SEMI-SUPERVISED ESTIMATION OF DRIVING BEHAVIORS USING ROBUST TIME-CONTRASTIVE LEARNING

Takuma Kuroki* Osamu Shouno† Junichiro Yoshimoto*‡
* Nara Institute of Science and Technology, Nara, Japan
† Otemon Gakuin University, Osaka, Japan
‡ E-mail: juniti-y@is.naist.jp Tel/Fax: +81-743-72-5984/5989

***Abstract*—Estimation of driving behaviors is an elemental technology in a driving support system for a vehicle. For realizing intelligent estimation of driver behaviors, many studies have explored the use of machine learning methods mainly in a supervised fashion that require a large amount of labeled driving data. In this study, we hypothesize that the time-contrastive learning (TCL) could be helpful for reducing the number of labeled data for the supervised learning and numerically tested it using a public data set. For this purpose, we constructed three models to estimate driving behaviors from vehicle dynamics: 1) a naive linear classifier implemented by linear discriminant analysis (LDA) model; 2) an LDA classifier combined with a feature extraction process by the original TCL; 3) the same as 2) except the robust version of TCL was employed instead of the original TCL. The results were not supportive to our hypothesis: Model 1) showed better performance than the other models when very few labeled data was available; and two models with TCL outperformed the other without TCL for a considerable number of labeled data. We conclude discussions on some limitations of this study and open issues for the future.**

I. INTRODUCTION

Estimation of driving behaviors is an elemental technology in a driving support system for a vehicle. Even in encountering drowsy driving and road rage, successful estimation of driving behaviors enables the system to shift to automatic driving mode swiftly and protect the drivers from critic accidents. Driver behavior analysis is also beneficial for car insurance industry to assess potential risks of their customers for providing fair insurance premiums.

For realizing intelligent estimation of driver behaviors, many studies have explored the use of machine learning methods mainly in a supervised fashion that require labeled driving data [1], [2], [3], [4], [5], [6], [7], [8]. For improving performance, a supervised approach usually requires a large amount of data recorded in realistic driving situations. However, the acquisition of labeled data for events or behaviors is often very costly or even infeasible.

In recent years, unsupervised learning has drawn increasing attention from the computer vision community because it can utilize unlabeled data efficiently for learning latent features, which can be a good basis for transfer learning to various tasks. Thanks to this property, unsupervised learning can be very beneficial when a large amount of data is available but only a quite small fraction of them are labeled. An interesting approach of this direction in the computer vision is unsupervised feature learning by maximizing distinction between image instances [9]. A similar approach for time series is the time-contrastive learning (TCL) that learns features by maximizing discrimination among all segments in a time series [10]. Theoretically, TCL combined with linear independent component analysis (ICA) is equivalent to the nonlinear ICA. In this study, we hypothesize that the analogy can be transferred to a context of the estimation of driving behaviors, and numerically tested it using a public data set.

II. METHODS

A. Problem Setting

We assume a situation where we have time-series data with the length of T, $D \equiv \{\boldsymbol{x}(t) \in \Re^P\}_{t=1}^T$, measured by some device attached to a vehicle. Herein, we simply refer to $\boldsymbol{x}(t)$ as a measurement (or measurement vector) with the dimensionality of P. For a specific subset of the meaturements (i.e. a labeled data set) $D_l \subseteq D$, any datum $\boldsymbol{x}(t) \in D_l$ is associated with a class label $y(t) \in \{1, \ldots, K\}$ indicating a human-annoted category of driving behaviors (e.g. nomal, drowsy, etc.). On the contrary, none of other measurements (i.e. unlabelded data) $\boldsymbol{x}(t) \in D - D_l$ are associated with the class label at all.

The goal of this study is to investigate whether or not we can construct a model that can precisely estimate a class label $\hat{y}(t)$ corresponding to a new measurement $\hat{x}(t)$, based on information about a given time-series D and their associated class labels. Accordingly, the problem falls into a class of semi-supervised learning except the features are given by time-series.

B. Data Set

1) UAH-DriveSet: Instead of real driving experiments, we employed a public data set named *UAH-DriveSet* [7] as a benchmark for our study. The data was collected from six different drivers (five males and one female), who were asked to perform three different behaviors (normal, drowsy and aggressive) on two types of public roads (motorway and secondary road). During the driving, a large amount of variables were captured and processed by all the sensors and capabilities of a smartphone and the data contains more than 500 minutes

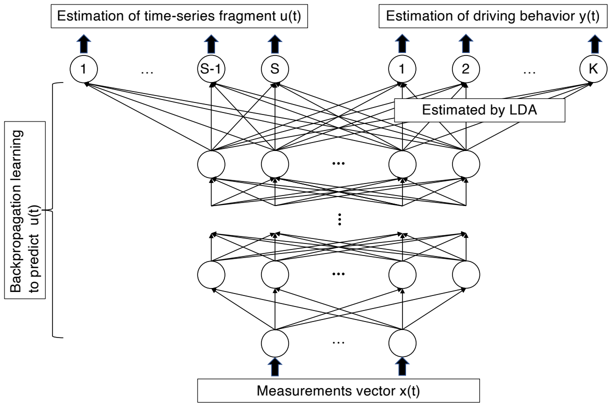


Fig. 1. Overview of two-step model for estimating the driving behaviors.

of naturalistic driving with its associated raw data and additional semantic information (see more details at http://www.robesafe.uah.es/personal/eduardo.romera/uah-driveset). In this study, we extracted only four variables of `speed`, `roll`, `pitch` and `yaw`, from the data set to assume a difficult situation.

2) Preprocessing: In *UAH-DriveSet*, `speed` was measured by a GPS sensor with the sampling rate of 1 Hz while the others (i.e. `roll`, `pitch` and `yaw`) were measured by an accelerometer with the sampling rate of 10 Hz. To resolve the inconsistency of the sampling rate, time-series of `speed` was up-sampled by ten times with linear interpolation so that all variables had 10 Hz frequency after the preprocessing. Additionally, in order to reduce possible effects of high dynamic ranges of the features, we linearly normalized the values for each variable so that the distribution of each variable have zero mean and unity standard deviation.

Finally, we defined $\boldsymbol{x}(t)$ as a collection of preprocessed values of (`speed`, `roll`, `pitch`, `yaw`). Each class label $y(t)$ associated with $\boldsymbol{x}(t) \in D_l$ was defined as follows:

- $y(t) = 1$ if normal driving was performed at time t.
- $y(t) = 2$ if drowsy driving was performed at time t.
- $y(t) = 3$ if aggressive driving was performed at time t.

C. Estimation of Driving Behaviors

For estimating the driving behaviors based on both labeled and unlabeled data, we considered a two-step model consisting of unsupervised feature extraction followed by supervised classification. In this section, we explain our implementation of those steps (See the overview in Figure 1).

1) Feature Extraction by TCL: In the first step, non-linear independent component analysis was executed to transform a measurement $\boldsymbol{x}(t)$ into a task-independent feature vector $\boldsymbol{z}(t)$ that can maximize the information content after the transformation.

For this pupose, the time-contrastive learning (TCL) [10] or Robust TCL [11] was employed in this study. TCL was typically implemented by a multi-layer neural network with multiclass logistic regression. The training data for TCL were prepared in the following procedure. First, the measurement time-series D was segmented into S fragments with the equal length, each of which was indexed by $s = 1, \ldots, S$. For each $\boldsymbol{x}(t) \in D$, its corresponding class label is given by $u(t) = s$ if $\boldsymbol{x}(t)$ was assigned to the s-th fragment. The neural network was trained so as to approximate the mapping from $\boldsymbol{x}(t)$ to $u(t)$ by maximizing the cross entropy loss:

$$L = -\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{S} \mathbb{I}(u(t), s) \ln P(s|\boldsymbol{x}(t)),$$

where $\mathbb{I}(\cdot, \cdot)$ is the indicator function that returns 1 if two arguments are the same and 0 otherwise; and $P(l|\boldsymbol{x}(t))$ is the output of the neural network expressing the probability that $\boldsymbol{x}(t)$ belongs to the s-th fragment. After the training, the values computed in the last hidden layer of the neural network were used as the feature vector $\boldsymbol{z}(t)$ corresponding to $\boldsymbol{x}(t)$. Theoretically, $\boldsymbol{z}(t)$ can be regarded as a mixing signal formed by linear combination of statistically independent signals (See theoretical details in [10]). Thus, $\boldsymbol{z}(t)$ is expected to contain the same information content as the the independent signal.

The original TCL mentioned above is vulnerable to outliers, which often happen in the real driving situation. To address the issue, we also considered the use of Robust TCL [11] as an alternative. The main difference between the original and Robust TCLs is the objective function to train the neural network. In Robust TCL, the objective function is replaced by the γ-cross entropy, defined as follows:

$$L = -\frac{1}{\gamma} \ln \left[\frac{1}{T} \sum_{t=1}^{T} \frac{\sum_{s=1}^{S} \mathbb{I}(u(t), l) r(s, \boldsymbol{x}(t))^{\gamma}}{\left(\sum_{s'=1}^{S} r(s', \boldsymbol{x}(t))^{\gamma+1} \right)^{\frac{\gamma}{\gamma+1}}} \right],$$

where $r(s, \boldsymbol{x}(t))$ is the internal potential of the s-th output unit in the neural network such that

$$r(s, \boldsymbol{x}(t)) = \exp(\boldsymbol{w}_s^\top \boldsymbol{z}(t) + b_s)$$

and

$$P(s|\boldsymbol{x}(t)) = \frac{r(s, \boldsymbol{x}(t))}{\sum_{s'=1}^{L} r(s', \boldsymbol{x}(t))}.$$

$\boldsymbol{w}_s$ is the weight vector between the last hidden layer and the s-th output unit, and b_s is the bias of the s-th output unit.

2) Classification: In the second step, a linear classification was executed to transform the feature vector into the corresponding driving behavior. For this purpose, we employed a multiclass linear descriminant analysis (LDA) model, for which the training data set consisted of a collection of feature vectors $\boldsymbol{z}(t)$ associated with the labeled data $\boldsymbol{x}(t) \in D_l$ and their corresponding class labels $y(t)$.

III. Results

To test the efficiency of feature extraction by TCL, we performed numerical experiments and compared three conditions:

1) *Naive-LDA*: The model was trained by LDA without any feature extraction process to estimate a class label of driving behavior $y(t)$ directly from a given measurement

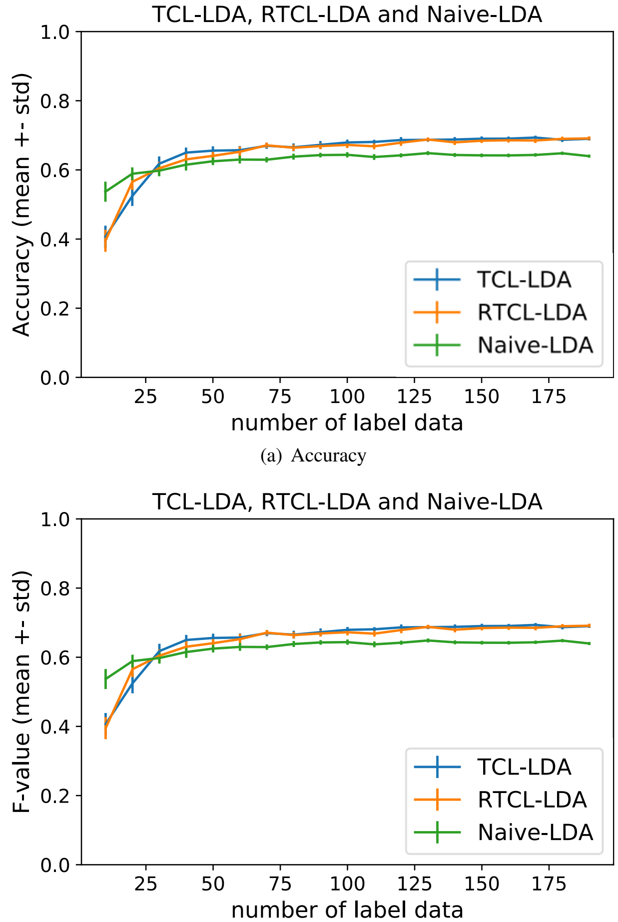
$\boldsymbol{x}(t)$. In other words, the feature extraction was done by the identify function such that $\boldsymbol{z}(t) = \boldsymbol{x}(t)$.

2) *TCL-LDA*: The model was trained by LDA to estimate a class label of driving behavior $y(t)$ from a feature vector $\boldsymbol{z}(t)$, which is the output of a neural network trained by the original TCL for a given measurement $\boldsymbol{x}(t)$.
3) *RTCL-LDA*: The model was the same as *TCL-LDA* except Robust TCL was employed instead of the original TCL.

For implementation of TCL, we employed a simple three-layer neural network model with a single hidden-layer consisting of 250 units. The number of fragments to generate the training data set for TCL was set at $S = 200$.

To evaluate the performance achieved in these conditions, we randomly assigned 80% of samples in *UAH-DriveSet* to the training data set and the remaining 20% to the test data set. Among the training data set, only n samples were randomly selected and treated as the labeled data by associating them with their corresponding class labels of driving behaviors; and the remaining m ($m \gg n$) samples were treated as the unlabeled data in order to assume that the availability of labeled data was extremely limited. Additionally, n varied from 10 to 200 with increment of 10 to check the improvement by increase in the number of labeled data. The accuracy and the F1 score for the test data set was used as the performance indices of each condition, and averaged over 100 independent runs to reduce the sampling randomness.

Figure 2 shows the results of the performance comparison among three models. Naturally, the performance was improved more as the number of labeled data increased for all models. In terms of comparison amond the models, *Naive-LDA* showed better performance than the others when the number of labeled data was 10. However, the difference almost disappeared around 30 labeled data and two models including TCL (i.e. *TCL-LDA* and *RTCL-LDA*) outperformed *Naive-LDA* when 70 or more labeled data were available. The difference between *TCL-LDA* and *RTCL-LDA* was quite little for any condition.



(a) Accuracy

(b) F1 score

Fig. 2. Performance comparison among three models: *Naive-LDA*, *TCL-LDA* and *RTCL-LDA*. (a) accuracy and (b) F1 score were represented as the performance indices. The solid lines and the error bars represents means and standard deviations over 100 independent runs.

IV. Discussions & Conclusions

In this study, we hypothesized that nonlinear ICA implemented by the original or its robust version of time-contrastive learning could be helpful for reducing the number of labeled data to realize an intelligent estimation of driving behaviors from vehicle dynamics, and numerically tested it using a public data set named *UAH-DriveSet*.

For this purpose, we constructed three models to estimate driving behaviors: 1) a naive linear classifier implemented by linear discriminant analysis (LDA) model; 2) an LDA model combined with a feature extraction process by the original TCL; 3) the same model as 2) except Robust TCL was employed instead of the original TCL. However, the results were not supportive to our hypothesis: Model 1) showed better performance than the other models when very few labeled data was available; the effectiveness of TCL became clearer as the number of labeled data increased; and finally, two models with feature extraction with TCL outperformed the other without any feature extraction process for a considerable number of labeled data. The difference between the original and Robust TCLs was quite little for any condition.

A reason why the naive LDA classifier without feature extraction was better in a scenario with very few labeled data may be that the distributions of measurements corresponding to different types of driving behaviors are overlapped with each other in *UAH-DriveSet* so that only a half of data were challenging to distinguish the corresponding behaviors. This can be speculated by the result that the naive LDA classifier achieved more than 50% accuracy (much better performance that the chance level) though only 10 labeled data was available.

However, the performance improvement of the naive LDA was saturated soon even if the number of labeled data increased, implying that the data requiring intelligent estimation could have nonlinear boundary between different driving behaviors. On the contrary, nonlinear feature extraction by TCL could transform the boundary so that the switching points from one behavior to another was emphasized, resulting in better improvement as the available labeled data increased. To confirm the speculation, we need to visualize and analysis the feature space constructed by TCL in the future study.

The result showing comparable performance between the original and Robust TCLs was also different from our hypothesis. This could be because *UAH-DriveSet* is so clean that there are very few outliers in the data set. The data were collected in real driving experiments but the participants were asked to mimic various driving behaviors, which may be different from the natural situation. The cloud-of-things via smartphone attached to volunteers' cars could be helpful for data collection in more natural situation as well as improve its efficiency: This is also a possible direction for future studies.

Acknowledgements

The authors thank Dr. Hiroaki Sasaki at the Future University Hakodate, Japan, for his valuable comments and providing the python package for Robust TCL.

References

[1] E. Ohn-Bar, S. Martin, A. Tawari, and M. M. Trivedi, "Head, eye, and hand patterns for driver activity recognition," in *2014 22nd International Conference on Pattern Recognition*, 2014, pp. 660–665.

[2] S. Martin, S. Vora, K. Yuen, and M. M. Trivedi, "Dynamics of driver's gaze: Explorations in behavior modeling and maneuver prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 2, pp. 141–150, 2018.

[3] L. Wei, S. C. Mukhopadhyay, R. Jidin, and C.-P. Chen, "Multi-source information fusion for drowsy driving detection based on wireless sensor networks," in *2013 Seventh International Conference on Sensing Technology (ICST)*, 2013, pp. 850–857.

[4] T. Nguyen, S. Ahn, H. Jang, S. C. Jun, and J. G. Kim, "Utilization of a combined EEG/NIRS system to predict driver drowsiness," *Scientific Reports*, vol. 7, no. 1, 43933, 2017.

[5] V. Vaitkus, P. Lengvenis, and G. Žylius, "Driving style classification using long-term accelerometer information," in *2014 19th International Conference on Methods and Models in Automation and Robotics (MMAR)*, 2014, pp. 641–644.

[6] M. Van Ly, S. Martin, and M. M. Trivedi, "Driver classification and driving style recognition using inertial sensors," in *2013 IEEE Intelligent Vehicles Symposium (IV)*, 2013, pp. 1040–1045.

[7] E. Romera, L. M. Bergasa, and R. Arroyo, "Need data for driver behaviour analysis? presenting the public uah-driveset," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016, pp. 387–392.

[8] K. Saleh, M. Hossny, and S. Nahavandi, "Driving behavior classification based on sensor data fusion using lstm recurrent neural networks," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 1–6.

[9] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3733–3742.

[10] A. Hyvarinen and H. Morioka, "Unsupervised feature extraction by time-contrastive learning and nonlinear ica," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016.

[11] H. Sasaki, T. Takenouchi, R. Monti, and A. Hyvarinen, "Robust contrastive learning and nonlinear ica in the presence of outliers," in *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, ser. Proceedings of Machine Learning Research, J. Peters and D. Sontag, Eds., vol. 124. PMLR, 03–06 Aug 2020, pp. 659–668.