

対話アシスタントにおけるローカル検索意図発話の分析

馬縹 美穂 和田 英之 橋本 力 颯々野 学
ヤフー株式会社

{mimatsun, hwada, chashimo, msassano}@yahoo-corp.jp

1 はじめに

Yahoo!音声アシスト¹ (以下「音声アシスト」と呼称)は2012年からスマートフォン上の対話アシスタントとして実際のユーザに利用されている。入力されたユーザの発話は内部で様々なタスクや雑談などの意図(ドメイン)を判定された上で応答を生成するための処理を行っているが、発話の処理にはまだ様々な段階において課題がある。本稿では音声アシストのログデータを題材にローカル検索意図判定に着目した時の課題について述べる。

ローカル検索は図1のように特定の地域に存在する施設を検索して提示する機能である。音声アシストでユーザの発話に含まれる意図の割合はCortana[1]と同様、雑談や発話をそのままクエリとするウェブ検索が半数以上を占めるものの、ローカル検索は音声アシストに限らず、話し言葉を処理してタスクを達成する必要のあるシステムにおいて重要な機能だと考えている。たとえば話し言葉によるウェブ検索²では、発話意図によっては単純な検索結果ではなく図1のように発話を更に解釈した上で直接的に結果を表示する必要がある。2017年の音声アシストで発話されたデータ中では、このように直接結果を提示すべき機能としてローカル検索は天気予報に次いで2番目と乗換案内の提示よりも多く使われている。また、Yahoo!カーナビ³のように実世界の施設に直接関連するサービスで話し言葉の言語理解に取り組む場合は、インターネット上で完結するサービスよりもローカル検索の必要性は大きい。

しかし、ローカル検索意図を判定する上ではまだ課題が多い。天気予報や乗換案内など、地名や予報用語のような限られた語彙から意図を判定している機能に比べると、ローカル検索は様々な施設、商品およびサービス名から意図を判定する必要があるため、他の



図1: 音声アシストにおけるローカル検索。

意図との曖昧性が発生しやすい。たとえば、音声アシストには次のような意図が不明瞭な発話が入力されることがある。

- ラーメン
- ポトフ

このような発話に対して、ログ中の共起頻度を調べることで意図が推定できる場合はある。音声アシストへの発話ログから「ラーメン」および「ポトフ」と共起する語を調べると、「ラーメン」は地名や「〇〇屋」のような店舗名と共起している例、「ポトフ」は「作り方」や「献立」のような語と共起している例が各語で半数以上を占めるため、同じ料理名であっても「ラーメン」はローカル検索、「ポトフ」はレシピ検索の意図を非明示的に含んでいると推定できる。ただし、これらの語は発話ログ中に比較的頻出しているため推定可能だが、ログ中に頻出しない語や今後新しく出現する語も数多く存在するため、語彙的な性質も合わせて考慮する必要がある。

また、実際のユーザが使用するアプリケーションで

¹<https://v-assist.yahoo.co.jp/>

²<https://about.yahoo.co.jp/pr/release/2014/02/04a/>

³https://map.yahoo.co.jp/blog/archives/20181220_carnavi_onsei.html

収集した発話データを音声アシストの改善に用いる上での課題も存在する。まず、入力される発話には天気予報や乗換案内の表示など特定のタスク達成を目的とするものから雑談まで様々な種類の意図が含まれるため、アプリケーション上でのローカル検索意図判定を改善する上では他の意図との兼ね合いも考慮する必要がある。また、ユーザは自由に発話を行っているため対話処理システムでは想定していない発話が入力される場合もある。

本稿では、主に携帯端末で利用される対話アシスタントにおいてローカル検索意図を自動判別するための第一段階として、入力された発話がローカル検索意図を含むか否かの判定基準について調査した結果および対話アシスタントが実際のサービスとして使われる上で対処する必要のある問題について報告する。

2 ローカル検索意図の判定基準

2.1 ローカル検索意図の概要

問題の整理のため、本稿ではローカル検索意図を以下のように定義する。

ある地域に存在する施設を訪問して目的を達成するためにその施設に関する情報を得ようとする意図。

ここで、施設とは一定の境界線で区切られた人間による構造物を指す。具体例には空港、公民館、病院などの公共的施設から商店、百貨店、コンビニなどの商業施設まで該当するが、道路などの境界線が明確ではない建築物や山など人間の手によらない地形は含まない。

入力された発話にローカル検索意図が含まれるか推定する上では、場所についてのクエリであるかどうかの程度明示されているかどうかで判定基準が異なる。以下、明示度合いごとに判定基準を述べる。

2.2 明示的意図

以下の例の下線部は施設を表現する語とよく共起するパターンである。これらのパターンが特定の施設名とともに発話されれば、その発話はローカル検索意図を持つ可能性は高いと判断できる。

- 東京駅近くのラーメン
- 病院を探して（現在地近くの病院を検索）

- カフェに行きたい（現在地近くのカフェを検索）

このようなパターンは比較的限定されており、発話中のログから検索することである程度容易に獲得できると考えられる。

2.3 非明示的意図

前節のように場所と共起するパターンが含まれていない場合でも施設や施設に関連する語が出現することもある。この時、検索対象が物理的な施設かどうかは明らかではなく、一般的な知識を得るためのウェブ検索や雑談を意図している可能性も十分に考えられる。それでも、以下の例のように訪問することで達成できる目的が存在する施設が含まれる発話は、適切に検索条件を補ったローカル検索を提示することでユーザが非明示的に要求する情報が提供できる可能性がある。

- イタリアン食べたい（現在地近くのイタリア料理店を検索）
- 髪切りたい（現在地近くの理容店を検索）
- フレンチ（現在地近くのフランス料理店を検索）

「イタリアン食べたい」や「髪切りたい」という目的は、発話と関連付けられる「イタリアン」や「理容店」によって達成できる。また、「フレンチ」という単独の発話の場合、文脈が省略されているため必ずしも外食することが目的であるとは限らないが、一般的には「料理する」よりは「どこかで食べる」方が目的になりやすいと考えられるため、ローカル検索意図を持つと推測している。

ただし、明示的なパターンよりもローカル検索かどうかの意図は曖昧であることから、「特定の地域にある施設を訪問して目的を達成したい」というユーザの要望に沿うかどうかの判定はより慎重に行う必要がある。

まず、検索対象は特定の種類の施設を曖昧性無く表現し、その施設を訪問する目的が存在する必要がある。たとえば「ソフトバンク」は店舗として存在すると同時に野球チームを表す語ともなっているため、どちらの情報を提示するべきかはこれだけでは判断することができない。また、「マンション」や「ビル」は訪問することで達成できる目的が明らかではないため、ローカル検索によって訪問するための情報を提示するのは適切ではない。

更に、目的は訪問することで達成可能で、かつ施設の訪問が目的達成の上で最適である必要がある。「習

判定	基準	具体例
明示的, 非明示的意図共に妥当	訪問で達成できる目的が明確かつ最尤な施設の説明になる語	カフェ, マッサージ
明示的意図ならば妥当	一般的な施設を説明する語	学校, 配送, 弁護士
明示的意図でも妥当ではない	一般的には施設と関連付かない語	iPhone, 地球, 東京都

表 1: 施設ごとのローカル検索意図判定の妥当性の基準.

い事」や「弁護士」は、一度訪問すれば目的が達成できる外食などは異なり目的の達成のためには継続的な訪問が必要となることが多い。また、「宅配したい」や「宅配ピザを取りたい」のように、目的地を直接訪問するよりもウェブ上の窓口を提示する方が目的を達成するのに適切な発話も存在する。このような場合は地理的な距離よりも提供されるサービスの詳細が重要であることから、通常のウェブ検索などで検索対象についての一般的な情報を提示した方が適切だと考えられる。

2.4 今後の課題

表 1 に本稿で扱った施設を示す表現のローカル検索意図が妥当となる基準を示す。2 節では、訪問することで達成できる目的が存在するかどうかによってローカル検索意図判定が行える例について説明した。ただし、単独の発話ではこのようなカテゴリに分類できないような語も存在する。たとえば「カレー」は外食する場合と料理する場合の両方が同程度に存在し、音声アシスト中のログ中で共起する単語も双方の例が拮抗している。

今後、このような例についても周りの文脈など単独の発話以外の情報を参照することでユーザの発話意図を推定することを目指している。たとえば、「カレーが食べたい」という発話の直後の発話を見ると、以下のように直前の発話を更に説明するような発話が見られる場合がある。

- 名古屋でお願いします。

この時、地名が「カレー」の詳細な検索条件だと考えられることからこの「カレーが食べたい」の意図はローカル検索である可能性が高い。このように前後の発話を参照したデータを集めることで、単独の発話だけでは意図の判定が難しい発話についても意図を推定することを考えていきたい。

3 実世界のローカル検索意図判定における課題

実際の音声アシストは音声認識の段階である程度の誤りが発生しうる他、背後のシステムを把握していない一般的なユーザはシステムでは想定していない発話を行う場合があり、ローカル検索意図に限らず問題となりうる。本節では実際にユーザの発話を処理する上で対処すべき問題について述べる。

3.1 音声認識誤り

たとえば、発話ログ中には以下のように音声に欠損がある例が一定数存在する。

- 近くのイタリア（「近くのイタリアン」の誤認識）
- 近くのビジネス本（「近くのビジネスホテル」の誤認識）
- お店（「～をして」などの誤認識）

このような誤りがあると、ローカル検索意図の判定自体に失敗したり、正しく判定したとしても正しい検索対象を解釈することができない。

その一方で、欠損があるように見える発話でも本当に欠損しているとは限らない。

- ラーメン食べた

音声アシストを含む既存の対話アシスタントではこのような発話に対しても「食べたい」の末尾が欠損しているとみなし、ラーメン店を案内する。しかし、音声アシストに入力される発話はタスク型だけではなく雑談が含まれる可能性もあり、実際に「ラーメン食べた」はただのユーザが食べたものの報告である可能性の方が高い。正確に欠損を補完する程度を見積もるには、集計された文字情報の数だけではなく、実際の入力である音声も調査しておく必要がある。

問題の種類	発話例
音声認識誤り	近くの自由 (近くのジーユー), 近くのビジネスモデル (近くのビジネスホテル)
存在しない施設の検索	近くのビデオ屋 (近くの DVD レンタル店), 近くのミドリ電化 (近くのエディオン)
雑談との区別が不明瞭	朝ごはん食べた
検索条件が不明瞭	甘い物食べたい, 楽しい所行きたい

表 2: 音声アシストに入力される発話のうち, 解釈が容易ではないもの. 本来の解釈が推測できる場合は括弧内に示す.

3.2 ユーザの挙動

一般的なユーザは音声アシストに対して UI の背後で動いているシステムを意識せずに発話しているため, 入力した発話と実際のユーザの要求が一致しているとは限らない. たとえば, ユーザは以下のように表層的な意味では存在しない施設を検索することがある.

- 近くのレンタルビデオ
- サンクスのトイレ

現在日本にはビデオテープのレンタルはほぼ存在しないにもかかわらず, かつて提供されていた媒体の名前で検索するユーザは一定数存在する. また, コンビニ系列のサークル K サンクスは 2018 年の時点で株式会社ファミリーマートに吸収合併され, 店舗としては全て閉店するかファミリーマート系列に統合されているが, ユーザはかつてはサークル K サンクスであったファミリーマートについて古い名称で検索を試みる場合がある. このような発話にも対応するためには, 対話システム側でクエリの置き換えに対応するとともに, 参照する地図情報が過去の施設情報にも対応している必要がある.

また, ローカル検索は店舗に含まれる説明文やキーワードに検索条件が含まれているかどうかで行うことになるが, ユーザは検索条件が自明なクエリで発話するとは限らない.

- 今すぐ食べたい
- 冷たいものが食べたい

このように検索条件が不明瞭な発話に対しては, 音声アシストを含め既存の対話アシスタントはカテゴリを更に絞り込むための質問か, 通常のウェブ検索を返却している. 曖昧な意図であってもユーザの要望にすぐ答えられるような対話アシスタントを目指すならば「今すぐ」や「冷たいもの」のような曖昧な条件を予め具体的な条件に関連付けておく必要がある.

3.3 課題の解決に向けて

ユーザの要求が不明瞭な発話に対して, その背後にあるユーザの要求を理解するアプローチの 1 つとしては, 発話と要求が正しく関連付けられたデータを獲得して学習するという方針がある. そのようなデータの獲得方法は, 人手によるタグ付けの他に, 音声アシストという対話処理を行うシステムの中で達成できる可能性はある. たとえば, 検索条件が不明瞭な発話については音声アシストがユーザに詳細なカテゴリを更に尋ねている. この訊き返しを続けユーザが回答した結果が蓄積することでそのような不明瞭な発話についても適切な条件が関連付けられるかもしれない. このような獲得方法を用いることで, 人手によるタグ付けに対してコストがほとんどかからず, 継続的にデータを収集できると考えられる.

4 おわりに

本稿では, 対話アシスタントの一機能であるローカル意図発話について, ローカル検索意図と判定される基準や実際のシステムの問題点について調査を行った. 今後は, 今回の調査で得られた知見を元にローカル検索意図の自動判別に取り組んでいくことを予定している.

参考文献

- [1] Jiepu Jiang, Ahmed Hassan Awadallah, Rosie Jones, Umut Ozertem, Imed Zitouni, Ranjitha Gurunath Kulkarni, and Omar Zia Khan. Automatic online evaluation of intelligent assistants. In *Proceedings of WWW*, pages 506–516, 2015.