

Development of a System to Analyze Emotions through the Data from Social Media using AWS

Azam Nawaz Birajdar, Bhagirathi Dupargude, Prof. S. G. Kulkarni,
Brahmdevdada Mane Institute of Technology, Belati, Solapur.
azamnawaz1029@gmail.com, bdupargude.bmit@gmail.com, shrigk@rediffmail.com

Abstract:- There is a large volume of data generated by internet users. Web become a platform for expressing ideas and sharing emotions. Social networking portal like Twitter, Facebook are popularity as they allow people to share and express their emotions about any topics, have discussion in their own circle, or post messages. There has been lot of work in the stream of emotion analysis of twitter data. This paper look into mainly on emotion analysis of twitter data which is helpful to analyze the information in the tweets where particular topic highly trend in some sort of topic. In this paper, we provide a approach to analyses techniques for emotion analysis by AWS and their services like EMR and S3 storage and twitter data streams. We have also put a point on general challenges and applications of Emotion Analysis on Twitter.

Keywords:- Twitter, Emotion analysis, Machine learning, Amazon web services (AWS), Mapreduce, Big data, Hadoop, S3 bucket, EMR.

1. INTRODUCTION

If you want to construct opinion towards a particular think there are basic two approaches which are driven.

First approach is Information driven and second approach is Data driven. The Information driven approach having the information but not meet the appropriate goal because of the information transfer man to man or peer to peer. So there is no surety of information. There is no statistical figures in information driven because of that we not be consider accuracy in information driven. Indeed Data driven approach is most accurate as compare to information driven. In the data driven approach we have the statistical figure which helps to opinion building in accurate way. As data driven meets approximate accuracy which leads to be into good results. nowadays Internet has changed the way people express their views, opinions. It done through twitters tweet, online forums, product review websites portals, social media, etc. Nowadays, millions of people are active at social network sites like Facebook, Twitter, LinkedIn, SnapChat, Instagram etc. to express their emotions, opinion and share views about their daily lives. Through the online communities, we get an data to trending where we get audit tails to meet and get approach to data driven. Social media is producing a large volume of emotion data in the form of tweets, status updates, comments, reviews, etc. Moreover, social media provides an opportunity to construct decision or opinion to meet the goal of data driven approach. Traditionally People mostly depend upon information driven content over the social meet ups for decision making. For e.g. if someone wants to start a tuition classes then they firstly look up for different persons opinions, the person provides in way of information driven, As we

discuss information driven is all about intuition The content generated by information driven for not proper for a normal user for decision making. So there is a need to in way approach of data driven so, emotion analysis used to way to get into data driven.

Emotion analysis tells user whether the statistic about staring tuition class the is satisfactory or not before they start it. We analysis data to understand about their really getting proper return on byte or not. The same way we analyse on tweets of twitter weather users are in favour or not by using treading words

2. EMOTION ANALYSIS

Emotion analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language Processing (NLP). Emotion analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". It's also referred as subjectivity analysis, opinion mining, and appraisal extraction.[4]

The words opinion, sentiment, view and belief are used interchangeably but there are differences between them.[4]

- Opinion:** A conclusion open to dispute (because different experts have different opinions)
- View:** subjective opinion
- Belief:** deliberate acceptance and intellectual assent
- Sentiment:** opinion representing one's feelings

An example for terminologies for Emotion Analysis is as given below,

<SENTENCE> = The story of the movie was weak and boring

<OPINION HOLDER> =<author>

<OBJECT> = <movie>

<FEATURE> = <story>

<OPINION >= <weak><boring>

<POLARITY> = <negative>

Emotion Analysis is a term that include many tasks such as emotion extraction, emotion classification, subjectivity classification, summarization of opinions or opinion spam detection, among others. It aims to analyze people's sentiments,

, attitudes, opinions emotions, etc. towards elements such as, products, individuals, topics ,organizations, and services. [4]

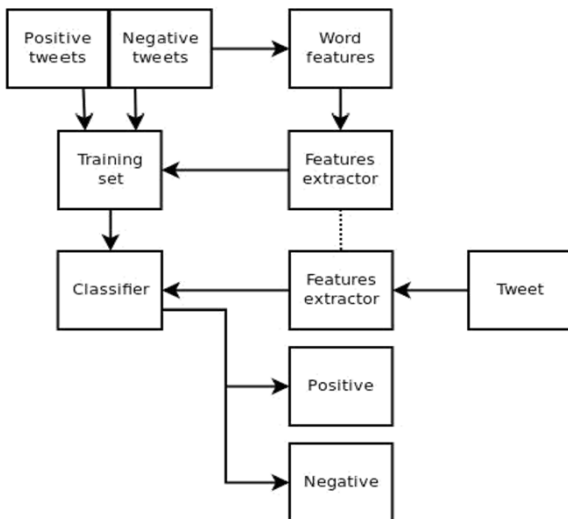


Fig.1. Emotion Analysis Architecture

Following are the phases required for emotion analysis of twitter data[4]

2.1 Amazon Web Services (AWS)

With Amazon Web Services, you pay only for the resources you use. Instead of maintaining a cluster of physical servers and storage devices that are standing by for possible use, you can create resources when you need them. AWS also supports popular tools like Hadoop, and makes it easy to provision, configure, and monitor clusters for running those tools.

Challenges	Amazon web services	Benefits
Data sets can be very large.	Amazon Simple Storage Service	Amazon S3 can store large amounts of data, and its capacity
Storage can become expensive	Amazon Simple Storage Service	feature set to keep its costs low.
Data corruption and loss can	Amazon Simple Storage Service	highly redundant and secure
Maintaining a cluster of physical servers to process data is expensive and time-consuming.	Amazon Elastic Compute Cloud (Amazon EC2)	virtual Amazon EC2 server, you pay for the server only while the application is running, and you can increase the number of servers within minutes, not hours or days to meet the processing needs of your application.

2.2 Features Amazon Web Services (AWS)[2]

- **Create Buckets** – Create and name a bucket that stores data. Buckets are the fundamental container in Amazon S3 for data storage.
- **Store data in Buckets** – Store an infinite amount of data in a bucket. Upload as many objects as you like into an Amazon S3 bucket. Each object can contain up to 5 TB of data. Each object is stored and retrieved using a unique developer-assigned key.
- **Download data** – Download your data or enable others to do so. Download your data any time you like or allow others to do the same.
- **Permissions** – Grant or deny access to others who want to upload or download data into your Amazon S3 bucket. Grant upload and download permissions to three types of users. Authentication mechanisms can help keep data secure from unauthorized access.
- **Standard interfaces** – Use standards-based REST and SOAP interfaces designed to work with any Internet-development toolkit.[2]

2.3 Getting Set Up

Sign Up for Amazon Web Services (AWS) when you create an AWS account will automatically be signed up for all AWS services. You pay only for the services that you use. If you have an AWS account already, skip and go the next step. If you don't have an AWS account, use procedure to create one.

Create a Key Pair

You'll need to create a key pair to connect to Amazon EC2 instances. For security reasons, EC2 instances use public/private key pair, rather than a user name and password, to authenticate connection requests. The public key half of this pair is embedded in the instance, so you can use the private key to log in securely without a password. In this step, you'll use the AWS Management Console to create a key pair. Later, you'll use this key pair to connect to the Amazon EC2 instances that are used in the tutorials.

To generate a key pair

- Open the Amazon EC2 console at <https://console.aws.amazon.com/ec2/>.
- In the top navigation bar, in the region selector, click **Asia Pacific (Sydney)**.
- In the left navigation pane, under **Network and Security**, click **Key Pairs**.
- Click **Create Key Pair**.
- Type **mykeypair** in the new **Key Pair Name** box and then click **Create**.
- Download the private key file, which is named **mykeypair.pem**, and keep it in a safe place. You will need it to access any instances that you launch with this key pair

Important

If you lose the key pair, you cannot connect to your Amazon EC2 instances. Information about key pairs, see *Getting an SSH Key Pair in the Amazon Elastic Compute Cloud User Guide*.

3. APPROACHES FOR EMOTION ANALYSIS

There are mainly two techniques for emotion analysis for the twitter data:

3.1 Machine Learning Approaches

Machine learning based approach uses classification technique to classify text into classes. There are mainly two types of machine learning techniques[4]

3.1.1. Unsupervised learning:

It does not consist of a category and they do not provide with the correct targets at all and therefore rely on clustering.[4]

3.1.2. Supervised learning:

It is based on labeled dataset and thus the labels are provided to the model during the process. These labeled dataset are trained to get meaningful outputs when encountered during decision-making.[4]

The success of both this learning methods is mainly depends on the selection and extraction of the specific set of features used to detect sentiment.[4]

The machine learning approach applicable to emotion analysis mainly belongs to supervised classification. In a machine learning techniques, two sets of data are needed:[4]

1. Training Set
2. Test Set.

3.2 Amazon EMR approach

Sentiment data is often derived from social media services and similar user-generated content, such as reviews, comments, and discussion groups. The data sets thus tend to grow large enough to be considered "big data."

3.2.1. EMR

Amazon EMR integrates open-source data processing frameworks with the full suite of Amazon Web Services. The resulting architecture is scalable, efficient, and ideal for analyzing large-scale sentiment data, such as tweets over a given time period. AWS CloudFormation stack that provides a script for collecting tweets. You'll store the tweets in Amazon S3 and customize a mapper file for use with Amazon EMR. Then you'll create an Amazon EMR cluster that uses a Python natural language toolkit, implemented with a Hadoop streaming job, to classify the data. Finally, you'll examine the output files and evaluate the aggregate sentiment of the tweets.

4. IMPLIMENTAION EMOTION ANALYSIS TASKS

Emotion analysis is a challenging interdisciplinary task which includes natural language processing, web mining and machine learning. It is a complex task and can be decomposed into following tasks, viz:

- a Twitter Developer Account
- Amazon S3 Bucket for the Amazon EMR Files
- Collect and Store the Emotion Data

- Customize the Amazon EMR
- Create an Amazon EMR Cluster
- Examine the Emotion Analysis Output

Twitter Developer Account

Go to <https://dev.twitter.com/user/login> and log in with your Twitter user name and password. If you do not yet have a Twitter account, click the Sign up link that appears under the Username field. If you need to retrieve your Twitter developer credentials at any point, you can go to <https://dev.twitter.com/apps>

Amazon S3 Bucket for the Amazon EMR Files

Amazon EMR jobs typically use Amazon S3 buckets for input and output data files, as well as for any mapper and reducer files that aren't provided by open-source tools

To create an Amazon S3 bucket using the console

- Sign in to the AWS Management Console and open the Amazon S3 console at <https://console.aws.amazon.com/s3/home>.
- Click **Create Bucket**.
- Enter a name for your bucket, such as **mysentimentjob**.
- Leave the **Region** set to **US Standard** and click **Create**.
- Click the name of your new bucket in the **All Buckets** list.
- Click **Create Folder**, then type **input**. Press **Enter** or click the check mark.
- Repeat this step to create another folder called **mapper** at the same level as the input folder.
- Click **Actions**, then click **Make Public**. Click **OK** to confirm that you want to make the folders public.

Note: To meet Hadoop requirements, Amazon S3 bucket names used with Amazon EMR are restricted to lowercase letters, numbers, periods (.), and hyphens (-).

Collect and Store the Emotion Data

AWS CloudFormation template to launch an instance, then use the tools on the instance to collect data via the Twitter API. You'll also use a command-line tool to store the collected data in the Amazon S3 bucket you created.

Customize the Amazon EMR

When you create your own Hadoop streaming programs, you'll need to write mapper and reducer executables as described in Process Data with a Streaming Cluster in the *Amazon Elastic MapReduce Developer Guide*. For this

we've provided a mapper script that you can customize for use with your Twitter search term.

Create an Amazon EMR Cluster

Reflects changes made to the Amazon EMR console in November 2013. If your console screens do not match the images in this guide, switch to the new version by clicking the link that appears at the top of the console. Amazon EMR allows you to configure a cluster with software, bootstrap actions, and work steps. For this tutorial, you'll run a Hadoop streaming program. When you configure a cluster with a Hadoop streaming program in Amazon EMR, you specify a mapper and a reducer, as well as any supporting files. The following list provides a summary of the files you'll use For the mapper, you'll use the file you customized in the preceding step For the reducer method, you'll use the predefined Hadoop package aggregate. For more information about the aggregate package, go to the Hadoop documentation

Examine the Emotion Analysis Output

When your cluster's status in the Amazon EMR console is **Waiting: Waiting after step completed**, you can examine the results.

To examine the streaming program results

- Go to the Amazon S3 console at <https://console.aws.amazon.com/s3/home> and locate the bucket you created in Step 2: Create an Amazon S3 Bucket for the Amazon EMR Files (p. 6). You should see a new output folder in your bucket. You may need to click the refresh arrow in the top right corner to see the new bucket.
- The job output will be split into several files: an empty status file named **_SUCCESS** and several **part-xxxxx** files. The **part-xxxxx** files contain Emotion measurements generated by the Hadoop streaming program.
- To download an output file, select it in the list, then click **Actions** and select **Download**. Right-click the link in the pop-up window to download the file. Repeat this step for each output file.
- Open the files in a text editor. You'll see the total number of positive and negative tweets for your search term, as well as the total number of tweets that did not match any of the positive or negative terms in the classifier (usually because the subject term was in a different field, rather than in the actual text of the tweet).

5. CHALLENGES IN EMOTION ANALYSIS

Emotion Analysis is a very challenging task. Following are some of the challenges faced in Emotion Analysis of Twitter.

1. Identifying subjective parts of text:

Subjective parts represent sentiment-bearing content. The same word can be treated as subjective in one case, or an objective in some other. This makes it difficult to identify the subjective portions of text.

For example:

1. The language of the Mr.Dennis was very crude.
2. Crude oil is obtained by extraction from the sea beds. The word „crude“ is used as an opinion in first example, while it is completely objective in the second example.

2. Domain dependence:

The same sentence or phrase can have different meanings in different domains. For Example, the word „unpredictable“ is positive in the domain of movies, dramas ,etc, but if the same word is used in the context of a vehicle's steering, then it has a negative opinion.

3. Sarcasm Detection:

Sarcastic sentences express negative opinion about a target using positive words in unique way..

6. APPLICATIONS OF EMOTION ANALYSIS

Emotion Analysis has many applications in various Fields.

1.Applications that use Reviewsfrom Websites:

Today Internet has a large collection of reviews and feedbacks on almost everything. This includes product reviews, feedbacks on political issues, comments about services, etc. Thus there is a need for a emotion analysis system that can extract sentiments about a particular product or services. It will help us to automate in provision of feedback or rating for the given product, item, etc. This would serve the needs of both the users and the vendors.[4]

2. Applications as a Sub-component Technology

A emotion predictor system can be helpful in recommender systems as well. The recommender system will not recommend items that receive a lot of negative feedback or fewer ratings.[4]

In online communication, we come across abusive language and other negative elements. These can be detected simply by identifying a highly negative emotion and correspondingly taking action against it.[4]

3.Applications in Business Intelligence

It has been observed that people nowadays tend to look upon reviews of products which are available online before they buy them. And for many businesses, the online opinion decides the success or failure of their product. Thus, Emotion Analysis plays an important role in businesses. Businesses also wish to extract emotion from the online reviews in order to improve their products and in turn their reputation and help in customer satisfaction .[4]

4.Applications across Domains:

Recentresearches in sociology and other fields like medical, sports have also been benefitted by Emotion Analysis that show trends in human emotions especially on social media.[4]

5.Applications In Smart Homes

Smart homes are supposed to be the technology of the future. In future entire homes would be networked and people would be able to control any part of the home using a tablet device. Recently there has been lot of research going on Internet of Things(IoT). Emotion Analysis would also find its way in IoT. Like for example, based on the current emotion or emotion of the user, the home could alter its ambiance to create a soothing and peaceful environment.[4]

Emotion Analysis can also be used in trend prediction. By tracking public views, important data regarding sales trends and customer satisfaction can be extracted.[4]

7. CONCLUSION

In this paper we provide study of e techniques for emotion analysis using Amazon EMR approaches, Collect live data from twitter and analyze it using Elastic Map Reduce. Collect data using Cloud Formation template provided by amazon.

Store the data in S3. Access the server setup above and collect data: as much data about any word of interest to you. Perform MR (mapreduce) analytics

REFERENCES

- [1] A.Pak and P. Paroubek. „Twitter as a Corpus for Emotion Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320-1326
- [2] Amazon Simple Storage Service Developer Guide (API Version 2006-03-01) [AWS Documentation](#) » [Amazon Simple Storage Service \(S3\)](#) » [Developer Guide](#) » Introduction to Amazon S3

-
- [3] Go, R. Bhayani, L.Huang. "Twitter Emotion Classification Using Distant Supervision". Stanford University, Technical Paper, 2009
 - [4] Vishal Kharde and S.S.Sonavane Sentiment "Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11, April 2016
 - [5] Bifet and E. Frank, "Emotion Knowledge Discovery in Twitter Streaming Data", In Proceedings of the 13th International Conference on Discovery Science, Berlin, Germany: Springer, 2010, pp. 1-15.
 - [6] Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Emotion Analysis of Twitter Data", In Proceedings of the ACL 2011 Workshop on Languages in Social Media, 2011, pp. 30-38
 - [7] Dmitry Davidov, Ari Rappoport. "Enhanced Emotion Learning Using Twitter Hashtags and Smileys". Coling 2010: Poster Volume pages 241{249, Beijing, August 2010
 - [8] Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.H., Liu, B.: Predicting flu trends using twitter data. In: 2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 702–707, April 2011
 - [9] Socher, Richard, et al. "Recursive deep models for semantic compositionality over a emotion Treebank." Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2013.
 - [10] Meng, Xinfan, et al. "Cross-lingual mixture model for sentiment classification." Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics Volume 1, 2012
 - [11] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. "Lexicon based methods for emotion analysis". Computational linguistics, 2011:37(2), 267-307.
 - [12] Li, S., Xue, Y., Wang, Z., & Zhou, G. "Active learning for cross-domain emotion classification". In Proceedings of the Twenty-Third International joint conference on Artificial Intelligence (pp. 2127-2133). AAAI Press, 2013
 - [13] Bollegala, D., Weir, D., & Carroll, J.. Cross-Domain Sentiment Classification using a Emotion Sensitive Thesaurus. Knowledge and Data Engineering, IEEE Transactions on, 25(8), 1719-1731, 2013
 - [14] Pang, B. and Lee, L. "A sentimental education: Emotion analysis using subjectivity summarization based on minimum cuts". 42nd Meeting of the Association for Computational Linguistics [C] (ACL-04). 2004, 271-278.
 - [15] R. Parikh and M. Movassate, "Emotion Analysis of User-Generated Twitter Updates using Various Classification Techniques", CS224N Final Report, 2009