

Review of Anomaly Detection and Removal Techniques from Autonomous Car Data

Abdul Sami Mohammed^{1*}, Ahmed Damdam¹, Ahmed S Alanazi¹ and Yousef A Alhamad¹

¹ Prince Mohammad Bin Fahd University, Saudi Arabia

*Corresponding author

Abstract

Artificial Intelligence and the development of computers has brought a huge change in our daily lives. Simple tasks that required humans to use precious time and energy are now being done by computers. One aspect of this is using AI in self-driving cars. By introducing such technologies in the world, it can be said that travelling has become a whole new experience. However, Artificial Intelligence cannot be considered perfect. It has its own faults and these faults can lead to undesired circumstances such as car accidents. One of the major faults in Artificial Intelligence is the presence of anomalies in data. Anomalies are values that are either too extreme to be considered, or are just not needed. Detecting these anomalies is a major task and requires algorithms that efficiently extract them. This paper discusses a few of these techniques and further highlights which ones can be used in autonomous self-driving cars.

Keywords: Anomaly Detection; Autonomous Car; Sensor Data; Algorithms

I. Introduction

In the present day, electric cars have become increasingly more efficient and compelling than traditional gas-powered cars. As electric cars don't have an engine maintaining them is a lot easier and cheaper. However, the batteries situation is below average, as many of them are lithium-based, it is very inefficient and dangerous for electric cars to keep that type of material so it is necessary to find a better material for the batteries. The technology of electric cars is phenomenal, as the industry remove most of the mechanical part of the car, most of the development is done on its sensory and self-driving AI which allow the car more autonomous and precise on its driving ability. Overall, the status of the electric cars is getting better as more development is done on its technology, it will soon perform better than gas-powered cars.

(a) Technological Limitations:

– **Driving Range:** electric cars range are much lower compared to gas-powered cars which hold electric cars back from mass adoption by the public. One of the main factors for choosing between a gas-powered and an electric car is how far can the car go with a full tank. In most cases gas-powered cars have a longer range because gas last longer than electric and the infrastructure of many nations revolve around gas-powered car.

– **Batteries:** In most electric cars the battery is made of lithium-ion which can burst into flame if it gets too hot and has a short service life. This cause people to avoid electric car since it is will endanger the driver's lives and the company can be sued for the endangerment of lives. Lithium-ion batteries can't last long which inhibits the electric car's performance and is thrown away at it end of its lifespan which causes pollution due to having hazardous materials.

– **Charging:** to charge an empty electric car it will take about 4-8 hours to fully charge and about 30 min-1 hour with a supercharger, while a gas-powered car will only take 1-3 min for a full tank. Many people will be reluctant to buy electric cars as it takes too long to charge and since the battery is lithium-ion, they must recharge frequency to be able to use it constantly.

(b) Ideas for the Future of Electric Cars:

- **New Batteries:** the batteries we use now in electric cars are very inefficient in terms of power output. They tend to run out very quickly without giving a good mileage. For this reason, new batteries need to be invented which can act as better driving forces for cars without having to be charged every now and then.

- **Wireless and Bidirectional charging:** as more and more electric products are being introduced to wireless charging, it won't be long before electric cars do to. Bidirectional charging can also be a great development in which charge can be transferred from on car to another. Combining these two ideas can bring about a huge breakthrough.

- **Data Generators:** Electric cars are big moving data generators which means their cameras and sensors are constantly reading data. If somehow, this data can be stored and classified, it can be a helpful resource for Artificial Intelligence research and projects which require massive datasets.

2. Literature Review

In a paper written by Dr. Manoharan, an improved safety algorithm for Artificial Intelligence enabled processors in self driving cars is proposed. The paper focuses on computer vision, sensor fusion, localization, path planning and control to see how the artificial intelligence interacts with the self-driving car. A new algorithm for trajectory adjustment is introduced and tested for improved safety. The advantages of self-driving cars are also mentioned (Improved safety, reduced carbon emission etc.). This papers ensures a highly effective and a safer driving on roads that include other vehicles and pedestrians. The decision of using self-driving cars would avoid unnecessary congestion on roads and also reduce the number of accidents.

In a research paper by R. Vinothkanna, a survey was conducted on novel estimation approach of motion controllers for self-driving cars. The motion planning was one of the most challenging tasks in autonomous driving cars. In this research paper, the novel planning method was compared and from doing so it was concluded that the suitable algorithm was between realistic unusual traffic and complex target. Both of them were real time traffic conditions. Their suggesting idea was that the autonomous control is estimating the modified version of action based coarse trajectory planning.

In a paper by S.E.Viswapriya, it was discussed how existing lane detecting systems in self-driving cars are still having issues, which could lead to future accidents. The surveys his team conducted explain why they are

using the Hough transform technique and how they will implement this into self-driving cars. His team proposed that new lane detecting systems should be made and have elements of Artificial Intelligence integrated with the systems. The way they did it is by using cameras to help the AI make histograms, which will help it calculate lane width, curvature, and off-center. His team believes this will reduce accidents, help the AI detect lane better, and reduce power consumption. The result is the new lane detecting systems were able to calculate the radius of the road, direction, and off-center.

In a paper by Shinji Kawakura, it is shown how they tried to make a deep, learning-based, self-driving car system to transport or deliver different things such as bottles, harvested onions, etc. To do this, they used an NVIDIA artificial intelligence-oriented board inside of a car-shaped robot. They made many experimental trials inside of a warehouse with many scattered items all over the floor to test out the small car's ability to move around the scattered objects. In the end, their results were good, but not perfect. The robot would get its wheels stuck after trying to avoid certain obstacles, it wouldn't detect obstacles quick enough to avoid them, and it would detect obstacles too early, and so on. Therefore, they are thinking of fixing and changing various factors with the programming and the exterior design such as tires and covers.

3. Background

Artificial intelligence (AI) has taken the automotive industry by storm to drive the development of Level 4 and 5 autonomous vehicles. Why do you think AI has become so popular now, even though it has been around since the 1950s? Quite simply, the reason for the explosion of AI is the vast amount of data available to us today. With the help of connected devices and services, we are able to collect data in every industry, driving the AI revolution. While efforts are underway to rapidly improve sensors and cameras to generate data with autonomous vehicles, Nvidia unveiled its first AI computer in October 2017, enabling Deep Learning, computer vision and parallel computing algorithms. AI has become an essential part of automated driving technology, and it's important to know how it works in autonomous and connected vehicles.

(a) What is artificial intelligence?

John McCarthy, a computer scientist, coined the term 'artificial intelligence' in 1955. AI is defined as the ability of a computer program or machine to think, learn and make decisions. In common usage, the term means a machine that mimics human cognition. With AI, we make computer programs and machines do what humans do. We feed these programs and machines a huge amount of data that is analyzed and processed to ultimately think logically and perform human actions. The process of automating repetitive human tasks is just the tip of the AI iceberg. Medical diagnostic equipment and autonomous cars have used AI with the goal of saving human lives.

(b) Do self-driving cars use artificial intelligence?

As technology evolves, the automobile industry has taken advantage of new discoveries to provide new methods to make the user's life easier (driver). Using artificial intelligence to develop self-driving automobiles is one of them. A self-driving car (also known as an autonomous car or driverless car) is a vehicle that travels to locations without the need for a human driver by utilizing a variety of sensors, radars, cameras, and artificial intelligence. Many companies have started to manufacture self-driving cars, such as Tesla, Audi, BMW, Ford, and many more. These businesses put their cars through a series of testing to guarantee that they are safe to drive on the

road. To be considered completely autonomous, an automobile must travel paths to a predetermined destination without the need for human interaction.

(c) The growth of AI in the automotive industry

The market for AI in the automotive industry is estimated at \$783 million in 2017 and is expected to reach nearly \$11,000 million by 2025, at a CAGR of about 38.5%. IHS Markit predicted that the installation rate of AI-based systems in new vehicles will increase by 109% by 2025, compared to an 8% adoption rate in 2015. AI-based systems will become standard in new vehicles mainly in these two categories:

1. Human-machine infotainment interface, including voice and gesture recognition, eye tracking and driver monitoring, virtual assistance, and natural language interfaces.
2. Advanced driver assistance systems (ADAS) and autonomous vehicles, including camera-based vision systems, radar-based detection units, driver state assessment, and sensor fusion of engine control units (ECUs).

Deep learning technology, a technique for implementing machine learning (an approach to achieving AI), is expected to be the largest and fastest-growing technology in the automotive AI market. It is currently used in speech recognition, voice search, recommendation engines, sentiment analysis, image recognition, and motion detection in autonomous vehicles.

(d) How does AI work in autonomous vehicles?

Let us first look at the human perspective of driving, where we use sensory functions such as vision and hearing to observe the road and other vehicles on the road. When we stop at a red light or wait for a pedestrian to cross the street, we use our memory to make these quick decisions. Through our years of driving, we have become accustomed to paying attention to the little things we often encounter on the road - that could be a better route to the office or just a big obstacle in the road.

Self-driving car frameworks are powered by artificial intelligence. Self-driving car engineers use massive amounts of data from image recognition systems, as well as AI and neural networks, to build self-driving vehicle frameworks. The neural networks recognize patterns in the data, which is then fed into the AI algorithms. Images from self-driving car cameras are among the data. The neural networks learn to detect traffic lights, trees, checks, pedestrians, road signs, and other elements in any given driving environment.

We are building autonomous vehicles that drive themselves, but we want them to drive like human drivers. That means we need to equip these vehicles with the sensory functions, cognitive functions (memory, reasoning, decision-making and learning) and executive abilities that humans use to drive vehicles. The automotive industry has been continuously evolving in recent years to do just that.

To give an example, Google has begun to develop self-driving cars, which employ a combination of sensors, light detectors, and technology such as GPS and cameras to combine all of the data generated by those systems around the vehicle, and the artificial system predicts what those objects will do next.

According to Gartner, by 2020, 250 million cars will be connected to each other and their surrounding infrastructure through various V2X (vehicle-to-anything communication) systems. As more information is fed into IVI (in-vehicle infotainment) or telematics systems, vehicles will be able to capture and share not only internal system status and location data, but also changes in their environment, in real time. Autonomous vehicles will be equipped with cameras, sensors, and communication systems that will enable the vehicle to generate vast amounts of data that, when combined with AI, will enable the vehicle to see, hear, think, and make decisions just as human drivers do.

(e) AI perception-action cycle in autonomous vehicles

A repetitive loop called the Perception Action Cycle occurs when the autonomous vehicle generates data from its environment and passes it to the intelligent agent, which in turn makes decisions and enables the autonomous vehicle to perform specific actions in the same environment. The following figure illustrates the flow of data in autonomous vehicles:

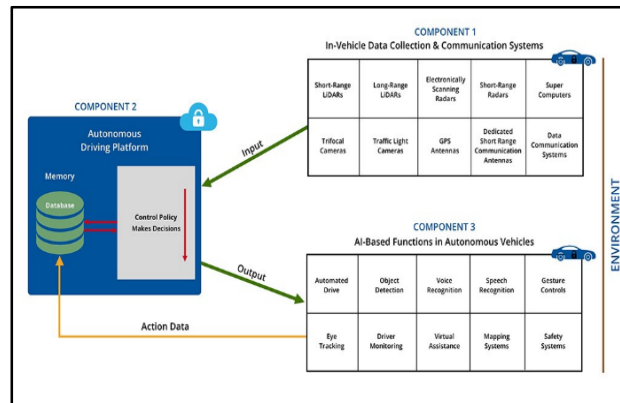


Fig. 1 illustrates the flow of data in autonomous vehicles

(f) How hardware components interact with software

In a paper by Mindy Support [6], they discuss how machine learning will enable self-driving cars to become a reality. They begin by describing how self-driving cars make decisions, which are based on object detection and categorization algorithms. The team uses data annotation to assist the machine learning system in making the best selections possible while navigating the roads. The car must be able to observe its surroundings in order to make decisions. Cameras, radar, and Lidar are the three key sensors used by self-driving cars, according to the team. When vision is poor, radar is best, while Lidar is comparable to radar but uses light waves instead of radio waves.



Fig.2 depicts visual data read by a Lidar

(g) Data Annotation

They believe that data annotation plays a crucial role in self-driving cars. By providing raw training data that has been processed using various annotation approaches, it will assist the AI system to comprehend what it needs to learn. The goal is to reduce serious accidents caused by human mistake. Self-driving cars require a machine

learning algorithm in order to be taught using real-world datasets. The algorithm is supplied instructions on how to interpret the incoming data when it is supervised. This enables the algorithm to evaluate training data against a fully labeled dataset, making supervised learning more effective for classification.

When it gets unsupervised learning with no instructions on how to process it, it is left to its own devices to figure out what to do. Self-driving cars are made possible by machine learning algorithms. They enable an automobile to acquire data from cameras and other sensors about its surroundings, understand it, and decide what actions to take. Cars can even learn to accomplish these activities as well as people thanks to machine learning.

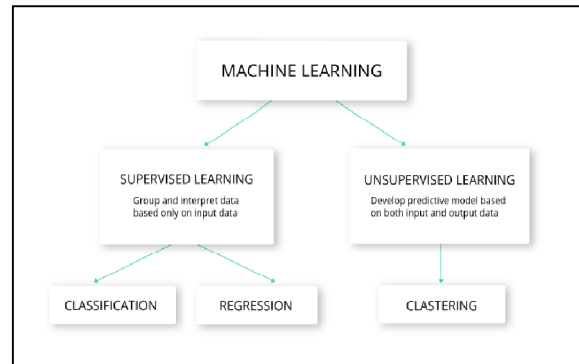


Fig. 3 shows an overview of the different approaches to machine learning

(h) Pros and cons

The biggest benefit touted by proponents of self-driving vehicles is security. According to a U.S. Department of Transportation (DOT) and National Highway Traffic Safety Administration (NHTSA) accurate prediction of traffic deaths for 2017, 37,150 people perished in engine vehicle car crashes. According to the National Highway Traffic Safety Administration, 94 percent of actual accidents are caused by human mistakes or poor decisions, such as drunk or inattentive driving. Self-driving vehicles remove such risk factors from the equation; yet, they are impotent against other elements that cause crashes, such as mechanical difficulties.

The financial benefits of autonomous vehicles could be enormous if they can reduce the number of accidents. According to the National Highway Traffic Safety Administration, injury has an economic impact of \$57.6 billion in lost working environment usefulness, \$594 billion in the death toll, and diminished personal satisfaction due to wounds.

In principle, if self-driving vehicles were used on the streets for the most part, traffic would flow smoothly and there would be less congestion. Tenants could do beneficial exercises while driving to work in fully automated vehicles. Individuals who are unable to drive due to physical limitations could gain new autonomy through self-driving cars and be able to work in fields that require movement.

4. Anomaly Detection

An anomaly is something that deviates from the normal and is an unexpected value in a set of values. In machine learning, it is of interest to know what these anomalies are and to get rid of them so that accuracy can be achieved to the highest degree.

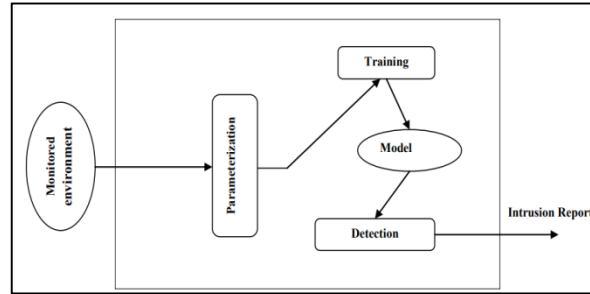


Fig.4 shows a generic anomaly based network intrusion detection system (A-NIDS) architecture

There are plenty of anomaly detection techniques but we will be discussing the most common supervised methods in this paper.

(a) Supervised Anomaly Detection

In supervised anomaly detection, datasets including both normal and anomalous data are used to develop the predictive model. In theory, supervised methods give a higher accuracy in detection as they have access to more information. However, the technical issues that come along with this make them less accurate. Issues include shortage of training data and training sets containing noises that can increase the false alarm rates. The most commonly used supervised methods are Bayesian Networks and Decision Tree, k-Nearest Neighbors, and Supervised Neural Networks.

1. K -Nearest Neighbor (k-NN)

This technique is one of the conventional techniques for classifying samples. It calculates the estimated distances between several points on the input vectors, and then allocates the unlabeled points to the class of its K-nearest neighbors. In the procedure of generating k-NN classifier, (k) is an important constraint and various (k) values can cause various performances. If k is very huge, the neighbors, which are used for prediction, will consume great classification time and affect the prediction precision.

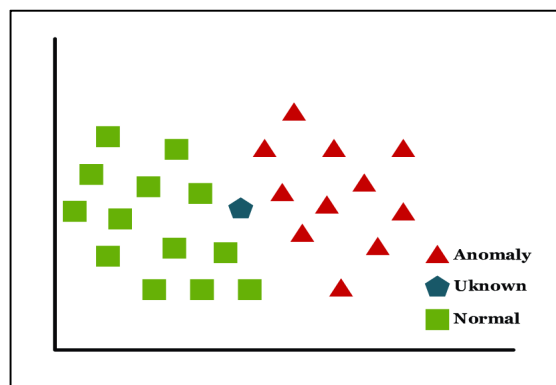


Fig.5 shows a graph that depicts the K-nearest neighbor classification principle

2. Bayesian Network (BN)

Bayesian is defined as a model that translates probabilistic relationships among variables of interest. This method is generally used for intrusion detection in combination with statistical schemes. It has several benefits,

including the ability of encoding interdependencies between variables and of predicting events, as well as the capability to integrate both prior knowledge and data.

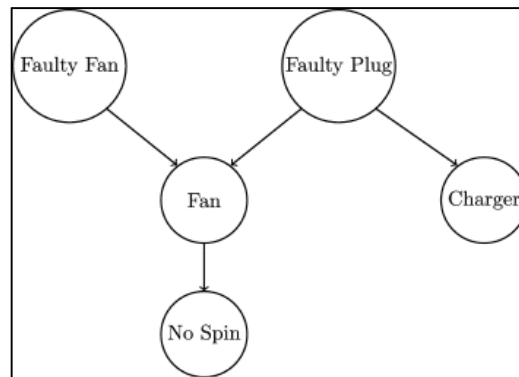


Fig.6 shows a simple example of a Bayesian network structure

3. Supervised Neural Network (NN)

The NNs learning calculate different users and daemons behavior in systems. If they are properly planned and executed, NNs have the capability to address many problems encountered by rule-based tactics. The main advantage of NNs is their tolerance to inexact data and uncertain information, and their ability to conclude solutions from data without having prior knowledge of the uniformities in the data. This, in combination with their ability to simplify from learning data, has made them a proper approach to ID. In order to apply this approach to ID, data representing attacks and non-attacks have to be introduced to the NN to adjust repeatedly network figures during the training phase.

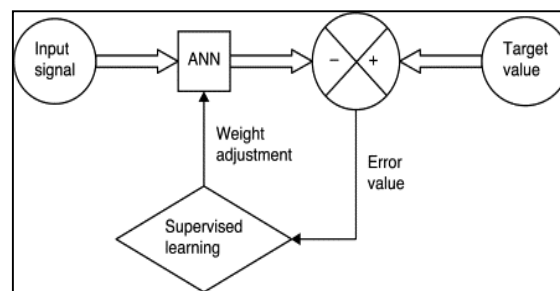


Fig.7 shows the process of Supervised Neural Networks in order to achieve error values (or anomalies)

4. Decision Tree (DT)

Decision Trees are defined as powerful and common tools for classification and prediction. A decision tree is a tree that has three main components: nodes, arcs and leaves. Each node is branded with a feature attribute, which is most informative among the attributes not yet measured in the path from the root. Each arc out of a node is labeled with a feature value for the node's feature, and each leaf is labeled with a category or class. A decision tree can then be used to classify a data point by starting at the root of the tree and moving through it until a leaf node is reached. The leaf node provides the classification of the data point. ID3 and C4.5 developed by Quinlan are the most common implementations of the Decision Tree.

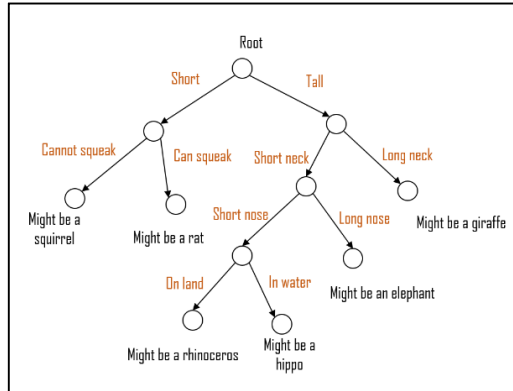


Fig.8 shows an example of Decision Tree being used to classify animals

5. Analysis

For our purpose of finding anomalies in data from autonomous cars, our values are quantitative. These are mostly readings from a sensor. For readings like these, K-Nearest neighbor is the most suitable supervised anomaly detection algorithm. Basically, it creates a scatter graph based on how the user pre-classifies the data. For example, infrared sensor readings could be: 1 meter is too close, 10 meters is the upper limit etc. By giving these values, the algorithm figures out which class the values belongs to once they have been fed into the model. If any of the values seem to be too extreme (i.e. too small or too large), the algorithms detects it as an anomaly and warns the user that data with an unpredicted value has been encountered.

Once the anomaly is detected, removing it is a pretty simple task. We simply do not include it in the sensor readings that are sent to the main processing unit of the car so it doesn't act abnormally by reacting to such anomalies.

6. Conclusion

In this paper, we looked at the basics of Artificial Intelligence and how it has become a part of many research studies as well as being used in the industry. We focused on how AI works in self-driving autonomous cars. By looking into previous research we noticed that there is a lack of studies being done on how to deal with anomalies that can be gathered from the sensor readings from the autonomous cars. We looked at several Supervised Anomaly Detection Algorithms which used different sophisticated techniques to detect anomalies. We inferred that K-nearest Neighbor is the best algorithm since it works best for quantitative data which is exactly what we are getting from the self-driving cars. By actively getting rid of such anomalies while the car is being driven, a much safer system of autonomous vehicles is introduced.

References

- [1] B. Xinwei Fang, An Improved Sensor Calibration with Anomaly Detection and Removal, White Rose Research, York, 2020.
- [2] S. B. Md Asri Ngadi, Machine Learning Techniques for Anomaly Detection:, Research Gate, Benghazi, Libya, 2013

- [3] S. U. Markus Goldstein, A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data, Plos One, Fukuoka, 2016
- [4] S. Gadam, Artificial Intelligence and Autonomous Vehicles,"19 4 2018. [Online]. Available: <https://medium.datadriveninvestor.com/artificial-intelligence-and-autonomous-vehicles-ae877feb6cd2>. \ [Accessed 15 5 2022]
- [5] EE, "Do self-driving cars use Artificial Intelligence(AI)?, 23 10 2019. [Online]. Available: <https://www.eescorporation.com/do-self-driving-cars-use-ai/>. [Accessed 5 15 2022].
- [6] M. San, How Machine Learning in Automotive Makes Self-Driving Cars a Reality, 12 2 2020. [Online]. Available: <https://mindy-support.com/news-post/how-machine-learning-in-automotive-makes-self-driving-cars-a-reality/>. [Accessed 15 5 2022]