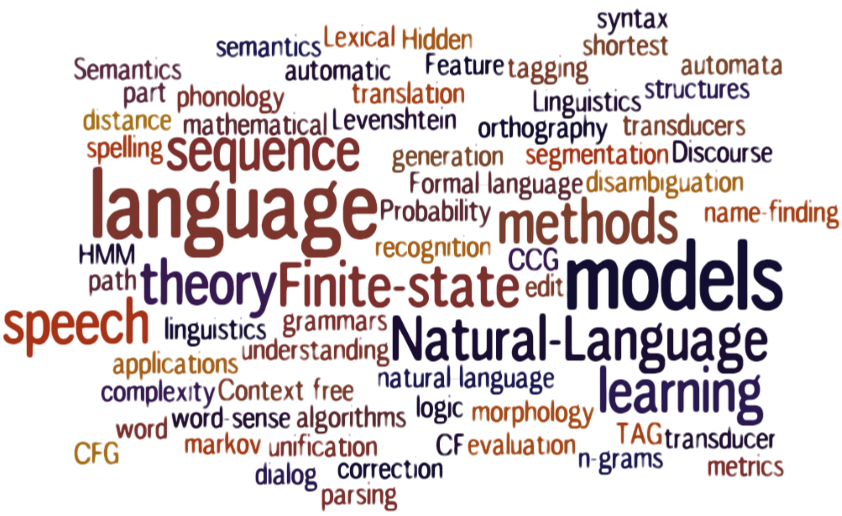
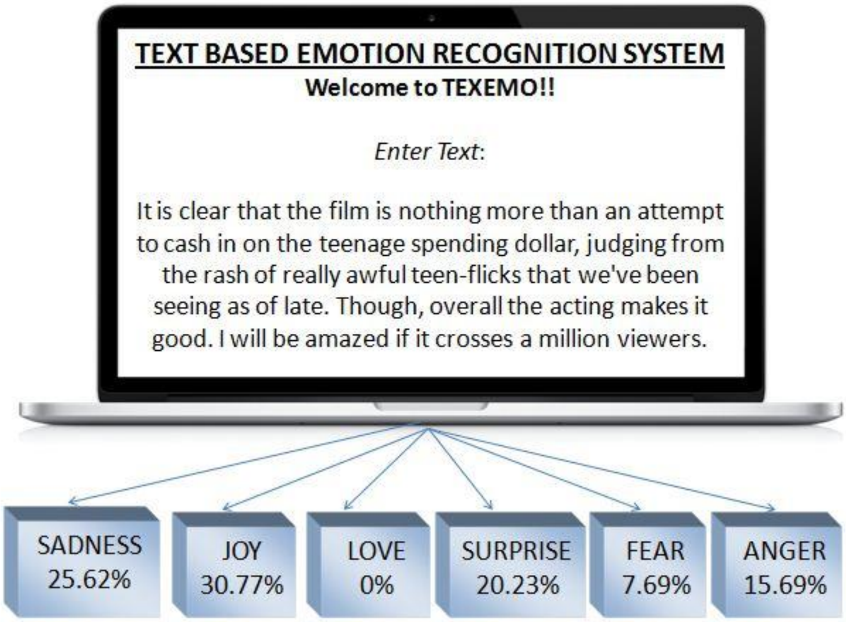
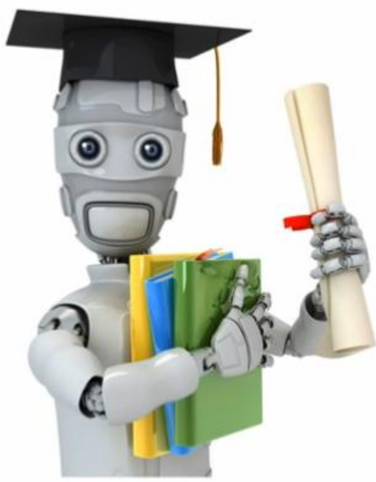
CMU NITK WINTER SCHOOL 2014
MULTIMEDIA ANALYSIS TRACK
PROJECT TEXEMO







PROJECT TEXEMO:

DHRUV GOEL

SATISH PALANIAPPAN

SKAND ARORA

VISION

In this era of artificial intelligence, human beings are designing machines that can perform any general task that humans are expected to perform. We are making machines that are intelligent and smart just like human beings are. However, one thing that separates machines from human beings is their ability to understand and identify various emotions. Despite so much progress in the field of machine learning and natural language processing, machines have still not become capable of detecting emotions clearly.

The most understandable forms through which emotions can be recognized are speech and visual imagery. A person's way of speaking, his/her tone, facial expressions and gestures are the best forms of expressing emotions. A lot work is being done in recognizing emotions through these modalities, yet we haven't come close to developing machines that can detect emotions successfully.

Most of the information in this internet era exists in the form of text. Be it digital encyclopaedias, websites, search engines; text constitutes around 70% of the information available to us. Hence, our aim is to make machines detect emotions hidden in every piece of textual information available.

OBJECTIVE

Emotions are universal and extend beyond boundaries of language, literature, religion, age, etc. Communicating with people is not just about transmitting information or a message but also expressing your emotions. Through our project we aim to make machines detect emotions from any form of text, which constitutes about 70% of information available to us. Given any form of text, the machine will identify the specific emotions of happiness, sadness, anger, surprise, fear or love that the text expresses. In a world full of blog posts, tweets and emails, the implications for businesses to be able to identify emotions contained in written communications are clear: a greater understanding of their users' behaviours and their desires. Understanding emotions exposes us to an array of possibilities: personalized information generation, like advertisements, search results and so on, development of powerful human-computer interaction machines and evolution of more intuitive and emotionally characterized text to speech systems.

LITERATURE REVIEW

The field of emotion recognition has attracted many researchers from computer science, psychology, cognitive sciences, linguistics, and so on. Emotions can be expressed by a person through facial expressions, speech, gestures, and textual conversations. Speech and facial recognition have been the most preferred modalities for researchers all around the world for detecting emotions as they convey emotion more clearly without any disambiguates. Comparatively, very less work has been done in the field of emotion recognition through text. Within text, considerable amount of work has been done in sentiment analysis of movie reviews, twitter feeds and so on. However, nothing significant has been achieved in detecting raw emotions of happiness, sadness, anger, fear, etc in text. The main reason for this is that there are no standard classifications of all human emotions due to complex nature of human minds and any emotion classifications can be seen as 'labels' annotated afterwards for different purposes. Since there is not any standard emotion-word hierarchy, focus is on related research about emotion in cognitive psychology domain.

Researchers have used techniques such as keyword spotting, assigning probabilistic affinities to various emotions, use of learning based methods and knowledge based artificial neural networks [1], rule based system to extract semantics related to specific emotions [2] , detecting emotions based on the cause triggering them [3] and development of situational personalized emotion model [4]. However, all the present researches do not take into account explicit emotions and analyze only a specific part of the psychological reason that leads to emotion in texts. They have not analyzed lack of keywords, disambiguate, lack of linguistic information and do not take the context into account. The corpus used in the research has been specifically developed such that they suit the main theme of the paper. None of the systems can produce satisfactory results for texts with disambiguates or lack of linguistic information.

INTRODUCTION

Psychologists all around the world have been trying to identify a standard way of classifying emotions. However, till now no one has been able to answer the question "How many different emotions we have?" A reason for why it is difficult is because our experiences are so complex and involve so many different factors, so distinguishing one emotion from another is a lot like drawing lines of sand in the desert. It can be hard to determine where one emotion ends or another begins. Even when we analyze a simple emotion like "happiness" or "anger," we know from everyday experience that these emotions come in many different degrees, qualities, and intensities. In addition, our experiences are often comprised of multiple emotions at once, which adds another dimension of complexity to our emotional experience.

Paul Ekman and his colleagues conducted a cross-cultural study in 1972, in which they concluded that the six basic emotions are anger, disgust, fear, happiness, sadness, and surprise. Ekman explains that there are particular characteristics attached to each of these emotions, allowing them to be expressed in varying degrees. Each emotion therefore acts as a discrete category rather than an individual emotional state. Sylvan Tomkins proposed that there were eight basic emotions: interest, surprise, enjoyment, distress, fear, shame, contempt, and anger. He proposed that these were innate and expressed through physical reactions such as facial expressions. Carroll Izard proposed that there were ten basic emotions: interest, joy, surprise, distress, anger, disgust, contempt, fear, shame and guilt. Each emotion has its own neural network in the brain, and corresponding behavioural response. Fischer, Shaver and Carnochan proposed that there was a hierarchy of emotions. They contended that we have two positive emotions - love and joy, three negative emotions - anger, sadness, and fear. Our other emotions stem from these basic emotions.

We consider 6 basic emotions in our system, namely: Joy, Sadness, Love, Anger, Fear and Surprise. After analyzing all the different models proposed by psychologists we came to a conclusion that these are the most basic emotions which compose/combine to form what we call advanced emotions. For example: the emotion of excitement can be a combination of joy and surprise whereas the emotion of disgust can be a combination of anger and sadness.

In our project we develop a hybrid system which combines a rule based approach and a learning based approach. We have used a number of techniques to achieve our objective. It is important to take syntactic structure of the words into account while detecting emotions from a sentence. For this purpose Continuous Bag of Words (CBOW) and Skip-Gram architecture models are most commonly used. The bag of words model is a simplifying representation model in which a text is represented as the bag of its words disregarding grammar and even word order but keeping multiplicity. An n-gram model is a type of probabilistic language model for predicting the next item in such a sequence in the form of an $(n-1)^{th}$ order Markov model.

Consider the sentence "Hi Fred how was the pizza?"

Continuos bag of words: 3-grams {"Hi Fred how", "Fred how was", "how was the”, ...}
Skip-gram 1-skip 3-grams: {"Hi Fred how", "Hi Fred was", "Fred how was", "Fred how the", ...}

We use the concept of word2vec[5]. The word2vec tool developed by Google provides an efficient implementation of the continuous bag-of-words and skip-gram architectures for computing vector representations of words. These representations can be subsequently used in many natural language processing applications and for further research. The word2vec tool takes a text corpus as input and produces the word vectors as output. It first constructs a vocabulary from the training text data and then learns vector representation of words. The word2vec tool enables us to find words that are semantically close to each other. The cosine distance between various vectors is calculated and the vectors with minimum cosine distances between them are said to be semantically similar.

Given any sentence, the first task is to consider individual words to analyze the emotional content. This is achieved by a process called tokenizing. We consider the tokens in our case to be separated by whitespace characters, such as a space or line break, or by punctuation characters.

Often we have words that convey the same meaning and emotion but are expressed in different forms. For example, in English, the verb 'to walk' may appear as 'walk', 'walked', 'walks', 'walking'. The base form, 'walk', that one might look up in a dictionary, is called the lemma for the word. So, we need to perform an operation called lemmatization to consider only the lemma for the word.

On analysis, we found that certain parts of speech contribute majorly to the emotion expressed. Hence, by using a tool called POS (Part of Speech) Tagger we can group the given words in a sentence into their respective parts of speech. On POS Tagging, the words get grouped into various categories as shown in Fig 1. We identified categories which contribute and those which do not contribute to expressing emotions in text. These are shown in Fig 2.

The words finally obtained can be converted into vector form and a supervised learning based system is developed which analyzes the composition of various emotions in the given text and the emotion with highest composition is considered to be the overall emotion expressed by the text.

POS Tag	Description	Example
CC	coordinating conjunction	and
CD	cardinal number	1, third
DT	determiner	the
EX	existential there	*there* is
FW	foreign word	d'hoevre
IN	preposition/subordinating conjunction	in, of, like
JJ	adjective	big
JJR	adjective, comparative	bigger
JJS	adjective, superlative	biggest
LS	list marker	1)
MD	modal	could, will
NN	noun, singular or mass	door
NNS	noun plural	doors
NNP	proper noun, singular	John
NNPS	proper noun, plural	Vikings
PDT	predeterminer	*both* the boys
POS	possessive ending	friend's
PRP	personal pronoun	I, he, it
PRP$	possessive pronoun	my, his
RB	adverb	however, usually, naturally, here, good
RBR	adverb, comparative	better
RBS	adverb, superlative	best
RP	particle	give up
TO	to	*to* go, *to* him
UH	interjection	uhhuhhuhh
VB	verb, base form	take
VBD	verb, past tense	took
VBG	verb, gerund/present participle	taking
VBN	verb, past participle	taken
VBP	verb, sing. present, non-3d	take
VBZ	verb, 3rd person sing. present	takes
WDT	wh-determiner	which
WP	wh-pronoun	who, what
WP$	possessive wh-pronoun	whose
WRB	wh-abverb	where, when

Figure 1: POS Tags

PARTS OF SPEECH TO BE CONSIDERED	**PARTS OF SPEECH NOT TO BE CONSIDERED**
Adjectives-(superlative, comparative)	Cardinal Numbers
Adverbs-(superlative, comparative)	Determiners
Verbs (root verbs)	Foreign Words
Personal/Possessive Pronouns	Preposition
Particles (eg: give *up*, broke *down)*	List marker
Interjections (Alas, Hooray)	Modal
Punctuations	Pre-determiner
Conjunctions	Proper Nouns
Nouns	Stop Words

Figure 2: Parts of Speech

IMPLEMENTATION

Given any input text or the text our system crawls through the web and finds out, we intend to derive the composition of different emotions in the text. For this we pre-process the text before proceeding, After a thorough analysis of what contributed to the emotion that was expressed by a piece of text, as mentioned above we found out that not all words contribute equally to the emotion distribution of the sentence. We tokenized and lemmatized the given text; POS tagged the individual words and removed certain words known as stop words that didn't contribute to the emotion of a sentence. We then classified these words in to 3 different classes:-

Primary Class: Verb, Adjective, Adverb

Secondary Class: Conjunctions, Punctuations, Pronouns

Tertiary Class: Nouns (Name, Place, Animal, Thing).

Primary class contributes more to the emotion when compared to the Secondary Class whereas Tertiary class rarely contributes to the emotion.

After classifying these words into different classes we represent these words in vector forms using the Word2Vec tool which maps the fed words on to a 300 dimension vector space such that semantically equivalent words get closer vector representations in that space. For example, for a word like "Joy" the Word2Vec tool makes sure all semantically equivalent words to "Joy" like Happy, Smile, Excitement, etc. get closer vector representations.

We used three different methods in trying to classify the text in to emotions:

1. **N-Gram Probabilistic Model and Naive Bayes Classifier**: Where we took into consideration the probability distributions of unigrams in a given emotionally tagged phrase and then this information was used to train the Naive Bayes Classifier. Further this work is to be extended to bigrams and trigrams too.
2. **K-means and SVM**: We had 3million word vectors obtained from Word2Vec trained dataset and these vectors were clustered in to groups using K-Means clustering algorithm. Once we got these clusters, we took every annotated sentence in our dataset and built a histogram over these clusters on them. With these histograms as feature vectors for those annotated sentence we trained a Multiclass SVM using RBF kernel over it.
3. **LDA and SVM**: We tried to use the LDA to derive topics out of the given emotional text, we fed about 200k emotional documents in to the LDA and made it learn topics from those documents, which are eventually related to emotions. Then we took a sentence and found out the topic distribution over that sentence for the topics that we discovered. These topic distribution vectors were used as feature vectors for every annotated sentence and we trained a Multiclass SVM using RBF kernel over it.

RESULTS

We have got varied results for each approach, we had the Naive Bayes Classifier working pretty well in detecting joy with an accuracy of 80% but at the same time it was poor for detecting anger. Likewise The K-Means and SVM did well in detecting anger and fear as an emotion. The results thus obtained were as follows:

EMOTION	TRAINING DATA SIZE	TESTING DATA SIZE	ACCURACY
JOY	1107	540	76.59%
DISGUST	611	540	16.60%
SADNESS	821	540	36.67%
ANGER	903	540	27.60%
GUILT	1075	540	60%
FEAR	741	540	40.87%

Figure 3: Results for Naive Bayes Classifier

EMOTION	% DATASET 1	% DATASET 2	After POS (SET2)
Joy	80.49	69.64	59.87
Anger	24.11	23.01	32.55
Disgust	18.41	22.20	19.20
Sadness	23.40	23.21	30.98
Fear	28.11	36.26	33.64
Surprise	14.44	17.73	24.52

Figure 4: Results for K-means and SVM Classifier

FUTURE WORK

The present research was performed on structured language. In our daily life, most of our communication involves use of short forms especially on social media platforms which is filled with unstructured language. At present, our system cannot recognize words like LOL, ROFL or short forms like ‘Plz’. Hence, our first step will be to make our system compatible with unstructured language.

We plan to make a web/mobile application- TEXEMO which will be based on this system. It will be a fun based application that will enable users to input any text and find out the emotion/ emotion distribution that the text input contains. This will be more of an advertisement for our idea and innovation; however it will also provide us a good platform to analyze our system in a much broader way.

Due to the novelty of our concept, we did not have an existing dataset that we could use for our reference. The dataset used currently was classified incorrectly into different emotions. As a result, it gave poor results. To have a better dataset, we are planning to use crowd sourcing options. Crowd-sourcing will enable us to acquire a better dataset directly from the general public and might also provide us an interesting topic of research on how different people feel about emotions and their complex use in daily conversations.

Speech is one of the best forms of expressing emotions. However, modern day text to speech systems are more mechanized and are not able to convey the emotion behind the text. Through a system that understands and detects the emotion behind text, text-to-speech systems can be made much more intuitive rather than just a lame process where an emotionless computerized voice simply speaks up.

An interesting application can be analysis of the most powerful speeches in the world. By converting the speeches into texts we can analyze the flow of emotions in these speeches so as to understand what made them inspire millions of people to change the world in some or the other way.

By understanding the emotions behind the status updates/tweets/search queries of users on social media we can generate personalized services to them. For example if a person posts something like ‘I’m in a relationship’ then we can generate specific advertisements related to love cards, teddy bears etc for that person. Similarly we can generate relevant search queries depending upon the emotion in the text.

Automatic emoticon generation in social media and chat applications depending upon the emotion of the text that a user enters can be another useful application.

Under our present research we have considered 6 basic emotions. There are a number of emotions which are a combination of 2 or 3 basic emotions. These are known as advanced emotions. For example, the emotion of excitement is a combination of joy and surprise, the emotion of disgust is a combination of fear and anger.

An interesting challenge for our system at the moment will be to recognize sarcasm which is commonly used these days. So, that can be a possible area of research for us.

REFERENCES

1. Shivhare, Shiv Naresh, and Saritha Khethawat. "The Second International Conference on Computer Science, Engineering and Applications (CCSEA-2012)." *EMOTION* (2012).
2. Lee, S. Y. M., Chen, Y., Huang, C.-R. and Li, S. (2013), DETECTING EMOTION CAUSES WITH A LINGUISTIC RULE-BASED APPROACH. Computational Intelligence, 29: 390–416. doi: 10.1111/j.1467-8640.2012.00459.x
3. Lee, Sophia Yat Mei, Ying Chen, and Chu-Ren Huang. "A text-driven rule-based system for emotion cause detection." *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*. Association for Computational Linguistics, 2010.
4. Seol, Yong-Soo, Han-Woo Kim, and Dong-Joo Kim. "Emotion Recognition from Textual Modality Using a Situational Personalized Emotion Model."*International Journal of Hybrid Information Technology* 5.2 (2012).
5. https://code.google.com/p/word2vec/
6. Socher, Richard, et al. "Recursive deep models for semantic compositionality over a sentiment treebank." *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2013.
7. Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *the Journal of machine Learning research* 3 (2003): 993-1022.