SENTIMENT DIFFERENCE
— A NEW FEATURE SPACE BASED ON INFORMATION THEORY

Hang Gao

WHAT'S NEXT

- Introduce sentiment and polarity analysis
- Present and discuss experiment result of polarity classification using common information weighting schemes
- Propose a new feature space "Sentiment Difference" that outperforms all the above information weighting schemes.
- Discuss the performance of this feature space and make a conclusion.

WHAT IS SENTIMENT ANALYSIS

- Why? Increasing demand for information on opinions and sentiment, e.g company needs to understand people's opinion on a specific ad.

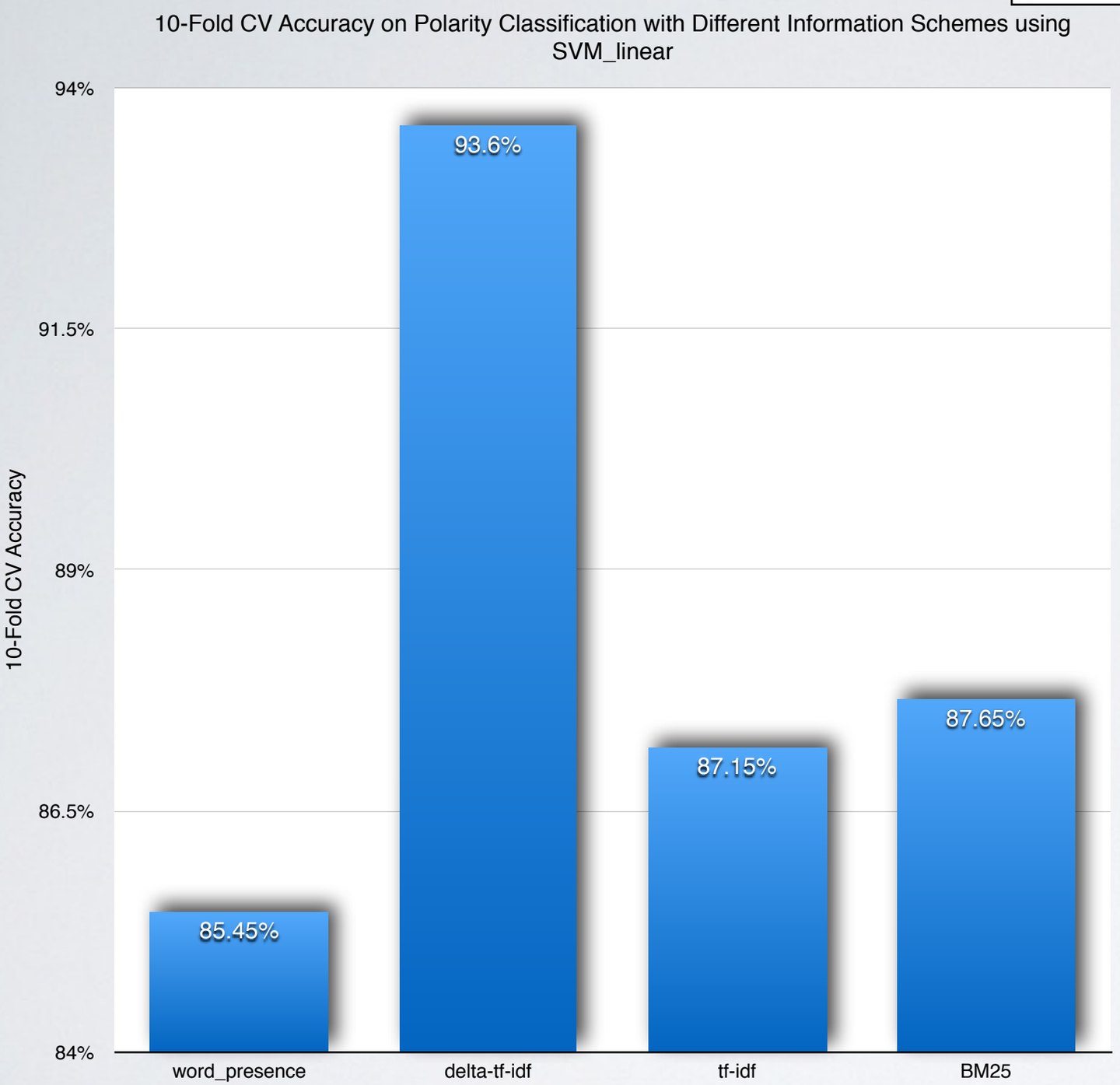
- Polarity classification, subarea of sentiment analysis, is a binary task of labeling an opinionated document as expressing either an overall positive sentiment or negative sentiment.(Pang and Lee)

POLARITY CLASSIFICATION WITH INFORMATION SCHEMES

- Dataset: Connell Movie Review dataset (Pang and Lee)
- Weighting schemes adopted: Word-presence, TF-IDF, Delta-TF-IDF and BM25.
- TF-IDF and BM25 are information schemes discussed in the class.
- Word-presence works by assigning weight 1 to a term if it occurs in a document and weight 0 if not.
- Delta-TF-IDF is proposed by Dr. Finin. It works by adopting the difference of IDF of different classes instead of over the entire corpus.

$$Weight(i,j) = (1 + \log tf_{i,j}) * \log(\frac{N_1 * df_{2,i} + 0.5}{N_2 * df_{1,i} + 0.5})$$

$$Weight(i,j) = TF(i,j) * (\log(\frac{N_1}{N_2} * \frac{df_{2,i}}{df_{1,i}})) = TF(i,j) * (\log(\frac{N_1}{N_2}) + \log(\frac{df_{2,i}}{df_{1,i}}))$$



Why Delta-TF-IDF works Best?

- Delta-TF-IDF pre-defines a boundary of sentiment and transform all data points to the two sides of that boundary.
    - Fact 1: N1 is equal to N2 in our dataset
    - Fact 2: Most words occur only once in a move review
- Delta-TF-IDF does not provide solution to multiple-class task.
- Its theoretical base is deficient.
- It is still limited by the framework of TF-IDF.

BEFORE WE START…

- Assumption 1: The sentiment trend of terms can be reflected by their distribution over the corpus.
- Assumption 2: The sentiment of a whole is a combination of the sentiment of the parts.
- Assumption 3: Task on sentiment can be solved by features on sentiment(close or semi-close world assumption).
    - It is doubtful whether information retrieved by classic IR weighting schemes is useful for sentiment task. For example, although word meaning is the information of interest, it is possible to be only noise and meaningless for sentiment task.

SENTIMENT DIFFERENCE

- Definition: The sentiment difference of a term is the normalized difference between its real distribution and its expected distribution of no sentiment, over the entire corpus.

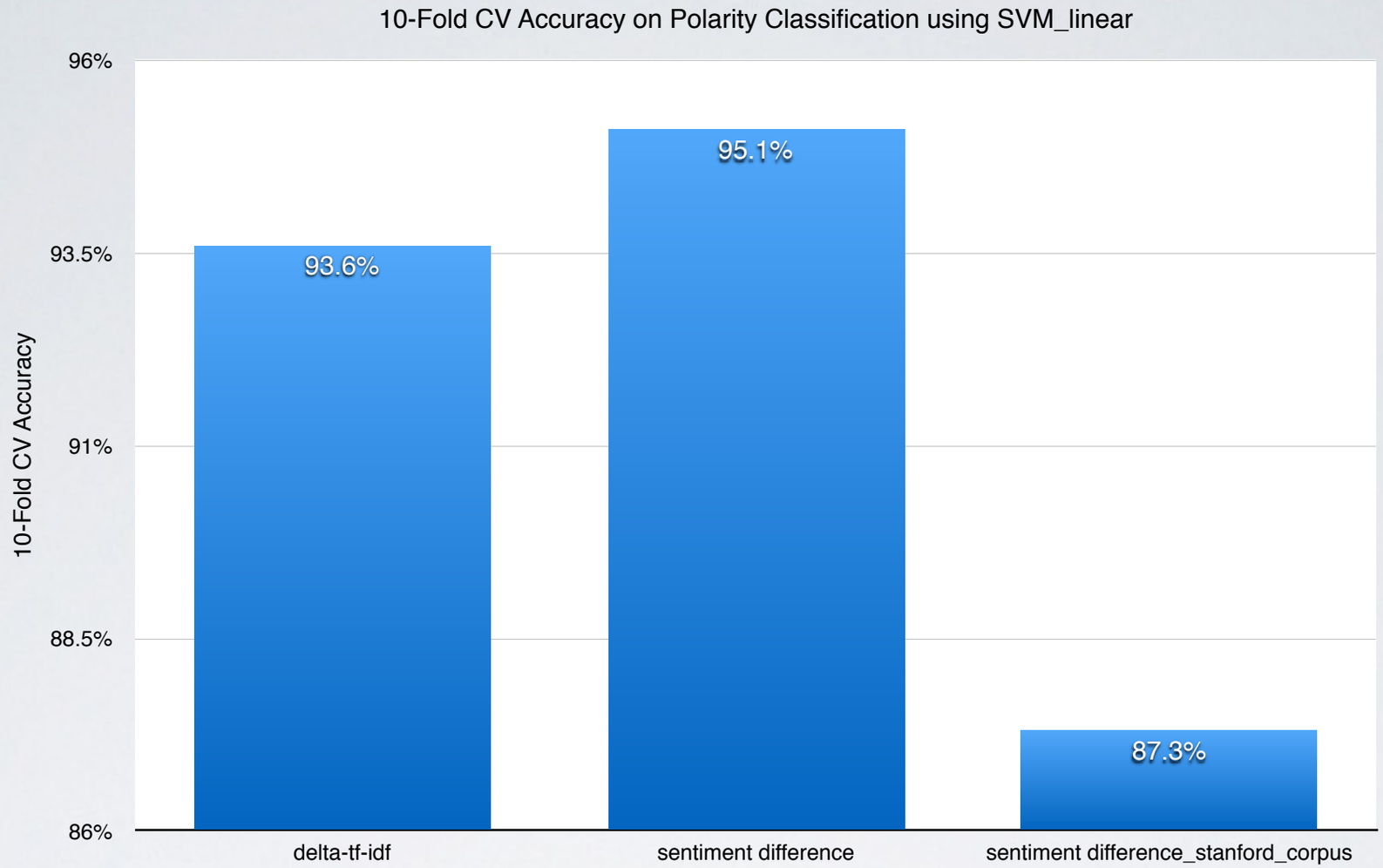
- For example, when considering polarity task(positive and negative), the sentiment difference of a term can be defined as:

$$N = Freq_{pos} + Freq_{neg}$$

$$SD = \frac{(Freq_{pos} - \frac{N}{2}) + (\frac{N}{2} - Freq_{neg})}{N}$$

SENTIMENT DIFFERENCE

- Note that SD defines a sentiment boundary, similar to Delta-TF-IDF.
- However, unlike Delta-TF-IDF whose boundary may vary due to the unbalanced corpus, the boundary of SD is always 0, which avoids problems caused by unbalance.
- The normalization step is introduced to overcome the bias problem caused by the usage frequency of terms.
    - If the un-normalized difference of both terms are K, the term with smaller overall occurrences tends to have greater sentiment contribution.



- The first and second series are the results of Delta-TF-IDF and SD performed on the dataset proposed by Connell while the third series is the result of SD performed on the dataset proposed by Stanford NLP group.

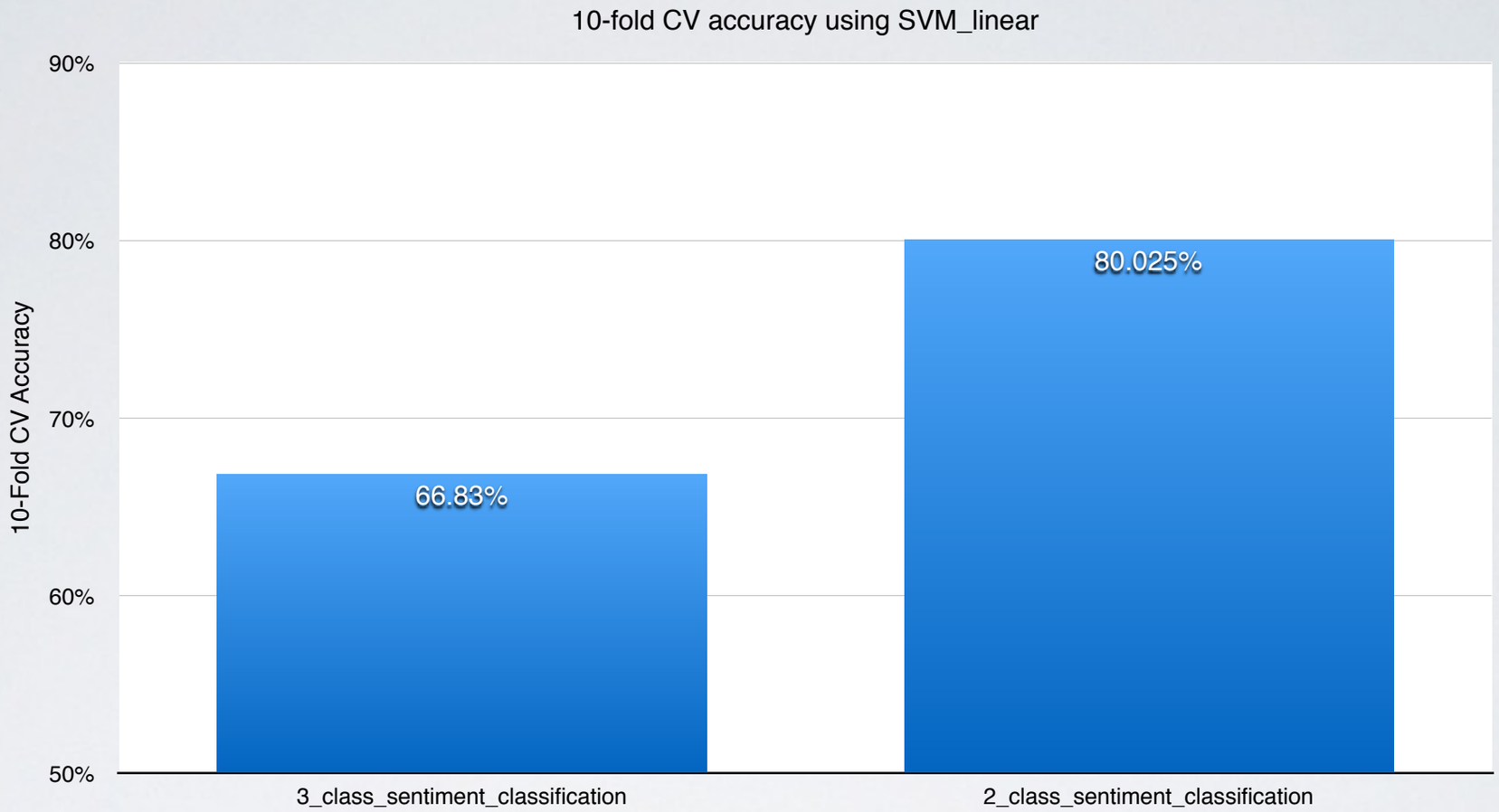
EXTEND TO MULTIPLE-CLASS CASES

- Definition: Let K be the number of classes, the Sentiment Difference of a term is itself a distribution of freedom K-1 that can be represented by a vector of length K-1 with each dimension of the form:

$$N = \sum_{i}^{K} Freq_i$$

$$SD_i = \frac{(Freq_i - \frac{N}{K}) + (\frac{N*(K-1)}{K} - \sum_{j \neq i}^{k} Freq_j)}{N}$$

where the boundary of each dimension is still 0



- The experiments are performed on dataset proposed by Stanford NLP group.
- Note that the baseline of 3-class classification is 33% and the baseline of 2-class classification is 50%.
- The performance can be improved by better use of SD and other models.

DISCUSSION

- Why no Term-Frequency? It is not clear whether the term frequency is a crucial information for sentiment classification. It is often the case that terms with no/little sentiment have higher or similar frequency than the terms with strong sentiment in a document. Thus if we take TF into consideration, it is possible to introduce noise. Prior experiments also prove it.

- Why no smoothing? There seems no need to introduce smoothing to the model. If a term never occurs in any document of a corpus, its SD is simply 0.

- Future work? The model never considers the structure of phrases, thus it is possible to mis-handle cases like negation and so on. However, we can alleviate the problem by applying the model to N-grams.

QUESTION?