2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Automatic drive annotation via multimodal latent topic model

Takashi Bando[1], Kazuhito Takenaka[1], Shogo Nagasaka[2], and Tadairo Taniguchi[2]

Abstract— **Time-series driving behavioral data and image sequences captured with car-mounted video cameras can be annotated automatically in natural language, for example, "in a traffic jam," "leading vehicle is a truck," or "there are three and more lanes." 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, using multimodal latent Dirichlet allocation, latent driving topics laid on each driving situation were estimated as a relation 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 latent-driving topics. The proposed method intuitively annotated more than 50,000 pieces of frame data, including urban road and expressway data. The effectiveness of the estimated drive topics was also evaluated by analyzing the performances of driving-situation classification. The topics represented the drive context efficiently, i.e., the drive topics lead to a 95% lower-dimensional feature space and 6% higher accuracy compared with a high-dimensional raw-feature space. Moreover, the drive topics achieved performance almost equivalent performance to human annotators, especially in classifying traffic jams and the number of lanes.**

I. INTRODUCTION

Recently, various advanced driver-assistance systems (ADASs) have been proposed; however, most of them assist a driver just before the vehicle runs into danger. To prevent further accidents, "multiplex assistance systems," which take several actions from well before an actual collision occurs, are needed. Securing additional time before a collision is useful for senior drivers to avoid accidents. However, securing sufficient additional time and providing intuitive support to the driver is difficult to do because of the diversity of complicated driving situations. Most conventional ADASs describe a driving situation by using simple physical-risk criteria such as the time to collision[1]. Such criteria have been useful in supporting a driver just before the vehicle runs into danger, and several extensions of such criteria have also been proposed, such as for handling uncertain situations[2]. However, several studies reported that these simple models, defined over physical-feature space, are not suitable for long-term prediction of driving behavior[3]. Conversely, human drivers recognize a driving situation more abstractly and achieve longer prediction against the complexity and diversity of driving situations.

Adjusting such a situation recognition strategy is key technology for creating a novel ADAS that can achieve fully cooperative driving between the driver and the vehicle. To do this, Taniguchi et al.[4] proposed an unsupervised method for segmenting driving behavioral data into semiotic symbols that represent driving situations, that is, a double articulation analyzer (DAA), and achieved longer prediction of symbolic driving situations. Even though their data-driven estimation is based only on the maximization of the generative probability of driving behavioral data, Takenaka et al.[5][6] reported that the estimated segmentation points correspond well to human recognition of changes in a driving context. DAA is an efficient driving-situation segmentation method; however, it has some drawbacks. It is a fully unsupervised method, and thus, extracted driving situations do not have intuitive situation labels for informing a driver. Moreover, more than 400 kinds of situations are extracted for 90 minutes of driving, which is too many for the driver to understand intuitively.

Bando et al.[7] proposed a framework for automatically translating driving situation symbols into "drive topics" in natural language. They used latent Dirichlet allocation (LDA) for clustering extracted driving situations into a small number of drive topics in accordance with the emergence frequency of the physical behavioral features observed in each driving situation. The labels for the extracted drive topics were also determined automatically by using the distributions of the physical behavioral features included in each drive topic. Because drive operations are executed on the basis of a driver's perception of environments, their approach factored environmental factors into drive topic estimation indirectly. A contrastive and more intuitive approach is to consider surrounding environmental information directly and to interpret a driving situation in natural language; annotating to behavioral data and image sequence captured with car-mounted camera in natural language.

In this paper, we first discretize time-series driving behavior into a symbol sequence that represents driving situations, and then, consider annotating them. We use DAA[4] for the segmentation of driving behavioral data, and the multimodal latent topic model for estimating multimodal drive topics that represents the relationship among the probabilistic distributions of driving behavioral features, image features, and human annotations observed in each driving situation. Pre-

*This work was not supported by any organization

[1]T. Bando and K. Takenaka are with Corporate R&D Div. 3, DENSO CORPORATION, Aichi, Japan {takashi_bandou, kazuhito_takenaka}@denso.co.jp

[2]S. Nagasaka and T. Taniguchi are with College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan s.nagasaka@em.ci.ritsumei.ac.jp, taniguchi@ci.ritsumei.ac.jp

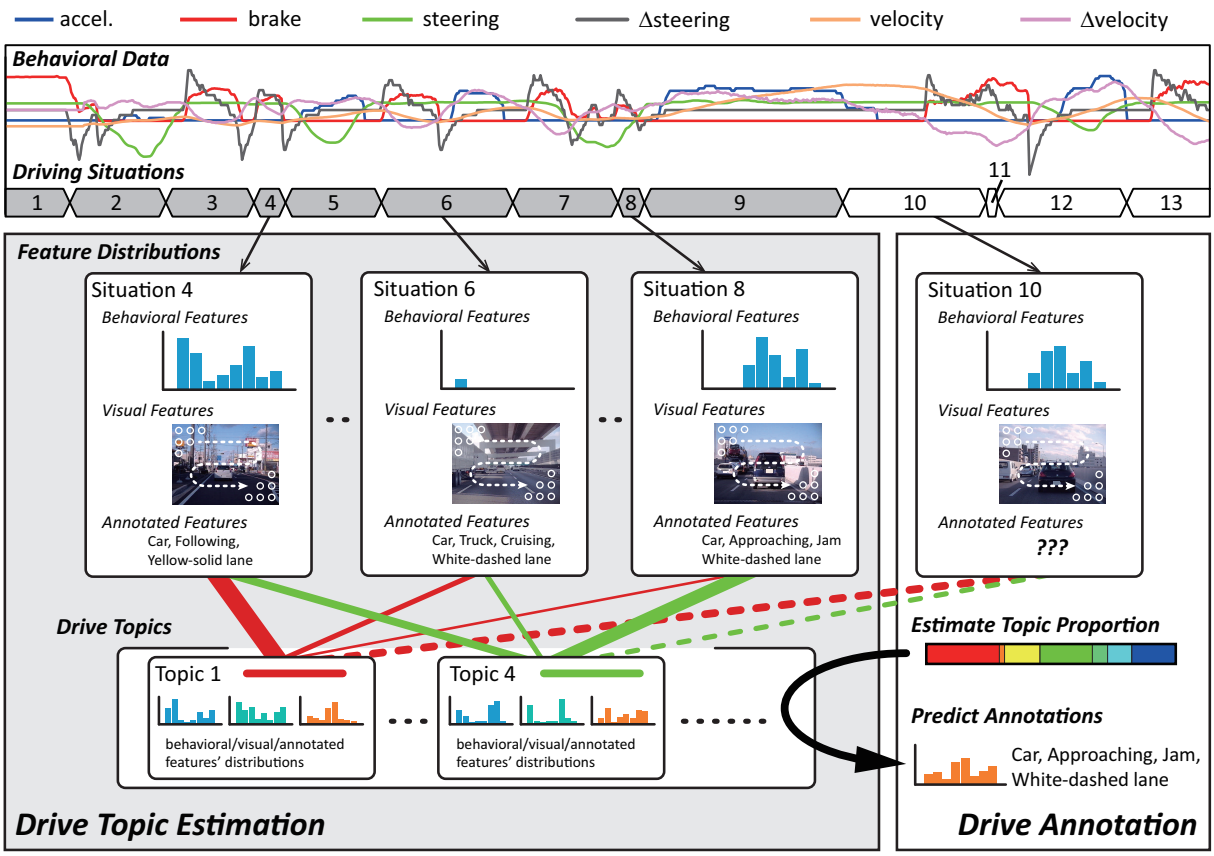

Fig. 1. Framework for drive annotation. Time-series driving behavioral data is discretized into sequence of driving situations. Then, distributions of observed behavioral, visual, and annotated features in each driving situation are modeled as mixtures of drive topics. The drive-topic proportion reflects abstractive driving context, and thus, the drive topics are useful for vehicles to recognize driving situation like a human driver.

dictive distribution of intuitive driving-situation annotations can be calculated via drive topics in accordance with the observed driving behavioral data and the image sequence around one's own vehicle.

Several situation annotation methods have been proposed in the image processing area, e.g., annotation estimation methods based on the latent topic model[8] and the deep Boltzmann machine[9]. Because these image annotation methods treat input images independently of each other, they could not consider the drive context, which is included in time-series driving behavior.

II. Automatic Drive Annotation

A. Overview

Figure 1 shows our framework for drive annotation. In this framework, time-series driving behavior is segmented by using DAA. By using a driver's operational signals, e.g., the angle of steering wheel, as input signals, DAA is able to extract the points of change in driving situations.

Since a driving situation has a high level of abstraction, annotations of driving situations have a large variance, even for human annotators. Conversely, the proposed framework annotates robustly because the variance of annotations are considered as the distributions of the multimodal features.

B. Double Articulation Analyzer

The first step of the proposed framework is the segmentation of continuous driving behavior into sequences of driving situations by using DAA[4]. DAA uses sticky hierarchical Dirichlet process-hidden Markov model (sHDP-HMM)[10] and nested Pitman-Yor language model (NPYLM)[11] to segment continuous driving behavioral data into sequences of driving situations. sHDP-HMM is a non-parametric extension of a conventional HMM that is widely used for time-series modeling. In addition to automatic determination of the number of hidden states, sHDP-HMM achieves more effective modeling of a continuous real-world data stream by introducing self-transition bias. sHDP-HMM segments the behavioral data into sequences of primitive elements of driving behavior on the basis of its locality in observed state space. NPYLM is a non-parametric parsing method used in the field of natural language processing. Note that NPYLM does not assume a preexisting dictionary; that is, it can parse sequences of extracted driving primitives into unknown driving situations and estimate the driving situation dictionary simultaneously.

C. Multimodal Latent Topic Model

Latent Dirichlet allocation (LDA)[12] is one of the most basic latent-topic models and is widely used for natural language processing. Recently, LDA extensions that can deal

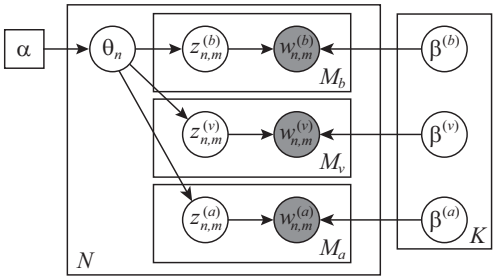

Fig. 2. Graphical model representation of multimodal latent Dirichlet allocation

with multimodal information, i.e., multimodal LDA (mLDA), have been proposed for annotating images[8] and grounding word meanings for robots[13]. Figure 2 shows a graphical model representation of mLDA. In the figure, the number of segmented driving situations is represented as N, and the number of drive topics is K. The behavioral features and visual features observed in the n-th driving situation are represented as $w^{(b)}_{n,m}$ and $w^{(v)}_{n,m}$, respectively, where, $M^{(b)}_n$ is the number of the behavioral features and $M^{(v)}_n$ is the number of the visual features. The human tags annotated for the n-th driving situation is represented as $w^{(a)}_{n,m}$, where $M^{(a)}_n$ is the number of the annotations. Here, mLDA assumes that the m-th feature $w^{(\cdot)}_{n,m}$ observed in the n-th driving situation is generated in accordance with the latent topic assignment $z^{(\cdot)}_{n,m}$, namely, the actual value w of the feature $w^{(\cdot)}_{n,m}$ is assumed to be generated in accordance with $\boldsymbol{\beta}^{(\cdot)}_{z_{n,m}}$. The assignment of latent topics is drawn from $\boldsymbol{\theta}_n$ with Dirichlet parameter $\boldsymbol{\alpha}$. Hence, mLDA assumes

$$\boldsymbol{\theta}_n \sim \mathrm{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha}), \tag{1}$$

$$z^{(\cdot)}_{n,m} \sim \mathrm{Mult}(z|\boldsymbol{\theta}_n), \tag{2}$$

$$w^{(\cdot)}_{n,m} \sim \mathrm{Mult}(w|\boldsymbol{\beta}^{(\cdot)}_{z_{n,m}}). \tag{3}$$

The log likelihood of the parameters is calculated as

$$\log p\left(\boldsymbol{w}|\boldsymbol{\alpha},\boldsymbol{\beta}\right) = \log \int \sum_{\boldsymbol{z}} p\left(\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{z}|\boldsymbol{\alpha},\boldsymbol{\beta}\right) d\boldsymbol{\theta}. \tag{4}$$

Because simultaneous optimization done with $\boldsymbol{\theta}, \boldsymbol{z}$ is actually intractable, several approximation methods, such as using variational approximation[12] or Gibbs sampling[14], have been proposed. For details on these approximation methods, please see the above references.

D. Drive Annotation

To estimate multimodal drive topics, the proposed framework uses driving behavioral features, visual features, and human-annotated tags.

Behavioral features: the features calculated from time-series driving behavioral data. The feature space consists of eight-dimensional driving behavior, i.e., throttle opening rate, brake master-cylinder pressure, angle of steering wheel, and vehicle velocity, and their differential values. The behavioral features are obtained as cluster indices of 1000 clusters[1], which are estimated by k-means in the eight-dimensional feature space. Note that one behavioral word is calculated per frame.

[1]Actually, the number of clusters do not affect storongly to drive topic estimation. Therefore, we set it appropriately.

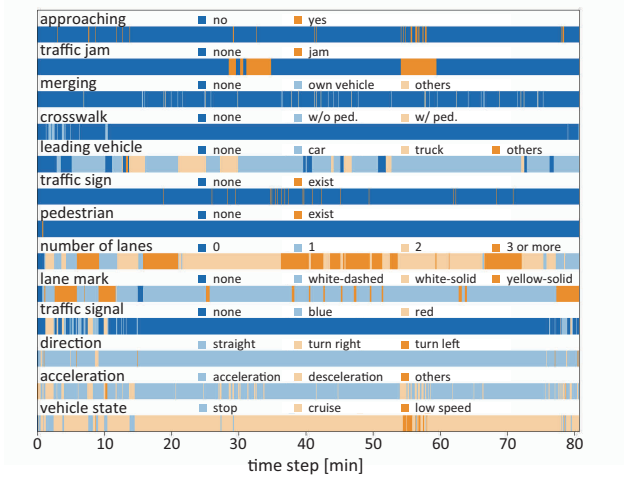

Fig. 3. Time-series representation of voting results of human annotators

Visual features: the features calculated from an image sequence captured by using a vehicle-mounted camera. We use scale-invariant feature transform (SIFT)[15] as visual features that were calculated every five pixels in captured 320×240 images, i.e., 3072 SIFT features calculated from one image[2]. The SIFT features were also clustered into 1000 clusters by k-means and then a distribution of indices is used as a bag of visual features.

Annotated features: the human-tag features annotated for vehicle behavior and the surrounding environments. An 80-minutes drive video was annotated by nine human annotators in terms of driving context such as the existence and vehicle type of the leading vehicle, the existence of traffic jams, and several pieces of environmental information, e.g., the number of lanes. Note that the annotated tags include large variances because of the variety of annotation criteria used by human annotators, e.g., the threshold of the inter-vehicle distance from the leading vehicle to judge whether there is the leading vehicle or not. The annotated items and voting results of the tags in each frame annotated by the nine annotators are represented in Fig. 3.

After learning the multimodal drive topics from the multimodal features, driving situation annotations in natural language are estimated from observed behavioral and visual features. Due to space limitations, the set of behavioral and visual features is replaced as $\boldsymbol{w}^* = \left\{\boldsymbol{w}^{(b)}, \boldsymbol{w}^{(v)}\right\}$; then, the probability of $\boldsymbol{w}^{(a)}$ given $\boldsymbol{w}^*$ is calculated as

$$p(\boldsymbol{w}^{(a)}|\boldsymbol{w}^*) = \int \sum_{\boldsymbol{z}} p(\boldsymbol{w}^{(a)}|\boldsymbol{z}^{(a)}) p(\boldsymbol{z}^{(a)}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{w}^*) d\boldsymbol{\theta}. \tag{5}$$

In other words, $\boldsymbol{z}^{(a)}$ is generated from the drive-topic pro-

[2]Because SIFT features observed in each frame much more than other modal features, weighting parameters that control contribution ratio of modalities should be introduced. We employ the mean number of word indices observed in each driving situation as discounting parameters to frequency of observed features.

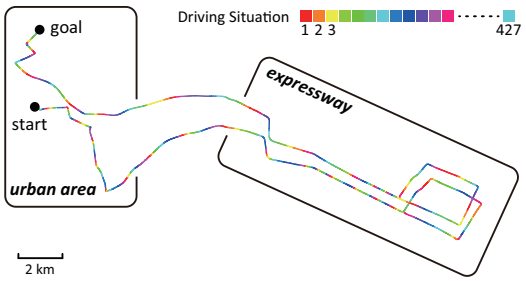

Fig. 4. Experimental course. Rainbow colors represent driving situations estimated by double articulation analyzer, i.e., changing in colors mean changing in driving situation. Note that the same colors do not represent same driving situation because the number of driving situation is too many to depicted in unique colors. Moreover, the vehicle position was estimated from the encoder's record, i.e., the trajectories deviated from the true position.

portion $\boldsymbol{\theta}$, which is estimated by using the observed $\boldsymbol{w}^*$. $p(\boldsymbol{\theta}|\boldsymbol{w}^*)$ can be calculated by marginalizing out the terms related to $\boldsymbol{w}^{(a)}$ from $p(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{\alpha}, \boldsymbol{\beta})$ in Eq. (4).

III. Experiments

For evaluation, driving behavioral data and image sequence were gathered over an experimental course over a duration of more than 80 minutes and a distance of about 70 km, which included urban roads and an expressway. The experimental course and segmentation results by DAA are depicted in Fig. 4. The driver started from the start point, and after driving in the urban area, he drove on an orbital expressway twice. Then, he return to the goal point in the urban area. There was heavy traffic twice on the expressway.

Multimodal drive topics were estimated by mLDA from the behavioral, visual, and annotated features. The extracted driving situations were divided into two groups: for training and for test. The latent drive topics were modeled by using the driving situations for training, and the predictive performance of situation annotations was evaluated by using the driving situations for test[3].

Figure 5 shows a time-series plot of the gathered behavioral data, segmentation results of driving situations, the estimated multimodal drive-topic proportion with $K = 10$, and several captured images and predicted annotations. The depicted annotations had maximum probabilities in the same tag items[4] in the case of $K = 10$ and $K = 100$. In the case of $K = 10$, diverse driving situations were represented as mixtures of small numbers of drive topics. Frequent annotations such as "cruise," "accelerate," and "car," for the leading vehicle tended to be generated with a small number of drive topics. In contrast, rare annotations such as "decelerate," "low speed," and "approaching to the leading vehicle" appeared when the number of drive topics was increased to $K = 100$. Another type of difference appeared, for example, the color of traffic signal and the number of lanes in the situations at 5:00 and 78:00. The main cause of this is that driving situations included several annotations in the case of the traffic signal changing its color from red to blue.

Nevertheless, while our framework can provide intuitive annotations as described above, the difficulty of quantitatively evaluating multimodal drive topics remains because of there are multiple true annotations in each driving situation. If the multimodal drive topics capture the driving features abstractly as well as do human drivers, it can extract features that well represent a human-recognized driving situation. Thus, in this paper, the multimodal drive topics were evaluated quantitatively as the feature extraction method for driving situation classification.

A conventional support vector machine (SVM) with multimodal raw features was used as a baseline for classifying the voted tags described in Fig. 3. The raw feature space of the SVM consists of 1000 dimensional behavioral features, 1000 dimensional visual features, and 12 dimensional annotated features, with the exception of objective annotation terms. The efficiency of multimodal drive topic representation was investigated by comparing it with the raw high-dimensional feature space. Figure 6 shows the clustering results for the three annotation terms of raw feature space, multimodal drive topics with $K = 10$ and $K = 100$, and the agreement rate of human annotators with the results of voting. The mean precisions were 72.11%, 56.00%, 78.22%, and 76.44% for the raw features, the drive topics with $K = 10$ and $K = 100$, and the agreement rate of the human annotators, respectively. In traffic jam classification, both the raw features and drive topic with $K = 100$ performed as well as did human annotators. Because traffic jams obviously influence driving behavior, traffic jam classification is easier than classification of the leading vehicles and number of lanes. In leading vehicle classification, conversely, all of the raw features and drive topics with $K = 10$ and $K = 100$ performed much lower than did human annotators. In particular, there are many instances of trucks being miss-classified as cars. In contrast, all of them could discriminate the existence of leading vehicles well compared with the leading-vehicle type classification. In the classification of the number of lanes, drive topics with $K = 100$ performed better than not only the raw features but also the human annotators. The main cause of this result is the large variance of human annotations, especially in the case of a large number of lanes; in this case, annotator have to predict the true number of lanes because the image captured in each frame does not include an entire shot of the road from end to end.

IV. Discussion

As described above, the number of lanes is actually a unique number that can be identified by observing the environment surrounding the vehicle. However, because the

[3]Because DAA uses Gibbs sampling, the estimated driving-situation sequence produced as output from DAA is stochastically distributed. Thus, ten driving-situation sequences were generated from one training data set, and mLDA was evaluated by using each sequence. The final evaluation results shown in Fig. 6 were calculated by averaging the evaluation results for each sequence.

[4]Because frequently emerged annotations have a high generative probability, e.g., "cruise" and "accelerate," most of the annotations predicted via generative probability were frequently emerged annotations. This is actually not suitable for prediction. Thus, we modify the generative probability by using the inverse document frequency $\mathrm{IDF}(w_i^{(a)}) = \log\left(N/|d : w_i^{(a)} \in d|\right)$ to bias rare annotations.

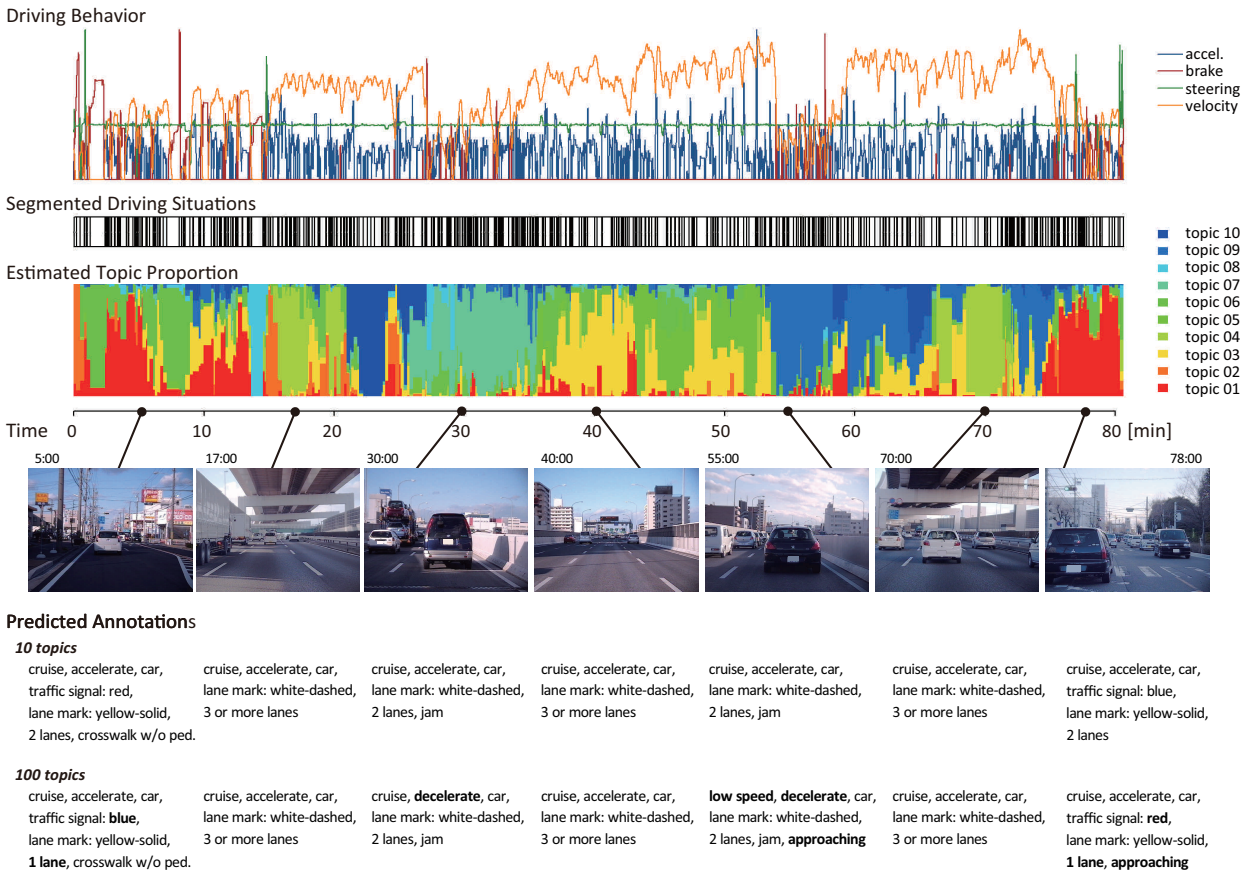

Fig. 5. Time-series representation of the gathered driving behavior, sequences of driving situations segmented by DAA, drive topic proportion estimated by mLDA, and several examples of captured images with predicted annotations. The bold texts in the case of $K = 100$ represent the different results between the cases of $K = 10$ and $K = 100$.

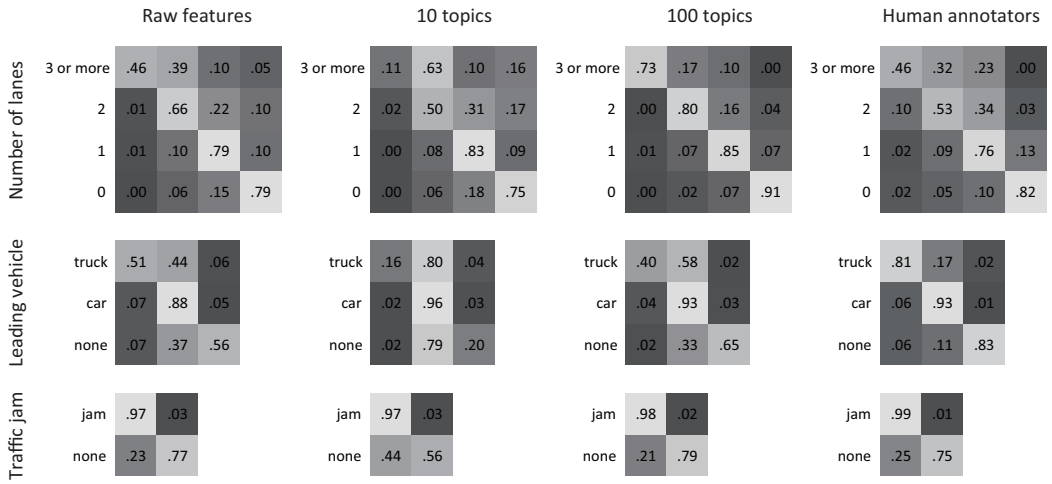

Fig. 6. Classification results of the voted tags by SVMs based on raw-feature space (left), 10 drive topics (middle-left), 100 drive topics (middle-right), and the agreement rates of human annotators with the results of voting(right). Voted tags represented as rows of each table, and estimated results of SVMs represented as columns. Numbers shown in each cell are rate of estimated tags, and colors of the cells mean degree of the rate.

image captured in each frame does not include an entire shot of the road from end to end and because there are several surrounding vehicles that occlude the road, classifying the number of lanes becomes a complicated task, even for the human annotators, as shown in Fig. 6. Multimodal drive topic achieve perform well treating diverse human tags as distributed features in lane number classification task.

These robust characteristics of mLDA lead to another engineering benefit. Various object recognition systems for ADASs nowadays have been developed, such as pedestrian and leading vehicle detection[16], and traffic sign recognition[17]. The proposed method will contribute to the

development of a novel type of ADASs; it gathers recognition results from intelligent vehicles and provides estimated information about a driving situation estimated from driving behavior and captured images to normal vehicles that have no special sensors. The different recognition systems have different levels of accuracy. Thus, in this case, the novel ADAS have to compete with the variance of automatic annotations because there are various intelligent vehicles with different recognition systems in a real environment.

Our multimodal drive topic proportion as a feature space of SVM increased classification accuracy about 6%. Compared with a 2012-dimensional raw-feature space, the 100-dimensional drive topic is an efficient representation of driving features; in other words, the drive topics and topic proportion include the drive context, and they are a low-dimensional representation of a complicated drive context. There are also various statistical driver models, such as the estimation of driver intention at an intersection[18][19], prediction of vehicle trajectory[20] and driver state estimation with smartphone sensors[21]. Most of them use HMM or Gaussian mixture models for modeling vehicle behavior or a driver's operation, and they do not model and process shared information among multimodal sensors by using a hierarchical model. Such higher level abstraction and extraction of contextual information by using the hierarchical statistical model achieves not only long-term prediction of driving behavior for safety driving assist systems but also intuitive retrieval of driving situations in natural language[22]. Annotating whole data in a large-scale database is actually impossible; however, our method will achieve automatic annotation without fully prepared annotations. It is useful character for giving instructions for novice drivers and contributing to further technical development of driving systems.

Future works include developing an actual on-line system that can treat large scale data with a distributed algorithm[23], and extending the model to nonparametric estimation of drive topics with the hierarchical Dirichlet process[24].

V. Conclusion

In this paper, an automatic driving-situation annotation method that uses driving behavioral data and image sequence captured with a car-mounted camera. The proposed method consists of a double articulation analyzer (DAA) and multimodal latent Dirichlet allocation (mLDA). DAA discretizes time-series driving behavioral data into semantic symbol sequences that represent driving situations in an unsupervised manner. Thereafter, on the basis of behavioral, visual features, and annotated human tags, drive topics laid on multimodal feature distributions are estimated by mLDA. Predictive annotations can be generated intuitively from behavioral data and image sequences via the drive topic proportion. The efficiency of multimodal drive topics was evaluated by using driving situation classification tasks. Compared with a high-dimensional raw-feature space, the drive topics reduced the dimensionality to 5% and improved classification performance about 6%. This is quite close to the performance of human annotators.

References

[1] D. Lee, "A theory of visual control of braking based on information about time-to-collision," *Perception*, vol. 5, no. 4, pp. 437–459, 1976.

[2] A. Berthelot, A. Tamke, T. Dang, and G. Breuel, "Stochastic situation assessment in adcanced driver assistance system for complex multi-objects traffic situations," in *Proc. of the IEEE/RSJ IROS*, 2012, pp. 1180–1185.

[3] W. Takano, A. Matsushita, K. Iwao, and Y. Nakamura, "Recognition of human driving behaviors based on stochastic symbolization of time series signal," in *Proc. of the IEEE/RSJ IROS*, 2008, pp. 167–172.

[4] T. Taniguchi, S. Nagasaka, K. Hitomi, N. Chandrasiri, and T. Bando, "Semiotic prediction of driving behavior using unsupervised double articulation analyzer," in *Proc. of the IEEE IV*, 2012, pp. 849–854.

[5] K. Takenaka, T. Bando, S. Nagasaka, T. Taniguchi, and K. Hitomi, "Contextual scene segmentation of driving behavior based on double articulation analyzer," in *Proc. of the IEEE/RSJ IROS*, 2012, pp. 4847–4852.

[6] K. Takenaka, T. Bando, S. Nagasaka, and T. Taniguchi, "Drive video summarization based on double articulation structure of driving behavior," in *Proc. of the ACM MM*, 2012, pp. 1169–1172.

[7] T. Bando, K. Takenaka, S. Nagasaka, and T. Taniguchi, "Unsupervised drive topic finding from driving behavioral data," in *Proc. of the IEEE IV*, 2013, pp. xxx–xxx.

[8] C. Wang, D. Blei, and L. Fei-Fei, "Simultaneous image classification and annotation," in *Proc. of the IEEE CVPR*, 2009.

[9] N. Srivastava and R. Salakhutdinov, "Multimodal learning with deep boltzmann machines," in *NIPS*, 2013, pp. 2231–2239.

[10] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "The sticky hdp-hmm: Bayesian nonparametric hidden markov models with persistent states," MIT Laboratory for Information and Decision Systems, Tech. Rep. 2777, 2007.

[11] D. Mochihashi, T. Yamada, and N. Ueda, "Bayesian unsupervised word segmentation with nested pitman-yor language modeling," in *Proc. of the ACL and the AFNLP*, vol. 1, 2009, pp. 100–108.

[12] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[13] T. Nakamura, T. Nagai, and N. Iwahashi, "Grounding of word meanings in multimodal concepts using lda," in *Proc. of the IEEE/RSJ IROS*, 2009, pp. 3943–3948.

[14] T. Griffiths and M. Steyvers, "Finding scientific topics," in *PNAS*, vol. 101, Apr. 2004, pp. 5228–5235.

[15] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. of the IEEE ICCV*, vol. 2, 1999, pp. 1150–1157.

[16] G. Stein, Y. Gdalyahu, and A. Shashua, "Stereo-assist: top-down stereo for driver assistance systems," in *Proc. of the IEEE IV*, 2010, pp. 723–730.

[17] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark," in *International Joint Conference on Neural Networks (submitted)*, 2013.

[18] G. S. Aoude, V. R. Desaraju, L. H. Stephens, and J. P. How, "Behavior classification algorithms at intersections and validation using naturalistic data," in *Proc. of the IEEE IV*, 2011, pp. 601–606.

[19] M. Liebner, M. Baumann, F. Klanner, and C. Stiller, "Driver intent inference at urban intersections using the intelligent driver model," in *Proc. of the IEEE IV*, 2012, pp. 1162–1167.

[20] J. Wiest, M. Hoffken, U. Kresel, and K. Dietmayer, "Probabilistic trajectory prediction with gaussian mixture models," in *Proc. of the IEEE IV*, 2012, pp. 141–146.

[21] H. Eren, S. Makinist, E. Akin, and A. Yilmaz, "Estiamting driving behavior by a smartphone," in *Proc. of the IEEE IV*, 2012, pp. 234–239.

[22] M. Naito, C. Miyajima, T. Nishino, N. Kitaoka, and K. Takeda, "A browsing and retrieval system for driving data," in *Proc. of the IEEE IV*, 2010, pp. 1159–1165.

[23] D. Newman and M. Welling, "Distributed Algorithms for Topic Models," *Journal of Machine Learning Research*, vol. 10, pp. 1801–1828, 2009.

[24] Y. Teh, M. Jordan, M. Beal, and D. Blei, "Hierarchical dirichlet processes," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1566–1581, 2006.