**LITERATURE REVIEW**

The process of emotion recognition involves the processing images and detecting the face then extracting the facial feature. Facial Expression Recognition consists of three main steps. In first step face image is acquired and detect the face region from the images and pre-processed the input image to obtain image that have a normalized size or intensity. Next is expression features are extracted from the observed facial image or image sequence.

Then extracted features are given to the classifier and classifier provides the recognized expression as output.

Face Detection and Pre-processing The face detection is the process of extracting the face region from the background. It means to determine the position of the face in the image. This step is require because images having a different scales. Input image having a complex backgrounds and variety of lightning conditions can be also quite confusing in tracking.

Face expression recognition tends to fail if the test image has a different lighting condition than that of the training images. For that facial point can be detected inaccurately for that pre-processing step is required. B.

Feature Extraction And Classification Selecting a set of feature points which represent the important characteristics of the human face. After the face has been located in the image, it can be analysed in terms of facial features. The features measure the certain parts of the face such as eyebrows or mouth corners. Various methods exist which can extract feature for expression based on motion of the feature such Active Appearance Model which is statistical model of shape and gray scale information.

The Features describe the change in face texture when particular action is performed such as wrinkles, bulges, forefront, regions surrounding the mouth and eyes. Image filters are used, applied to either the whole-face or specific regions in a face image to extract a feature vector. Principal Component Analysis, Local Binary Pattern

Fisher’s Linear Discriminator based approaches are the main categories of the approaches available.

After the set of features are extracted from the face region are used in classification stage. The set of features are used to describe the facial expression. Classification requires supervised training, so the training set should consist of labelled data. Once the classifier is trained, it can recognize input images by assigning them a particular class label. The most commonly used facial expressions classification is done both in terms of Action Units, proposed in Facial Action Coding System (FACS) and in terms of six universal emotions: happy, sad, anger, surprise, disgust and fear base systems lie outside of our discussion here.

**Techniques used for emotion recognition**

**A. Principal Component Analysis**

Principal Components Analysis (PCA) ia a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. The facial expression recognition using eigen faces in which PCA is used to extract features from input image. First of all they create training dataset to compare result. Once inputted face image is pre-processed and compare with training dataset which are already computed but based on the idea, they divided the training set into six basic classes according to universal expression(Happy, Surprise, Disgust, sad, Angry, Fear)

**B. Local Binary Pattern**

LBP based feature extraction method is used owing to its excellent light invariance property and low computational complexity. The neighbourhood values are threshold by the centre value and the result is treated as a binary number.

If the canter pixels value is greater than the neighbour’s value write 1, otherwise 0. In this way, it encodes the neighbourhood information very efficiently

**C. Active Appearance Model**

Active Appearance Model (AAM) is a statistical approach for shape and texture modelling and feature extraction. It has been extensively used in computer vision applications. AAM generates statistical appearance models by combining a model of shape variation with a texture variation. So the AAM creates the shape, texture combination model of training facial image sequence “Textures” are pixel intensities of the target image.

**D. Facial Action Coding System (FACS)**

Facial Action Coding System (FACS) was developed by Paul Ekman and Wallace Friesen in 1976 is a system for measuring facial expression. FACS is based on the analysis of the relations between muscle contraction and changes in the face appearance. The Face can be divided into Upper Face and Lower Face Action units. Action Units are changes in the face caused by one muscle or a combination of muscles. There are 46 AUs that represent changes in facial expression and 12 AUs connected with eye gaze direction and head orientation.

**E. Haar Classifier**

Haar classifier based method is chosen for face detection owing to its high detection accuracy and real time performance. Consists of black and white connected rectangles in which the value of the feature is the difference of sum of pixel values in black and white regions. The computational speed of the feature calculation is increased with the use of Integral image

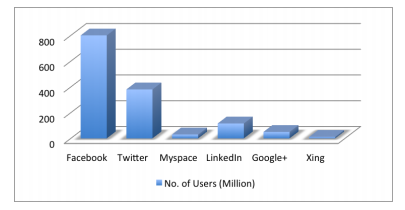
Sentiment analysis aims to identify and extract opinions and attitudes from a given piece of text towards a specific subject. There has been much progress on sentiment analysis of conventional text, which is usually found in open forums, blogs and the typical review channels. However, sentiment analysis of microblogs is considered as a much harder problem due the unique characteristics possessed by microblogs (e.g. short length of status updates and language variations). This report studies existing literature on sentiment analysis of microblogs, raises my research questions, presents the work that have been done in the first year, and finally outlines future plan for the remaining two years.

**The Phenomenon of Microblogs**

Microblogging is a network service, which allows users to post and broadcast messages to other subscribed users of the same service. Microblogging services differ from traditional blogging services in that their posts are brief (typically 140 - 200 characters). The first microblogging service was tumblelogs, which appeared in 2005. Later years have shown a birth of different microblogging websites and services such as Twitter, Tumbler, Jaiku and Pownce (2007) and Plurk (2008). Other social media tools like Facebook, MySpace, LinkedIn, and XING also provide microblogging services, which are known in this case as status updates. Recent statistics as shown in Figure below show that Twitter and Facebook are now considered as the most popular social networks and microblogging services. While Twitter has 200 million users, Facebook has 800 million active users. 600 tweet messages and 700 status updates are sent and published every second.

Twitter is an online microblogging service, which was created in March 2006. It enables users to send and read text-based posts, known as tweets, with the 140-character limit for compatibility with SMS messaging. Twitter allows users to subscribe (called following) to other users’ tweets. A user can forward or retweet other users’ tweets to his followers (e.g. “RT @username [msg]” or “via @username [msg]”).

Facebook is an online social network, launched in February 2004. Once users register with Facebook, they can create their own personal profiles, construct their friendships networks by adding other users as friends, share and exchange short textual updates known as status updates with the 420-character limit. Moreover, users can join groups with common interests. These groups are usually organized by private or public parties.



**Motivation**

The emergence of social media combined with microblogging services’ easy-to-use features have dramatically changed people’s life with more and more people sharing their thoughts, expressing opinions, and seeking for support on such open social and highly connected environments. Monitoring and analysing opinions from social media provides enormous opportunities for both public and private sectors. For private sectors, it has been observed that the reputation of a certain product or company is highly affected by rumours and negative opinions published and shared among users on social networks. Understanding this observation, companies realize that monitoring and detecting public opinions from microblogs leading to building better relationships with their customers, better understanding of their customers’ needs and better response to changes in the market.

For public sectors, recent studies show that there is a strong correlation between activities on social networks and the outcomes of certain political issues. For example, Twitter and Facebook were used to organise demonstrations and build solidarity during Arab Spring of civil uprising in Egypt, Tunisia, and currently in Syria. One week before Egyptian president’s resignation the total rate of tweets about political change in Egypt increased ten-fold. In Syria, the amount of online content produced by opposition groups in Facebook increased dramatically.

Another example is the UK General Election 2010. It has been shown that activities at Twitter are a good predicator of popularities of political parties. Thus tracking and analysing users’ activities on social media are they key to understanding and predicting public opinions towards certain political event.

**3D Problems**

Much work has been done on sentiment analysis. Most of this work focuses on extracting sentiments from text found in traditional online media, such as open forums, blogs and peer-to-peer networks. However, applying previous approaches to detect sentiment from microblogs poses new challenges due to several unique characteristics possessed by microblogs. These challenges can be categorised into three main categories as follows:

**Data**

One common characteristic shared between many microblogging services is the short length of their update messages. While Facebook has a limit of 420 characters for status updates, Twitter has a 140-character limit. Another characteristic is language variations. Users post from different media including their personal laptops, cell phones and tablets in such, they tend to use a large variety of short forms and irregular words. These two characteristics induce significant data sparseness and thus affect the performance of typical sentiment classifiers learned from such noisy data.

**Domain**

Microblogs like Twitter and Facebook are open social environments where there are no restrictions on what users can tweet about and in which domain. This differs from previous work on sentiment analysis, which focused on a specific domain of interest such as product reviews. Most of previous work on microblog sentiment analysis follows the supervised machine learning approaches (see Chapter 2 for a literature survey). However, supervised classifiers require training labelled data which is impractical and time-consuming to get. Also models trained on one domain might face a serious loss in performance when shifting to another domain. To overcome these drawbacks, an automated method using emoticons (called distant supervision) was proposed. However, learning sentiment classifiers from noisy labels may hinder the overall performance. Also it is not possible to capture the sentiment of instances (e.g. tweets) with no associated emoticons. Moreover, different emotions maybe associated with the same instance, which also makes distant supervision infeasible process.

**Dynamics**

Microblogging services in general operate on the data stream fashion where data is transferred, viewed and discarded immediately. This raises new problems for sentiment classification. First, classifiers should work with limited resources of time and space. Second, the dynamic nature of the data means that we need to deal with imbalanced classes where training instances in some classes are significantly less than those in other classes. For example, a training corpus may contain more positive tweet messages then negatives ones. This also differs from the previous work which assumes the balance between negative and positive instances.

**Research Objectives**

It can be realized from the aforementioned problems that sentiment analysis of Microblogs faces the following challenges:

• The short length of status updates coupled with their noisy nature makes the data very sparse to analyses using standard machine learning classifiers.

• The lack of labelled data needed for classifiers training.

• Open nature of microblogs poses an open-domain problem where classifiers should work

in a multi-domain environment.

• The streaming fashion of microblogs where data arrives at a high speed. This means

data should be processed in real time and classifiers should adapt quickly with the newly

arrived data.

Thus, the main research question of my research can be formulated as follows: “How to build a domain-independent sentiment classifier learned from short textual sparse data, which is able to operates in a streaming fashion, and adapt dynamically to the new data.”

We can look at our research problem from a software engineering perspective, where the research question here can serve as the main functional requirement of the system. This actually helps us to frame our work and formulate our research objectives as follows:

**Obj1.** Data sparsity should be alleviated; this implies that data should be pre-processed before it is getting fed into classifier training.

**Obj2.** Sentiment classifiers should be trained in a domain-independent way. In simple words, they should be able to provide a similar performance when a domain shift occurs.

**Obj3.** Sentiment classifiers should be able to operate on the data streaming paradigm of microblogs. This means they should have the ability to work with limited resources of time and space.

**Obj4.** The problem of imbalanced sentiment distribution (sentiment drift) should be considered when building sentiment classifiers. This means classifiers are expected to work with imbalanced numbers of training instances in different classes.

**Obj5.** Classifiers should be easily adapted to work with different microblogging services. In my study here, Twitter was used as a case study. The reasons behind this choice are:

(1) Twitter has been used as case study by almost all previous work on sentiments analysis of microblogs. Thus conducting our experimental work on tweets data allows us to compare our approach to existing approaches.

(2) Twitter is more flexible in its privacy policy than other microblogging services. For example, Twitter provides a set of APIs, which makes collecting and analyzing data an easier process compared to other microblogging services such as Facebook.

It is worth mentioning that the approaches that will be proposed in my study are equally applicable to other microblogging services since the problems to be addressed are common to most other microblogging data.