Suzhou·2023
29th ITS World Congress

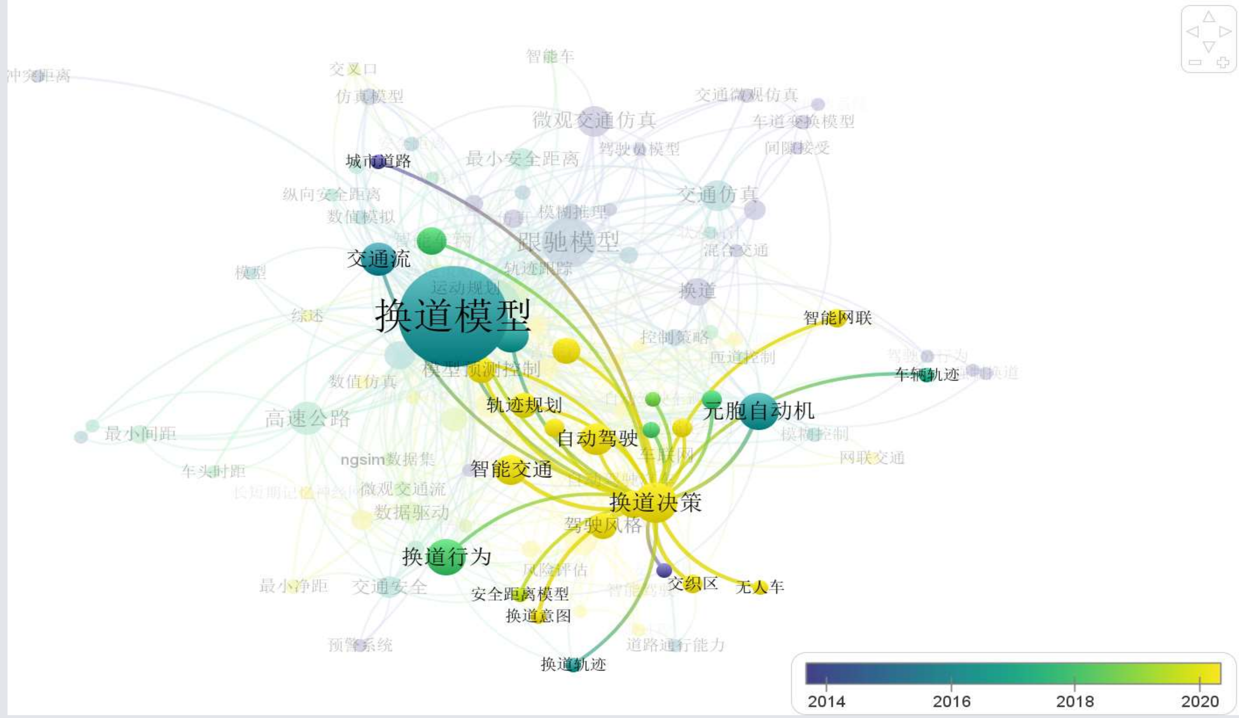
A Review of Data-driven Lane-changing Decision Modeling for Connected and Automated Vehicles

Zhengwen Fan, Shanglu He, Xinya Zhang, Yingshun Liu
Nanjing University of Science and Technology, Nanjing, China

Abstract

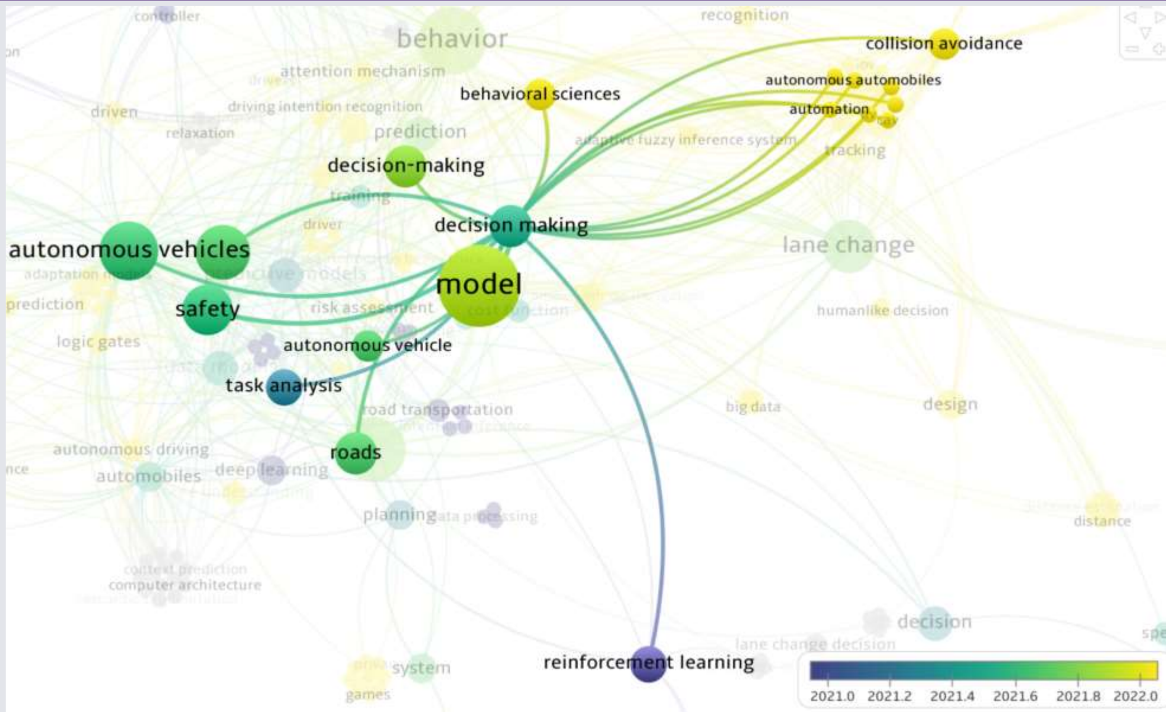
This study reviews data-driven lane-changing decision (LCD) modeling for connected and automated vehicles (CAVs). Co-occurrence knowledge graphs of the lane-changing decision in both Chinese and English are shown in this study. We summarize existing research into two tables from two perspectives: data sources and modeling methods. Three future research directions, including the need for comprehensive datasets, innovative data-driven methods, and standardized testing are identified.

Knowledge Graphs-based Analysis



Keywords:
- 换道模型
- 元胞自动机
- 智能交通

Fig 1 Co-occurrence knowledge graph of the LCD unit in Chinese literature



Keywords:
- lane-changing model
- data-driven
- autonomous vehicles

Fig 2 Co-occurrence knowledge graph of the LCD unit in English literature

Summary:
- Literatures are collected from 1998 to 2022, including 385 Chinese literatures and 248 English literatures.
- A significant recent trend in LCD research has centered on the analysis of vehicle trajectories gathered from Intelligent Transportation Systems (ITS).
- Machine learning approaches have been increasingly spotlighted as a focal point in the realm of LCD research.

Mainstream Datasets and Features

Table 1 Commonly used datasets for data-driven lane-changing models

Data Source	Data Collection	Country	Scenario	Number of Maps	Length of Collected Road Section(m)	Collection Hours	Number of Lanes per Direction	Road User Type	Key Fields	Features
NGSIM	Road side cameras	the U.S.	Highway, merging section	3	500-640	1.5	5~6	car, truck	instantaneous speed and acceleration	The best known, most widely used and earliest published public trajectory dataset
HighD	Drone	Germany	Highway, merging section	6	400-420	16.5	2~3	car, truck	The minimal distance headway (DHW) The minimal time headway(THW) The minimal time to collision (TTC)	·Official pre-processed data including: surrounding vehicles, THW and TTC, vehicle size and classification of driving behavior including lane changes. ·Highway without speed limit ·5600 complete lane-changing tragectories
inD	Drone	Germany	Unsignaled intersection	4	80*40~140*70	10	2~3	car, truck, pedestrain, bicycle	The type of road users, and their horizontal and longitudinal speed and acceleration	·Four different recording locations ·Different intersection types ·Typical positioning error <10 cm ·HD maps in lanelet2 are provided ·Visualization of recorded trajectories
rounD	Drone	Germany	Roundabout	3	140*70	6	1~2	car, truck, van, pedestrain, bicycle, motorcycle	Accurate visualized trajectories, the type of road users, the direction of every trajectory concerning adjacent time steps, speed and acceleration	·Officially released data pre-processing and visualization tools: https://www.github.com/ika-rwth-aachen/drone-dataset-tools ·The only dataset providing the exact location of the recording sites and gives geo-referenced coordinates
exiD	Drone	Germany	Highway entrance and exit	7	420	16.1	2~4	car, truck, van	Velocity and acceleration in the x-y and the radial-latitudinal direction, the width of current lane, whether to change lanes, the DWH, THW, TTC and relative speed on current lane	·High traffic volume ·Rich merging scenarios ·Different speed limit scenarios (no speed limit, 120km/h and 100km/h)
INTERACTION	Drone and road side cameras	China the U.S. Germany Bulgaria	Merging, Lane-changing, unsignaled intersection, Signaled intersection, Roundabout	12	—	2 7 1 6	—	car, pedestrain, bicycle	the coordinates, size, horizontal and longitudinal speed, yaw rate	·Complexity of the behavior(negotiations, inexplicit right-of-way, irrational behavior and aggressive maneuvers) ·Diversity of the scenarios(unsignalized and signalized intersections, roundabounts with stop/yield signs, as well as zipper merging and lane changes in urban and highway scenarios.) ·HD-map with full semantics(lane connections, turn directions traffic rules, etc.)
Ubiquitous Traffic Eyes	Drone	China	Highway, merging section	6	140~430	0.8	5	car, truck	the DWH, THW and TTC	·Time accuracy 0.1s, position accuracy 0.01m ·Achieved 100% vehicle detection after manual correction

Summary:
- NGSIM dataset has served as the largest and fundamental resource for microscopic traffic research, forming the cornerstone for data-driven studies. Though there are some errors due to the limitations of camera quality and image processing technology at that time.
- Recently released open-source high-precision vehicle trajectory datasets perform better in terms of data volume, accuracy, and the variety of traffic scenarios.
- The appearance of new datasets helps to solve the problems of the NGSIM dataset and also presents new opportunities and challenges for the validation and generalization testing of existing data-driven LCD models

Modeling Methods and Evaluation

Table 2 Commonly used methods for data-driven lane-changing models

Types	Machine Learning Methods	Year	Inputs	Outputs	Evaluation
LCD models based on traditional machine learning	Neural Networks	1994	$x(t)$ $v(t)$ $\Delta S(t)$	Target lanes and coordinates	Correct classification rate of 70%
	BP Neural Networks	2014	x Δv	Target lanes	Leftward lane change prediction accuracy of 94.6%
	BP Neural Networks	2022	v ΔS	Whether to change lane	Overall accuracy of 96.5%
	SVM	2016	$x,\Delta v$ gaps	Whether to change lane	Non-merging section-94% Merging section-78%
	Bayesian Networks	2015	v Δv ΔS	Whether to change lane	Lane change recognition accuracy of 88.7%
LCD models based on deep learning	Deep Learning (DBN)	2019	v Δv ΔS	Whether to change lane	Prediction accuracy of up to 99.32%, significantly better than the comparison group of BP neural network-based and rule-based models
	Deep Learning(CNN)	2020	$x,v,a,\Delta S,\Delta T$ driving style	Target lanes	Prediction accuracy of 98.66%
	Deep Learning(LSTM)	2020	$x,\Delta v,\Delta S$ vehicles size, lane lateral offset M	Coordinates of the next time sequence	By introducing lane lateral offsets, the accuracy and generalization capability of the proposed model can be improved by about 10%
	Deep Learning(LSTM)	2021	$x,\Delta v,\Delta S$ vehicles size	Coordinates of the next time sequence	In the accuracy test, the model was reduced by 31% compared to the GRU comparison group In the mobility test, the MSE was reduced by 39.7%
	Deep Reinforcement Learning(DQN)	2018	v ΔS	Target lanes	Significant improvement in decision-making performance and traffic capacity compared to the rule-based model
	Deep Reinforcement Learning(D3QN)	2022	$v,\Delta v,a$ the total number of lane changes N the number of dangerous lane changes N_1	Target lanes	24% increase in driving speed compared to original data
Integrated LCD models	Rule-based+SVM	2022	$v,\Delta v,a,\Delta S$ the neccisity, safety degree and benefits of lane-changing	Target lanes	Prediction accuracy improved by 10.78% after augmentation
	Bayesian Networks+Decision Trees	2014	Δv ΔS d	Whether to change lane	Prediction accuracy of 79.3% and 94.3% for lane-changing and no lane-changing
	Bayesian Network+BP Neural Networks	2015	v the distance to lane lines steering angle	Whether to change lane	Prediction accuracy of 91.4%, improved 6% compared to a merely BP NN based model
	Imitation Learning(XGBoost) +Reinforcement Learning(DDPG)	2021	$v,\Delta v,a,\Delta S$ Adjacent lanes' passable status	Target lanes and coordinates	Significant improvements in safety, traffic efficiency, comfort and speed of strategy learning compared to reinforcement learning alone

Note： x refers to position, v refers to velocity, a refers to acceleration, d refers to driving distance, ΔS refers to DHW, ΔT refers to THW, Δv refers to relative velocity, t refers to current time.

Summary:
- Generally, there are two types of data-driven LCD models: traditional machine learning-based LCD models and deep learning-based LCD models. Nowadays, a new kind of integrated models is emerging.
- Traditional machine learning-based models include neural networks, support vector machines, and Bayesian filters. Deep learning-based models include DBN, CNN, LSTM, and DRL.
- Integrated models include: LC rules + machine learning models and multiple machine learning fusion models

Conclusion

This paper provides a state-of-the-art review of the data-driven lane-changing decision models for CAVs from the perspective of mainstream trajectory databases and commonly used LCD modeling methods. Three research orientations for the development of future data-driven CAVs LCD models are outlined:
- **Data:** datasets collected in mixed traffic flow with both micro-driving trajectory and driver characteristics
- **Modeling Methods:** to achieve a balance between lowering model complexity and increasing model prediction accuracy and interpretability
- **Verification and Testing:** to test and verify the model from both microscopic and macroscopic viewpoints, as well as how to establish more thorough evaluation indicators or procedures.

: matthew_fan@njust.edu.cn, she17@njust.edu.cn. TEL： (+86) 15927692569

Acknowledgements： This research was supported by the National Key Research and Development Program of China (Grant No.2019YFE0123800), the National Natural Science Foundation of China (Grant No. 52102380), China Postdoctoral Science Foundation (Grant No. 2021T140325,No.2018M642257), Fundamental Research Funds for the Central Universities (Grant No.330920021140).

南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY