

Sentimental Analysis Of Twitter Data Using Machine Learning

Mr. S. Niresh Kumar¹, Dr. V. Devarajan², Naveen Kumar K³

¹Assistant Professor, Dhanalakshmi Srinivasan College of Engineering and Technology

²Professor, Dhanalakshmi Srinivasan College of Engineering and Technology

³Student, Dhanalakshmi Srinivasan College of Engineering and Technology.

ABSTRACT: Social networks are the main resources to gather information about people's opinion and sentiments towards different topics as they spend hours daily on social medias and share their opinion. In this technical paper, we show the application of sentimental analysis and how to connect to Twitter and run sentimental analysis queries. We run experiments on different queries from politics to humanity and show the interesting results. We realized that the neutral sentiment for tweets are significantly high which clearly shows the limitations of the current

1. INTRODUCTION

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or

instruction, in order to look for patterns in data and make better decisions in the future

based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

2. MATERIALS AND METHODS

A. Machine Learning Methods

Machine learning algorithms are often categorized as supervised or unsupervised.

B. Supervised machine learning algorithms

It can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly. In contrast,\

C. Unsupervised machine learning algorithms

They are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

D. Semi-supervised machine learning algorithms

Fall somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training – typically a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabeled data generally doesn't require additional resources.

E. Reinforcement machine learning algorithms

Learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

Machine Learning undoubtedly helps people to work more creatively and efficiently. Basically, you too can delegate quite complex or monotonous work to the computer through Machine Learning - starting with scanning, saving and filing paper documents such as invoices up to organizing and editing images. In addition to these rather simple tasks, self-learning machines can also perform complex tasks. These include, for example, the recognition of error patterns. This is a major advantage, especially in areas such as the manufacturing industry: the industry relies on continuous and error-free production. While even experts often cannot be sure where and by which correlation a production error in a plant fleet arises, Machine Learning offers the possibility to identify the error early this saves down times and money. Self-learning programs are now also used in the medical field. In the future, after "consuming" huge amounts of data (medical publications, studies, etc.), apps will be able to warn a in case his doctor wants to prescribe a drug that he cannot tolerate.

3. RESULTS AND DISCUSSIONS

Traffic Predictions: We all have been using GPS navigation services. While we do that, our current locations and velocities are being saved at a central server for managing traffic. This data is then used to build a map of current traffic. While this helps in preventing the traffic and does congestion analysis, the underlying problem is that there are less number of cars that are equipped with GPS. Machine learning in such scenarios helps to estimate the regions where congestion can be found on the basis of daily experiences. Online Transportation Networks: When booking a cab, the app estimates the price of the ride. When sharing these services, how do they minimize the detours? The answer is machine learning. Jeff Schneider, the engineering lead at Uber ATC reveals in an interview that they use ML to define price surge hours by predicting the rider demand. In the entire cycle of the services, ML is playing a major role.

The sentiment analysis over twitter is very effective way to observe the feelings of public towards their brands, political news, business, and movies etc. There are several methods to classify the datasets on the blogs. The datasets on twitter is added up with a feature that specifies the positives/negatives on it. Twitter messages can lead more efficient way of extracting the data. It provides the best way of social communication on an issue, which helps to gather common opinion.

1. Used to review the related websites, review the movies/ products.
2. To know about the sensitive matters. process requires skilled professional
3. To know the current trends and fashions.
4. Public opinion on political leaders, business leads, Actors et

Social Networking sites are the resources which contains huge data. For example, Twitter produces millions of bytes of the data. These data can be used for business or social purpose. Analyzing data from these social networking Web sites is one of the new buzzword for many business strategies. Election campaigns, World health issues, Technical concepts, inventions, Entertainment, Natural resources can be effectively handle by using sentimental analysis. Our proposed work evaluates sentimental analysis of Twitter data using Stand ford NLP Libraries implemented in SaaS (cloud) which will handle all current affairs in the world. Cloud implementation will give process efficiency, result growth and improvement in time to market. The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Google, YouTube, Bit Torrent, etc. Python programming is widely used in Artificial Intelligence, Natural Language Generation, Neural Networks and other advanced fields of Computer Science. Python had deep focus on code readability & this class will teach you python from basics.

- It provides rich data types and easier to read syntax than any other programming languages
- It is a platform independent scripted language with full access to operating system API's
- Compared to other programming languages, it allows more run-time flexibility

You can create scalable Web Apps using frameworks and CMS (Content Management System) that are built on Python. Some of the popular platforms for creating Web Apps are: Django, Flask, Pyramid, Plone, Django CMS. Sites like Mozilla, Reddit, Instagram and PBS are written in Python. The support of Numpy makes the task more easier. Numpy is a highly optimized library for numerical operations. It gives a MATLAB-style syntax. All the OpenCV array structures are converted to-and-from Numpy arrays. So whatever operations you can do in Numpy, you can combine it with OpenCV, which increases number of weapons in your arsenal. Besides that, several other libraries like SciPy, Matplotlib which supports Numpy can be used with this. So OpenCV-Python is an appropriate tool for fast prototyping of computer vision problems. Anaconda is a free and open source, easy to install distribution of Python and R programming languages. Anaconda provides a working environment which is used for scientific computing, data science, statistical analysis and machine learning.

Anaconda Navigator is a desktop graphical user interface (GUI) included in the Anaconda distribution and easily manage conda packages, environments and channels without the use of command-line commands. It is available for Windows, macOS and Linux.

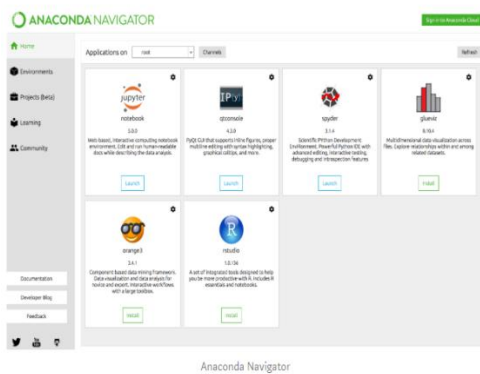


Fig. 3.1 Applications Provided In Anaconda Distribution

The Anaconda distribution comes with the following applications along with Anaconda Navigator.

1. Jupyter Lab
2. Jupyter Notebook
3. Qt Console
4. Spyder Glueviz

5. Orange3

6. RStudio

7. Visual Studio Code

> Jupyter Lab: This is an extensible working environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

> Jupyter Notebook: This is a web-based, interactive computing notebook environment. We can edit and run human-readable docs while describing the data analysis.

> Qt Console: It is the Py Qt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips and more.

> Spyder: Spyder is a scientific Python Development Environment. It is a powerful Python IDE with advanced editing, interactive testing, debugging and introspection features.

> VS Code: It is a streamlined code editor with support for development operations like debugging, task running and version control.

> Glueviz: This is used for multidimensional data visualization across files. It explores relationships within and among related datasets.

> Orange 3: It is a component-based data mining framework. This can be used for data visualization and data analysis. The workflows in Orange 3 are very interactive and provide a large toolbox.

> R studio: It is a set of integrated tools designed to help you be more productive with R. It includes R essentials and notebooks.



Fig. 3.2. Anaconda 5.3

A. DESIGN

Data with emoticons which categorizes tweets to positive, negative and neutral based on the tweeters opinion.

The dictionary collected from the web with English translations.

In Twitter, the tweets are posted by the celebrities, politicians of different country, business tycoons, and entrepreneurs. The set of text in twitter can be as below:

1. Positives texts like happiness, enjoy etc. with emoticons like
2. Negative texts like sadness with emoticons
3. They just right the fact without any emoticons.

Preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

B. Black Box Testing

The black box testing is the type of testing in which the internal components of the software is hidden and only the input and output of the system is the key for the tester to find out a bug. It is therefore a simple type of testing. A programmer with basic knowledge can also process this type of testing. It is less time consuming as compared to the white box testing. It is very successful for software which are less complex are straight-forward in nature. It is also less costly than white box testing.

4. CONCLUSION

Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. A characteristic of these large data sets is a large number of variables that require a lot of computing resources to process. Feature extraction is the name for methods that select and /or combine variables into features, effectively reducing the amount of data that must be

processed, while still accurately and completely describing the original data set. The definition of an **emoticon** is a sideways facial expression created using keyboard **symbols** that represent an attitude or a facial expression in e-mail. An example of an **emoticon** is a happy face made by typing a colon and parenthesis. The functional tests help in providing the systematic representation that functions tested are available and specified by technical requirement, documentation of the system and the user manual.

REFERENCES

1. Collective intelligence & sentimental analysis of twitter data by using Stanford NLP libraries with software as a service (SaaS), Hase Sudeep Kisan ; Hase Anand Kisan ; Aher Priyanka Suresh, 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)
2. Discovering Public Opinions by Performing Sentimental Analysis on Real Time Twitter Data, G. Kavitha ; B. Saveen ; Nomaan Imtiaz, 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)
3. Sentiment Analysis of Tweets Using Machine Learning Approach, Megha Rathi ; Aditya Malik ; Daksh Varshney ; Rachita Sharma ; Sarthak Mendiratta, 2018 Eleventh International Conference on Contemporary Computing (IC3)
4. Sentiment Analysis on Twitter Using Streaming API, M. Trupthi ; Suresh Pabboju ; G. Narasimha, 2017 IEEE 7th International Advance Computing Conference (IACC)
5. Twitter Sentiment Analysis, 3-Way Classification: Positive, Negative or Neutral, Mestan Firat Çeliktüğ, 2018 IEEE International Conference on Big Data (Big Data)