Evaluation of Sentiment Polarity Prediction using a Dimensional and a Categorical Approach

**Ioana MURESAN, Adriana STAN,
Mircea GIURGIU, Rodica POTOLEA**

Technical University of Cluj-Napoca,
Speech Processing Group, http:/speech.utcluj.ro

Contents

1. Introduction

- In the context of Text to Speech Synthesis;
- Current trends to improve the SoA TTS are focused on:
  - Text → expressive speech synthesis system;
  - Audio → render natural expressive speech signal;
- Expansion of the social networks → large amount of emotional data (text and speech) is available;
- The basic concept: each word has an inherent emotional state that vary upon context.

1. Introduction – models for sentiment analysis

- **Dimensional model - based on psychological studies (continuous val.)**
  - **Valence** - measures how pleasant an emotion may be
  - **Arousal** - measures the intensity of the emotion
  - **Dominance** - represents the controlling and dominant nature of the emotion

- **Categorical model – subsets of words around certain associated labels (discrete val.)**
  - **Ekman's six basic emotions** (anger, disgust, fear, joy, sadness and surprise)
  - **subjectivity** (subjective vs. objective)
  - **polarity** (positive vs. negative vs. neutral)
  - **stubbornness** (opinionated vs. non-opinionated)

- **BENEFIT: dimensional --**use existing **emotional thesaurus (**WordNet-Affect, ConceptNet, SentiWordNet) and **statistical NLP techniques.**

2.Objectives

- to evaluate the effectiveness of lightly supervised machine learning techniques for **sentiment polarity prediction** .

- to reduce the number of supervised steps involved in a common sentiment polarity prediction framework by **unsupervised techniques**.

- to prepare the tools and datasets for the long-term goal of sentiment polarity prediction in an **unsupervised** and **language independent** manner for text-to-speech applications.

3. Related work (1)

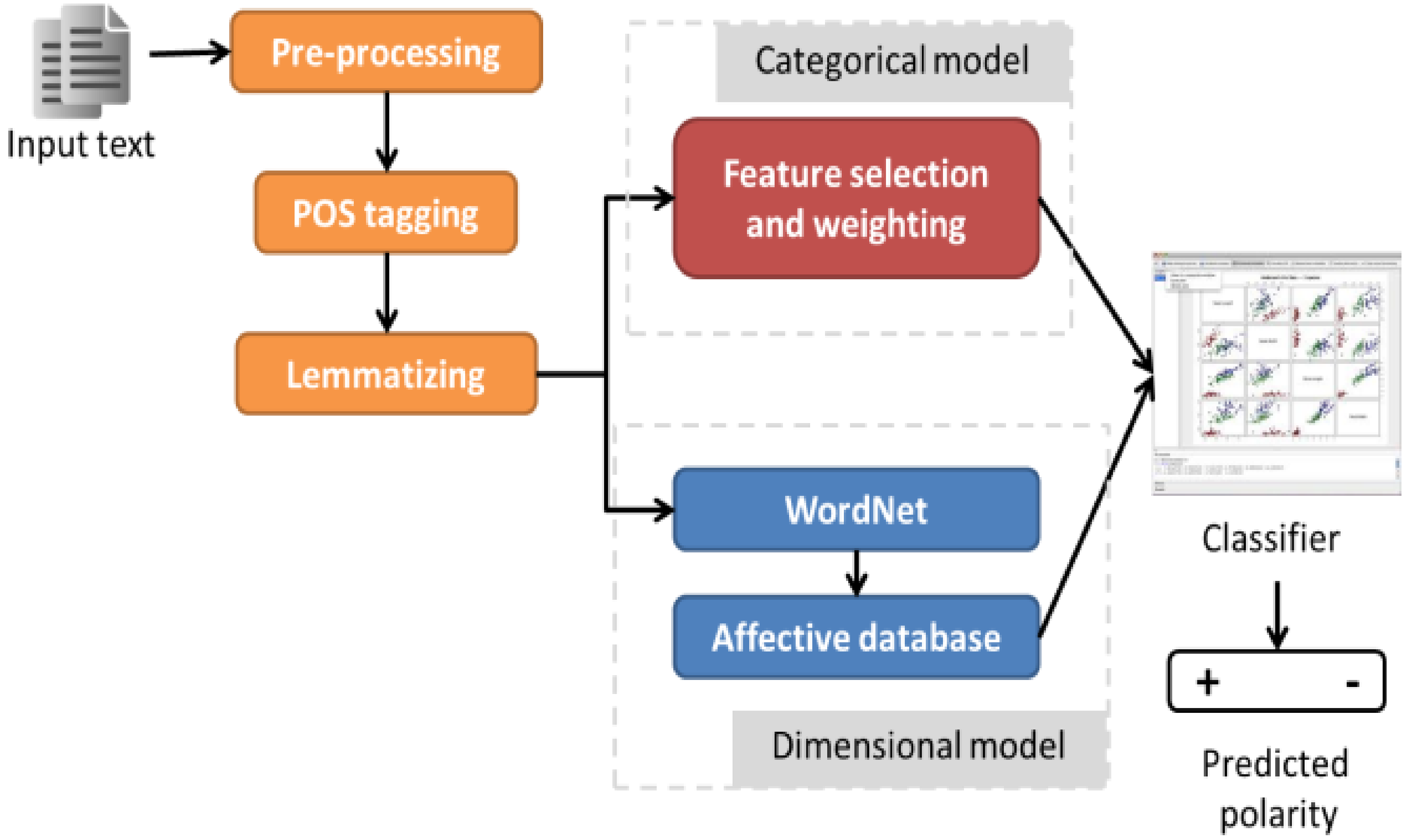
- Hatzivassiloglou, Turney (2003) – exploring the semantic orientation of adjectives and adverbs (PMI, LSA to infer the semantic orientation);
- Guohong Fu (2010) - determining **static polar words** and **dynamic polar words** using fuzzy sets and Chi-Square;
- Velikovich (2010) - building large **polarity lexicons** semi-automatically from web using graph propagation algorithm
- Chang Cui (2006) - determine the impact of **high order n-grams** (n=3,4,5,6) using Chi-Square for predicting sentiment polarity of a large amount of web data, 200,000 online reviews
- Mac Kim (2010) - recognizing the four **affective states**:
  - **dimensional** representation, ANEW(Affective Norms of English Words
  - **categorical** representation, using VSM and dimensionality reduction techniques : LSA (Latent Semantic Analysis), PLSA (Probabilistic Latent Semantic Analysis), NMF (Non Matrix Factorization)
- Trilla (2013) - categorizing sentences into **+** and **--** by combining NLP tools for feature reduction and weighting.

3. Related work (2)

Our approach:

- reduce the number of supervised steps for sentiment polarity prediction;
- propose a feature selection / weighting approach for both the dimensional and categorical approaches
- evaluate the system on standard datasets

4. Proposed method (overview)



4. Processing flow (Common steps)

Pre-Processing:

- punctuation sign stripping, lower case conversion;

POS tagging for each word:

- emotional words (adj, adv, vbs) are determined;
- computation is reduced by discarding words without an affective value;

Lemmatizing:

- helps normalize the term variety;
- improves the building of bag-of-words (BOW)

2 Models: Categorical Model (BOW concept),

Dimensional Model (normative affective database)

4. Processing flow (Categorical model)

Determine a set of words/features which represent each category:

- Problem: end up with large set of features → feature selection
- Feature selection: scoring the features by a metric, select first *k* features
- The metric: Chi-Square
- discrimination between relevant / non-relevant terms → feature weighting = *Relevance Factor (RFt,c)*

4. Processing flow (Categorical model)

A. Feature Selection and Weighting

Chi-Square:

$$x^2(t,c) = \frac{N \times (AD - CB)^2}{(A+C) \times (B+D) \times (A+B) \times (C+D)} \quad (1)$$

Relevance Factor:

$$RF_{t,c} = log_2(1 + (A+B))log_2 \frac{A}{\max(1, B\ )} \quad (2)$$

, where

- ***t*** is the term
- ***c*** the text instance category
- **N** the number of text instances.
- ***A*** is the number of times *t* and *c* co-occur
- **B** the number of times *t* occurs without *c*
- **C** be the number of times *c* occurs without *t*
- **D** be the number of times neither *t* nor *c* occurs

4. Processing flow (Categorical model)

Classifier features

- **Score** per **text instance**, computed as: $$\frac{\sum_{i=1}^{n} \text{category}(i) * x^2(i,c) * RF_{i,c}}{\sum_{i=1}^{n} x^2(i,c) * RF_{i,c}} \quad (3)$$

, where n is the number of words, and category (i) = sgn(AD-CB);

- Number of **negative words**
- Number of **positive words**
- Maximum **weight** of **negative** words
- Maximum **weight** of **positive** words

4. Processing flow (Dimensional model)

Map the emotions in 3 dimensions (VAD):

- Valence; Arousal; Dominance;
- Dataset: “ANEW – **A**ffective **N**orms for **E**nglish **W**ords, 13.915 lemmas), Bradley, 1999;
- NLTK for lemma, WordNet used for synonyms;

Additional processing:

- Polarity → Cluster the VAD space in 3 clusters (+, -, neutral);
- Beyond the words → text fragments: a) VAD mean, b) # of + and – words; c) max A of + and – words;
- Words not found in the dataset → use their synonyms from the WordNet

4. Processing flow (Classifiers)

Naïve Bayes:

- probabilistic classifiers for supervised learning;
- conditional independence (does not fully holds for text).

Support Vector Machines (SVM):

- discriminative classifiers;
- optimal hyperplane for maximum distance between classes

C4.5:

- decision tree – information entropy;
- maximize the information gain in each node.

☺ ALL IN WEKA TOOLS !!

5. Evaluation data sets (1)

Semeval 2007 - news headlines: anger, disgust, fear, sadness and joy;

(-)Bombers kill shoppers / (+) Kate is marrying Doherty

ISEAR - human experiences and reactions: anger, disgust, fear, joy, sadness, shame, guilt;

(-) When I did not speak the truth / (+) During the period of falling in love

Fairy Tales - emotional sentences: angry-disgust, fearful, happy, sad and surprised

(-) It feels sad / (+) When Jemima alighted he quite jumped

Movie Review- movie reviews: negative and positive;

(-) What a script, what a story, what a mess / (+) You will find this movie extremely funny

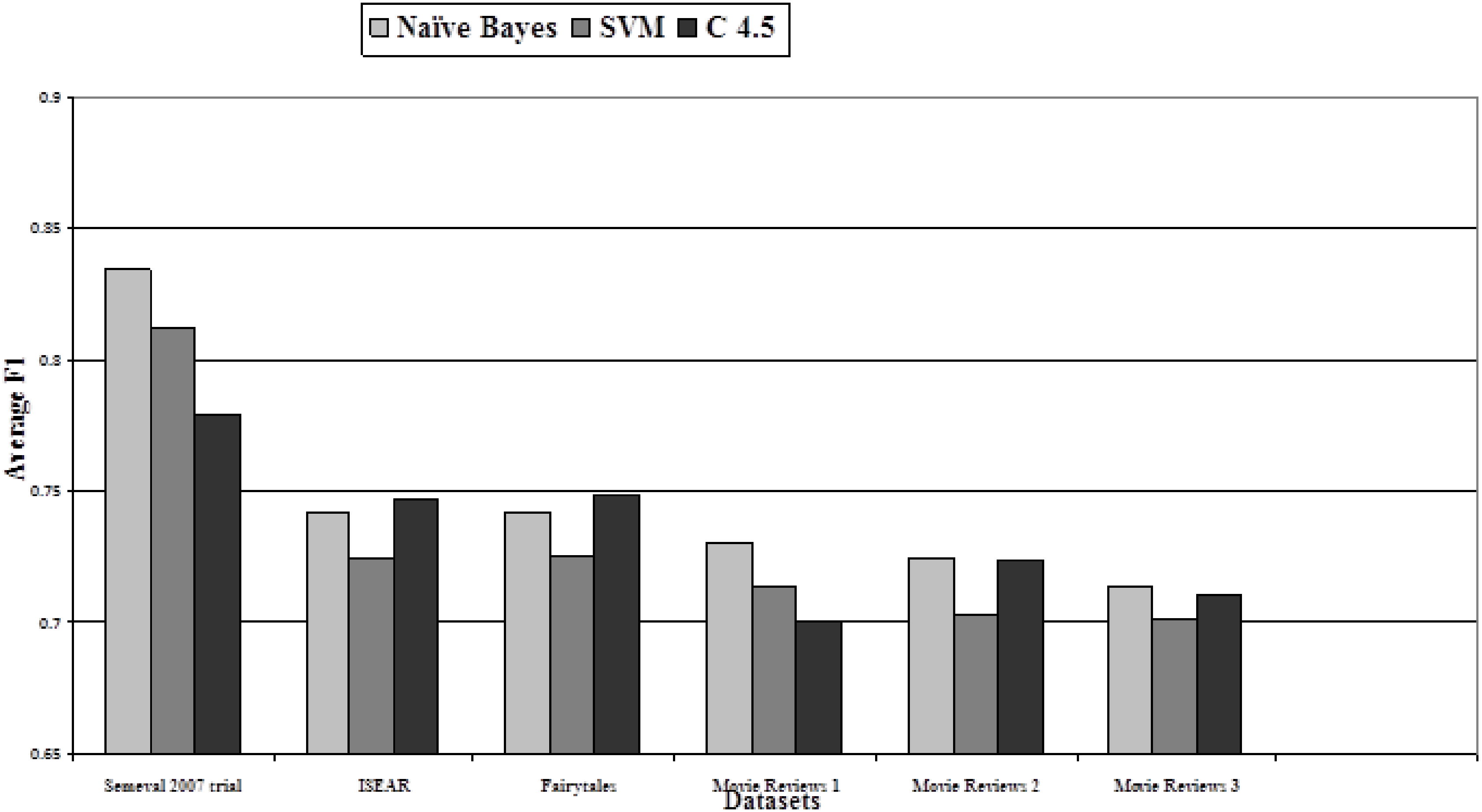
5. Evaluation data sets (2)

Dataset	Negative	Positive	Average number of words
Semeval 2007	82	62	6
ISEAR	6578	1094	20
Fairy-tales	648	559	24
Movie Reviews	12500	12500	227
Movie Reviews 1 (max 50 words/review)	321	352	40
Movie Reviews 2 (max 75 words/review)	877	1071	54
Movie Reviews 3 (max 100 words/review)	1548	1738	68

6. Evaluation results – Dimensional model (1)

	Semeval 2007	ISEAR	Fairy-tales	Movie Reviews 1	Movie Reviews 2	Movie Reviews 3
Naïve Bayes	**0.834**	0.742	0.742	**0.725**	0.698	0.690
SVM	0.812	0.725	0.726	0.722	**0.701**	0.687
C 4.5	0.779	**0.747**	**0.749**	0.708	0.689	**0.694**

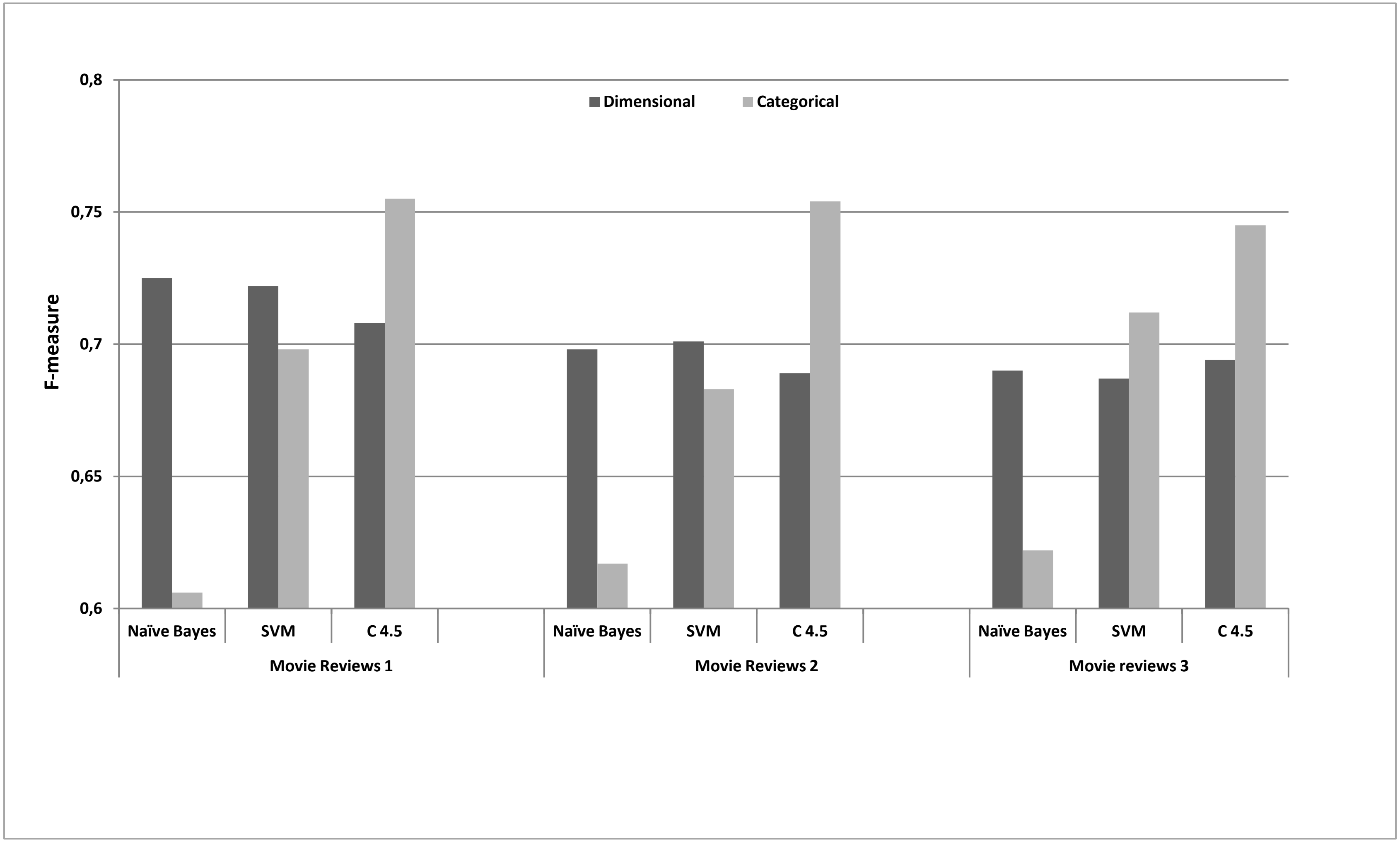
6. Evaluation results – Dimensional model (2)



6. Evaluation results – Categorical model (1)

Classifier	Average F1		
	Movie Reviews 1	**Movie Reviews 2**	**Movie Reviews 3**
Naïve Bayes	0.606	0.617	0.622
SVM	0.698	0.683	0.712
C 4.5	**0.755**	**0.754**	**0.745**

6. Evaluation results - Comparative results



7. Conclusions

- both dimensional and categorical approaches achieve comparable results with previous studies, reaching a maximum F-measure of **0.83**.
- no direct correlations between **datasets** and **classifier performance** for the **dimensional approach**.
- variation in performance due to: **text genre** in each dataset, **amount** of words/instances
- **C 4.5** performs best for the **categorical method** attaining an average 0.75 F-measure for all datasets.
- Possible reason: some **underlying features** that can not be captured by these simplified models.

8. Future work

- extending unigram features to higher order **n-grams** to capture larger contexts and thus performing a more precise classification.

- creating features staring from a set of emotional seed-words using web text data and statistic metrics to create **an unsupervised normative dataset.**

Thank you for your attention!

Acknowledgement:

The research leading to these results has received funding from the EC FP7 Programme, under the grant agreement No. 287678 (Simple4All), www.simple4all.org