iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** I **Month of publication:** January 2018

DOI: http://doi.org/10.22214/ijraset.2018.1309

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 6 Issue I, January 2018- Available at www.ijraset.com

A Survey on Vehicle Navigation Using Natural Spoken Language

Neethu V M[1], Shemitha P A[2]
[1]M.tech Student, [2]Assistant Professor, Department of Computer Science and Engineering
IES College of Engineering Thrissur, Kerala, India

Abstract: Vehicle navigation using natural language processing mainly introduces interface relating to human and vehicle that the vehicle can understand human language in a natural way. So mainly two methods are used for understanding human language, which are Automatic Speech Recognition (ASR) and Natural Language Processing (NLP). ASR means converting the audio input to text and NLP retrieve the navigation data from the text. NLP is done by using Deep Neural Network (DNN). It is to finding the sentence level sentiment analysis and for context extraction. For this DNN use Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM) architecture. In previous works double articulation analyser (DAA) is used to make these predictions i.e. Instead of collecting input data from human as audio, early systems directly collect input data from vehicle using double articulation method and process the collected data as time series for the prediction of future driving behaviour. From these time series parse the actual meaning of data using HDP hidden Markov Model (HDP-HMM) and Nested Pitman-Yor Language Model (NPYLM).
Keywords: Automatic Speech Recognition, Natural Language Processing, Deep Neural Network, Recurrent Neural Network, Hidden Markov Model

I. INTRODUCTION

Driving is a difficult task because of the need to make correct decisions rapidly. Each decision a driver makes is important since it directly impacts traffic safety. Manoeuvres involving changes in speed and steering wheel angle are the most influential factors concerning safety. Any change in speed or steering wheel can make the driving situation unsafe. Here integrate artificial intelligence with conversational environment and applying this expertise into a vehicle environment. So the auto makers and suppliers can create a highly customise solutions that will offer personalised experience for both drivers and passengers while keeping hands on the wheel and eyes on the road. Vehicle navigation using natural spoken language system is a voice based human interface to understand driver's spoken language in a natural way. it consists of mainly two stages Automatic Speech Recognition (ASR) and Natural Language Processing (NLP). ASR converts the given input data to text. NLP stage is used to retrieve the navigation associated data and it is done by using Deep Neural Network (DNN) which consists of Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM). According to its previous work it is mainly based on the sequence prediction of driving behaviour using double articulation analyser (DAA). It collects the input data directly from vehicle not voice based. These input data are collected as time series and then process these time series data to retrieve actual meaning using Double Articulation Analyser for the prediction of future driving behaviour, which is done by using Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM) and Nested Pitman-Yor Language Model (NPYLM).

II. LITERATURE SURVEY

The term "driver behaviour" is used to signify the driver's intent as it pertains to the most probable next manoeuvre. The goal of this work is listed as follows: First is, Understanding drivers' behaviours; finding relationships between driving parameters such as speed, turn signals and steering wheel as inputs as necessary to assess a probability concerning the next manoeuvre. Second is, devising a powerful model for driving assistance systems able to predict the next driving manoeuvre before any specific or unintended manoeuvre begins. Most drivers are familiar with manoeuvres such as passing, changing lanes to the left or right, starting, stopping, and turning left or right. In this thesis, the focus is placed on predicting a subset of canonical manoeuvres such as speed and steering wheel angle. Vehicle navigation using natural language processing mainly introduces interface relating to human and vehicle that the vehicle can understand human language in a natural way. So here mainly two methods are used for understanding human language. First are Automatic Speech Recognition (ASR) and Natural Language Processing (NLP). ASR means converting the audio input to text and NLP retrieve the navigation data from the text. NLP is done by using Deep Neural

Network (DNN). It is to finding the sentence level sentiment analysis and for context extraction. For this DNN use Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM) architecture.

In order to understand driver behaviour, data recordings were needed to make important observations on drivers. Beauchemin et al. (2012) used car-mounted video cameras to capture surrounding forward traffic, the driver's head pose and gaze direction, and driving data from the vehicle's internal network. These data sources were recorded with 16 drivers over a pre-determined course inside the city of London. In total, 3TB of data was collected over more than 450 kilometres. In sum, stereo data from forward pointing cameras, vehicular attitude from the CANbus interface, and ocular movements from the drivers were recorded.

In predictive systems, data captured in real time is usually analysed immediately. In this thesis, captured data is considered as a time-series and analysed after its collection. In order to achieve a predictive model, first have to find a model that can be adapted to time-series driving data. A good number of methods have been presented to model driving time series data. Among these, review those shown to be best-in-class. In order to predict driving manoeuvres, Hidden Markov Models (HMM) have often been used to extract information from time-series data. HMMs model hidden discrete states as x and observations as y. HMMs treat every observation as a statistically independent entity. A next generation approach developed and named Auto Regressive Hidden Markov Models (ARHMM) allows some stochastic dependencies to exist between observations, as shown in Fig 1. The current observation is dependent to a past observation, as there is a correlation between the two. For this reason, ARHMM can be used with dynamic behaviours since it is powerful enough to model feedback systems

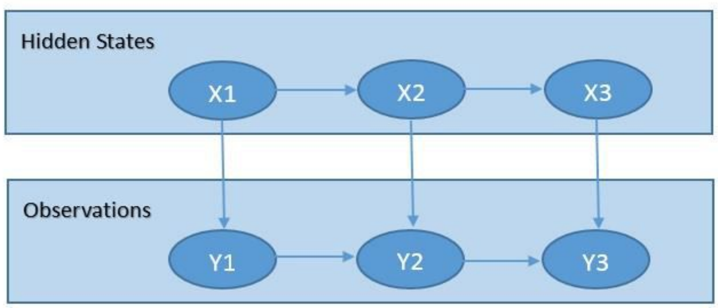

Fig. 1 Auto Regressive Hidden Markov Model. The hidden state X3 is not Visible to the observer, but we can say that there is a dependency between hidden States (X3 and X2) because of the existing dependency between observations (Y2 and Y3).

Both HMM and ARHMM encounter serious problems because they need to have a fixed number of states a priori. In order to address these problems, a novel and efficient method, called Beta Process Auto Regressive Hidden Markov Model (BP-AR-HMM) was proposed by Fox E.B. et al. (2011). This is a robust model that can simultaneously handle multiple related time-series data and automatically determine the number of states. Hamada et al. (2013) applied a BP-AR-HMM model to their driving dataset. Three different driving operations (gas pedal opening rate, brake pressure, and steering wheel angle) were considered. Four estimated state sequences were obtained. In each state, operations had different behaviours from the other states. Hamada et al. (2013) obtained good results. To model long-term contextual information, the Double Articulation Analyser (DAA) was proposed by Taniguchi et al. (2012). They suggested that contextual information and human driving behaviour possess a double articulation structure.

The term "double articulation structure" was first presented in order to analyse a speech stream. A speech stream data possesses a dual layer of information that can be decomposed into several meaningful linguistic units, and each unit can be divided into meaningless elements. Meaningless elements called phonemes are at the lowest level of speech organization. Morphology, syntax, and semantics give the meaning to phonemes and they are the higher levels of speech organization. In order to understand long-term human action, it has to be decomposed into short-term chunks. To extract long-term human action chunks, Taniguchi. T, and Nagasaka S. (2011) presented a framework that included both a language model, Nested Pitman-Yor Language Model (NPYLM), and sticky Hierarchical Dirichlet Process Hidden Markov Model (sHDP-HMM). Sticky HDP-HMM is an augmented version of HDPHMM (Fox, E. B. 2011) in which the number of states is not predefined. The main difference between HDP-HMM and sticky HDP-HMM is that HDP-HMM tends to give high posterior probability to states with rapid switching while sticky HDP-HMM fixes this problem. If we look at our spoken language, it also has a double articulation structure. For example, a sentence can be decomposed into single letters then these single letters can be chunked into words. Letters individually do not have any meaning, however words do. Human action time-series have the same pattern. It is obvious that at each time, point data do not have any meaning individually, but when some of them come together, they form a meaningful segment. By expanding this idea, several successive segments of time series data form a meaningful sequence of a specific human action. A meaningful sequence of a specific human action is assumed to be a word. Fig 2 shows a graphical representation of the NPYLM model. NPYLM takes a text as input. The text must be a set of successive letters separated by spaces. The length of the text has an important effect on the final result in that we should make sure that the selected letter sequence is large enough to represent each letter frequency properly.

Sticky HDP-HMM and NPYLM are able to extract meaningful chunks from the continuous time series-data. Compared to conventional HMM and simple NPYLM, Taniguchi et al.'s (2012) method is a better approach as it considers the incompleteness of observed sentences and improves the long-term prediction performance. Taniguchi et al. (2012) first applied their method to a sentence data model. Their predictor method assumed that a number of words and states are unknown. In order to have a trained generative language model, a word model had to be trained. Thus, 100 artificial sentences consisting of 50 words each were generated. Then, for evaluating prediction performance, they used one of the sentences as test data and the rest of the sentences for training. Following this, they changed the test sentence by omitting the last parts of the sentence.

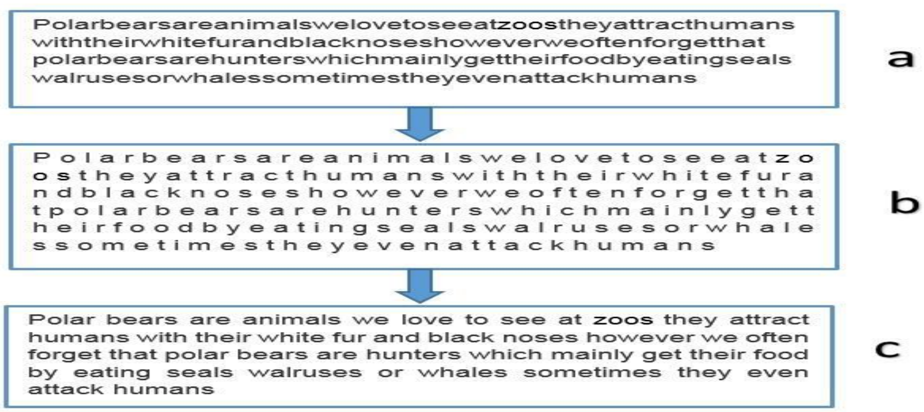

Fig. 2 An example of the NPYLM procedure. (a) Shows the coarse text which is a mixture of different letters, and hard to read. (b) is a pre-processing level, which is preparing the coarse text for main processing. (c) is a readable text which is the result of applying the NPYLM process.

In 2013, Bando et al. (2013) proposed a new framework that can automatically translate driving data into sequences of "drive topics" in natural language. In this framework, brake pressure, opening rate, steering wheel angle, and velocity are considered as physical features. Each of these had different frequencies for each word. They used Latent Dirichlet Allocation (LDA) to cluster extracted driving situations based on the existing frequency of physical behavioural features that are observed in each driving sequence. The distribution of the physical behavioural features included in each drive topic was used for automatic driving word labelling. Here "accelerating" and "high speed", are the driving topics which are assigned to each driving words and this considered as its results. Although DAA and LDA are unsupervised methods, this framework creates human-understandable tags. Being independent of any human-created tags is one of the greatest benefits of this method.

In 2016 an automaker company created a vehicle dialogue system using natural language processing called Dragon Drive System. The artificially intelligent automotive assistant does more than just listen. The Dragon Drive platform uses AI and natural language understanding to go way beyond interpreting simple commands. It understands human voice and meaning to analyse what's being said and deliver a response back to human through this navigation system. Dragon Drive's entire stack is powered by innovative machine learning and contextual reasoning to create an AI platform optimized for the connected car. Voice is the most efficient and safest way to connect the driver and passengers to the car. AI elevates the car transportation and transforms it into an intelligent automotive assistant giving the driver instant access to information, services and content. For more than a decade, Nuance has leaded the industry in creating reliable intelligent systems that personalize the driving experience. With just a spoken phrase, Nuance's automotive assistant keeps drivers more connected than ever to work and life, with a platform that drives interoperability between in-car systems and the rapidly growing Internet of Things. Dragon drive system is currently using vehicle dialogue system based on natural language processing. Dragon Drive isn't just the shine on the wheels – it's a deeply integrated solution that is customized to meet the needs of your brand and more importantly – your customers. It's not a self-contained experience. Dragon Drive is an agnostic platform that transforms into a branded solution reflecting the millions of hours you have invested in the vehicles you make. That's the difference. It's your brand and your experience – you should own it. Nuance's conversational and cognitive solutions are as much a part of your brand experience as the leather you use for your seats or the casting you use for your alternator [8]. Alexandre Alahi et al. (2016) introduced a human trajectory prediction in crowded spaces, which explains how pedestrians follow different trajectories to avoid obstacles and accommodate fellow pedestrians. Any autonomous vehicle navigating such a scene should be able to foresee the future positions of pedestrians and accordingly adjust its path to avoid collisions. This trajectory prediction problem can be viewed as a sequence of generation task, where we are interested in predicting the future trajectory of people based on their past positions. Following the recent success of Recurrent Neural Network (RNN) models for sequence prediction tasks, propose an LSTM model which can learn general human movement and predict their future trajectories. Recently Recurrent Neural Networks (RNN) and their variants including Long Short Term Memory (LSTM) and Gated Recurrent Units have proven to be very successful for sequence prediction tasks: speech recognition caption generation machine translation, image/video classification , human dynamics to name a few. A recurrent neural network (RNN) is often used to process sequential inputs like speech and language, element by element, with hidden units to store history of past elements. A RNN can be seen as a multilayer neural network with all layers sharing the same weights, when being unfolded in time of forward computation. It is hard for RNN to store information for very long time and the gradient may vanish. Long short term memory networks (LSTM) and gated recurrent unit (GRU) were proposed to address such issues, with gating mechanisms to manipulate information through recurrent cells

III.CONCLUSIONS

Vehicle navigation dialogue system using natural spoken language is a system, which can be implemented in our vehicle. In early systems Natural Language Processing is based on the Double Articulation Analyser, which means input data are taken from the vehicle and collect the data as time series and from these time series data predict the future driving behaviour using HDP-HMM and NPYLM models. So there will be a driving agent in our vehicle and it responds back to the driver about the future driving behaviour based on the data from current vehicle behaviour. In future navigation systems we can implement the vehicle navigation system based on Automatic Speech Recognition (ASR) and Natural Language Processing (NLP). Here the NLP stage is based on the Deep Neural Networks (DNN) such as Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM). Which are based on the sentence level sentiment analysis and context extraction. Input data is voice and this voice is converted to text using ASR and from this text collect the meaningful words related to navigation associated using DNN.

REFERENCES

[1] Yang Zheng, Yongkang Liu, and John H.L, Hansen, Fellow,"Navigation Orientated Natural Spoken Language Understanding for Intelligent Vehicle Dialogue", IEEE Intelligent Vehicles Symposium (IV) June 11-14, 2017, Redondo Beach, CA, USA.
[2] Tadahiro Taniguchi, Shogo Nagasaka, Kentarou Hitomi, Naiwala P. Chandrasiri, "Sequence Prediction Of Driving Behavior Using Double Articulation Analyzer" IEEE Transactions On Systems, Man, And Cybernetics: Systems, VOL. 46, No. 9, September 2016
[3] Bando, T., Takenaka, K., Nagasaka, S., & Taniguchi, T."Unsupervised drive topic finding from driving behavioral data". In Intelligent Vehicles Symposium (IV), 2013 IEEE (pp. 177-182). IEEE june 2013.
[4] Bando, T., Takenaka, K., Nagasaka, S., & Taniguchi, T. "Automatic drive annotation via multimodal latent topic model". In Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference (pp. 2744-2749). IEEE 2013.
[5] Beauchemin, S. S., Bauer, M., Kowsari, T., & Cho, J. "Portable and scalable vision-based vehicular instrumentation for the analysis of driver intentionality". Instrumentation and Measurement, IEEE Transactions on, 61(2), 391-401, 2012.
[6] Fox, E. B., Sudderth, E. B., Jordan, M. I., & Willsky, A. S." A sticky HDP-HMM with application to speaker diarization". The Annals of Applied Statistics, 1020-1056 (2011).
[7] Tadahiro Taniguchi, Shogo Nagasaka, **"**Double Articulation Analyzer for Unsegmented Human Motion using Pitman-Yor Language Model and Infinite Hidden Markov Model" , SI International IEEE 2011.
[8] https://www.nuance.com/mobile/automotive/dragon-drive.html, "Bringing intelligence to the driver, to the road and the beyond" 2016.
[9] Alexandre Alahi, Kratarth Goel , Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, Silvio Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces", IEEE Conference on Computer Vision and Pattern Recognition 2016
[10] Ryunosuke Hamada, Takatomi Kubo, Kazushi Ikeda, Zujie Zhang, Tomohiro Shibata, Takashi Bando ,Masumi Egawa, "Towards Prediction Of Driving Behavior Via Basic Pattern Discovery With Bp-Ar-Hmm" Nara Institute of Science and Technology, Japan IEEE 2013
[11] Yuxi Li**,** Deep Reinforcement Learning: An Overview, EBook 2017
[12] Mark Gales,.M., & Young, S, The Application of Hidden Markov Models in Speech Recognition, Ebook 2008.
[13] Oluwatobi Olabiyi, Eric Martinson, Vijay Chintalapudi, Rui Guo "Driver Action Prediction Using Deep (Bidirectional) Recurrent Neural Network", Intelligent Computing Division Toyota InfoTechnology Center USA, jan 2017