

興味推定による自動車内音声対話情報提供システム

橋本 拓観† 伊藤 慶明† 小嶋 和徳†

岩手県立大学大学院 ソフトウェア情報学研究科

1 はじめに

長距離自動車運転時に一人である場合、刺激が少なく眠気や退屈により運転に集中できず安全な運転が脅かされる。運転手が音声対話システムと会話することで眠気や退屈感を払拭できれば、安全運転に繋がると考える。近年、スマートフォンに Siri 等の音声認識機能が付与され、雑談や「こんな事を聞いたらどうなるか」などにも多く用いられている[1]。一般の音声応答システムを実際に使用するとユーザの好みの情報にすぐに到達しないことも多い。

本稿では長距離自動車運転時に眠気や退屈感を感じた時に、運転手（ユーザ）が興味を持つ情報をシステムが積極的に提供することで、眠気や退屈感を払拭するシステムの提案を行う。ユーザは運転中に目や手を離すことができないため、ユーザとシステムは対話は音声で行う。ユーザが対話の内容に興味を持つように、対話内容はこれまでのユーザの好みをプロファイリングし、ユーザが興味を持つニュースを自動的に選択し提供する。

今回はユーザが興味を持つニュースの分析には LDA を用い、パソコン上で動作するプロトタイプシステムを構築し、シミュレーション実験を通して、提案システムの実用性・課題を検討する。

2 システム概要

ニュースは Web から常時最新のものを取得しておく。ユーザは本システムを利用しながら聞いたニュースを基にユーザの興味を分析し、プロファイリングする。ユーザが利用するときにはそのプロファイリングにもとづいて LDA を利用し、ユーザのトピックモデルと各ニュースのトピックモデルとの類似度を求める。これにより利用するにつれてユーザが興味を持つような類似度の高いニュースを提供できるようになる。以降で今回の実装方式を説明する。

2.1 ニュース

今回使用するユーザに提供するニュースは Web から取得した最新の Yahoo!ニュース[2]を使用した。Yahoo!ニュースはタイトル、内容、カテゴリ、日付を保持している。カテゴリは国内、国際、経済、エンタメ、スポーツ、IT、科学、地域の 8 種類ある。過去のニュースのデータをハードディスクに保存しておき、新しいニュースを追加する。

2.2 システムの流れ

システムは Web から最新の Yahoo!ニュースを随時取得し、各ニュースの各単語について LDA を用いてトピックを推定する。ユーザが指定した条件に一致するニュースをニュースリストとして、ニュースのタイトルを 1 件ずつ読んでいく。そのニュースの内容を読むかどうかユーザは応答し、読んだ場合はそのニュースをプロファイリングに利用する。

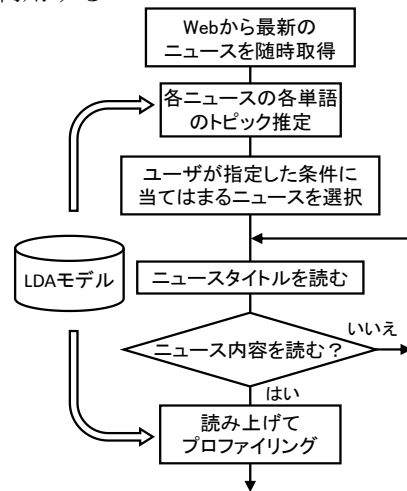


図 1 システムの流れ

2.3 ニュースリスト

ユーザが指定した条件に該当するニュース群をニュースリストとし、「おすすめ」、「カテゴリ」、「最新」、「検索」の 4 種類ある。

- (1) おすすめ：プロファイリングによって予測されたユーザが興味を持つようなニュース
- (2) カテゴリ：指定したカテゴリのニュース
- (3) 最新：最新の全ニュース
- (4) 検索：指定した検索語がタイトルか内容に含まれるニュース

An Information Providing System through Speech Interaction in a Car Using Interest Estimation
 Hashimoto Takumi†, Itoh Yoshiaki†, Kojima Kazunori†
 †Iwate Prefectural University Graduate School of Software and Information Science

ユーザは(2), (3), (4)の条件でニュースをシステムから提供された場合プロファイリングが行われ, (1)のユーザが興味を持つニュースの提供に利用される。

2.4 音声認識技術・音声合成技術

ユーザとシステムが音声対話を行うために音声認識技術と音声合成技術を使用する。音声認識技術には Microsoft の音声認識エンジン, 音声合成技術には Microsoft の音声合成を使用した。

3 LDA

1つの文章が潜在的に持っている意味をトピックという。Latent Dirichlet Allocation (LDA)[3]とは自然言語処理で用いられるトピックモデルで, 1つの文章が複数のトピックから成ると仮定したモデルである。

文章ごとの学習データから最終的に各単語がどのトピックに分類されるかをクラスタリングしていき, 各単語はどのトピックに属するかの尤度が求められる。トピック数は事前に指定しておく。

3.1 特徴ベクトルの抽出

各単語の最も高い尤度を有するトピックをその単語のトピックとする。文章から各単語のトピックをカウントし正規化したベクトルを特徴ベクトルとした。今回1件のニュースを1つの文章として考え, 単語は名詞のみを使用し MeCab[4]によって単語分割した。

3.2 プロファイリング

本稿でのプロファイリングとはユーザが興味を持つようなニュースを分析することを示す。ユーザが聞いたニュース全体に対してトピックの重みがついた, 即ちユーザの好みを表す特徴ベクトルが生成される。

3.3 提供候補のニュース

システムが Web から取得した各ニュースについてそれぞれのニュース内の単語から特徴ベクトルを抽出する。

3.4 ユーザ好みのニュースの抽出

提供候補のニュース群のうちシステムがおすすめとしてユーザに提供するためには, このニュースの順位をつける必要がある。ユーザ好みの順に提供するためプロファイリング結果の特徴ベクトルと提供候補のニュースの特徴ベクトルとのコサイン類似度を求める。類似度が高いニュースをユーザが興味を持つニュースとして提供する。

4 実験

4.1 実験条件

31,301件のYahoo!ニュースをクラスタリングし, LDA モデルをトピック数 128 で作成する。実験対

象のデータは1ヶ月のYahoo!ニュース 2,290件について, 1番目の被験者はサッカーのニュース 54件, 2番目の被験者は大企業のニュース 34件, 3番目の被験者はスマートフォンのニュース 17件に興味を持っているという設定で3人の被験者による実験を行う。システムがおすすめとして提供したニュースについて評価を行う。1ヶ月のニュースを5分割し, 5分割中の4つをプロファイリングのための学習データ, 1つを評価データとしてクロスバリデーションを行った。実験結果として Precision-Recall カーブを描きユーザが興味を持つニュースが上位で出現するかを確認する。

4.2 結果・考察

図2は各被験者の興味を持つニュースについて5回の平均の Precision-Recall カーブを示す。システムはユーザに上位候補のニュースから1件ずつ順に提供していくため, ユーザが興味を持つニュースが早く提供されるべきである。実験結果より被験者 1, 3 については興味を持つニュースを早く提供することができていることが確認できた。実験では興味のあるニュース数が少ないため, ニュースの数を増やし, 被験者を増やして複数のパターンで実験を行う必要があると考える。

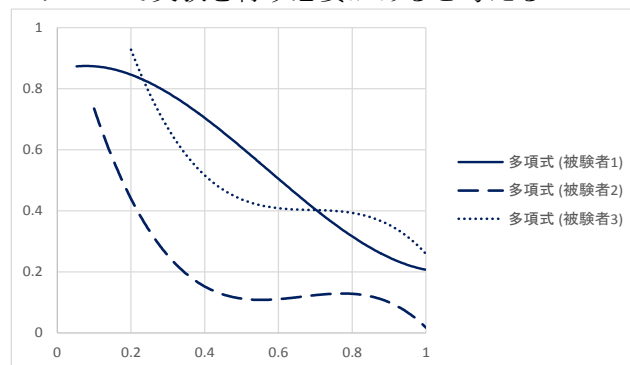


図2 被験者の Precision-Recall カーブ

5 おわりに

本稿では長距離自動車運転時にユーザが眠気や退屈感を払拭し安全運転を行うために, ユーザが興味を持つ情報をシステムが提供するシステムの提案を行った。実験内容を整理して新たな実験を行い本システムの有効性を検証していく。

参考文献

- [1] 吉村健, "しゃべってコンシェルと言語処理", 情報処理学会研究報告, Vol.2012-SLP-93, No.4, (2012)
- [2] Yahoo!ニュース, <http://news.yahoo.co.jp>, (2015/12/21)
- [3] David M. Blei, Andrew Y. Ng, Michael I. Jordan, "Latent Dirichlet Allocation", Journal of Machine Learning Research 3, pp993-1022, (2003)
- [4] MeCab, <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>, (2015/12/21)