**Summary of Important Papers on Sarcasm Detection using NLP**

**Prepared by:** Hari Thapliyal

**Reviewed by:** Dr. Anil Vuppala.

**Date: 27-Aug-20**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Paper** | **Language, Text Type** | **Features** | **Model** | **Metrics** |
| 1 | Irony Detection in Twitter: The Role of Affective Content  (Faŕias et al., 2016) | English,  Twitter | colon, exclamation, question, punctuation mark, words in tweet, character length of tweet, verbs, nouns, adjectives, adverbs, number of uppercase letters in tweet, number of emoticons in tweet, degree of similarity in the tweet, number of hashtags, number of mentions. | SVM, DT, NB | Acc: 73-96% depends upon datasets and classifier. |
| 2 | Natural Language Processing Based Features for Sarcasm Detection: An Investigation Using Bilingual Social Media Texts  (Suhaimin et al., 2017) | English, Twitter | Lexica1 features – Unigram, Pragmatic features- Punctuation marks, Hashtag, Prosodic features – Interjection, Syntactic features -Part of Speech, Idiosyncratic features - Idiosyncratic | Non Linear SVM | Acc: 82.5% |
| 3 | Semantics-aware BERT for Language Understanding  (Zhang et al., 2020b) | English, Normal Text | Semantic Role Labeller | CNN | Acc: 94.6% on Large dataset of SST2 |
| 4 | Multi-Rule Based Ensemble Feature Selection Model for Sarcasm Type Detection in Twitter  (Sundararajan and Palanisamy, 2020) | English, Twitter | 20 features. Noun and verb count, positive intensifier, negative intensifier, bigram, trigram, skip gram, unigram, emoji sentiment, sentiment score, interjections, punctuators, exclamations, question mark, uppercase, repeat words count, positive word frequency, negative word frequency, polarity flip, and parts of speech tagging. These are grouped in 3 categories Linguistic Features, Sentiment Features, Contradictory Features | Random Forest, Naive Bayes, Support Vector Machine, K-Nearest Neighbour, Gradient Boosting, AdaBoost, Logistic Regression, and Decision Tree. | Acc: 86.61% to 99.79% Depending upon the type of sarcasm. Final classifier is RF |
| 5 | Sarcasm Detection in Typo-graphic Memes  (Kumar and Garg, 2019) | English,  Instagram Images | Number of negative words, number of positive words, POS tag, hashtag | SVM, Logistic Regression, GBoost, Random Forest, Decision Tree, RNN | Acc: 73.25% to 87.95% depending upon the classifier used. |
| 6 | Sentiment Analysis in a Resource Scarce Language: Hindi  (Jha et al., 2016) | Hindi,  Movie Reviews | POS Adjective | Naive Bayes, Multinomial NB, SVM, Maximum Entropy | Acc: 92.2% to 100% depending upon unigram or bigram feature and classifer |
| 7 | Sarcasm detection on twitter : A Behavioral Modeling Approach  (Rajadesingan et al., 2015) | English,  Tweet | Created 335 SCUBA features in following categories,  Sentiment Score, Sentiment Transition between past and present, Sarcasm as a complex form of expression, emotion (mood, frustration, affects and sentiments), language familiarity, sarcasm familiarity, environment familiarity, written expression related, structural variation | Logistic Regression | Acc: 83.46% |
| 8 | Lexicon-Based Sentiment Analysis in the Social Web  (Asghar et al., 2014) | English, Tweet | Emoticon score, Lexicon score, SentiWordNet Score, Slang Score | Rule based | Acc: 95.24% |
| 9 | Harnessing Context Incongruity for Sarcasm Detection  (Joshi et al., 2015) | English, Tweet | a) Lexical features- unigram using chi-square test, (b) Pragmatic- emoticons, punctuation marks, capital words, (c) Explicit congruity- related to polarity changed, and (d) Implicit incongruity features. | LibSVM with RBF kernel | F1: 61% |
| 10 | Contextualized Sarcasm Detection on Twitter  (Bamman and Smith, 2015) | English, Tweet | Author Features, Addressee Feature, Audience features: Historical data of author and addressee, Response Feature, Tweet Features, Environment Features | Logistic regression | Acc: 85.1% |
| 11 | Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews  (Turney, 2002) | English, Opinion Survey of Products | POS | Rule Based PMI calculator | Acc: 74.39% |
| 12 | Sentiment Analysis of Hindi Review based on Negation and Discourse Relation  (Mittal and Agarwal, 2013) | Hindi, Movie Reviews | Semantic orientation & polarity values | Rule Based | Acc: 80.21% |
| 13 | BHAAV- A Text Corpus for Emotion Analysis from Hindi Stories  (Kumar et al., 2019) | Hindi, Short stories | Tokens using Classical language tool kit, unigram, bigram, FastText embedding, TF-IDF | SVM linear kernel, LR, RF, Shallow CNN + Bi-Directional LSTM | Acc: 62% |
| 14 | Towards Multimodal Sarcasm Detection: An Obviously Perfect Paper  (Castro et al., 2020) | English, Clips of YouTube, TV Shows, Transcription | Text features: lexical and pragmatic features, stylistic feature, incongruity, situational disparity, hashtag  Speech features: Acoustic patterns that are related to sarcastic behavior  Speaker related feature | SVM RBF kernel and a scaled gamma | F1: 71.8% |
| 15 | Context-based Sarcasm Detection in Hindi Tweets.  (Bharti et al., 2018) | Hindi, Tweets | POS, Used SentiWordNet features, Word Polarity Score for tweet and context | Rule Based | Acc: 87% |
| 16 | A Sentiment Analyzer for Hindi Using Hindi Senti Lexicon  (Sharma et al., 2014) | Hindi, Movie Reviews, Product Reviews | POS, Hindi SentiWordNet, Word Polarity Score | Rule Based | Acc: 85 to 89.5% |
| 17 | A Transformer-based approach to Irony and Sarcasm detection  (Potamias et al., 2020) | English, Irony/SemVal-2018-Task,  Reddit SARC2.0 politics,  Riloff Sarcastic Dataset | Embedding: ELMo , USE, NBSVM, FastText, XLnet, BERT base cased, BERT base uncased, RoBERTa base model, UPF, ClaC, DESC | RCNN-RoBERTa base model | Acc: 85% to 94% depending upon dataset |
| 18 | Detecting Sarcasm is Extremely Easy ;-)  (Parde and Nielsen, 2018) | English,  Tweet, Amazon product reviews | Contains twitter indicator,  Twitter-based predicates and situations,  Star rating, Laughter and interjections,  Specific characters, Polarity, Subjectivity  PMI, Consecutive characters, All-caps bag of words | Naïve Bayes | F1: 59% (Twitter)  F1: 78% (Amazon) |
| 19 | CARER: Contextualized Affect Representations for Emotion Recognition  (Saravia et al., 2020) | English, Tweets | BoW, char n-gram, TF-IDF, Word2Vec, fastText(ch), word-cluster, enriched patterns, Twitter-based pre-trained word embeddings and reweight them  via a sentiment corpus through distant supervision | CNN | Acc: 81% with CARER |
| 20 | A Corpus of English-Hindi Code-Mixed Tweets for Sarcasm Detection  (Swami et al., 2018) | Hindi-English, Tweets | Word N-gram, Char N-gram, Sarcasm score of each token, emoticons, chi-square to select feature | SVM with RBF, SVM with Linear, RF | Acc: 78.4% with RF |
| 21 | Harnessing Online News for Sarcasm Detection in Hindi Tweets  (Bharti et al., 2017) | Hindi, Tweets | POS | Rule Based | Acc: 79.4% |
| 22 | The perfect solution for detecting sarcasm in tweets #not  (Liebrecht et al., 2013) | Dutch, Tweets | POS (Adjective, Adverb) - Intensifier | Ruled Based | AUC: 77% |
| 23 | A2Text-net: A novel deep neural network for sarcasm detection  (Liu et al., 2019a) | English, Tweet, News Headlines, Reddit | Punctuation, POS, chi-square test to selected variables. | DNN, LSTM, SVM, RF, LR, GRU, A2Text | F1: 71% - 90% depending upon dataset with A2Text classifer |
| 24 | Sarcasm as contrast between a positive sentiment and negative situation  (Riloff et al., 2013) | English, Tweet | Word N-Gram, POS | LibSVM with RBF | F1: 51% |
| 25 | Sarcasm Detection in Hindi sentences using Support Vector  (Desai and Dave, 2016) | Hindi, various online sources (using polarity levelled corpora) | Unigram, Positive Score, Negative Score, Hashtag, Emoticons, Polarity, TFIDF | LibSVM with RBF | Acc: 84% |
| 26 | Twitter as a Corpus for Sentiment Analysis and Opinion Mining  (Pak and Paroubek, 2010) | English, Twitter | N-Gram, POS | Multinomial Naive Bayes, SVM, CRF | Not Mentioned |
| 27 | Exploring the fine-grained analysis and automatic  detection of irony on Twitter  (Van Hee et al., 2018) | English, Tweet | POS, Work unigram, bigram, Character tri and fourgram, number of character, number of punctuation, presence of punctuation, number of hashtags, interjections, tweet length, number of emoticons, hastags/word ratio, number of NE, tweet overall polarity, difference of highest positive word polarity, highest negative word polarity, cluster wise word2vec. Overall 4 group of features- Lexical, Sentiment, Semantic, Syntactic | SVM, LSTM | Acc: 67.54% (SVM)  Acc: 68.27% (LSTM) |
| 28 | Exploiting Emojis for Sarcasm Detection  (Subramanian et al., 2019) | English, Twitter, Facebook | Word vector + emoticon vector | GRU | F1: 89.36% (Twitter)  F1: 97.97% (facebook) |
| 29 | A novel automatic satire and irony detection using ensembled feature selection and data mining.  (Ravi and Ravi, 2017) | English, Newswire, Satire news articles, Amazon | (Linguistic, Semantic, Psychological, unigram) in named LIWC features (L), TAALES Features(T), Unigram Features(D), feature subset ensemble, feature selection (IG, GR, Chi, CORR, TSTAT) | SVM (Liner, RBF, Sigmoid, Polynomial), LMT, LR, RF, NB, BN, MLP | F1: 96.58% (L+T+D features) + GR feature selector + SVM RBF Classifier |
| 30 | Automatic Satire Detection: Are You Having a Laugh?  (Burfoot and Baldwin, 2009) | English, Newswire and Satire news articles | Headline Features, Profanity Features, Slang Features, Binary Unigram Features, Unigram, | SVM | F1: 79.8% |
| 31 | Semi-supervised recognition of sarcastic sentences in twitter and Amazon  (Davidov et al., 2010) | English, Twitter, Amazon | Meta Tag (User, Company, Product, Title, Author, Link, HashTags, Meta Tag based content Matching, Punctuation, | KNN | F1: 78% Amazon  F1: 83% Twitter |
| 32 | Identifying Sarcasm in Twitter: A Closer Look. In  (González-Ibáñez et al., 2011) | English, Twitter | Lexical Features: unigram, dictionary based (Linguistic Processes (e.g., adverbs, pronouns), Psychological Processes (e.g., positive and negative emotions), Personal Concerns (e.g, work, achievement), and Spoken Categories (e.g., assent, non- fluencies)) + WordNet Affect + interjection + punctuation  Pragmatic Features: positive emotions like smily, negative emotions like frowning face. ToUser like @Name  χ2 test to select features | SVM | Acc: 55.59% to 75.78% depending upon tweet format. |