Machine Learning and Sentiment Analysis

Using Public Data

5038-47626

May 3, 2018

Contents

1 Abstract

Machine learning is a way of doing data analysis that allows for the extraction information which requires context. This is impossible to program in absolutes. There are many uses for this technology, including sentiment analysis. In sentiment analysis, the machine-learning algorithm is fed text. It then returns what sentiment the text expresses. The algorithm is written to model a brain. The brain has synapses that connect at junction points. These junction points only allow the next synapse to fire if the previous signal meets conditions. These synapses, also known as tensors, meet at junctions. A weight is then multiplied by the input signal and then is checked to see if

it meets a threshold. If so, then the signal carries on to the next tensor. This is called a neural network, which has to be trained. Random values are selected for the thresholds and weights. Data is then fed through the network, and the results are checked against an answer key. Each junction then gets adjusted. The data is fed through it again. If the data is more accurate, the process is repeated on the new network, otherwise, the previous network is used. Large amounts of data are required to make the network accurate. A way to source data is to find it in public places. Twitter has thousands of posts tagged by hashtags that can act like answer keys for the emotion within them. Thus, it can be used as the perfect place to source data from.

2 Research Question

How to apply machine learning to sentiment analysis in order to estimate the sentiment expressed within a text?

3 Introduction

Machine learning is a fast growing technology allowing for the processing of vast quantities of information quickly. Previous approaches to data processing required certain conditions that must be met to process the data, and these conditions had to be expressed in terms of absolutes. These approaches have limitations when it comes to processing data. For example, by reading a sentence, the brain can figure out what emotion has been expressed in the sentence. This refers to a topic in artificial intelligence called sentiment analysis. It is impossible to perform sentiment analysis using the traditional approaches, because emotions such as humor cannot be expressed absolutely.

Machine Learning models a brain(Vu et al.) in order to process data (text in the case of sentiment analysis) and draw a conclusion (the sentiment expressed in the text).(Dean, Patterson, and Young) There are many uses for sentiment analysis. Facebook, for example, uses it to decide what order to show your feed in. If it finds that you show positive sentiment towards a certain product or idea, it is weighted higher on your feed and vice versa. Google uses this to know what ads to serve you when on sites that use their ad service.

4   Neural Networks

4.1   The Model

Machine learning is a technique for data processing that models a brain. This means anything a human can understand, a machine can too. This model is called a neural network, since it models neurons in the human brain. Fig. 1 shows the model of a neuron. A neuron, just like a real one in a brain, takes inputs from multiple neighboring neurons, multiplies them with corresponding weights, and then adds up the results as shown in the figure. The sum is then compared against a threshold. If the threshold is met, the neuron generates an output signal. If not, then it simply suppresses the signal. Whatever the output is, it is fed into other neurons as their input and they go through the same process. This way a signal is propagated in a neural network, like the one shown in fig. 2, and the data is processed.
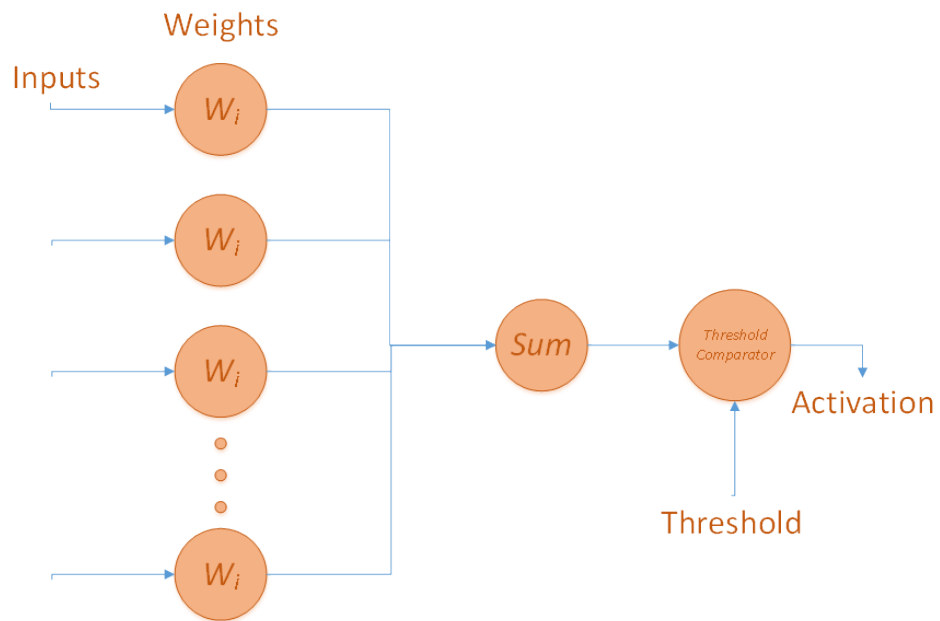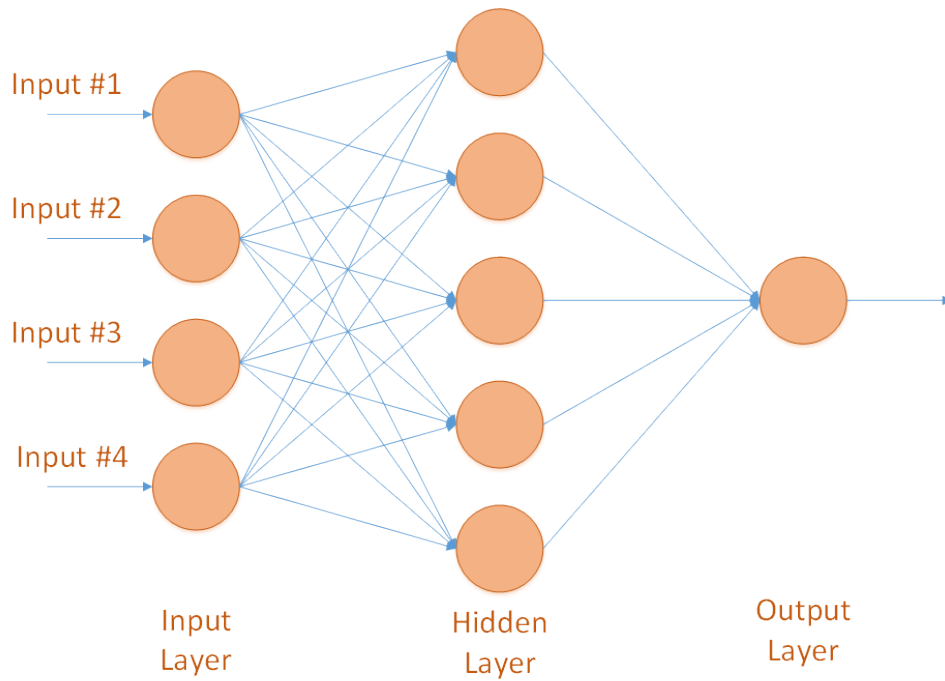
Figure 1: Mathematical model of a neuron.

Figure 2: A model of a neural network.

4.2 Training

Training the neural network requires modifying the weight and threshold of each junction until the network has desirable results. This requires large quantities of data. Ways of getting this include using user inputted data on a service, harvesting data from user posts, and captchas which ask you to click on pictures that include certain items. This data also requires an answer key. For example, a neural network to tell if a post on facebook is happy or sad should be trained with data that includes sample posts and a answer for whether it is happy or sad. These sets of data known as datasets can consist of millions of entries.

Each one of the entries within a dataset is fed through the network with random values for the weights and biases. These are then modified and then the entries are fed through again. If this modified network has more accurate results than the previous one, then the previous network is discarded and this new one is used. Otherwise, the previous one is used. This process repeats. This brute force method of training requires large amounts of compute power and time in order to have a well trained and accurate network.

5 Sentiment Analysis using Neural Networks

5.1 Sentiment Analysis

Use cases for machine learning are hard to find. Since machine learning is not 100% accurate, just like humans, if there is another way to solve the problem it is most likely a better method(Lovelace, Fridman-Rojas, and Long). So using machine learning to solve basic arithmetic is not a good use case. An example of a good use case, however is sentiment analysis.
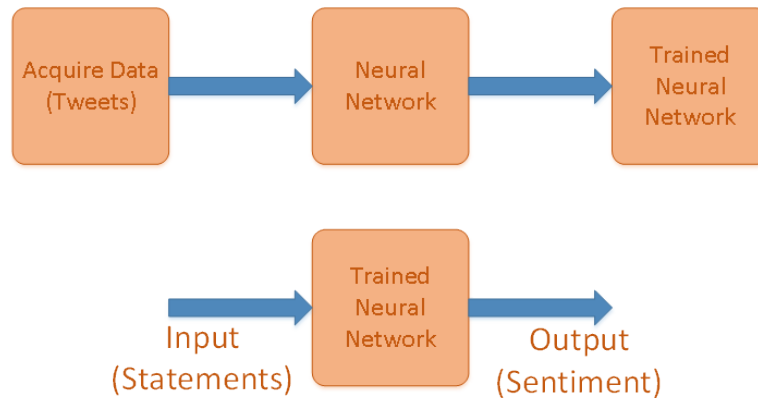
Figure 3: The process of building a neural network.

Sentiment analysis is a special use case of neural networks. Sentiment (the view or attitude) of a piece of text is a hard problem to express in traditional programming and data science. There is no absolute way to express humor, happiness, or any other emotion. However, humans can understand and comprehend this. This is a problem suitable for machine learning.

5.2 Collecting Data

Data collection is a major part of machine learning. One method to get data is by pulling data from a public source like Twitter. Twitter is a good source since the users tag there posts using hashtags. These hashtags can act like answer keys. For example, if a tweet uses the hashtag happy, it means the tweet represents happy sentiment.

5.3 Training

Training the neural network for sentiment analysis also has a major difficulty to overcome. Neural networks require large amounts of computation power to train. GPUs (graphics processing units) which are used for training neural networks are very expensive. This makes training large neural

networks very expensive. Training is also very time consuming since it has such a brute force approach and diminishing returns the longer you train. Refer to fig. 3 for a visual diagram of this process.

Works Cited

Dean, Jeff, David Patterson, and Cliff Young. "A New Golden Age in Computer Architecture: Empowering the Machine-Learning Revolution". *The Journal of Neuroscience* (2017). Print.

Lovelace, Robin, Ilan Fridman-Rojas, and Rob Long. "New tools of the trade? The potential and pitfalls of Machine Learning and DAGs to model origin-destination data". *White Rose University Consortium* (2017). Print.

Vu, Mai-Anh T., et al. "A Shared Vision for Machine Learning in Neuroscience". *The Journal of Neuroscience* (2017). Print.