

Sentiment Analysis for Visuals using Natural Language Processing

Hari Iyer
Pursuing B.E.,
Department of
Computer Engineering,
Dwarkadas J. Sanghvi
College of Engineering
Mumbai, India

Mihir Gandhi
Pursuing B.E.,
Department of
Computer Engineering,
Dwarkadas J. Sanghvi
College of Engineering
Mumbai, India

Sindhu Nair
Assistant Professor,
Department of
Computer Engineering,
Dwarkadas J. Sanghvi
College of Engineering,
Mumbai, India

ABSTRACT

In this paper, the aim is to build a hybrid Word Sense Disambiguation(WSD) technique, which is acutely focused on text associated with a certain form of visual. Natural language processing helps establish a context among the data elements that are aggregated to establish a certain meaning. Analyzing transcripts of visuals being uploaded in real-time saves resources and time required to sort content based on genres or emotions. The training data lays a foundation to rate the polarities of elements, on top of which the dictionary expands as an when new content is supplied to the apparatus. Third-party intelligence is combined with the dictionary to experience growth even when the consumer usage is idle. All these entities are mutually intertwined to ensure maximum utility and output.

General Terms

NLP, WSD.

Keywords

Natural Language Processing, Third-party intelligence, Training Data, Polarity, Word Sense Disambiguation.

1. INTRODUCTION

The past decade saw an explosion of visual content on the web. Popular databases like YouTube, DailyMotion, etc. experience continuous uploading of video data all over the world. It is a resource consuming task to analyze this data for sentiments[10]. The distinguishing feature of paper-based knowledge as opposed to audio is representation of visual data. The path to bridge this gap is to analyze the emotion using a dynamic approach (possibly WSD)[2]. There are softwares to accomplish this goal, like TalkWalker, Semantria, and Salesforce Radian6*. These softwares interpret text word by word, and wrap them into windows to prepare runs of queries, as many as possible, in order to satisfy all the possibilities of the intended message to be conveyed. The video content can be converted into suitable subtitle format using Automatic Speech Recognition (ASR)[6]. Expressions will be traversed from left to right and after each jump, the corresponding word will be analyzed for their polarity. For instance, in “Wonderful birds exist!”, the word “Wonderful” will be encountered first. The back-end dictionary will be looked up for the element to find out the sentiment score of the former is a hit. In failure of which, the entire process is recursed in anticipation of a hit. Such figures are calculated for sentences as a whole, which gives an idea of the sentiment, opinion, and emotion. This concept, if applied to text that is mapped with visual data like video frames and static images, standards can be set for the content that is approved to be uploaded onto the database.

With this approach, a lot of resource time can be saved and can be put to other applications. Apart from sentiments, this process can cater to e-commerce markets to analyze product utility, cost, and recognition can be carried out using attributes. There are two types of sentiment analysis techniques, supervised and unsupervised. In supervised analysis, each and every video feature is attributed to labels used to distinguish frames as basic elements to operate upon. On the contrary, unsupervised analysis uses a chunk of text describing the visual at length. Words and phrases all over the text need to be pooled together to make sense of the video as a single entity. In this paper, it can be seen how NLP techniques[1] can be used to achieve this goal.

*www.socialmediaexaminer.com/6-social-media-monitoring-tools/

2. DATA SURVEY AND RELATED WORK

2.1 Content genre

All the media formats that can be analyzed for their transcripts are eligible for this operation. This covers a wide spectrum of plausible use cases across domains. Popular media formats include mp3, 3gp, and mp4, which can be made to support ASR, so that the data contained can be mined for keys to be analyzed for polarities. This is a generic technique, regardless of the existing data format and its semantic contents. The principle juxtaposing these different kinds of electronic information format is the association between frames/instances and the labels of the elementary units, which is assumed to be the fundamental entities of the raw data assembly.

2.2 Existing approaches

YouTube, a service offered by Google Inc. has an inbuilt video to subtitle converter which is used to interpret video conversations without hearing dependencies. Such techniques are employed to accomplish the first step of extracting text based format of the visual. This text is now to be mined for target points out of which some sense can be developed and interpreted. Existing intelligence analyzing tools include PaperRater, EasyBib, and WordNet.** There are open source, free and commercial variants of such tools. Summarization of text using such resources also reduces the workload of the next operational functionality elements to follow. Also, the execution time is reduced considerably.

The selected chunks of text now pass through a series of filters before the final sentiment analysis algorithm takes charge of the situation. It includes the syntactic and stylistic analysis of the data. Syntax implies the expression's

structural science and it's linguistic category. On the other hand, a stylistic analysis includes the subjective perception understood from the tone and accent of the extracted media-text abstract[4].

Thus, the sentimental elements are analyzed, provided the above steps are executed in a systematic manner.

**<http://www.thetoptens.com/best-alternatives-youtube/>

3. THE BACK-END DICTIONARY

Predefined data needs to be stored on the system, which can act as a reference for adjudicating polarities of hybrid phrases and expressions. There are two approaches to this problem, static and adaptive, with the latter being more productive and preferred. The following features demonstrate the working of these techniques.

3.1 Structure

The database initially contains the experimental data that is a base for polarity rating purposes. Although the database client may vary from instance to instance, the methodology of storing data stands uniform. The following schema indicates the proposed system for content storage:

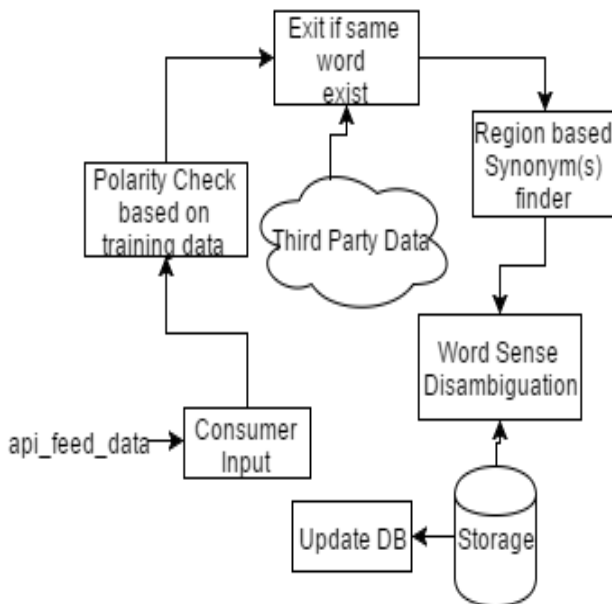


Figure 1: Structural Data Flow

3.2 Adaptiveness

This phenomenon plays a very important role in improving the overall efficiency of understanding, memory utilization, use-case orientation, and response-time with respect to database administration. Considering training data as a platform, the dictionary must grow mainly as a consequence of usage. This improves the chance of finding an existing match faster to the forthcoming query requests. Shifting focus from data management to semantics, understanding and linking terms can leverage the existing data chunk to predict and integrate future possibilities. Homogeneous content, if clustered logically, promotes parallel updation, even when the client side usage is suspended. This application is extended to region-specific requests, which pave the way for Geo-distributed and language-dependent interpretation[#]. A pronunciation-independent system is achieved here, which makes global utility possible. Multiple runs of queries, that are received in the form of windows,

may contain an idiom or phrase as a whole[7]. Thus, an adaptive dictionary can help recognize such cases and the polarity is be calculated based on specific entries of a particular word, since there may be homonyms. A major part of these adaptivity prerequisites is fulfilled by external resources, which is discussed in the forthcoming point.

[#]<https://support.google.com/webmasters/answer/6144055?hl=en>

3.3 Third-party data integration

Big Data, which is spread across enterprises and open-source initiatives across domains, is used as an ore for mining out relevant data combined with capabilities of powerful utilities. The structured chunk of data is used combined with tools that can interpret that data, and a connection is established between the dictionary-database and foreign data fetched. Word2vec^{##} is a program that converts text data into word vectors. This mathematical approach stands useful while deciding polarity rating, since context is retained even after instance-based operations are carried out. References can be levied onto the result by magnifying or compressing the base context of the incoming remote data. The following illustration explains the working.

Example:

$$\text{vec}(\text{"Tiger"}) = 0.6575$$

$$\text{vec}(\text{"Cub"}) = 0.3825$$

$$\text{vec}(\text{"Cow"}) = 0.1342$$

A tiger's small one is known as a cub, and a cow's counterpart is to be found. So, the following operation is used to determine a magnitude, which contextually refers to that of the word "calf".

$$\begin{aligned} &\text{vec}(\text{"Tiger"}) - \text{vec}(\text{"Cub"}) + \text{vec}(\text{"Cow"}) \\ &= 0.6575 - 0.3825 + 0.1342 = 0.4092 \\ &\approx \text{vec}(\text{"Calf"}) \end{aligned}$$

Social media is an exploding space of sentimental data, wherein emotions are expressed on a large scale. Thus, the above mentioned approach can be applied on a global basis for expression sentiment analysis. The data, being open source, can be pulled into the system using Application Programming Interfaces (API).

^{##}Word2vec: <https://code.google.com/p/word2vec/>

4. SENTIMENT DETECTION

A hybrid approach to sentiment analysis, if designed helps to pool in multiple existing solutions to adjunct various entities of the process. The broad classification of emotion scrutiny methodologies branch out as lexicon and machine learning based designs. A layer, which has machine intelligence as the base, a robust and self-prospering dictionary on top of it, combined with Big Data capabilities, provides a platform for writing flexible and dependable algorithms for data interpretation and approval, based on smart sensing of the content at the first encounter itself. This paves the way for a swift and a automated resource that serves enterprise applications across domains.

4.1 Data collection and training

Research has proved that a language's meaning and context interpretation is closely attainable using mathematical processes. As a prerequisite for this approach, a large chunk

of raw data is analyzed for the polarity of the words contained in it. Polarity rating references are preset, 1 and -1 being the most positive and negative words respectively. The data under consideration is pulled from open source real time social networking databases. Although machine trains the data in the long run, human analysis is required in the early stages to provide a reference regarding polar inclination of elements. For instance, reviews have become a part and parcel of online forums, blogs, and media. The human will be given data across the spectrum for inspection, so that a wide range of words will be covered [5]. Video reviews to product ratings, consumer feedback to literature: all the zones are attributed to sentiments. Also, free online dictionary data can be fed in for registration.

The work done in this phase finds application in the following algorithm.

4.2 The Algorithm

A stratified bottom-up approach is adopted in order to enable cost and space based testing between state transitions. The algorithm, layer-by-layer can be stated as follows:

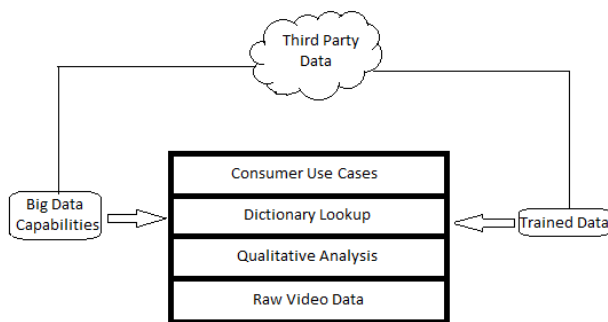


Figure 2: Hybrid Sentiment Analysis Model

4.2.1 Raw video data

Visuals, being the elementary form of data as an input to the arrangement, are captured on a real-time basis. Collection of frames is stored in a buffer, and breakpoints are calculated as stated:

// Called by default to initialize date elements:

```

__constructor()
{
    /* Keep track of blanks vs fills: */
    counter_i = 0, counter_j = 0;
    /* Initialize buffer with visual frames: */
    buffer [size] = VIDEO_DATA;
    api_feed_data = "";
    /* Initialize API for speech to text: */
    Annyang an = new Annyang();
}
// Constructs transcript of video data:
__append(counter_i)
{
    /* Convert and append most recent block of data:
    */

```

```

while(counter_j <= counter_i){
    /* Calling routine calc() on Annyang object */
    api_feed_data + = an.calc(buff
[counter_j]);
    counter_j++;
}
}
// Expression separator:
__segregation(counter_i)
{
    /* Check for line break. */
    if(buffer [counter_i] == blank_note)
        __append(counter_i);
    if(!end_of_video_frames)
        __segregation(++i);
}
main()
{
    /* Execution starts here.*/
    __segregation(counter_i);
    return;
}

```

Thus, the algorithm works to convert the raw video data to text. Next, the data is analyzed for its semantics, language, and meaning.

\$Annyang Library: <https://www.talater.com/annyang/>

4.2.2 Qualitative analysis

The text fetched from the visual, being unsupervised, contains emotions, meaning, and the context in a hybrid format. It is now to be analyzed for some sense, a flow of content, and catchwords from the transcript that will help summarize the bulky text into a short, but precise version of the visual. There are many text evaluating tools in the market, but what matters more than the quality of the tool, is the mechanism in place to send and receive the elementary data format and to link two adjacent request instances. The cognitive workload of the system needs to be acutely focused on phrases and terms that reveal out a sentiment. Using heuristics, a new approach to minimize the dictionary's scope that needs to be looked up for a particular routine call. A virtual map of the traversal plan is formulated and updated based on the probability of each use cases' occurrence with respect to history. The working is as follows:

// Server side execution of the following algorithm:

// Called by default to initialize date elements:

```

__constructor()
{
    /* Initialize map as a single dimension array: */
    map[SIZE] = {0};

```

```

data: /*
    /* Overall sentiment of an expression and history
    current_exp_rating = 0;
    /* Word by word traversing buffer: */
    logger[SIZE] = {0};
}
// Develops the overall sentiment score container for
//operations.
__mapmaker()
{
    i = 0;
    /* explode() separates words using space as
    reference.*/
    __toArray = explode(api_feed_data, " ");
    while (__toArray(i))
    {
        logger[i] =
        polarity(toArray[i]);
        if(end_of_line)
            logger[i+1] = "";
        i++;
    } // End loop
    i=0, j=0, k=0;
    /* The below loop runs for various combinations of
    logger[], and the training data consists of decisive words on
    which nouns depend for their polarities. */
    while (logger[i]){
        /* Compress expression stats: */
        if (logger[i] == ""){
            while (k<i){
                /* Calculate aggregate polarity[9].*/
                map[j][j] =  $\sum$  logger[k..i-1];
                map[j++] /= length(k,i-1);
            } // End loop
            k = i+1;
        } // End If
        i++;
    } // End loop
}

```

The above operation gives us the overall polarity both, on a word and expression basis. This hierarchy is useful for cross verification of the sentiments, which will be associated in the forthcoming steps.

4.2.3 Dictionary lookup:

There are two predominant cases where the dictionary data is used for interpretation. They are explained as follows:

- logger[] entry:
 - Here, the polarity condition is checked to rate the word as positive(1) or negative(0). The dictionary table for words ranges from 0 to 1, because of which direct matching is possible. The polarity of these words individually may hamper the overall expression sense[3]. The mean value of isolated sentences is calculated, as the algorithm stated.
- Expression placement:
 - When each and every word's polarity rating has been retrieved, an aggregate is calculated in order to minimize the context ambiguity of two consecutive words that may strengthen or weaken a sentiment. Word sense disambiguation (WSD) is thereby given a modified approach, which comes as a part of natural language processing[6]. Here, the aggregate polarity value is looked up in the table, and the sentence is assigned the emotion associated with the polarity value.

An example of the above implementation is shown below:

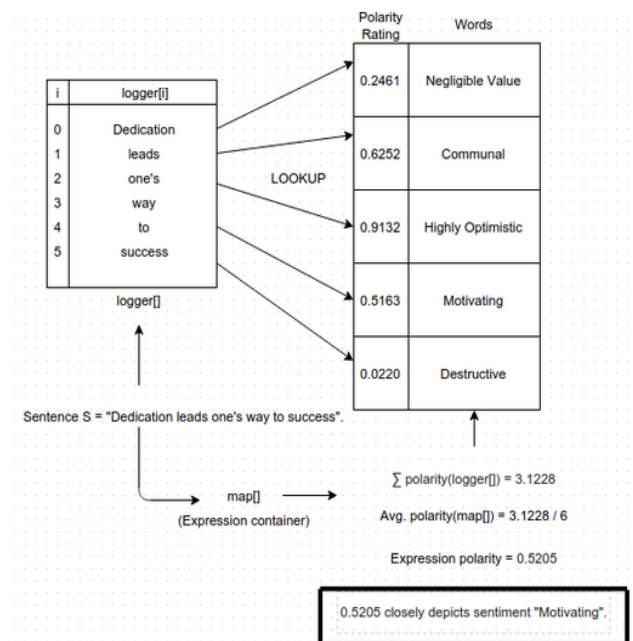


Figure 3: Dictionary Lookup

4.2.4 Consumer use cases

The cognitive computing protocol stack shown in figure 1 mentions consumer use cases as the topmost layer. This is because, all the applications and real-world design of enterprise utilities needs to have all the functionalities that are embedded in the system. Appdata^{\$\$} that is capable of evolving into an ecosystem of it's own is the perfect fit as an input to the arrangement. Customer surveys and feedbacks too play an important role in the adaptivity constraints to be updated or appended.

^{\$\$}<https://www.youtube.com/yt/press/statistics.html>

5 CONCLUSION AND FUTURE WORK

An amalgamated approach is proposed, to analyze sentiments of visual data, which helps in the automation of real-time data segregation or decision making with respect to the data being uploaded. Quicker access, easy interpretation, and platform-independence are the goals of this concept. The algorithm will be tested and applied to various open-source knowledge graphs, and a schema will be provided to the approved data chunk. The Blind's Eye, an e-learning platform for the blind, is willing to implement this model for providing a larger scope of data access for the student use cases. Thus, for a wide range of applications, the designed algorithm provides leverage for optimization.

6 REFERENCES

- [1] Manisha Kanakaraj and Ram Mohana Reddy Guddeti, "Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis using NLP Techniques," in IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015), February 7-9, 2015.
- [2] M. Rajani Shree and Dr. Shanbhavi B.R, "Performance Comparison of Word Sense Disambiguation Approaches for Indian Languages", in IEEE International Advance Computing Conference (IACC), 2015.
- [3] Lakshmish Kaushik, Abhijeet Sangwan and John H. L. Hansen, "Automatic Sentiment Extraction from Youtube Videos", in IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2013.
- [4] Ms Kranti Ghag and Dr. Ketan Shah, "Comparative Analysis of the Techniques for Sentiment Analysis", at ICATE 2013.
- [5] Jose Costa Pereira, Jordi Luque and Xavier Anuera, "Sentiment Retrieval on Web Reviews Using Spontaneous Natural Speech", in the 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP).
- [6] Ankit Kumar, Mohit Dua, Arun Choudhary, "Implementation and performance evaluation of continuous Hindi speech recognition", at the Electronics and communication systems (ICECS), 2014.
- [7] Zohreh Madhoushi, Abdul Razak Hamdan, Suhaila Zainudin, "Sentiment Analysis Techniques in Recent Works", at the Science and Information Journal in 2015.
- [8] Haruna Isah, Paul Trundle, Daniel Neagu, "Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis", at the Computational Intelligence (UKCI), 2014 14th UK Workshop.
- [9] Farhan Hassan Khan, Usman Qamar, M.Younus Javed, "SentiView: A Visual Sentiment Analysis Framework" at the International Conference on Information Society (i-Society 2014).
- [10] Monisha Kanakaraj , Sowmya Kamath S, "NLP based Intelligent News Search Engine using Information Extraction from e-Newspapers" at the 2014 IEEE International Conference on Computational Intelligence and Computing Research.