Tweets About Self-Driving Cars: Deep Sentiment Analysis Using Long Short-Term Memory Network (LSTM)

Anandi Dutta and Subasish Das

Abstract Due to the extensive growth of social media usage, sentiment analysis using social media data such as Twitter is an important task. The current study presents an empirical investigation of consumer sentiment toward self-driving cars or autonomous vehicles (AVs) based on the acquired self-driving car-related tweets. Information retrieval in social media is a complex task that requires technical insights. We used a hierarchical attention-based long short-term memory network (LSTM), a popular deep learning tool, to classify sentiment-specific document representations. The findings show that favorable attitudes toward AVs are associated with technological advantages and safety improvements, while more negative attitudes are associated with self-driving car-related crashes, media coverage, and deployment uncertainty. The results show that the estimated accuracy of LSTM is 85%. Our study indicates the necessity of examining big social media data in understanding the perceptions of end-users toward autonomous vehicles.

Keywords Autonomous vehicles · Sentiment analysis · Opinion mining · Deep learning

1 Introduction

Emerging technologies such as self-driving cars or automated vehicles (AVs) are significantly transforming the transportation system. Companies like Tesla and Google are competing to attract more people to purchase AVs. Vehicle autonomy is one of the most controversial new technologies, and it is unfamiliar to many consumers. In this context, the future market and strategic environment for deployment have had

A. Dutta
The University of Texas at San Antonio, San Antonio 78249, USA
e-mail: anandixd@gmail.com

S. Das (✉)
Texas A&M Transportation Institute, Bryan 77807, USA
e-mail: s-das@tti.tamu.edu

D. Gupta et al. (eds.), *International Conference on Innovative Computing and Communications*, Advances in Intelligent Systems and Computing 1165,
https://doi.org/10.1007/978-981-15-5113-0_40

a significant foundation laid by customer perception. As AVs become more prevalent, authorities, automakers, and companies must investigate their consequences and impacts [1]. To improve consumer knowledge and to facilitate a better understanding of the associated benefits of AVs, researchers conduct robust measurement with the help of sentiment mining.

Social media is a major platform for users to express their opinions and concerns regarding AV technologies and other related subjects. Analysis of these sentiments can enable automakers to monitor their product life-cycle and obtain a data-driven perspective on consumer attitudes and feedback. Many studies conducted analysis on the public's perception of AVs using different survey tools such as polls, and telephone interviews [2, 3]. Most of these studies have categorized different types of users' opinions based on age, gender, location, and other socio-demographic characteristics. For example, some studies have concluded a totally different level of acceptance in men than in women. Despite the large number of studies in this field, there is limited number of studies on social media mining and self-driving cars. There is a need for an in-depth study on people's acceptance level and perceptions regarding AVs.

In the present study, we collected approximately 38,000 unique tweets associated with self-driving cars. We used a hierarchical attention-based long short-term memory network (LSTM) to predict customer sentiments from the text narratives of the collected tweets. The intent of this paper is to address two research questions: (1) **RQ1**: What is the general sentiment trend of the end-users toward self-driving cars? (2) **RQ2**: Are deep learning tools adequate in classifying sentiments from self-driving car-related tweets?

2 Related Work

In recent years, AV research has begun to focus more on the public perception of AVs and the factors that can influence the likelihood of potential users adopting AVs. In a recent study, Greig [4] conducted two surveys to determine public sentiment toward the adoption of AVs. The study found that only one-fifth of the respondents had positive feelings toward AVs, and a majority of them remained skeptical of AVs' benefits. In another study, Zmud et al. [5] conducted an online survey in Austin, Texas, to determine potential user intent toward AVs. The researchers classified the respondents into four intent-to-use classifications. The 'very unlikely' category contained 18% of respondents; 32% of respondents were in the 'somewhat unlikely' category; 36% were in the 'somewhat likely' category; 14% were in the 'very likely' category.

A recent study by Merat et al. [6] focused on the social-psychological factors that can influence people's acceptance of SAE Level 4 shared AVs (SAVs). The study found that the most effective pathway to the widespread acceptance and adoption of AVs is an incremental and iterative stage-by-stage process that provides users with hands-on experience throughout every stage. In another study, Regan et al. [7]

analyzed survey data from 5263 participants to measure public awareness of AVs in Australia and determine people's knowledge of AVs and the likelihood of them adopting automated cars. In a similar Australian study, Greaves et al. [8] conducted an online survey of 455 people to investigate consumer sentiment toward AVs. The study found that young male respondents, less frequent drivers, and people open to the idea of sharing their car were associated more highly with favorable attitudes toward AVs. Negative attitudes were associated more with older female respondents, those who drive more frequently, and those who are less open to sharing their car.

Liljamo et al. [9] examined the results from a large citizen survey of 2036 people to determine the magnitude and type of concerns that people have in regard to the adoption of AVs. The study found that the key perspectives that affect public approval of AVs are traffic safety and ethical perspectives. Another study was conducted by Nazari et al. [10]. This study jointly modeled public attention in both private AVs and multiple shared autonomous vehicles (SAVs). In 2016, Sener et al. [11] conducted a survey to examine the factors influencing people's intent to use AVs in the future; the survey included people in various Texas cities, including Dallas, Houston, and Waco. The factors that had the most significant effect were attitudes toward self-driving vehicles, execution expectancy, apparent protection, and community impacts. Nordhoff et al. [12] conducted a survey with 10,000 participants to determine the relationship between socio-demographic characteristics and people's acceptance of AVs. Asgari and Jin [13] employed a structural equations model with latent variables to analyze consumers' willingness to pay and adoption (WTPA) for different AV categories. The model simultaneously regressed adoption and WTPA measures for multiple variables, including demographic and socioeconomic factors, car ownership and usage, and personal opinions or preferences.

3 Data Description

With approximately 500 million tweets per day, Twitter provides real-time textual contents with a wide range of themes and topics. We used the open-source R software package 'twitteR' [14] to collect relevant data. For tweet collection, users need to collect data via Open Authorization (OAuth), an authentication mechanism that allows Web tools to provide the user applicability to a Web service without allowing an end user's credentials to the client itself, was required for all Twitter-related data collection.

To gather the relevant data, several keywords have been used during the data collection process. The keywords include *'driverlesscar,' 'driverlessvehicle,' 'driverless,' selfdrivingcar,' 'selfdrivingvehicle,' 'autonamtedcar,' 'automatedvehicle,' 'autonomouscar,' and 'autonomousvehicle.'* The dataset was collected over a span of four months in 2019 (March 12, 2019–July 16, 2019). A total of 79,214 unique tweets were collected. These tweets were retweeted 211,090 times and are associated with approximately 37,000 Twitter handles.

4 Methodology

4.1 Long Short-Term Memory (LSTM) Networks

In 1977, Sepp Hochreiter and Jürgen Schmidhuber proposed long short-term memory (LSTM), which is a modified version of recurrent neural network (RNN) model. Unlike conventional neural feed-forward networks, LSTM has response links that provide a 'general-purpose computer' capable of calculating anything a Turing machine can [15, 16]. By presenting Constant Error Carousel (CEC) units, LSTM manages the problems regarding vanishing and exploding gradient. In theory, the conventional RNN models can handle the long-term dependencies in the input sequences; however, the dilemma emerges when the model is in computational nature. While preparing a typical RNN model applying the method of back-propagation, the gradients that are back-propagated tend to zero or infinity due to the calculations associated with the method that require precise numbers. Therefore, the flow may be altered, resulting in a different outcome than expected [17].

4.2 Our Approach

A natural language processing or NLP tool that concentrates on distinguishing negative and positive opinions, evaluations, and emotions expressed in text data is referred to as sentiment analysis. The common approach of performing sentiment analysis is to compare the presence of a word in a document and assignment sentiments based on some established sentiment lexicons [18]. There exist several existing sentiment lexicons with sentiment scores assigned to a group of words. For example, Affective Norms for English Words (ANEW) was one of the most common sentiment lexicons. 'AFINN,' which was used in this study, performs better than ANEW [19].

The usual practice in machine learning and deep learning is to divide the dataset into several groups: training, test, and validation datasets. It is important to note that we developed the model based on the training data and the performance of the model will be tested and validated by the other two datasets. The training for model building, the validation data for out-of-sample error measurement and model selection, and the test data are used for final model performance evaluation. We used several deep learning R packages to perform the analysis [20, 21]. Before running the deep learning models, we performed word embeddings [22]:

- **Word embeddings**: The *word2vec* is used to convert a sentence into its vector representation. In the present approach, for a sentence $S = \{w_1, w_2, \ldots, w_n\}$, where n represents the count of words in S. Here, each word, w_i, is related with a D-dimensional vector embedding, $X_i \in R^d$. The word vector representations are concatenated in their individual order, $X_{1:n} = X_1 \oplus X_2 \oplus \ldots \oplus X_n$, in which $\oplus$ is the concatenation operator. Additionally, each sentence is padded with zero-vectors to a fixed length.

For the deep sentiment analysis, we used the complete set of 37,999 unique tweets. These tweets were divided into three datasets: training (60%), validation (20%), and testing (20%). The sentiment scores are divided into three categories: positive, negative and impartial. Some of the parameters of the model development are: (1) Layer embedding: Output dimension = 32, (2) Embedding layer dropout rate = 0.5, (3) Layer LSTM: units = 256, dropout = 0.2, recurrent dropout = 0.2, (4) Layer Output: units = 3, activation = 'softmax.'

For each of the datasets, ratio of positive and negative sentiment is approximately 3:1. This finding answers **RQ1** by indicating that positive sentiments are higher toward AV-related tweets.

5 Results and Discussions

The loss and accuracy values for different epochs are shown in Fig. 1. We used several performance measures to examine the model performances. True positive (TP) and false positive (FP) represent measures of accurate and inaccurate classifications per real classification category, respectively. True negative (TN) and false negative (FN) are measures of accurate and inaccurate rejections per real classification category, respectively. Some of the common measures are the following:

- Recall or sensitivity $= \dfrac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$
- Specificity $= \dfrac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$

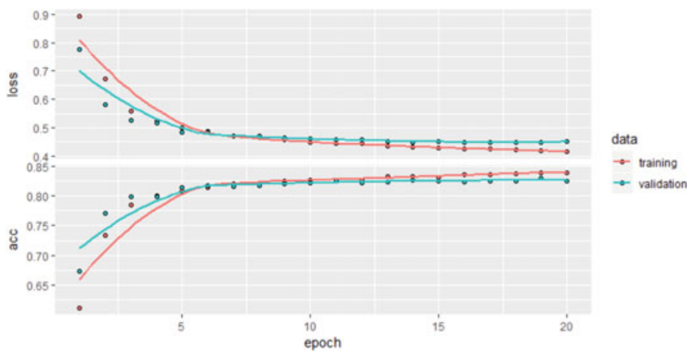

Fig. 1 Performances of the model for training and validation data

- Prevalence $= \dfrac{\mathrm{TP} + \mathrm{FN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$
- Positive Predictive Value $= \dfrac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$
- Negative Predictive Value $= \dfrac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}}$
- Accuracy $= \dfrac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$
- Balanced Accuracy $= \dfrac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \times 0.5 + \dfrac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \times 0.5$
- Detection Rate $= \dfrac{\mathrm{TP}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$
- Detection Prevalence $= \dfrac{\mathrm{TP} + \mathrm{FP}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$

Tables 1 and 2 list all performance measures calculated from the LSTM outputs. The results show that the accuracy is 85.5% for train data (95% confidence range: [85.1, 85.9%]), 82.6% for validation data (95% confidence range: [81.7, 83.4%]), and 82.8% for test data (95% confidence range: [81.9, 83.7%]). The findings answer RQ2 by providing evidence that deep learning tools are adequate in classifying sentiments from self-driving car-related tweets.

6 Conclusions

In this paper, we presented an application of deep learning-based method for sentiment classification of self-driving car-related tweets. This study demonstrates the capability of obtaining hidden trends about the consumer opinions and sentiments from Twitter posts with high accuracy. This paper introduces the concept of applying deep learning tools to classify AV-related sentiments and specifically outlines how the findings benefit the decision-making process of the future marketplace of AVs. Moreover, the paper reveals future customer needs in the context of AV deployment. The current framework developed in this study contributes to the ongoing state-of-the-art studies associated with application of sentiment analysis to understanding the perception of people regarding AVs.

We applied deep learning tools to unearth the nature of public understanding toward AV technology in the form of attitudes, sentiments, and opinions. The current approach used in this study indicates that text narratives in self-driving car-related tweets have significant key attribute measures that can be used in classifying different types of sentiments. This study has some limitations. For example, all words are

Table 1 Performance measures of train, validation, and test data

Data	Measures	Negative	Neutral	Positive
Train	Sen	0.8135	0.3552	0.9547
	Spe	0.9393	0.99056	0.759
	PPV	0.8451	0.80404	0.8622
	NPV	0.9252	0.93374	0.9139
	Pre	0.2894	0.09829	0.6123
	DR	0.2354	0.03491	0.5845
	DP	0.2786	0.04342	0.678
	BA	0.8764	0.67288	0.8569
Validation	Sen	0.7734	0.29918	0.9321
	Spe	0.9173	0.98879	0.7253
	PPV	0.7904	0.73986	0.845
	NPV	0.9094	0.92976	0.8693
	Pre	0.2874	0.09632	0.6163
	DR	0.2222	0.02882	0.5745
	DP	0.2812	0.03895	0.6799
	BA	0.8453	0.64398	0.8287
Test	Sen	0.7675	0.31773	0.9339
	Spe	0.9244	0.98564	0.7264
	PPV	0.8071	0.6935	0.8452
	NPV	0.906	0.9339	0.873
	Pre	0.292	0.09276	0.6153
	DR	0.2241	0.02947	0.5746
	DP	0.2776	0.0425	0.6799
	BA	0.8459	0.65169	0.8302

Note: *Sen* sensitivity, *Spe* specificity, *PPV* positive predictive value, *NPV* negative predictive value, *Pre* prevalence, *DR* detection rate, *DP* detection prevalence, *BA* balanced accuracy

weighted same for the sentiment scores. No specific weights were provided for high influential and relevant words associated with the current format of datasets. The same holds for negative words. Another potential improvement would be related to the exposure of sentiment fluctuations over time. Currently, the study is limited to four months in 2019, which is a short temporal window. Future research is needed to improve the proposed approach and its effectiveness to future marketplace of AV technologies.

Table 2 Accuracies and p-values by train, validation, and test data

Data	Measures	Values
Train	Acc	0.8549
	95CI	(0.8503, 0.8595)
	NIR	0.6123
	p-value	<2.2e−16
	K	0.7098
	p-value (MT)	<2.2e−16
Validation	Acc	0.8255
	95CI	(0.8168, 0.834)
	NIR	0.6163
	p-value	<2.2e−16
	K	0.6485
	p-value (MT)	<2.2e−16
Test	Acc	0.8282
	95CI	(0.8195, 0.8366)
	NIR	0.6153
	p-value	<2.2e−16
	K	0.654
	p-value (MT)	<2.2e−16

Note: *Acc* accuracy, *95CI* 95% confidence interval, *NIR* no information rate, p-value = p-value [Acc > NIR], K kappa, p-value (MT) mcnemar's test p-value

References

1. R. Sadiq, M. Khan, Analyzing self-driving cars on twitter (2018)
2. L.M. Hulse, H. Xie, E.R. Galea, Perceptions of autonomous vehicles: relationships with road users, risk, gender and age. Saf. Sci. **102**, 1–13 (2018)
3. D.J. Fagnant, K. Kockelman, Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transport. Res. Part A Policy Pract. **77**, 167–181 (2015)
4. N.J. Greig, Driver less cars-the view of the consumer? (2017)
5. J. Zmud, I.N. Sener, J. Wagner, Self-driving vehicles: determinants of adoption and conditions of usage. Transport. Res. Rec. **2565**(1), 57–64 (2016)
6. N. Merat, R. Madigan, S. Nordhoff, Human factors, user requirements, and user acceptance of ride-sharing in automated vehicles (2017)
7. M. Regan, M. Cunningham, V. Dixit, T. Horberry, A. Bender, K. Weeratunga, A. Hassan, Preliminary findings from the first Australian national survey of public opinion about automated and driverless vehicles. Australia and New Zealand Driverless Vehicle Initiative
8. S. Greaves, B. Smith, T. Arnold, D. Olaru, A.T. Collins, Autonomous vehicles down under: an empirical investigation of consumer sentiment (2018)
9. T. Liljamo, H. Liimatainen, M. Pöllänen, Attitudes and concerns on automated vehicles. Transport. Res. Part F Traffic Psychol. Behav. **59**, 24–44 (2018)
10. F. Nazari, M. Noruzoliaee, A.K. Mohammadian, Shared versus private mobility: modeling public interest in autonomous vehicles accounting for latent attitudes. Transport. Res. Part C Emerg. Technol. **97**, 456–477 (2018)

11. I.N. Sener, J. Zmud, T. Williams, Measures of baseline intent to use automated vehicles: a case study of texas cities. Transport. Res. Part F Traffic Psychol. Behav. **62**, 66–77 (2019)
12. S. Nordhoff, J. de Winter, M. Kyriakidis, B. van Arem, R. Happee, Acceptance of driverless vehicles: results from a large cross-national questionnaire study. J. Adv. Transport. **1–22**, 2018 (2018)
13. H. Asgari, X. Jin, Incorporating attitudinal factors to examine adoption of and willingness to pay for autonomous vehicles. Transport. Res. Rec. J. Transport. Res. Board **2673**(8), 418–429 (2019)
14. J. Gentry. twitteR: R based twitter client
15. S. Hochreiter, J. Schmidhuber, Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
16. H.T. Siegelmann, E.D. Sontag, On the computational power of neural nets, in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* (ACM, 1992), pages 440–449
17. A beginner's guide to LSTMs and recurrent neural networks. https://skymind.ai/wiki/lstm. Accessed 2019-07
18. J. Silge, D. Robinson, tidytext: Text mining and analysis using tidy data principles in r. J. Open Sour. Softw. **1**(3), 37 (2016)
19. F.A. Nielsen, A new anew: evaluation of a word list for sentiment analysis in microblogs (2011)
20. J.J. Allaire, F. Chollet, R interface to 'keras'. r package version 2.2. 4 (2018)
21. J.J. Allaire, Y. Tang, tensorflow: R interface to 'tensorflow'. r package version 1.10. 0.9000 (2019)
22. M. Kamkarhaghighi, M. Makrehchi, Content tree word embedding for document representation. Expert Syst. Appl. **90**, 241–249 (2017)