

自然言語で記述された運転文脈を活用する逆強化学習ベース運転行動予測

Inverse Reinforcement Learning-based Driving Behavior Prediction Utilizing Natural Language Context

○学 近藤 一冨 (東京科学大学) 正 下坂 正倫 (東京科学大学)

Kazusa KONDO, Institute of Science Tokyo, kondo@miubiq.cs.titech.ac.jp

Masamichi SHIMOSAKA, Institute of Science Tokyo, simosaka@miubiq.cs.titech.ac.jp

As autonomous driving technologies become widespread, context-aware driving behavior modeling on complex residential roads is increasingly demanded by society. While context-dependent behavior modeling has been acknowledged as effective, its application has been limited to discrete representations of external factors such as weather and time of day, failing to sufficiently reflect the diversity of real-world driving environments. We propose LLM-BiIRL, a method leveraging large language model (LLM) embeddings for bilinear inverse reinforcement learning, enabling consideration of driving situation diversity that conventional discrete representations cannot capture. Using cross-validation with driving data collected on multiple residential roads, we compare against conventional approaches to verify the effectiveness of LLM-based embedding representations obtained from textual descriptions of external factors, clarifying both the validity of the proposed method and effective text formats for describing external factors.

Key Words: Imitation learning, Behavior prediction, LLM embedding

1 序論

近年、高齢運転者による交通事故が多発しており、ヒューマンエラーを起因とする交通事故の抑制に貢献する運転行動分析が注目されている [1]。複雑な運転行動を人が書き下したルールで定義することは困難であるため、データから最適な行動規範をモデリングできる逆強化学習への期待が高まっている。既存研究は生活道路における危険を予測するための熟練ドライバの振舞いに注目しており、逆強化学習を用いた運転行動モデリングを提案している [2, 3, 4, 5]。また、大規模言語モデル (LLM) を応用する研究が急速に進展しており、Masked IRL [6] のように言語を逆強化学習に組み込む手法や、GPT-Driver [7] のようにセンサ情報をテキストに変換して LLM に軌跡を直接予測させる手法も提案されている。

運転行動は天候、時間帯、道幅、一方通行か否かといった外的要因によって変化する。例えば、晴れの昼間の道路と雨の夜間の道路では、同じ場所であっても運転者の行動は異なる (図 1)。このような外的要因を考慮することで、より人間らしい運転行動モデリングが可能になる。

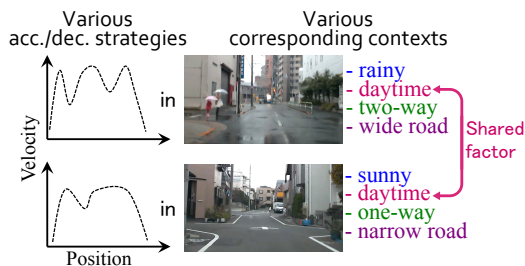


Fig.1: 文脈を考慮した運転データセットと共有文脈要因の適用概念。データが共有する文脈要因を活用することで、より柔軟かつ頑健な予測が可能となる。

Nishi ら [3, 4] は外的要因を考慮した運転行動モデリングを提案しており、「雨である/ない」「夜である/ない」「道路が狭い/広い」「一方通行である/ない」といった外的要因を one-hot エンコーディングで表現し、双線形逆強化学習 (BiIRL) に適用している。しかし、この手法にはいくつかの課題がある。第一に、考慮する外的要因の増加に伴う計算量の増大である。外的要因の数が増えるにつれ、外的要因ベクトルの次元が指数的に増加する。

第二に、考慮できる外的要因の柔軟性の欠如である。「天候」「時間帯」等、考慮できる外的要因が事前に固定されており、標識や歩行者、地域的特性といった新しい要因を入力として追加することが困難である。第三に、外的要因の捉え方が離散的である点である。「0/1」のバイナリ表現では、「小雨/大雨」「道幅が狭い/広い」等の連続的な区別ができない。

本研究では、これらの課題を解決するために、外的要因をテキスト化し、大規模言語モデル (LLM) でエンベディングすることで外的要因ベクトルを生成する手法 (LLM-BiIRL) を提案する。本研究の主な貢献は、外的要因をテキスト化し LLM でエンベディングすることで、考慮する要因が増えてもベクトルの次元数を固定できる手法を実現した点にある。また、ドライブレコーダーの画像等を用いることで、様々な外的要因に柔軟に対応可能な枠組みを構築し、外的要因を離散的でなく連続的に捉えることを可能にした。さらに、実車データを用いた実験により、提案手法が従来手法 (MEIRL および BiIRL) よりも高い予測精度を達成することを明らかにしている。

2 外的要因を考慮した逆強化学習

本章では、関連研究を整理した後、本研究の基盤となる双線形逆強化学習および本研究の位置づけについて述べる。

2.1 関連研究

IRL による運転行動モデリング。 逆強化学習 (IRL) は、専門家の示範データから報酬関数を推定することで、人間らしい行動をモデリングする手法として広く研究されている [8]。自動運転・運転支援の文脈では、Shimosaka ら [2] が生活道路における熟練ドライバの運転行動を IRL で学習する手法を提案し、Hirakawa ら [5] がブースティングを組み合わせることで精度を向上させている。

文脈要因を考慮した IRL。 単一のモデルでは天候・時間帯・道路状況などの外的要因による行動の変化を十分に捉えられない。Nishi ら [3, 4] はこれらの外的要因を one-hot エンコーディングで表現し、双線形逆強化学習に組み込む手法 (BiIRL) を提案した。しかしこのアプローチには、考慮する外的要因の増加に伴う次元の指数的増大、考慮できる要因が事前に固定されることによる柔軟性の欠如、離散表現では連続的な値の違いを捉えられないという課題がある。

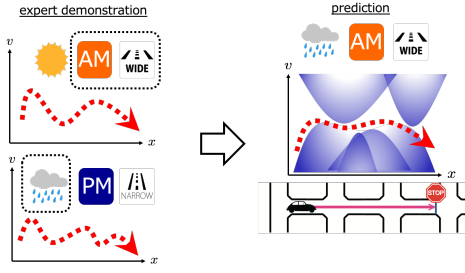


Fig.2: 双線形逆強化学習 (BiIRL) の概要図。

LLM を用いた行動モデリング研究。近年、大規模言語モデル (LLM) を応用する研究が急速に進展しており、言語を逆強化学習に組み込む Masked IRL [6] は示範データと言語から報酬関数を推定する手法であり、GPT-Driver [7] はセンサ情報をテキストに変換し LLM に軌跡を直接予測させる手法である。しかし、これらの研究では BiIRL の外的要因ベクトル $\mathbf{d}$ の生成方法そのものを LLM エンベディングに置き換えるアプローチは取られていない。

本研究の位置づけ。本研究は、BiIRL の枠組みにおける外的要因表現を、従来の離散的 one-hot 表現から LLM による連続的な Semantic Embedding へと置き換えることに新規性がある。これにより、次元爆発を回避しつつ、事前に定義されていない外的要因にも意味的な類似性を通じた汎化が期待できる。

2.2 逆強化学習の枠組み

多くの逆強化学習は何らかの動的システムが存在していることを前提とし、そのシステムにおける意思決定問題をマルコフ決定過程 (MDP) として定式化している。動的システムは、ある時刻 t における状態 $s_t \in \mathcal{S}$ と動作 $a_t \in \mathcal{A}$ が与えられると、次の時刻における状態 s_{t+1} への遷移は確率モデル $P(s_{t+1} | s_t, a_t)$ で表現できる。MDP は4つの要素の組 $\langle \mathcal{S}, \mathcal{A}, T, R \rangle$ で表される。ここで、 $\mathcal{S}$ と $\mathcal{A}$ は状態と動作の有限集合を、 $T: \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ は遷移関数を、 $R: \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ は報酬関数をそれぞれ示している。運転行動モデリングでは、状態 $s \in \mathcal{S}$ は道路上で離散化された車両位置を、動作 $a \in \mathcal{A}$ は隣接位置への移動方向をそれぞれ表す。

逆強化学習は通常、不完全な MDP $\langle \mathcal{S}, \mathcal{A}, T \rangle$ 、すなわち報酬関数 R を未知数として取り扱う。熟練者の教示データから報酬関数を学習することで、熟練者の行動を模倣する方策を獲得することが目的である。

最も基本的な逆強化学習手法として、Maximum Entropy IRL (MEIRL) [8] がある。MEIRL では、報酬関数を式 (1) のように特徴関数の線形結合として定義する。

$$R(s) = \mathbf{w}^\top \mathbf{f}(s) \quad (1)$$

ここで、 $\mathbf{w}$ は学習される重みベクトル、 $\mathbf{f}(s)$ は状態 s における特徴ベクトルである。

2.3 双線形逆強化学習

外的要因を考慮した運転行動モデリングを実現するために、Nishi ら [4] は双線形逆強化学習 (BiIRL) を提案している。BiIRL では、外的要因を表すベクトル $\mathbf{d}$ を導入し、報酬関数を式 (2) のように定義する。

$$R(s) = \mathbf{d}^\top \Theta \mathbf{f}(s) \quad (2)$$

ここで、 $\mathbf{d}$ は外的要因を表すベクトル、 Θ は学習される重み行列、 $\mathbf{f}(s)$ は状態 s における特徴ベクトルである。

従来の BiIRL では、各外的要因を図 3 に示すようなバイナリ (例: 「晴れ/雨」、「昼/夜」等) で表現し、それに固定値 1 を加えた拡張ベクトル $\mathbf{c}'_e = [1, \mathbf{c}_e^\top]^\top$ のクロネッカー積をとることで、各要因の組み合わせおよび共通性を表現する外的要因ベクトルを生成している。本研究では、外的要因ベクトル $\mathbf{d}$ の生成方法を LLM エンベディングに置き換えることで、上記の課題を解決する。

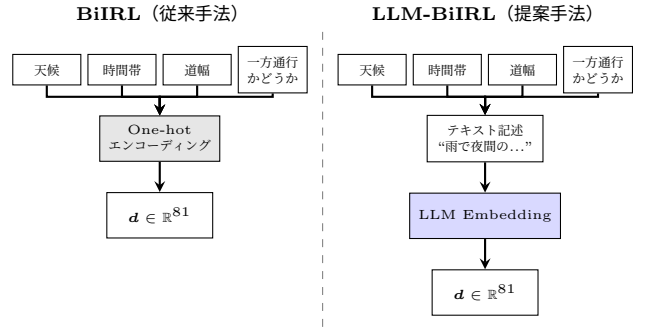


Fig.3: BiIRL (従来手法) と LLM-BiIRL (提案手法) の比較。従来の one-hot (離散) 表現を、LLM による Semantic Embedding (連続) 表現に置き換えることが本研究の核心である。

3 LLM エンベディングを用いた双線形逆強化学習

本研究は外的要因をテキスト化し、LLM でエンベディングすることで外的要因ベクトルを生成する手法 (LLM-BiIRL) を提案する。本章では、提案手法の概要と数理モデルについて述べる。

3.1 提案手法の概要

従来の BiIRL では、天気や道路状況などの外的要因を図 3 に示すような離散的な one-hot ベクトルとして表現していた。これに対し提案手法 LLM-BiIRL は、BiIRL の外的要因ベクトル $\mathbf{d}$ の生成方法を LLM エンベディングに置き換えることが本質的な差異である。外的要因をテキストに変換し、LLM エンベディングモデルで連続的なベクトル空間に写像することで、従来の離散 (one-hot) 表現が抱えていた3つの課題——次元の指数的増大・要因の固定化・離散表現の限界——を同時に解決できる。また、エンベディング次元はモデルごとに固定されているため、考慮する外的要因の情報量が増加しても重み行列 Θ のサイズを一定に保つことができる。学習アルゴリズムは BiIRL [4] と同様であり、報酬関数を $R(s) = \mathbf{d}^\top \Theta \mathbf{f}(s)$ として、双線形逆強化学習の枠組みで重み行列 Θ を最適化する。

3.2 外的要因のテキスト化とエンベディング

外的要因をテキスト化する手段として、本研究では2つのアプローチを採用する。第一はメタデータに基づくテキスト生成である。これは実際の天気・時間帯・道幅等の情報を直接テキスト化するもので、構造文 (例: 「天気: 晴れ / 道幅: 狭い」) や自然文 (例: 「天気は晴れで、道幅の狭い道路である。」) といった形式を検討する。第二は画像を元にしたテキスト生成である。図 4 に示すドライブレコーダーの画像を OpenAI 社のマルチモーダル LLM である GPT-4o に入力し、天気・時間帯・道幅等の外的要因を構造文や自然文形式でテキスト化する。また、周辺の駐車車両や歩行者等のメタデータで得ることのできない周辺情報も画像から動的に取得できる点が特徴である。生成されたテキストは OpenAI 社の text-embedding-3-large モデルに dimensions=81 を指定して入力することで、直接 81 次元の外的要因ベクトル $\mathbf{d} \in \mathbb{R}^{81}$ を得る。



Fig.4: 表 1 のテキスト生成に使用したドライブレコーダー画像。

Table 1: テキストの具体例

種別	粒度	形式	テキスト例
メタデータ	二値	構造文	天気: 晴れ / 時間帯: 昼 / 道路: 一方通行 / 道幅: 狭い
		自然文	天気は晴れであり、時間帯は昼頃。一方通行の狭い道路である。
	段階	構造文	天気: 晴れ / 時間帯: 昼 / 道路: 一方通行 / 道幅: やや狭い
		自然文	天気は晴れであり、時間帯は昼頃。一方通行のやや狭い道路である。
	数値	構造文	天気: 降水量 0.0mm / 時間: 12:31 / 道路: 一方通行 / 道幅: 6.2m
		自然文	天気は降水量 0.0mm で、時間は 12:31 である。一方通行の道幅 6.2m の道路である。
画像	二値	構造文	天気: 晴れ / 時間帯: 昼 / 道路: 一方通行 / 道幅: 狭い
		自然文	天気は晴れであり、時間帯は昼頃。一方通行の狭い道路である。
	段階	構造文	天気: 曇り / 時間帯: 昼 / 道路: 一方通行 / 道幅: やや狭い
		自然文	天気は曇りであり、時間帯は昼頃。一方通行のやや狭い道路である。
	4つの外的要因以外に関する文		道路の両側には住宅が立ち並んでおり、左側には駐車された車と自転車が見られます。右側にはフェンスがあり、歩行者の通行が予想されます。道路標識があり、進行方向を示しています。見通しは比較的良好ですが、交差点が近いので、歩行者や自転車の飛び出しに注意が必要です。特に左側の駐車車両付近では、車幅に注意しながら慎重に運転する必要があります。

4 性能評価

4.1 目的

本章では、提案手法である LLM-BiIRL が従来手法 (MEIRL および BiIRL) と比較して、運転行動の予測精度において優れていることを示す。また、テキストの形式 (構造文/自然文) や、量的情報 (降水量, 時間帯, 道幅等) のテキストでの表現粒度が予測精度に与える影響についても分析する。これにより、LLM エンベディングによる外的要因表現が実際の運転行動モデリングにどの程度寄与するかを明らかにする。

4.2 実験設定

実験には、生活道路 20 コースにおける熟練ドライバの走行データ (計 50 デモ) を使用し、5-fold 交差検証により評価を行った。評価指標は値が小さいほど精度が高い Modified Hausdorff Distance (MHD) を用いる。比較対象として用意した MEIRL は外的要因を一切考慮しないベースラインであり、BiIRL は天候・時間帯・道幅・一方通行かどうかの 4 要因を one-hot エンコーディングにより外的要因ベクトル $\mathbf{d} \in \mathbb{R}^{81}$ として表現する従来手法である。提案手法である LLM-BiIRL では、これらの外的要因をテキスト化し、OpenAI 社の text-embedding-3-large を用いて従来手法と同じ 81 次元のベクトルを生成して評価を行う。

LLM-BiIRL に入力するテキストとして計 12 種類を検討した。具体的には、メタデータから直接生成するテキスト (6 種類) と、ドライブレコーダー画像から GPT-4o を用いて生成するテキスト (6 種類) である。

メタデータベースでは、「構造文/自然文」の 2 形式と、量的情報のテキストでの表現粒度 (「二値表現/段階表現/数値表現」) の組み合わせで 6 種類を生成した。画像ベースでは、「構造文/自然文」×「二値表現/段階表現」の 4 種類に加え、駐車車両や歩

行者の存在など基本 4 要因以外の周辺情報のみを記述したテキスト (1 種類) と、「基本 4 要因以外の周辺情報のみを記述したテキスト」と「基本 4 要因を記述したテキスト」を結合したテキスト (1 種類) を含む 6 種類を検討した。

ここで「二値表現」は「狭い/広い」「雨/晴れ」等のように 1 つの閾値で二値に分類した表現、「段階表現」は「やや狭い/普通/やや広い」等のように複数の閾値で多段階に分類した表現、「数値表現」は降水量や道幅等の実測値をそのまま記載した表現を指す (表 1 参照)。

4.3 実験結果

Table 2: LLM-BiIRL におけるテキスト種類別の予測精度

カテゴリ	テキスト種類	Test MHD
メタデータ	構造文 (二値)	0.992±0.098
	構造文 (段階)	0.993±0.104
	構造文 (数値)	1.064±0.098
	自然文 (二値)	0.961±0.068
	自然文 (段階)	0.970±0.077
	自然文 (数値)	1.049±0.088
画像ベース	構造文 (二値)	1.010±0.110
	構造文 (段階)	1.025±0.107
	自然文 (二値)	1.014±0.086
	自然文 (段階)	1.016±0.112
	周辺情報のみ	1.200±0.132
	自然文 (段階) + 周辺情報	1.052±0.076

Table 3: 従来手法との予測精度 (MHD) の比較

手法	Test MHD	Train MHD
MEIRL (最良)	1.126±0.086	1.085±0.025
BiIRL (最良)	1.009±0.076	0.829±0.012
LLM-BiIRL (最良)	0.961±0.068	0.849±0.014

Table 4: 構造文と自然文の予測精度の比較

テキスト形式	Test MHD	改善率
構造文 (二値)	0.992±0.098	-
自然文 (二値)	0.961±0.068	3.1%
構造文 (段階)	0.993±0.104	-
自然文 (段階)	0.970±0.077	2.3%
構造文 (数値)	1.064±0.098	-
自然文 (数値)	1.049±0.088	1.4%

Table 5: 表現粒度での予測精度の比較 (メタデータ)

粒度	構造文	自然文
二値	0.992±0.098	0.961±0.068
段階	0.993±0.104	0.970±0.077
数値	1.064±0.098	1.049±0.088

Table 6: メタデータと画像ベースの予測精度の比較

入力ソース	二値	段階
メタデータ	0.961±0.068	0.970±0.077
画像ベース	1.010±0.110	1.016±0.112

Table 7: 周辺情報の有無による予測精度の比較

外的要因の情報	Test MHD
周辺情報 (4 つの基本的外的要因以外) のみ	1.200±0.132
4 つの基本的外的要因 (段階) + 周辺情報	1.052±0.076
4 つの基本的外的要因 (段階) のみ	1.016±0.112

4.4 全テキスト種類における予測精度比較

表 2 に、実験で用いた全 12 種類のテキストにおける予測精度の詳細を示す。全体としてメタデータに基づくテキストが、ドライブレコーダー画像から生成されたテキスト（画像ベース）よりも一貫して高い予測精度を達成した。特に、メタデータの情報を自然文形式で二値表現で記述したテキストにおいて、最も良好な精度である MHD 0.961 が得られている。これは、天候や道幅といった運転行動に直接的な影響を及ぼす情報を、画像解釈によるノイズを介さずに LLM が適切に処理できたためと考えられる。

4.5 従来手法との優位性検証

表 3 は、各手法における最良の予測精度（MHD）を比較したものである。提案手法である LLM-BiIRL（最良）は、テストデータにおいて MHD 0.961 を達成し、外的要因を考慮しない MEIRL (1.126) や外的要因を one-hot エンコーディングで表現する従来の BiIRL (1.009) を上回った。

この結果から、LLM エンベディングによる外的要因の表現が、運転行動モデリングの精度向上に有効であることが示唆された。図 3 に示すように、BiIRL が外的要因を one-hot エンコーディングで離散的に表現するのにに対し、LLM-BiIRL ではテキストを介した LLM エンベディングにより連続的な表現を実現しており、より豊かな運転文脈を反映できることがその要因と考えられる。

4.6 テキスト記述形式と表現粒度の影響分析

表 4 に、構造文と自然文の記述形式による精度の差異を示す。実験したすべての条件において、自然文形式が構造文形式を上回る精度を示しており、最大で約 3.1% の改善が確認された。これは、単なるキーワードの羅列（構造文）よりも、文章としての文法的・意味的な連続性を持つ自然文の方が、LLM が文脈的な繋がりをエンベディングベクトルにより適切に反映できたためであると推察される。この傾向は表記粒度（二値表現、段階表現、数値表現）によらず一貫して観察されており、文章形式の優位性は LLM の文脈理解能力に由来するものと考えられる。

表 5 は、メタデータに基づくテキストにおける表記粒度が精度に与える影響を比較したものであり、詳細な「数値表現」を含むテキストよりも、「二値表現」や「段階表現」を用いたテキストの方が高い精度が得られた。これは、LLM にとって詳細すぎる数値情報は予測におけるノイズとなりやすく、むしろ「狭い」「広い」といった抽象的かつ文脈的な表現の方が、走行環境の特徴を適切に埋め込み空間に写像できたためと考えられる。

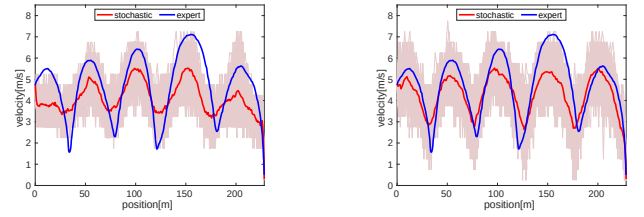
4.7 入力メディアと周辺情報の感度評価

表 6 および表 7 は、入力メディアの差異と基本要因以外の周辺情報の影響を示している。まず入力メディアに関しては、メタデータ由来のテキストが画像ベースのテキストと比較して一貫して高い精度を維持しており、画像からのテキスト生成において LLM による外的要因の解釈に誤りが含まれたためと推察される。

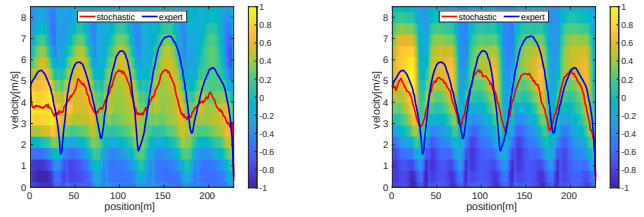
また、周辺情報（駐車車両や歩行者等）のみを入力した場合は精度が大きく悪化したものの、基本要因に周辺情報を加えた際の精度の悪化は限定的であった。この事実、LLM エンベディングが主要な走行文脈と補助的な情報ある程度区別し、モデル全体としての堅牢性を維持できる能力を有していることを示唆している。

4.8 予測精度の定性的解釈

BiIRL に対する LLM-BiIRL の全デモ平均 MHD 改善幅は 0.048 であった。この数値が実際の予測軌跡においてどのような差異として現れるかを視覚的に確認するため、50 デモの中から BiIRL に対する改善幅が平均値に近い代表的な走行ケース（改善幅 0.040）を選択した。図 5 で確認できるように、提案手法の方が熟練者の行動に対する再現性が高く、より忠実な軌跡を生成できている。また、図 6 に示すように、提案手法の方が報酬値の高い（黄色）領域に熟練者の走行軌跡が集中しており、適切な報酬関数が学習できていることが確認できる。



(a) 従来手法 (BiIRL) (b) 提案手法 (LLM-BiIRL)
Fig. 5: 速度プロファイルの予測比較 (青: 熟練者, 赤: 確率的予測).
BiIRL: MHD=1.678, LLM-BiIRL: MHD=1.637 (差 0.040).



(a) 従来手法 (BiIRL) (b) 提案手法 (LLM-BiIRL)
Fig. 6: 報酬関数を背景とした速度プロファイルの比較 (青: 熟練者, 赤: 確率的予測, 背景: 報酬値).

5 結論

本研究では、BiIRL の外的要因ベクトルを LLM エンベディングで置き換える LLM-BiIRL を提案した。従来の離散的 one-hot 表現を連続的な Semantic Embedding へと変換することで、次元爆発の回避・柔軟な要因追加・連続値の取り扱いを同時に実現した。5-fold 交差検証を用いた実験の結果、提案手法 LLM-BiIRL (MHD 0.961) が従来の BiIRL (1.009) および MEIRL (1.126) をともに上回る性能が確認された。特にメタデータを自然文・二値表現で記述したテキストが最良であった。今後の課題として、より多様な道路環境への適用が挙げられる。

参考文献

- [1] Mozghan Nasr Azadani and Azzedine Boukerche. Driving behavior analysis guidelines for intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):6027–6045, 2021.
- [2] Masamichi Shimosaka et al. Modeling risk anticipation and defensive driving on residential roads with inverse reinforcement learning. In *Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1694–1700, 2014.
- [3] Kentaro Nishi and Masamichi Shimosaka. Predicting driving behavior using inverse reinforcement learning with multiple reward functions towards environmental diversity. In *Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV)*, 2015.
- [4] Kentaro Nishi and Masamichi Shimosaka. Fine-grained driving behavior prediction via context-aware multi-task inverse reinforcement learning. In *Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2281–2287, 2020.
- [5] Yuki Hirakawa and Masamichi Shimosaka. 特徴量探索とパラメータ最適化の交互実行によるコンパクトな運転行動推定モデル. In 第 22 回計測自動制御学会システムインテグレーション部門講演会, 2021.
- [6] Minyoung Hwang et al. Masked IRL: LLM-guided reward disambiguation from demonstrations and language. arXiv:2511.14565, 2025.
- [7] Jiageng Mao et al. GPT-Driver: Learning to drive with GPT. In *Workshop on Foundation Models for Decision Making at NeurIPS*, 2023.
- [8] Brian D Ziebart et al. Maximum entropy inverse reinforcement learning. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI)*, volume 8, pages 1433–1438, 2008.