

音声対話によるカーナビゲーションシステム – ナイト 2000 への挑戦 –

The car-navigation system by a speech dialogue – The challenge to Night 2000 –

堀口 賞一^{*1}
Shoichi HORIGUCHI

藤本 拓^{*1}
Hiroshi FUJIMOTO

^{*1} NTTドコモ サービス&ソリューション開発部
Service & Solution Development Department, NTT DOCOMO Inc.

While attentions gathered for “Eyes Free” which Apple Inc. advocated in June, last year, NTT DOCOMO has released a car navigation service “Docomo Drive Net (DDN)” that can be operated by naturally speaking from December, 2012. This service applies “Shabette-Concier” services of our company to the car navigation system. “Knight Rider” is the special-effects teleplay broadcast in the U.S. for about four years after 1982. “Night 2000” which appears into it is the dream car by which spoke about man's words and special equipment was carried out, and is one targets, such as an artificial intelligence researcher, and a car navigation system developer. In this paper, we describe the intention understanding function which makes the core of naturally speaking in DDN and describe the subject for developing K.I.T.T. which is an artificial intelligence engine of “Night 2000”.

1. はじめに

「ナイトライダー」は 1982 年から約 4 年間、米国で放送された特撮テレビドラマである。その中に登場する「ナイト 2000」こそ、人間の言葉を話し特殊装備されたドリームカーであり、我々 AI 研究者やカーナビ開発者等の 1 つの目標と言っても過言ではない。ナイト 2000 の登場から約 30 年後 (2012 年 12 月)、NTTドコモは、ユーザの自然な発話によりカーナビ操作可能な「ドコモドライブネット (DDN) [NTT docomo 1]」の提供を開始した。

本稿では、当 DDN の中心的機能である意図解釈機能 [Tsuji 2013] について述べると共に、今後、ナイト 2000 の人工知能である K.I.T.T. (Knight Industries Two Thousand) の開発へ向けて取り組まなければならない課題について触れる。

2. DDN の概要、およびシステム構成

従来のカーナビは音声コマンド方式 [Kurata 2010] が多く、定型の発話文によりカーナビを操作していた。一方、DDN では、ユーザの自然な発話に対して、適切なカーナビ操作が可能である。例えば、「近くにコンビニある？」とか「トイレ行きたい」などと発話すると現在地付近のコンビニを地図上に表示するし、「お腹空いた」とか「喉渇いた」と言うと現在地付近のレストラン等を地図上に表示する。更に、「家に帰りたい」と言うと自宅までのルートを表示するといった具合である。

図 1 に DDN の機能ブロック図を示す。主に、クライアントアプリ、音声認識サーバ、意図解釈サーバ、コンテンツサーバの 4 つの機能ブロックから構成される。ユーザの自然な発話に対して、音声認識処理され、テキスト化された自然言語に対して、意図解釈 (適切なタスク判定とクエリ抽出) 処理され、クエリに基づきコンテンツを検索し、検索されたコンテンツをクライアントアプリがタスクに基づき表示する。以下、個々の機能ブロックについて述べる。

2.1 クライアントアプリ

クライアントアプリは、DDN のユーザインターフェースであり、android 上で動作するアプリケーションである。図 2 に示す通り

連絡先: 堀口 賞一, NTTドコモ, 神奈川県横須賀市光の丘
3-6 NTTドコモ R&D センタ, horiguchis@nttdocomo.co.jp

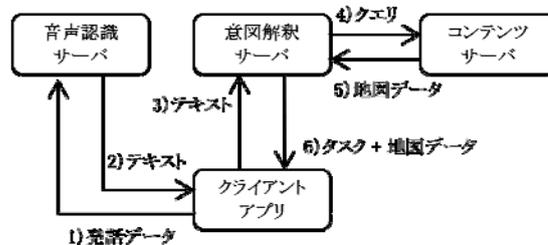


図 1; DDN の機能ブロック図



a) 発話表示 b) テキスト表示 c) ルート表示

図 2; クライアントアプリの画面

発話のタイミングを促すと共に、ユーザの自然な発話を集音し、音声認識サーバへ送る。

次に、音声認識処理された自然言語テキストを受け取り、意図解釈サーバに送る。

最終的に検索されたコンテンツを意図解釈サーバから受け取り、地図上に近隣の施設情報表示や、目的地までのルート表示、更にはカーナビの音量調整や地図の縮尺変更などを実施する。

2.2 音声認識サーバ

音声認識サーバは、クライアントアプリからユーザの自然な発話データ (16kHz サンプリング) を受け取り、自然言語テキストに変換し、そのテキストをクライアントアプリに返信する。このとき、発話される言葉データ (テキスト) を予め保持している言語モデ

表 1;DDN のタスク例

主なタスク	説明	主な発話例
目的地設定	目的地を設定し、案内を開始するタスク	東京に行きたい/横浜駅に行きたい
施設検索	施設を検索し、地図上に表示するタスク	近くの駐車場どこ?/コンビニ探して
自宅設定	自宅を目的地に設定し、案内を開始するタスク	家に帰る/自宅まで
案内中止	案内を中止するタスク	案内止めて/お疲れ
音量下げ	音量を下げるタスク	音量下げて/うるさい
地図拡大	地図を拡大するタスク	地図を大きくして/地図拡大

表 2;意図解釈機能が利用する主な辞書

辞書名	辞書サイズ	登録データ例
地名辞書	138,300	横浜、横浜駅、東京スカイツリー、富士山
施設名辞書	21,988	コンビニ、マクドナルド、駅、駐車場
住所辞書	729,644	東京都、横須賀市光の丘
路線辞書	973	山手線、京浜急行
揺らぎ辞書	40,891	(マック、マクドナルド)、(ファミマ、ファミリーマート)

ルと、発話環境を予め保持している音響モデルとが認識精度に大きな影響を与える。

2.3 意図解釈サーバ

意図解釈サーバは、クライアントアプリから自然言語テキストを受け取り、当テキストを予め定めたタスクに判別すると共に、タスクに応じて地図データを取得するためのクエリ抽出も実施する。さらに、判別されたタスクに応じて、コンテンツサーバに対してタスクに応じたクエリ検索を実施し、その検索結果をクライアントアプリに返す。

DDN で実施している主なタスク例を表 1 に示す。例えば、「東京駅に行きたい」という発話に対して、「目的地設定」というタスクを判別する。前述した「トイレに行きたい」という発話に対しては「施設検索」といった具合である。これらは、ユーザの発話を入力文章とした文章分類問題と見なすことが出来る。機械学習による文章分類手法は数多く提案されているが、DDN では SVM[Joachims 1998]を利用して判定モデルを構築し、タスク判定を実施している。

クエリ抽出は、表 2 に示す地名辞書、施設名辞書などを利用したワードマッチングと、Conditional Random Fields (CRF) [Lafferty 2001]を組み合わせて実現している。CRF を利用することで、辞書に未登録の単語や、複数の形態素の組合せにより生成される地名の抽出も可能になる。

意図解釈の処理フローを図 3 に示す。例えば、「横浜駅周辺のマップに行きたい」と言う自然言語テキストに対して、形態素解析をし、揺らぎ補正を行う。このとき、「マック」が「マクドナルド」へ、「行き」が「行く」へ補正される。次に、これら補正された形態素を素性とする特徴ベクトルからなる判定モデルに基づき、「施設検索」というタスクと、「横浜駅」と「マクドナルド」というクエリが抽出される。この結果、コンテンツサーバに対して、「横浜駅」と「マクドナルド」をクエリとして施設検索を実施する。

2.4 コンテンツサーバ

コンテンツサーバは、地図(緯度、経度)情報を基にした地名、施設名、およびリアルタイムな駐車場の空き情報などで構成されるコンテンツ DB 群である。意図解釈サーバからのタスクに応じた様々な検索に対して地図情報を返す。

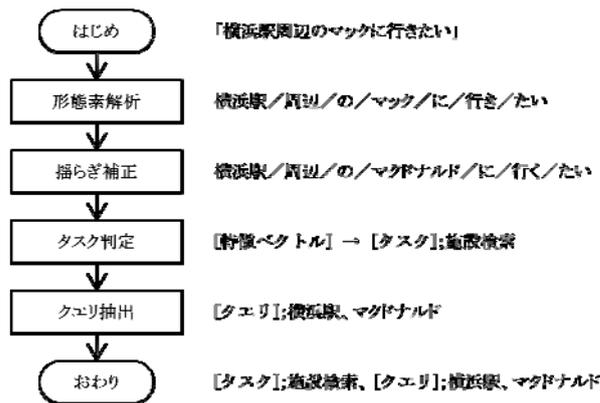


図 3;意図解釈の処理フロー

3. K.I.T.T.へ向けて

現在の DDN は、運転開始前にルート表示や周辺検索を実施することを想定しており、走行時の利用は想定されていない。今後、K.I.T.T.を目指すためには数多くの課題を解決する必要がある。以下に、主な検討課題を列挙する。

- ロードノイズ対策やオーディオ再生時のエコーキャンセラ等の走行時における音声認識精度の向上
- 対話によるスロット埋め込み、選択肢絞込みの実現
- 対話による同音異義発話への対応
- ダイナミックな状態遷移に応じたタスク判別の実現
- ドライバーの嗜好や発話パターン、行動パターンに応じたパーソナライズ化

その他、自動車メーカーと一体となった車両制御などの開発が必要であることは言うまでもない。

4. おわりに

DDN における意図解釈機能を中心に、今後、K.I.T.T.を目指すための主な課題を列挙した。近い将来、K.I.T.T.を超える人工知能搭載のカーナビを実現するためには、多くの AI 研究者達が共に協力し合い、諸課題をクリアする必要がある。

参考文献

[NTT docomo 1] NTT docomo: 「ドコモドライブ ネット」, http://www.nttdocomo.co.jp/service/information/drive_net/, accessed Apr.10, 2013.

[Tsujino 2013] 辻野孝輔, 柴藤稔, 磯田佳徳, 飯塚真也: 実サービスにおける音声認識と自然言語インタフェース技術, 人工知能学会誌, vol.28, No.1, pp75-81, 2013.

[Kurata 2010] 倉田岳人, 市川治, 西村雅史: “ユーザの発話傾向分析に基づく車載機器操作のための音声入力手法,” 電子情報通信学会論文誌, vol.J93-D, no.10, pp.2107-2117, 2010

[Joachims 1998] T.Joachims: Text Categorization with Support Vector Machines: Learning with Many Relevant Features, Proc. The 10th European Conference on Machine Learning, vol2, pp.137-142, 1998.

[J.Lafferty 2001] J.Lafferty, A.Mccallum, F.Pereira: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, Proc. The 18th International Conference on Machine Learning, pp.282-289, 2001