# SUMMARIES OF RELATED PAPERS

## Experiments with Mood Classification in Blog Posts

* classifying blog text according to the mood reported by its author
* dataset: online diary entries

**Motivation**

* more and more blogs - personal, highly opinionated journals
  + increasing interest from research community
  + analysing consumers’ attitudes for marketing purposes
* stylometric research – in particular, research regarding emotion and mood analysisin text – is becoming more common recently, in part due to the availability of new sources of subjective information on the web filtering search results by mood, identifying communities, clustering
* assisting behavioral scientists and improving doctor-patient interaction
* authorship attribution
* gender classification

*scientific challenge*: different styles and definitions of moods

**Research Questions:**

* in what way does mood classification in blogs differ from mood classification in other domains?
* how much data is needed for reliable training? how many features are required?
  + it has been observed that, for NLP tasks, continuously increasing training set size improves results consistently [M. Banko and E. Brill. Scaling to very very large corpora for natural language disambiguation.]

**Process**: machine learning approach; identifying a set of features to be used for the learning progress

**Approaches of mood categorisation:**

* Rubin et al. investigated discriminating terms for emotion detection in short texts
* Liu et al. present an effective system for affect classification based on large-scale “common-sense” knowledge bases

**Feature Set**

* Frequency counts
* Length-related
  + the total length in bytes
  + the number of words in the post
  + the average length of a sentence in bytes
  + the average number of words in a sentence
* Semantic Orientation Features
* Mood PMI-IR
* Emphasized words
* Special Symbols

**Evaluation**

* Effectiveness of identifying individual moods in a blog post; examine the effect of changes in the training set size in classification accuracy
* Partition the moods manually into two mood sets: positive vs. negative

-> does combining closely-related moods improve performance?