EVALITA 2011

Evaluation of NLP and Speech Tools for Italian

EVALITA 2011

SuperSense Tagging with a Maximum Entropy Classifier and Dynamic Programming

G. Attardi, L. Baronti, S. Dei Rossi, M. Simi

Dipartimento di Informatica – Università di Pisa

Introduction

- Annotating nouns, verbs, adjectives and adverbs Evaluation and results
  - Special case of chunking
- Tanl chunker
  - Flexible and customizable tagging tool
    - PoS tagging, Named Entity tagging, …
  - Maximum Entropy classifier for learning how to chunk texts
  - Dynamic programming in order to select sequences of tags with the highest probability

Tanl Chunker

- Features Types
  - Attributes Features
    - Attributes like PoS, Lemma, NE of surrounding tokens
    - relative positions w.r.t. to the current token
    - E.g: POSTAG -1 0
  - Local Features
    - Binary morphological features
    - Extracted from the analysis of the current word and the context in which it appears
    - E.g.: “previous word is capitalized”
  - Global Features
    - Properties holding at the document level
    - E.g.: if a word in a document was previously annotated with a certain tag, then it is likely that other occurrences of the same word should be tagged similarly
- Textual configuration file to specify features

Dataset

- Training corpus tuning sub-sets
  - Training set (~70% of the corpus)
    - used to train the models
  - Validation set (~20% of the corpus)
    - used to choose the best model
  - Test set (~10% of the corpus)
    - used to evaluate the performance

Baseline

- Basic configuration
  - No attributes features
  - Standard set of local features
    - Features of Current Word: first word of sentence and capitalized; first word of sentence and not capitalized; two parts joined by a hyphen
    - Features from Surrounding Words: both previous, current and following words are capitalized; both current and following words are capitalized; both current and previous words are capitalized; word is in a sequence within quotes
- 100 iterations of the Maximum Entropy algorithm
- F-score of 71.07 on the validation set

Tuning and feature selection

- Creation of many configuration files with different combination of features
  - About 300 positional permutations of the attributes features
  - Variation of other parameters
    - Number of iterations
    - Cutoff feature
    - Refine feature
  - Performance computed testing the model on the Validation set
  - Validation process
    - Same configuration file
    - New model trained on Training set + Validation set
    - Performance tested on the Test set
  - Best run on the validation set
    - F-score of 80.01
    - ~10 points higher than the baseline

Final Submission

- Four runs with the best and most balanced performance on the Validation and Test set
- Only closed task
  - No improvements using external dictionaries and gazetteers
    - E.g.: ItalWordNet (IWN)

Attributes Features

	Run 1-2	Run 3-4
FORM	0	0
POSTAG	-2 0 1 2	0 1
CPOSTAG	-1 0	-1 0
LEMMA	-1 0	0

Local Features

- Standard set for all the runs
- Additional set for run 3 and 4
  - Goal: improve performance on the classes with lowest F-score
    - verb.emotion, verb.possession, verb.contact and verb.creation
  - List of the most common non-ambiguous verbs in those classes
    - Obtained from the training set and
    - Added as local features for the current LEMMA
  - The list of verbs is the following:
    - verb.emotion: sperare, interessare, preoccupare, piacere, mancare, temere, amare
    - verb.possession: vendere, perdere, offrire, pagare, ricevere, raccogliere
    - verb.contact: porre, mettere, cercare, colpire, portare, cercare, toccare
    - verb.creation: realizzare, creare, produrre, aprire, compiere

Global Features

- Refine option: OFF
  - Performed well for tasks with a lower number of classes NER
  - Less relevant for SST
- Cutoff threshold: 0
  - Values > 0 showed no improvements
- Different numbers of training iterations
  - run 1: 100
  - run 2: 150
  - run 3: 200
  - run 4: 500

Final Results

	Accuracy	Precision	Recall	FB1
UniPI - run 3	**88.50%**	**76.82%**	**79.76%**	**78.27**
UniPI - run 2	88.34%	76.69%	79.38%	78.01
UniPI - run 1	88.30%	76.64%	79.33%	77.96
UniPI - run 4	88.27%	76.48%	79.29%	77.86

Discussion

- Maximum Entropy tagger achieves the best F1 results with a reduced number of iterations
  - E.g. between 100 and 200 iterations
  - Important information for future tunings of the tagger
    - Decreases the number of experiments to be performed
    - Positive effect on the execution time for training the system