

# マルチモーダル情報からの運転者の内部状態の推定

梶野 尊弘<sup>†</sup>

武田 龍<sup>‡</sup>

小路 悠介<sup>§</sup>

駒谷 和範<sup>‡</sup>

<sup>†</sup> 大阪大学 工学部電子情報工学科

<sup>‡</sup> 大阪大学 産業科学研究所

<sup>§</sup> 三菱電気株式会社

## 1. はじめに

人間は困惑していたり、指示を理解できていなかったりすると、それが様々な挙動に現れる。カーナビゲーションシステムにおいて、運転者からの明示的なコマンドだけでなく、それらの挙動をシステムが検出できれば、運転者に対して親切な案内が実現できる。例えば、ナビ画面を注視しているといった挙動から、指示を理解できていない状況をシステムが検出できれば、運転者にもう一度指示を行うなどできる(図1)。

本研究では、運転中に現れた様々な挙動から、運転者の内部状態を推定する。ここでは内部状態として、運転者がシステムからの指示を理解しているか否か(ルート理解度)を扱う。

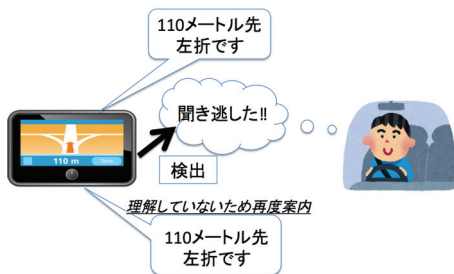


図1: 推定した内部状態に基づいて案内するシステム

## 2. 内部状態の推定

本研究では、機械学習を用いて、運転者の内部状態を推定する。内部状態推定の枠組みを図2に示す。入力とは様々なセンサ情報である。例えば、車速やハンドル角度などの車内センサ情報である。出力は、内部状態としてルート理解度を扱う。これは、システムから指示されたルートを、運転者が理解しているか否かを表す。

本稿では特に、推定に有効な特徴の発見という課題に取り組む。このために、特徴選択を行い、推定に有効な特徴について分析する。マルチモーダル情報を用いて状態推定を行う研究には、韻律情報と頭部の動作を用いて、ユーザが肯定的か否定的かを推定する研究[1]や、ユーザの動作や顔向き、韻律情報などを用いてロボットの応答義務を推定する研究がある[2]。カーナビという具体的な応用を視野に入れ、車載機器から得られる情報を含める点、また人手で付与された情報を特徴として用いる点で異なる。将来的に後者を自動付与することを含めて有効な特徴について分析する。

## 3. 実験条件

### 3.1 使用データ

運転者が、知らない道を人もしくはカーナビにより案内される際のデータを収集した。ここでは音声・画像収

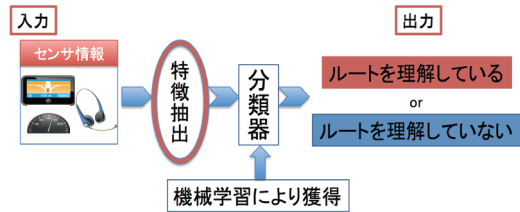


図2: ルート理解度推定の枠組み

録機器を用いて、ドライバーの動作、運転状況、道路状況を記録した。被験者は8名である。このうち、カーナビ案内による運転部分(一人当たり約30分)を対象とする。また、ルート理解度と、動作や道路状況を表す複数の特徴量を、人手で付与した。ルート理解度は、「完全に指示を理解している」、「中程度指示を理解している」、「指示を理解していない」の3段階で付与した。

データに付与されている情報は以下である。これらは100ミリ秒を1フレームとして記録されている。

1. 車内センサ  
運転者から車への入力される、各時刻でのブレーキペダル角度、ハンドルの回転角度など9個。
2. 運転者の身体センサ  
各時刻における心拍、体表温、脳波など14個。
3. カーナビゲーションシステム情報  
ルートへの接続リンク本数、案内地点までの距離など133個。
4. 人手で付与されたラベル  
運転者発話の有無、運転者の注視の有無など23個。

本稿では、「中程度指示を理解している」、「指示を理解していない」とされた区間はルート理解度が低い(low)とし、それ以外の箇所は高い(high)とした。このhighとlowを正解ラベルとし、これらを推定する二値分類として問題を定式化する。

### 3.2 特徴抽出

図2でのセンサ情報から特徴抽出する手法について述べる。ルート理解度は、1フレームごとに変動するものではないため、窓幅やシフト幅を設定する。ここでは、 $w$ フレームを窓とし、それを $s$ フレームずつシフトさせて特徴量を計算する。

窓内の正解ラベルや特徴量の値は以下のように計算する。正解とするルート理解度ラベルは、ルート理解度が低いとされた区間と、 $w$ フレームの半分以上が重なる場合にlowとした。特徴が数値の場合は、 $w$ フレーム中の1フレームあたりの平均値を用いる。特徴が文字列(ラベルや地名など)の場合は、 $w$ フレーム中の1フレームあたりの文字列の有無の率を用いた。文字列の違いは考慮していない。1フレーム毎に値が得られない特徴については、窓内の値の平均を用いた。以降では $w = 50$ 、 $s = 10$ として実験を進める。

Estimating Driver's Internal States from Multimodal Information: Takahiro Kajino, Ryu Takeda (Osaka Univ.), Yusuke Koji (Mitsubishi Electric Corp.), and Kazunori Komatani (Osaka Univ.)

表 1: ルート理解度推定性能

検証方法	class	適合率	再現率	F 値
単純 10-fold CV	low	.887	.938	.912
	high	.934	.881	.907
8-fold CV (user held-out)	low	.734	.476	.512
	high	.671	.839	.702

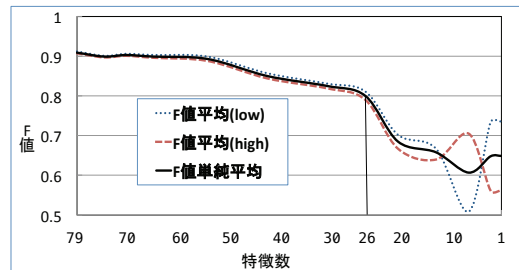


図 3: 利用する特徴数を減らした際の推定性能

### 3.3 重みの付与と用いる特徴

被験者 1 名あたり約 30 分 (約 18000 フレーム) のデータから、1 名あたり約 1795 個の特徴量と正解ラベルの組が得られる。ルート理解度ラベル (low) のラベル数は 534、ルート理解度ラベル (high) のラベル数は 13209 であった。ラベル数に偏りがあるため、ルート理解度ラベル (low) のデータに対し、ルート理解度ラベル (high) のデータ数との比を重みとして与え、学習と評価を行う。

また、ここで特徴間の相関を利用して、明らかに不要な特徴を除く。2 つの特徴間の相関係数  $|r|$  が 1 である場合、一方の特徴がもう一方の特徴で完全に表現できる。ここでは任意の 2 つの特徴間の相関係数を計算し、 $|r|$  が 0.99 以上の値を取る特徴の組の集合を 1 特徴に集約する。結果として、49 特徴が、11 個の集合へと集約できた。つまり、38 個の特徴を取り除いた。特に、値が文字列であり、その有無のみを用いた特徴において、値がほぼ同じ特徴となっているものが見られた。

以降では、この 38 特徴と、値が全く変化しない 61 特徴を除外し、79 の特徴を用いる。学習・評価には Weka (ver. 3.6.13) を利用した。分類器は線形 SVM とし、各特徴の値は正規化して用いる。推定性能を評価するための指標として、適合率、再現率、F 値を計算した。F 値は、適合率と再現率の調和平均である。

## 4. 実験結果

### 4.1 単純な交差検証による性能

被験者を分離しない条件での交差検証による、性能を示す。被験者 8 人分のデータを用いて、10 分割交差検証を行った (単純 10-fold CV)。また、ここではルート理解度ラベル (low) のデータに対し、24.7 の重みを与えた。表 1 上部にルート理解度推定の結果を示す。low, high いずれのラベルに対しても、0.9 以上の F 値が得られている。

### 4.2 有効な特徴の調査

まず、他の特徴との相関係数が大きい特徴を除いていくことで、より重要な特徴のみを残していく。これを、推定に用いる特徴数が 1 になるまで行った。図 3 に特徴数を減らしていった場合の推定性能を示す。全体に特徴数を減らすにつれ性能が低下しているが、特に大きく性能が低下したところで、重要な特徴が除かれたと推測できる。このことから、この二値分類問題には、特徴が 26 個程度必要であることが示唆されている。

次に、この 26 特徴を用いた際の、推定に用いる特徴の重要度を分析する。ここでは、SVM の識別平面を表す式における特徴の係数に着目する。係数の絶対値が大きいほど、ルート理解度の推定に重要であり、また例えば係数の符号が負ならば特徴の値が大きいほどルート理解度は low となる傾向がある。係数から重要とみなせる

特徴とその係数を以下に示す。

1. 運転者発話の有無 [-3.713]
2. 運転者の注視の有無 [-2.916]
3. 案内地点までの距離 [2.064]

運転者発話の有無や運転者の注視の有無の係数が負であることから、独り言が多いまたは、ナビ画面や前方を注視することが多いほどルート理解度は low となる傾向にある。一方、案内地点までの距離の係数が正であることから、目的地までの距離が長いほどルート理解度は high となる傾向にある。

### 4.3 被験者単位の交差検証による性能の評価

被験者ごとに学習データと評価データを明確に分離した条件での交差検証により、性能を評価する。4.1 節では、完全に学習データと評価データを分離できていない。8 人分のデータのうち、学習データは被験者 7 人分とし、評価データは学習データに含まれない被験者 1 人分として、交差検証 (8-fold CV) を実施した。重みは、各 fold における学習データと評価データのそれぞれにおいて、2 クラスのラベルの比として計算し付与した。利用する特徴は、4.1 節と同じ 79 特徴である。

表 1 下部にルート理解度推定の結果を示す。10-fold CV と比較して、性能は低い。10-fold CV では、学習するデータと評価するデータの中に、1 つの区間に対応するほぼ同一の特徴データが含まれており、その性能が不当に高かったためと考えられる。

## 5. おわりに

運転中に観測できるセンサデータと人手で付与されたラベルデータを用いて、運転者の内部状態 (ルート理解度) を推定する手法を提案した。推定に有効な特徴の調査では、SVM の識別平面の式における係数から、ルート理解度推定に重要な特徴を示した。今後、被験者単位で特徴の分析を実施した場合に、推定に有効な特徴がどう変化するか、検証する。

## 参考文献

- [1] 藤江真也, 江尻康, 菊池英明, 小林哲則: 肯定的/否定的発話態度の認識とその音声対話システムへの応用. 電子情報通信学会論文誌, Vol.J88-D2, No.3, pp.489-498, 2005.
- [2] 杉山貴昭, 船越孝太郎, 中野幹生, 駒谷和範: 多人数対話におけるロボットの応答義務の推定. 人工知能学会第 29 回全国大会, 2015.