Seminar
On

SENTIMENT ANALYSIS WITH MULTI – MODALITY

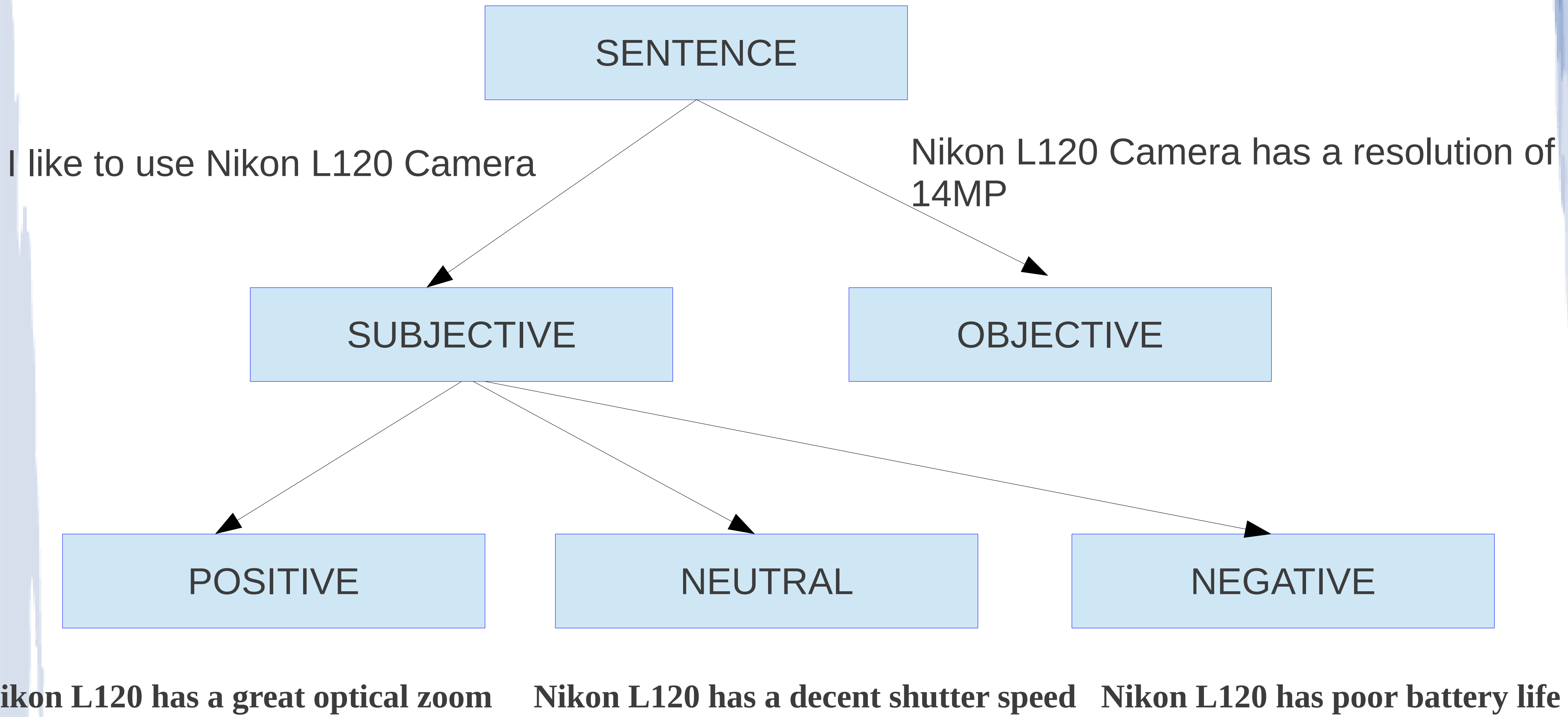
Presented By:

ANAMAY TENGSE
RAKSHA SHARMA
ADITYA SHROTRI

SUBJECTIVITY CLASSIFICATION AND SENTIMENT ANALYSIS

- **Subjectivity Classification:** Distinguishing factual sentences from those sentences which may be polarized or opinionated.
- **Sentiment Analysis:** It is a process to identify and extract the polarity of a document along with the intended emotional communication of the speaker.
- Sentences conveying some opinion are first separated from those stating simple facts. Then the sentences containing opinion are analyzed for polarity.
- However this is not strictly the rule. Many approaches choose to perform SA directly without explicit subjectivity classification.

SENTENCE CLASSIFICATION



SENTENCE

I like to use Nikon L120 Camera

Nikon L120 Camera has a resolution of 14MP

SUBJECTIVE

OBJECTIVE

POSITIVE

NEUTRAL

NEGATIVE

Nikon L120 has a great optical zoom

Nikon L120 has a decent shutter speed

Nikon L120 has poor battery life

OUTLINE

- Motivation for Sentiment Analysis
- Classical Methods for SA
- Limitations of Classical Methods
- Multimodality – A brief overview
- Audio, Visual and Textual Feature Selection
- Polarity Indicators
- Experimental Setup
- Experimental Methodology
- Results
- GEMEP Corpus
- Summary
- Conclusion
- Future Work
- References

MOTIVATION FOR SA

- Rise in Web 2.0 has lead to emergence of blogs reviews recommendations
- Social media are a huge untapped source of honest user opinion for various products and services
- Organizations want to process this information for improving their products
- SA provides a way to automate the process of filtering out the noise, understanding the conversations, identifying the relevant content and classifying it appropriately

Classical Methods for SA

- SA has been largely text based owing to the fact that most of available web content was plain text
- Documents are analyzed one sentence at a time.
- Sentences are scanned to find a subjectivity clue – such as a polar adjectives, adverbs and even some verbs and nouns
- A final score for the sentence is calculated after taking into account the polarity of each word and also negations
- Then the sentiment of the entire document is inferred by combining the score of each sentence.

Limitations of Classical Methods

- With the emergence of video sharing websites like Youtube, Vimeo etc. a large parallel source for mining opinions has emerged
- More and more people express their opinions via videos – everyday more than 10,000 videos are uploaded on Youtube alone
- Videos have more impact on public opinion than text
- Many new audio-visual features such as intonation, facial expressions, gestures etc. yield a sizable amount of sentiment related information which is not available in text data alone
- Simply transcribing all videos will not capture all relevant information
- Eg: Sarcastic comments like “Manmohan Singh is a very outspoken person!” are difficult to analyze purely from text. However, facial expressions and intonation can provide a clue as to the true sentiment of the speaker.

MULTI-MODALITY

- Multi Modality entails the use of multiple media such as audio and video in addition to text to enhance the accuracy of sentiment analyzers.
- Intuitively, analyzing data in multiple forms will help overcome limitations discussed earlier
- Eg: While uttering a sarcastic sentence, the speaker's tone and body language is markedly different. Eye rolling and dramatized intonation are common.
- In conjunction with transcription, auditory and visual streams can provide vital clues about the sentiments in a video

AUDIO-VISUAL AND TEXTUAL FEATURE SELECTION

Features available in each mode:

- **Audio**: Pitch, Pauses, Intonation, Energy distribution over a sentence, speed of utterance
- **Video:** Facial Expressions like smiles, frowns etc., Gestures, Posture, Gazes, Eye – Contact
- **Text**: Polar Words, Word Groups / Phrases, Character N-Grams, Phoneme N-Grams

Out of these, only a subset is used for optimal classification

Textual Indicators

The following few features have been found to be most effective:

- **Words:** *[Morency et. al., 2011]*
  - only a few words account to the sentiment of the sentence.
  - Adjectives, adverbs and in some cases, verbs and nouns.(Ex. Good, Bad, Love, Hate, Awesome, Great, etc.)
  - By assigning polarity scores to such words, and accounting for certain valence shifters (like 'not','hardly',etc.); a cumulative polarity is computed as:

$$\Sigma_w \, P_{word} * VS_{word}$$

(where VS_{word} is the valence shifter binding with the polar word)

Contd..

..contd

- **Character n-grams:***[Raaijmaker et. al.]*
  - Few groups of characters tend to appear together in similar sentimental words
  - Eg: 'ould' in modal verbs or 'ive' in adjectives and 'ly' in adverbs.
- **Phoneme n-grams:***[Raaijmaker et. al.]*
  - Similar to character n-grams, but instead we use phonemes.
  - This is based on an idea that certain phonemes are uttered only in words expressing a class of sentiments.

Auditory Indicators

- **Pauses**:*[Morency et. al., 2011]* Generally an utterance with more pauses indicates neutral sentiment
  - Pauses indicate thought which implies a factual / neutral sentence
- **Pitch:***[Morency et. al., 2011]* Variation in pitch indicates subjectivity
  - Low pitch indicates seriousness
  - High pitch indicates anxiety / excitement
- **Intensity (Energy):***[Morency et. al., 2011]* High energy utterances indicate emphasis on a word or phrase, giving information about polarity of the word or phrase.

Visual Indicators

- **Smile:***[Morency et. al., 2011]*
  - The most obvious feature which indicates a positive sentiment
  - New age cameras detect smiles and also assigns an intensity score to the smile
- **Gaze:***[Morency et. al., 2011]*
  - Orientation of speaker's face with the camera can be used to detect an eye contact or a gaze. (Available in modern cameras).
  - Eye contact indicates possible positive sentiment.
  - Whereas a look away might indicate a negative or neutral sentiment.

EXPERIMENTAL SETUP*[Morency et. al., 2011]*

- **GOAL** – Classify an audio visual clip into one of the three labels positive, negative or neutral
- Answer the question: Are the 3 modalities complementary and if so can they help each other in classification?
- Textual features
  - Extracted by automated transcription software.
  - Positive and negative lexicons were used along with Valence shifter lexicon
- Visual Features
  - Automated extraction by the commercial software OKAO
  - It returned scores on smile duration and gaze duration
- Auditory Features
  - OpenEAR software used to extract pitch and intensity at same frame rate as the video

EXPERIMENTAL METHODOLOGY

- HMMs are used as the probability model since they work well in dynamic processes eg: Speech recognition
- Each element of the Markov chain represented one spoken utterance
- For each utterance trimodal features were calculated which summarized the audio visual cues happening during that utterance
- HMM learns the relative importance of each audio visual cue and also the dynamic between the utterances
- 47 preprocessed Youtube clips were used as the dataset
- Leave - one - out testing was done
- Precision Recall and F – measure were calculated

RESULTS

	F1	**Precision**	**Recall**
Text only HMM	0.430	0.431	0.430
Visual only HMM	0.439	0.449	0.430
Audio only HMM	0.419	0.408	0.429
Tri-modal HMM	**0.553**	**0.543**	**0.564**

- Precision is the probability that predicted backchannels correspond to actual listener behavior.
- Recall is the probability that a backchannel produced by a listener in the test set was correctly predicted by the model.
- F1 is the F-measure with equal weight for Precision and Recall
- Tri-modal HMM showed significant improvement

GEMEP

- The **Geneva Multimodal Emotion Portrayal Corpus GEMEP***[Bänziger et. al.]*
  - large dataset consisting of more than 7,000 audio-video emotion portrayals,
  - 18 emotions – positive, negative and neutral
  - portrayed by 10 different actors with the help of a professional theatre director
- Pride, amusement, elation, relief, pleasure included instead of simply 'happiness'
- Similarly negative emotions like irritation, anxiety, rage, panic, fear
- Actors were given scenarios to 'perform' for each emotional state
- Cameras recorded
  - facial expressions and head orientations of the actors
  - body postures and gestures from the perspective of an interlocutor,
  - Body postures and gestures from the perspective of an observer

OVERCOMING DRAWBACKS OF EMOTION PORTRAYALS

- Portrayals reflect stereotypes, not genuine emotions
  - Acting techniques used stir genuine emotions in actors
  - Judgment studies used to select only the best portrayals
- Portrayals represent infrequent emotions
  - Very Intense emotions like rage, panic, depression are not encountered often in daily communication
  - Actors were instructed to portray variety of 'realistic' states while toning down extreme reactions
- Portrayals are decontextualized
  - Short scenarios were given to the actors before hand

SUMMARY

- Sentiment Analysis deals with determining the polarity of a document
- Multimodal SA, which encompasses audio-visual and textual mediums is necessary for the new face of the web
- Features Useful for SA-
    - Auditory: Pitch, Intonation, Intensity, Pauses, Speed
    - Video: Facial Expressions, Gestures, Eye contact
    - Textual: Polar Words, Character N grams, Phoneme N grams
- Trimodal SA performs better as compared to classical and unimodal SA
- New corpora like the Geneva Multimodal Emotion Portrayal Corpus may prove beneficial for SA in the future

CONCLUSION

- We believe that multimodality will also help in detecting whether a speaker is expressing his own opinion or merely parroting somebody else's views. In such cases a mere text based approach will fail, as the most important clues will be found in intonation and facial expressions
- Hence multimodality will find applications in a broader spectrum such as analyzing interviews, interrogations, lie detection etc.
- Multimodal SA is very much an open ended topic. Lots more research needs to be done as evident from the results of the discussed experiment

FUTURE WORK

- Feature selection for auditory and visual media can be data driven instead of model driven
- Since large corpora like the GEMEP are now available, such a data driven approach can be feasible
- The said experiment was run on a set of only 47 videos. However, running it on a large dataset like GEMEP may spring new challenges and results
- Increasing the spectrum of sentiment classes may provide valuable information, which is not captured currently
- Eg: Anger, Anxiety, Elation, Confidence etc. instead of Positive Negative and Neutral

REFERENCES

- Louis-Philippe Morency, Rada Mihalcea, and Payal Doshi, Towards Multimodal Sentiment Analysis: Harvesting Opinions from the Web, in Proceedings of the International Conference on Multimodal Computing (ICMI 2011), Alicante, Spain, November 2011.
- Bänziger, Tanja, and Klaus Scherer. "Using actor portrayals to systematically study multimodal emotion expression: The GEMEP corpus." Affective computing and intelligent interaction : 476-487.
- Raaijmaker,Truong, Wilson, Multimodal subjectivity analysis of multiparty conversation in EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing, Pages 466-474, Association for Computational Linguistics Stroudsburg, PA, USA ©2008
- http://en.wikipedia.org/wiki/Sentiment_analysis