約をかけずに一般の検索エンジンを用いる. 検索型 QA システムの詳細については [Higashinaka et al.2013] を参照されたい. 結果的に, リアルタイム性を有すると判定された場合は, 最近のリソースから回答が導かれ(「内海です」), 古いリソースから得られる回答はフィルタされる(「桑田です」).

質問応答において, 質問文を何かしらの基準で分類するアプローチは一般的に採用される方法であり, Ferrucci らは lexical answer types (LATS) を用い [Ferruci et al.2012], 東中らは固有表現タイプを用いている [Higashinaka et al.2013]. しかし我々の知る限り, 質問文のリアルタイム性に着目した研究は存在しなかった.

2.1 リアルタイム性判定のための手がかり語の考察

本節ではリアルタイム性判定の方法について述べる. リアルタイム性判定において重要な点は, ユーザの短い発話中から, いかにユーザの意図を包含する文字列を抽出するかである. そのような重要な語のリストが作成できた後, キーワードマッチングとして用いたり, 機械学習における一つの素性として用いることでリアルタイム性判定を実現できる.

我々はまず, リアルタイム性を示す手がかり語を, 大きく3つのタイプに分類した. 1つ目は「今日」や「最近」のように, 近い時間範囲を直接的に示す語であり, カレントワードと呼ぶ. 質問文にカレントワードを含む場合には, 語が明示する通り, リアルタイム性を伴うと考えて良いだろう. 2つ目は近い時間範囲で特定の単語が高い頻度で現れるバースト現象に対応するためのバーストワードである. これらバーストワードが含まれる場合にも, 同じくリアルタイム性のある質問の可能性は高いと言える.

そして3つ目が日替わり語である! 巨人の先発投手は誰?」のように, 質問文中に明らかなリアルタイム性を示す単語を含まない文であっても, 人間であればリアルタイム性を有すると判断可能な文が存在する.

これは「先発」という単語の参照先が, 短い時間間隔で変化することを, 人間が経験的に知っているからと考えられる. 一方! 仙台のランドマークはどこ?」の場合「ランドマーク」の参照先が頻繁に変わる対象ではないため, リアルタイム性はない, と判断できる.

リアルタイム性判定では, 以上3種類の語が含まれている場合にリアルタイム性有となるように, キーワードマッチベース, あるいは機械学習ベースの識別器を構成する. ただし, 明示的に過去の日時を示す発話が含まれる場合には, 非リアルタイムな質問であるとの判別を行うこととする(例:「昭和58年オールスター最終戦の先発は?」) カレントワードやバーストワードについては, 既存辞書やバースト時の検索ランキング等から得ることが比較的容易であるため, 以下の節では, 残る日替わり語の特徴や獲得方法について検討を加えていく.

2.2 日替わり語と従来の語クラスとの比較

日替わり語は単語の集合であるため, 既存の単語クラスに近いものがないか調査を行った. はじめに2700以上の単語クラスを含む日本語語彙大系 [池原 他 1997] において「高揚する, 引退する」といった用言の属するクラスである「属性変化」が挙げられる. 「属性変化」の概念には「日替わり語」の概念と重複する部分が存在し, 実際に「引退する」(例:「巨人を引退したのは誰」)は日替わり語として含めるべき単語と考えられるが, 一方で「先発」のような体言を含まない点で異なる. また「高揚する」のように個人的経験に依る語も, そもそも質問文として入力されづらいという特徴があり, 我々の収集したい日替わり語とは合致しないものであった. 実験節では, これらの違いを実証的に明らかにするため「属性変化」クラスをリアルタイム性の手がかり語とみなした場合の実験を行い比較を行った.

次に既存タスクとの類似を見る. TAC 2011 および TAC 2013 の Knowledge-Base Population (KBP) ¹ で取り扱われた “temporal slot filling” は比較的近いタスクである [Ji et al.2011]. 本タスクの目的は, 任意のエンティティに対し, 定義されたスロット(属性)に対する時間情報を埋めることである. 例えば, ある人物が任意の組織に所属した「開始時期」と「終了時期」を埋める, というタスクである. KBP では計8種類のスロットのみを対象としており², これらのスロット名(スロットの値ではなく)は日替わり語に似た部分

¹<http://www.nist.gov/tac/2013/KBP/>

²PERSON: spouse, PERSON: title, PERSON: employee of, PERSON: member of, PERSON: cities of residence, PERSON: state or province of residence, PERSON: countries of residence and ORGANIZATION: top employee member.

表 1: 日替わり語獲得のためのテンプレート。(TIME), (NE), (DCW) はそれぞれ日時表現, 固有表現, 日替わり語のスロット化される部分を示す。

Table 1: A list of templates for extracting Daily Changing Words. Upper line is original Japanese pattern and lower line is English translation. (TIME), (NE) and (DCW) indicate the slots of time, named entity, and Daily Changing Word, respectively.

Extraction template	Examples of extracted strings	DCW candidates
(TIME) の (DCW) は (NE)	(TIME=今日) の (DCW=ゲスト) は (NE=田中)	ゲスト
(TIME) 's (DCW) be (NE)	“(TIME=Today)’s (DCW=guest) is (NE=Tanaka).”	guest
(TIME) (DCW) する (NE)	(TIME=昨日) (DCW=発表) された (NE=iPhone)	発表
(TIME), (NE) be (DCW)	“(TIME=Yesterday), (NE=iPhone) was (DCW=announced).”	announce

もある。しかし日替わり語が短い期間で参照先の値が変化するのに対し, temporal slot filling のスロットでは, 変化の期間は特に考慮していない。temporal slot filling のスロットの取り得る値も踏まえた上で, 所望の期間で変化するスロット名に絞り込むことも可能だが, temporal slot filling のスロット名の粒度は 8 種類と粗いため, リアルタイム性を判定するための語を網羅できない。

別のアプローチとしては時制を扱う方法も考えられる。英語の場合であれば, 質問文が “Who is” と始まれば現在のことを尋ねていることが自明である³。しかし日本語の場合, 時制は動詞の活用形で表現される場合が多く, さらに時制の用法も厳密ではない。

先行研究では日本語の時制の分析, 特に第二言語習得者向けの研究が成されている。[Shirai and Kurono1998][Ishida2004]。また, 日本語機能表現辞書「つつじ」[松吉 他 2007] に基づいた意味ラベル推定の研究 [今村 他 2011] は, 時制の自動推定を可能とするものである。これらの研究は主に動詞と機能動詞に着目したものであるのに対し, 我々は動詞に限らず, 名詞等も含め, 直接的にリアルタイム性のある質問の発見につながる単語の収集を目的とする点で異なる。

3 大規模教師なしコーパスからの日替わり語の自動獲得

本節では教師なしコーパスからの, ルールに基づく日替わり語の自動獲得手法について述べる。具体例として, ここでは「ゲスト」という単語を含む 2 文を取り上げる。「昨日のゲストは田中さんでした」「今日のゲストは鈴木さんでした」この 2 文においては日時表現 (昨日, 今日) と固有表現 (田中さん, 鈴木さん) が差分として表れている。このような事例より, 日替わり語の周辺には, 多様な日時表現と, それに呼応し

て変化する固有表現が含まれる可能性が高い。そこで我々は, 日替わり語を獲得するために, 日時表現のスロットと固有表現スロットを含むテンプレートを設計し, これを大規模なテキストに適用する。具体的なテンプレートを表 1 に示す。上段は一般名詞を獲得するためのもので, 下段は動作性名詞を獲得することが可能である。日時表現と固有表現の抽出には固有表現抽出器を用いた [Suzuki et al.2006]。なお, 本稿では固有表現として, IREX で定義された固有表現のうち, “人名”, “組織名”, “地名”, “人工物名” の 4 種を扱う。また, 日替わり語が頻繁に対象を変えするという特徴を反映するため, テンプレートを大規模テキストに適用し, 日替わり語候補が得られた後, 固有表現, 時刻表現の種類が少ないものを, 日替わり語候補から除外した。

ここまでで自動獲得された結果を分析したところ, 「起床」のように, 音声対話における質問応答ではあまり尋ねられない単語が含まれていた。これは, 利用者が音声対話による質問応答において, 公共性のある情報を聞きやすい, という特徴によると考えられ, 我々はこの点についても獲得手法に反映させることを試みた。具体的には, 新聞記事を用いることで, 手がかり語候補の中から公共性のない語をフィルタリングを行う。新聞記事には公共性の高い内容が記述されていることから, 新聞記事において十分な頻度 (実験節では 10 回に設定) 出現する場合には, それらを公共性の高い語として採用し, 一方で頻度の低い候補については棄却するとした。

4 実験と考察

提案手法の有効性を確認するため, 本節では 2 通りの実験を行う。1 つ目は提案手法によって獲得された日替わり語の獲得精度についての検証, 2 つ目は日替わり語を実際にリアルタイムな質問の検出に用いた実験である。

日替わり語の獲得には, 教師なしコーパスとして 2012 年 2 月にクロールした 3500 万のブログ記事と,

³“Tell me X’s wife” と尋ねられた場合は日本語と同様の問題が生じる

表 3: リアルタイム質問の検出結果 “sys.=T” と “sys.=N” はシステム推定結果が，リアルタイム性有，リアルタイム性無であることをそれぞれ示す．“ref.=T” と “ref.=N” は正解が，リアルタイム性有，リアルタイム性無であることをそれぞれ示す．提案手法 (4 列目) とベースライン手法 (3 列目) の間で二項検定を行った結果， $p < 0.01$ で有意であった．

Table 3: Results for the detection of timeliness questions. “sys.=T” and “sys.=N” indicate the system estimation is timeliness question and non-timeliness question respectively. “ref.=T” and “ref.=N” indicate the annotated reference is timeliness question and not timeliness question. The difference in accuracy between the proposed method (fourth) and the baseline method(third) is significant by binominal test with $p < 0.01$.

	Keyword matching method				Machine learning method			
	Goi-Taikai (baseline)		DCW (proposed)		Supervised (baseline)		Supervised + DCW (proposed)	
	sys.=T	sys.=N	sys.=T	sys.=N	sys.=T	sys.=N	sys.=T	sys.=N
ref.=T	2	56	23	35	40	18	49	9
ref.=N	0	31	1	30	14	17	14	17
total	Prec.=1.000, Rec.=0.035, F val.=0.068		Prec.=0.958, Rec.=0.397, F val.=0.561		Prec.=0.740, Rec.=0.690, F val.=0.714		Prec.=0.778, Rec.=0.845, F val.=0.810	

表 2: 日替わり語の自動獲得例

Table 2: Extraction results of daily changing words.

獲得された日替わり語 (一般名詞) 全 44 語よりサンプリング Extracted Daily Changing Words (general noun) (samples from 44 correct extractions) ゲスト (guest), 先発 (starting pitcher), 試合 (game), 主役 (leading role), 話題 (topics) 日没 (sunset), 衣装 (costume), 司会 (MC)
獲得された日替わり語 (動作性名詞) 全 34 語よりサンプリング Extracted Daily Changing Words (verbal noun) samples from 34 correct extraction 発売 (launching), 発表 (announcement), 出演 (appearance), 公開 (release), 結婚 (marriage), 完成 (accomplishment), 来日 (visiting Japan)
獲得された誤り例 Extracted NOT Daily Changing Words 入手 (obtaining), 創業 (establishment)

フィルタリング用に 2012 年 6 月から 7 月にクロールした 30,465 の新聞記事を用いたテスト用質問文には，別途 86 の質問文を複数作業員で作成し，質問文作成者とは異なる複数作業員によって正解ラベルを付与したものをを用いた．質問文作成時，カレントワードとバーストワードは含めないこととし，またそれぞれの質問文は独立に作成され，質問文間の連続性はないとした．最終的に，リアルタイム性有り質問文 55 文と，リアルタイム性無し質問文 33 文を作成した．

4.1 日替わり語の獲得実験

表 2 に日替わり語の獲得結果の具体例を示す．提案手法により 82 個の日替わり語を獲得することができ，その正解率は 97.6% (80 / 82) であった．2 つの誤りは “創業” と “入手” であった．このような誤った獲得を抑えるためには，テンプレート中の日時表現のスロットを「今日」や「昨日」といった最近の日時を表す単語を含むように制約を加える方が考えられる．一方で獲得数が減少してカバレッジが下がるというデメリットも生じるため，今回は適用しなかった．

各処理ステップにおける獲得精度の詳細は以下の通りである．まず，獲得テンプレートだけを用いた日替わり語候補の獲得では，合計 154 語を獲得した．このうち 95 語が正解であり精度は 61.7% であった．次に，新聞記事を用いたフィルタリングにより，72 個の候補語がフィルタリングされた．このうち 59 語は日替わり語として不適切な語，13 語は適切な日替わり語で以下のようなものを含んでいた．“MVP”，“テレビ”，“ニュース” and “急逝”．フィルタリング単独の精度は 81.9% であった．

4.2 リアルタイムな質問の検出実験

リアルタイムな質問文の検出のために，本節では 4 つの手法を比較する．

1. 各質問文に日本語語彙大系の「属性変化」クラスの語を含むか否かのキーワードマッチ (2.2 節に上げたもの．ベースライン)，

表 4: リアルタイム質問検出結果の正解例及び不正解例。獲得した日替わり語を教師あり学習に用いた場合を示す。アスタリスク (*) は日替わり語を用いない場合とと比較して改善した例を示す。下線部は日替わり語を示す。

Table 4: Examples of correct and incorrect detection of timeliness questions using a supervised method with Daily Changing Word (DCW) in Table 3. The asterisk (*) indicates improved example by adding Daily Changing Words to the result of the supervised method from only the supervised method. The words underlined are DCWs.

	sys.=T	sys.=N
ref. =T	<p>ソニーが <u>発表</u> した新製品は？ -Tell me the new product <u>announced</u> by SONY? NTT の <u>終値</u> は？* -Tell me the <u>closing stock price</u> of NTT? * 次の iPad はいつ <u>発売</u>？* -Tell me the <u>release date</u> of next iPad? * 読売新聞朝刊、一面記事は何？ -Tell me the article on the front-page of Yomiuri's morning paper?</p>	<p>誕生花は何？ -Tell me the birth flower (of today)? アントラーズ戦の審判は誰？ -Tell me the umpier of the game of Antlers? 巨人のマジックは？ -Tell me the magic number of Giants?</p>
ref. =N	<p>一番長い氷河はどれ -Tell me the longest glacier? ドラえもん の誕生日はいつ？ -Tell me the birthday of Doraemon?</p>	<p>秋の花といえば -Tell me a flower of fall? 島根県出身の首相は何人 -Tell me the number of prime ministers elected from Shimane?</p>

- 各質問文に日替わり語を含むか否かのキーワードマッチ (提案手法),
- ロジスティック回帰モデルを用いた教師あり学習 (ベースライン),
- 教師あり学習に加え, 日替わり語の一致結果を追加 (提案手法).

日替わり語として, 前節で得られた日替わり語の他に, wikipedia に掲載される実験期間に放映されたテレビ番組名を含めた。これは当該コンテンツが毎日, 毎週変化するためである。語彙大系を用いた単語クラスの作成においては, 用言クラス「属性変化」に属する語 645 単語から, 末尾の「する」を除外して用いた。教師ありアプローチでは素性に Bag of words の素性を用いて LIBLINEAR の SVM [Fan et al.2008] による学習・評価を行った。

表 3 に各手法に基づく結果を示す。キーワードマッチベースの手法は (1, 2 列目) は高い適合率を示し, 再現率において提案手法は語彙大系を大きく上回った。教師ありアプローチにおいても, 日替わり語を考慮することで再現率を向上させることができている。80 語の日替わり語を追加するだけで, 再現率は 15% (0.690 から 0.845) 改善しており, 提案手法の効果が確認された。

提案手法 (4) によって検出された質問文の例を表 4 に示す。正解例 (表 4 で “sys.=T” and “ref.=T” の交差する箇所) のうちアスタリスクが付与されるものは, 日替わり語の追加により改善した事例である。下線部

の日替わり語を用いることで, 偽負例を改善することができた (18 例から 9 例に減少) Table 3)。残る偽負例 (“sys.=N” and “ref.=T” の交差) は “審判”, “マジック” and “誕生花” といったコーパス中での出現頻度の低い単語を含むものであった。全 14 の偽正例 (intersection of “sys.=T” and “ref.=N”) は全て教師あり学習によって検出されたものであり, 日替わり語からの湧き出しは発生しなかった。以上の結果から, 提案手法によって, 特に再現率の観点においてリアルタイム性を有する質問文の検出精度向上が成された。

5 おわりに

本稿ではリアルタイム性のある質問を判別するための手がかりとして, 新たな単語クラスである日替わり語を用いることと, 日替わり語の自動獲得手法, 及びリアルタイム性のある質問判別への適用について述べた。実験において, 82 個の日替わり語を自動獲得できることを確認し, それらが判別において有意な精度向上に寄与することを確認した。また, 既存の単語クラス, 語彙大系の語クラスを代用する場合に比べても精度が改善することを確認した。

今後の課題は, 獲得手法の洗練による日替わり語のより広範な拡張が挙げられる。また, 日替わり語の各々の参照先の値の変化期間に応じた細分類によって, より細かい粒度でのユーザ意図の把握が可能になると考える。

参考文献

- [Fan et al.2008] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874.
- [Ferruci et al.2012] David Ferruci, Anthony Levas, Sugato Bagchi, David Gondek, and Erik Mueller. 2012. Watson: Beyond Jeopardy! *Artificial Intelligence*.
- [Higashinaka et al.2013] Ryuichiro Higashinaka, Kugatsu Sadamitsu, Kuniko Saito, and Nozomi Kobayashi. 2013. Question answering technology for pinpointing answers to a wide range of questions. *NTT Technical Review*, 11(7).
- [池原 他 1997] 池原 悟, 宮崎 正弘, 白井 諭, 横尾 昭男, 中岩 浩巳, 小倉 健太郎, 大山 芳史, 林 良彦 1997. 日本語語彙大系.
- [Ishida2004] Midori Ishida. 2004. Effects of recasts on the acquisition of the aspectual form -te i-(ru) by learners of japanese as a foreign language. *Language Learning*, 54(2):311–394.
- [Ji et al.2011] Heng Ji, Ralph Grishman, and Hoa Trang Dang. 2011. Overview of the TAC2011 knowledge base population track. In *In Fourth Text Analysis Conference (TAC 2011)*.
- [Shirai and Kurono1998] Yasuhiro Shirai and Atsuko Kurono. 1998. The acquisition of tense-aspect marking in japanese as a second language. *Language learning*, 48(2):279–244.
- [Suzuki et al.2006] Jun Suzuki, Erik McDermott, and Hideki Isozaki. 2006. Training Conditional Random Fields with Multivariate Evaluation Measures. In *Proceedings of the COLING-ACL Conference*, pages 217–224.
- [今村 他 2011] 今村 賢治, 泉 朋子, 菊井 玄一郎, 佐藤 理史. 2011. 述部機能表現の意味ラベルタガー. In 言語処理学会第 17 回年次大会, pages 308–311.
- [松吉 他 2007] 松吉 俊, 佐藤 理史, 宇津呂 武仁. 2007. 日本語機能表現辞書の編纂. 自然言語処理, 14(5):123–146.