

## マルチモーダル潜在トピックモデルによる運転シーン認識

## Driving Scene Recognition based on Multimodal Latent Topic Model

坂東 誉司\*1

Takashi Bando

竹中 一仁\*1

Kazuhiro Takenaka

長坂 翔吾\*2

Shogo Nagasaka

谷口 忠大\*2

Tadahiro Taniguchi

\*1株式会社デンソー

DENSO CORPORATION

\*2立命館大学

Ritsumeikan University

Various driving support systems nowadays have been developed for safe and comfortable driving. To develop more effective driving assist systems, abstractive recognition of driving situation performed just like a human driver is important in order to achieve fully cooperative driving between the driver and vehicle. To achieve human-like annotation of driving behavioral data and image sequences, we first divided continuous driving behavioral data into discrete symbols that represent driving situations. Then, based on latent topic model proposed in Natural language processing area, latent driving topics laid on each driving situation were estimated as a relational model among driving behavioral data, image sequences, and human-annotated tags. Finally, automatic annotation of the behavioral data and image sequences can be achieved by calculating the predictive distribution of the annotations via estimated driving topics. The proposed method intuitively annotated more than 50,000 pieces of frame data, including urban road and expressway data.

## 1. はじめに

本論文では、時系列運転挙動データと車両前方を撮影した画像系列から、高文脈な運転の状況を自然言語で自動的にアノテーションする手法の提案を行う。従来の運転支援システムでは、例えば Time-to-Collision の様に比較的シンプルな物理的モデルで運転シーンを記述する。こうした手法は衝突直前の支援に対しては有効に働くことが示されて来た。これに対してドライバは、より抽象的に運転シーンを捉えることで、長期に渡って運転シーンの遷移を予測しながらスムーズな運転を行っていると考えられる。ドライバと車両の自然な協調運転実現の為に、システムも複雑な運転シーンをドライバの認識に近い形で抽象的に捉えることが必要である。

この問題に対して [Taniguchi 12] では二重分節解析器 (DAA) を利用して運転挙動データを分節化した。しかし、DAA は教師なし手法であるため、ドライバとの直感的なインタラクションの実現には、分節化された運転シーンに直感的なラベルを付与する必要がある。[Bando 13] は運転シーンの背景に潜在的に存在する運転のトピックを Latent Dirichlet Allocation (LDA) [Blei 03] を利用して推定し、各トピックに顕著な運転挙動を利用して運転トピックに対するラベルを自動的に付与していた。しかし、彼らの手法は運転挙動データのみから運転トピックの推定を行っている。自車周囲の環境情報も考慮した上で運転シーンを自然言語で解釈する方が、環境を認識して車両の操作を行うドライバにとって、より直感的である。

ここでは、運転挙動データと前方画像データに対してドライバと良く一致したシーンアノテーション法の構築を考える。

## 2. 運転シーン認識手法

図 1(A) に提案手法の枠組みを示す。提案手法ではまず、連続時系列の運転挙動データを DAA を用いて分節化する。この時、ドライバの操作入力を利用して分節化することでドライバの感じた運転シーンの変化点を効果的に抽出できる。ここでは

運転挙動データとして、アクセル・ブレーキペダル操作量、ステアリング操作量、車両速度と各々の微分という 8 次元の特徴空間を利用して運転シーンへの分節化を行った。

その後、各運転シーンを一つの文書と考へて、multimodal LDA (mLDA) による運転トピック推定を行う。mLDA は近年、画像へのアノテーション [Wang 09] やロボットの語彙獲得 [Nakamura 09] にも利用されている。ここでは運転挙動データに加えて、前方画像、前方画像に自然言語で付与したタグの 3 つのモダリティに跨がる運転トピックを学習する。提案手法で採用した各モダリティの特徴量は下記の通りである。

**Behavioral words ( $w^{(b)}$ ):** 時系列の運転挙動データから算出される特徴量。先述の 8 次元の特徴空間で k-means ( $K = 1000$ ) により挙動データを分割する。この時、各フレームのデータ点に割り当てられるクラスタインデックスを  $w^{(b)}$  とする。

**Visual words ( $w^{(v)}$ ):** 車両前方を撮影した画像から算出される特徴量。画像サイズを  $320 \times 240$  pixel に縮小後、5 pixel 毎の 3072 点で算出した SIFT 特徴を k-means ( $K = 1000$ ) により分割。このクラスタインデックスを  $w^{(v)}$  として利用する。

**Annotated words ( $w^{(a)}$ ):** 車両挙動や周辺環境に関するタグ特徴。前方画像を見ながら先行車の有無や種別、渋滞しているか、車線数など、大まかなタグ項目のみ指示して 9 人の被験者にタグを付与してもらった。このタグのヒストグラムを  $w^{(a)}$  とする。タグには、例えば、どの程度近づいた時に先行車として判断するか、といったバラツキが見られた。

## 3. 評価実験

市街路、高速道路を含む全長 69km のコースで取得した約 80 分のデータを利用して提案手法を評価する。データは 10fps で記録した。DAA によって分節化された運転シーンをランダムに訓練用とテスト用に分け、訓練用のシーンで運転トピックモデルの構築を行い、テスト用のシーンで運転シーンに付与されたタグを予測する。運転トピックの数は 100 とした。予測時には、観測された  $w^{(b)}$ ,  $w^{(v)}$  から対象シーンに含まれる運転トピックの混合比率を算出し、これと各運転トピックが持つ運転シーンタグの分布を利用して  $w^{(a)}$  を生成する。

連絡先: 坂東誉司, 株式会社デンソー, 愛知県刈谷市昭和町 1-1,  
Email: TAKASHI.BANDOU@denso.co.jp

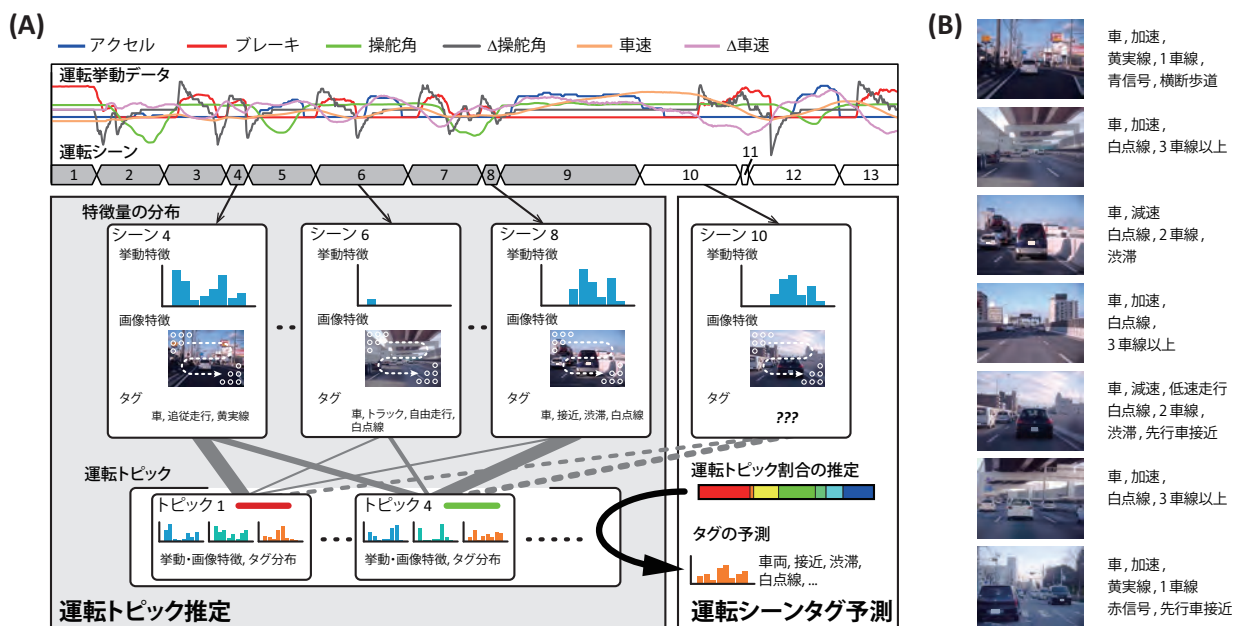


図 1: (A) 運転シーntag推定の枠組み。時系列運転挙動データは運転シーンの系列に離散記号化した上で、各運転シーン内での挙動特徴、画像特徴、運転シーntagの分布から運転トピックを推定する。運転シーntagの予測時には、対象シーン内の挙動特徴と画像特徴の分布から、対象シーンを構成する運転トピックの混合比率を算出し、これと各運転トピックが持つ運転シーntagの分布を利用してタグを生成する。(B) 予測された運転シーntagの例。各タグ項目中で最も確率が大きいものを示した。

但し、各フレームで1つ観測される  $w^{(b)}$ 、9人の被験者が各フレームに付与した9つのタグで構成される  $w^{(a)}$  に対して、各フレームで観測される  $w^{(v)}$  は3072点と非常に多い。運転時に撮影した前方画像は比較的緩やかに変化するため、観測される  $w^{(a)}$  が相対的に多いことは、運転トピックに強い時間方向のスムージングをかける効果を持ち、運転挙動やタグの寄与度が相対的に低下してしまう。そこで、各運転シーンに含まれるユニークな特徴の数で各モダリティを正規化して、各モダリティを均等に考慮するように工夫した。

また、運転シーntagの予測分布から単純にシーntagを生成すると、例えば、「加速」などの頻出タグが過度に生成されやすくなる。ここでは、逆文書頻度 (IDF) と呼ばれる指標を利用して生成確率を修正し、特定の運転シーンで特徴的に観測される運転シーntagをバイアスした。

図 1(B) に生成された運転シーntagの例を示す。先行車種別、車両の過減速状態、レーンマーク種別、車線数、信号の有無と色、横断歩道の有無、渋滞の有無、先行車接近の有無、の各タグ項目中で最も生成確率が大きいタグを示した。提案手法は、1枚の画像ではなく、時系列の運転挙動データと複数の画像系列運転シーンに対してそれを説明するタグを付与する。この為、例えば一番下の図に示す通り、手前に青信号、奥に赤信号が存在するシーンに対して「赤信号」タグのみを付与してしまうことなどが多々あるものの、全体的には直感的な運転シーntagの付与に成功している。提案手法に依れば、車両挙動データと車両前方の画像系列から抽出した低次の特徴量のみから、渋滞や先行車の接近等、比較的高次の運転コンテキストの推定が自然言語で可能となる。提案手法を応用すれば、様々なセンサに基づいた先行車や歩行者、レーンマーク等の物体認識技術を大規模に収集することで、カメラのみを搭載した車両でも運転シーンの認識を行える可能性があり、今後の運転支援システムの開発に有効であると期待される。

#### 4. おわりに

本論文では、ドライバと自然に協力できる運転支援システム構築のため、運転挙動、車両前方の画像、人が付与した運転シーntagの間に跨がる運転トピックを推定し、これを基に運転挙動と前方画像から運転シーntagを予測する手法を提案した。提案手法に依れば、車両挙動情報と前方画像から抽出した低次の画像特徴量のみから渋滞や先行車の接近、車線数と言った比較的高次の運転コンテキストが推定可能となる。

予測された運転シーntagからの運転を説明する自然文生成や、盛んに研究されている物体認識技術を導入したより大規模なデータでの定量的評価が今後の課題である。

#### 参考文献

[Bando 13] Bando, T., Takenaka, K., Nagasaka, S., and Taniguchi, T.: Unsupervised drive topic finding from driving behavioral data, in *the IEEE IV* (2013)

[Blei 03] Blei, D. M., Ng, A. Y., and Jordan, M. I.: Latent Dirichlet Allocation, *JMLR*, Vol. 3, pp. 993–1022 (2003)

[Nakamura 09] Nakamura, T., Nagai, T., and Iwahashi, N.: Grounding of word meanings in multimodal concepts using LDA, in *the IEEE/RSJ IROS* (2009)

[Taniguchi 12] Taniguchi, T., Nagasaka, S., Hitomi, K., Chandrasiri, N., and Bando, T.: Semiotic prediction of driving behavior using unsupervised double articulation analyzer, in *the IEEE IV* (2012)

[Wang 09] Wang, C., Blei, D., and Fei-Fei, L.: Simultaneous Image Classification and Annotation, in *CVPR* (2009)