# Lexical-semantic resources: yet powerful resources for automatic personality classification

**Xuan-Son Vu**, Lucie Flekova, Lili Jiang, Iryna Gurevych

1. DDM Lab, Umeå University, Sweden
2. UKP Lab, TU Darmstadt, Germany

Global Wordnet Conference, Singapore, 2018-01-10

# Outline

- Introduction
  - Motivation
  - Automatic Personality Profiling
- Methodology
  - System design, Feature Design & Algorithms
- Experiments
- Results and discussion
- Conclusions

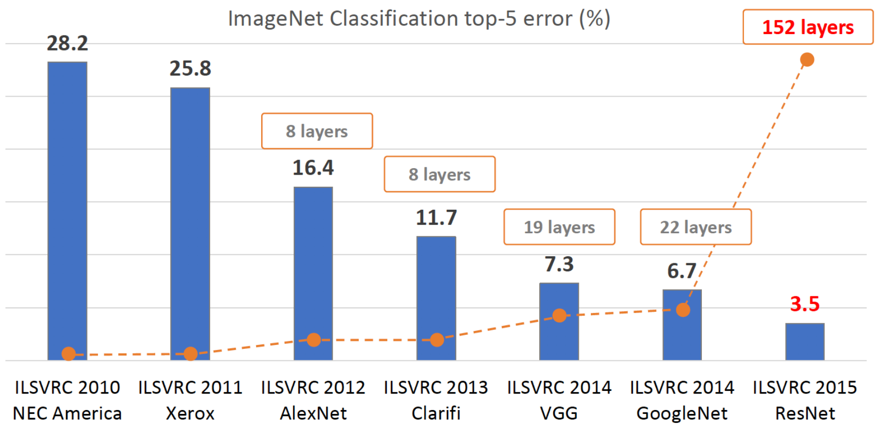
# 1. Introduction

# Introduction

- Motivation (1/2):
  - Study of personality and individual differences supporting for IR and Recommender systems.
    - Solve cold-start user problem (Flekova and Gurevych, 2015)

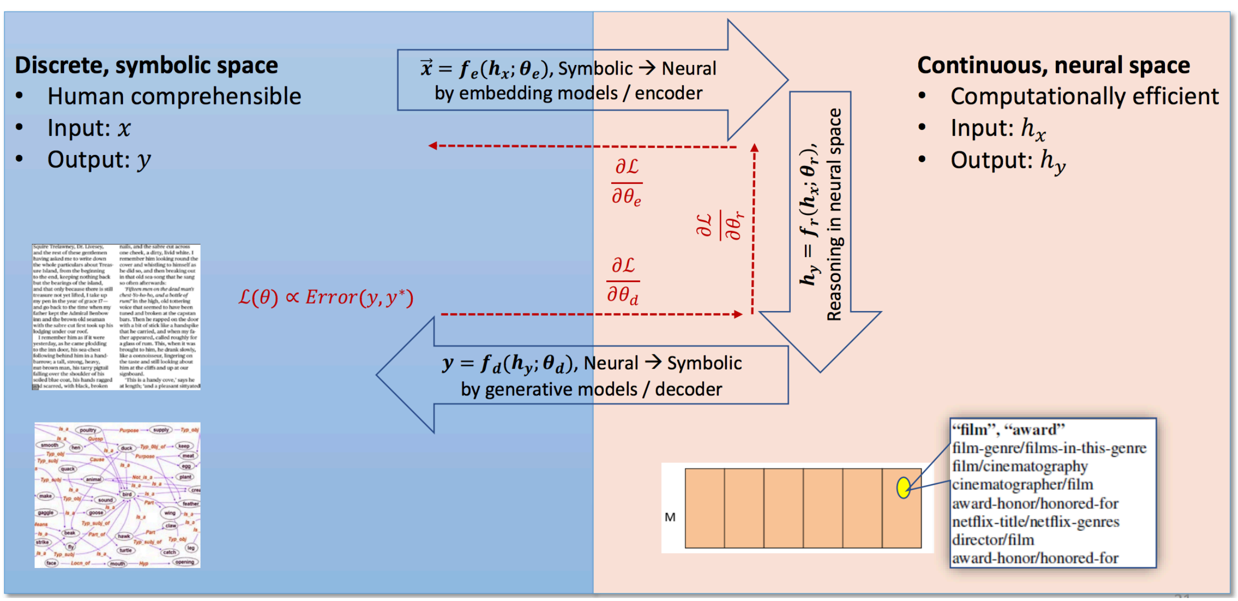

  - Applying to low-resource languages
    - No personality-specific resources are available
      - LIWC, MRC
    - Wordnet and SentiWordNet are more popular

(Jian Sun)



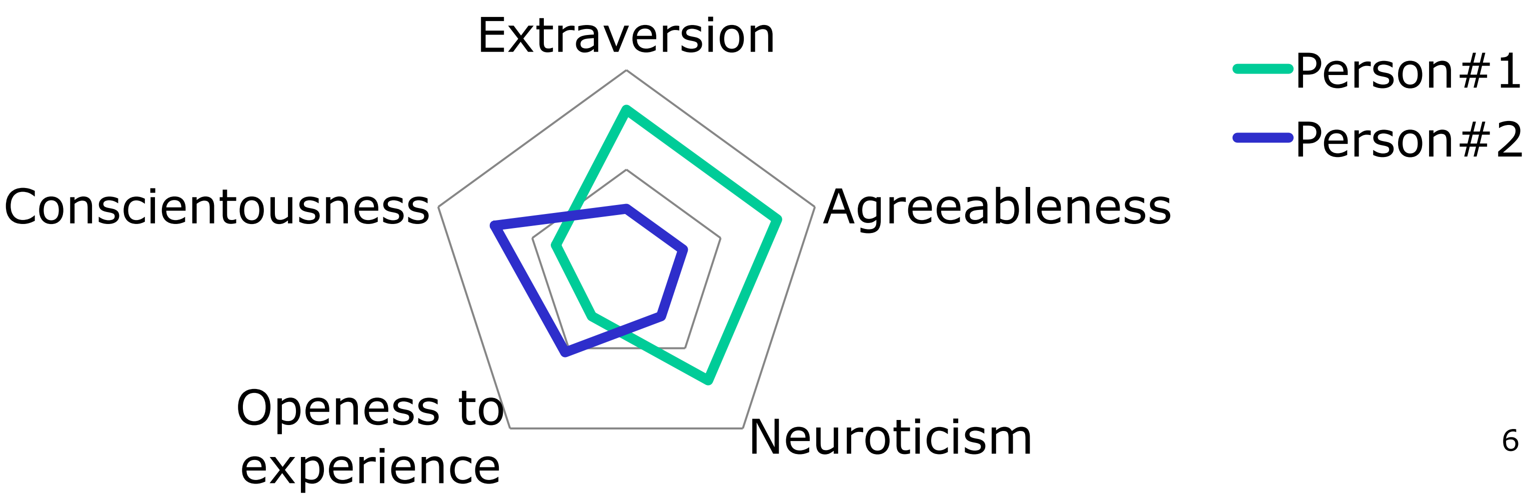

# Introduction

- Motivation (2/2):
  - Why emphasizing "**yet** powerful"?
    - Not a 'cool' paper when not using Deep Learning?
    - Small data
    - Reasoning



# Introduction

- The Big Five Factor Model of Personality (FFM)
  - Influenced many aspects of task-related individual behavior.
  - Has become standard in psychology over the last 50 years (Costa and McCrae, 2008)



# Introduction

- The Big Five Dimensions of Personality
  - Extraversion vs. Introversion

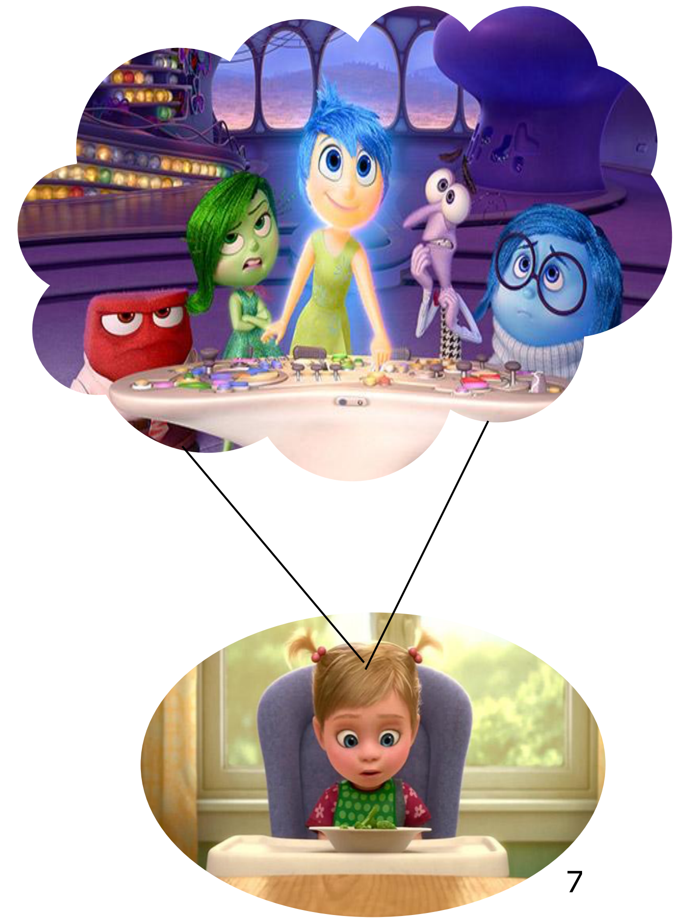
    (sociable, assertive, playful vs. aloof, reserved, shy)
  - Emotional stability vs. Neuroticism

    (calm, unemotional vs. insecure, anxious)
  - Agreeable  vs. Disagreeable

    (friendly, cooperative vs. antagonistic, faultfinding)
  - Conscientious vs. Unconscientious

    (self-disciplined, organized vs. inefficient, careless)
  - Openness to experience

    (intellectual, insightful vs. shallow, unimaginative)



Movie title: Inside out

# 2. Methodology

# Dataset Overview

1. Facebook status updates [1]
2. Stream-of-consciousness texts [2]
3. Transcripts of Youtube videos, annotated for personality [3]
4. User tweets [4]

[1] http://mypersonality.org/wiki/doku.php?id=wcpr13

[2] http://mypersonality.org/wiki/doku.php?id=wcpr13

[3] https://sites.google.com/site/wcprst/home/wcpr14

[4] http://www.uni-weimar.de/medien/webis/events/pan-15/pan15-web/author-profiling.html

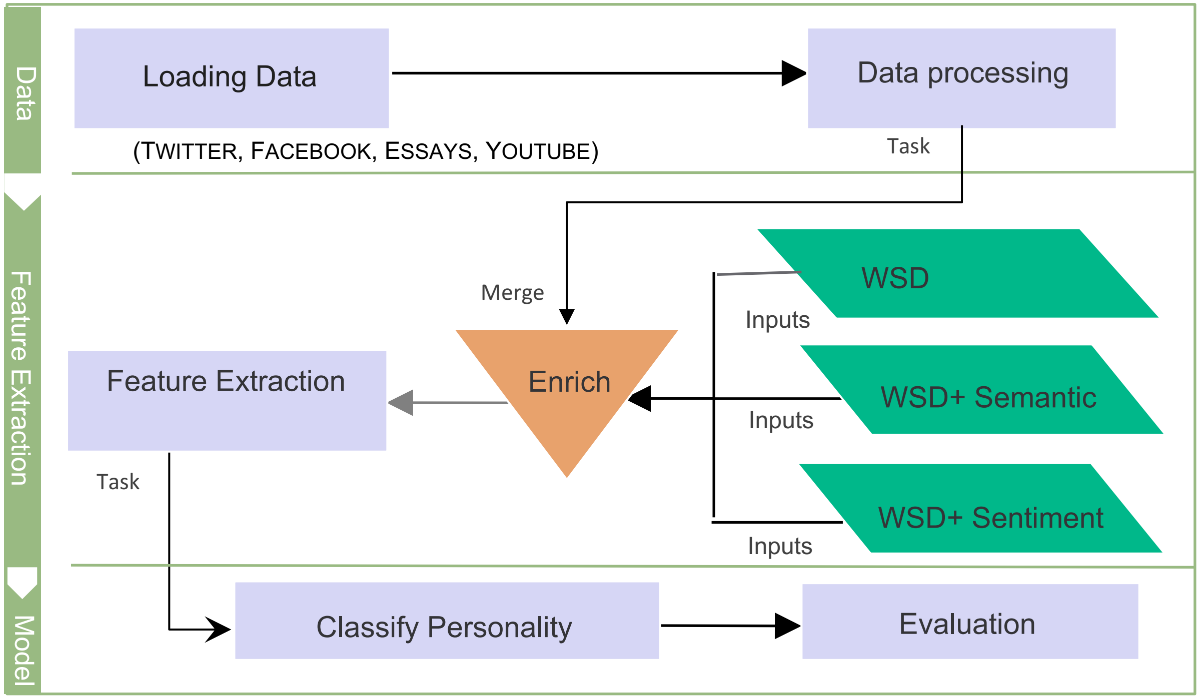
# Data statistics

| Dataset | #Sen | #Word | #Users | Non-standard words |
|---|---|---|---|---|
| TWITTER | 145.7 | 216.8 | 153 | 51.27% |
| FACEBOOK | 67.1 | 78.3 | 250 | 23.3% |
| ESSAYS | 48.8 | 15.3 | 2469 | 30.85% |
| YOUTUBE | 41.7 | 29.5 | 404 | 8.05% |

The number of sentences (#Sen), the number of words (#Word), and the number of users (#Users). Non-standard words may be either out -of-vocabulary tokens (e.g., tmrw for 'tomorrow') or in-vocabulary tok ens (e.g., wit for with in 'I come wit you').

# System workflow

- Implemented using UIMA framework & DKPro
  - Easy to add/remove different modules




https://dkpro.github.io/

# Feature Extraction

| ID | Description |
|---|---|
| WORD | Word-level features. |
| WN-WORD | Word-level features in which only words that present in WordNet are used. |
| WN-MFS | Sense-level features based on the most frequent sense algorithm. |
| WN-S-LESK | Sense-level features based on the Simplified Lesk algorithm. |
| S_SENSE | WordNet semantic label (or WordNet supersense) features. |
| SENTI | Three sentiment features including posscore, negscore, and neuscore. |

# Word Sense Disambiguation

- Why applying WSD?
  - “Neurotic” and “extrovert” people use the emotion words significantly differently.
    - Neurotic people use more 1st person single pronouns
      - While less positive emotional words.
  - “Openness” people use more abstract concepts
- How to apply WSD?
  - Current WSD systems perform an extremely poor performance on low frequent senses
    - Postma et al. (2016)

UMEÅ UNIVERSITET

# Word Sense Disambiguation

- Applying WSD to expand dataset with WN Lexicographer files (WN supersenses)
  - With two different WSD algorithms (i.e., MostFreq and SimLesk).
  - Adding by concatenating or as another feature

She loves Java coffee and Java programming

WSD.MostFreq.Lex.Concate

She love%1 **Java.noun.food** coffee and **Jav a.noun.communication** programming

WSD.MostFreq.Lex.Another

She love%1 **Java%2 noun.food** coffee and **J ava%3 noun.communication** program ming

WSD.Lesk.Lex.Concate

WSD.Lesk.Lex.Concate

- Applying sentiment lexical resource (i.e., SentiWordNet)

She loves Java coffee and Java programming

WSD.MostFreq.Lex.Another.Senti

She love%1 **Java%2 noun.food** coffee and **Ja va%3 noun.communication** programming #**posscore** #**negscore** #**neuscore**

WSD.Lesk.Lex.Another.Senti

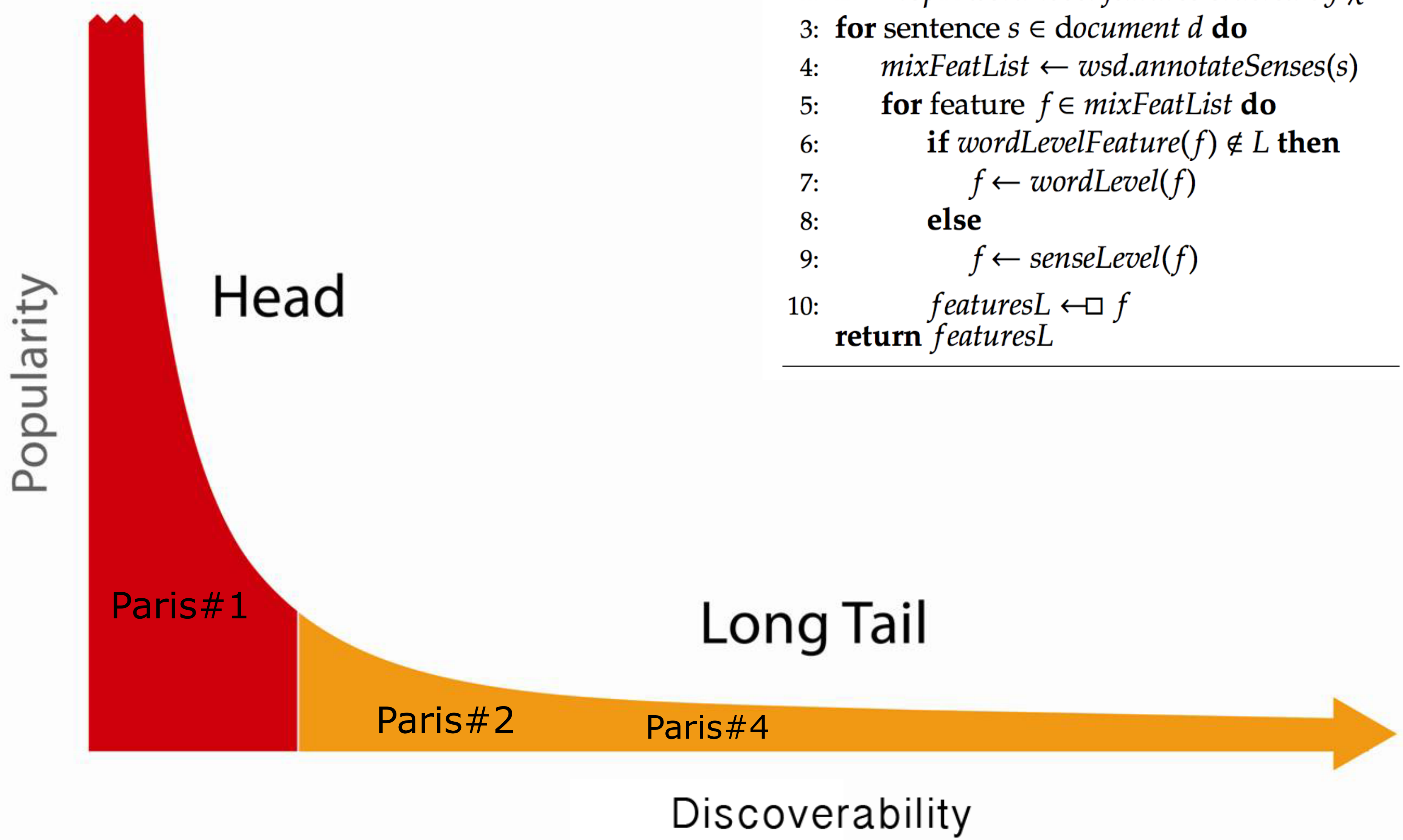
# 4. *Selective.WSD*

**Procedure 1** *Selective.WSD*

**Input:** a word-level document.

**Output:** a selective mixture of word-level and sense-level feature list.

```
featuresL ← initialize an empty list
L ← topK word-level features ordered by χ²
for sentence s ∈ document d do
    mixFeatList ← wsd.annotateSenses(s)
    for feature f ∈ mixFeatList do
        if wordLevelFeature(f) ∉ L then
            f ← wordLevel(f)
        else
            f ← senseLevel(f)
        featuresL ←□ f
return featuresL
```



# 3. Experimental Results

# WSD vs Non.WSD

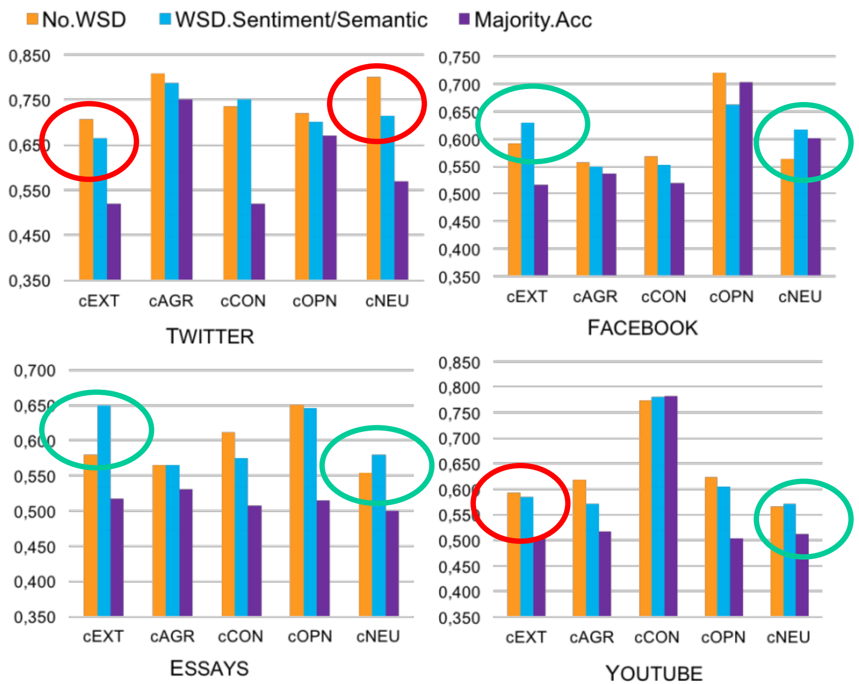



Figure 2: A comparison between not-using WSD (i.e, No.WSD) versus using WSD in a combination with sentiment/semantic features (i.e., WSD.Sentiment/Semantic) in the four datasets. The majority accuracy (i.e., Majority.Acc) is the accuracy when we predict all test instances to a major class.

# Which features work most?

- The restriction to WordNet only words
  - works in 10/24 ≈ 41% of the cases, especially on ESSAYS dataset
- WSD.Sentiment/Semantic
  - improves **extraversion** and **neuroticism** ¾ cases

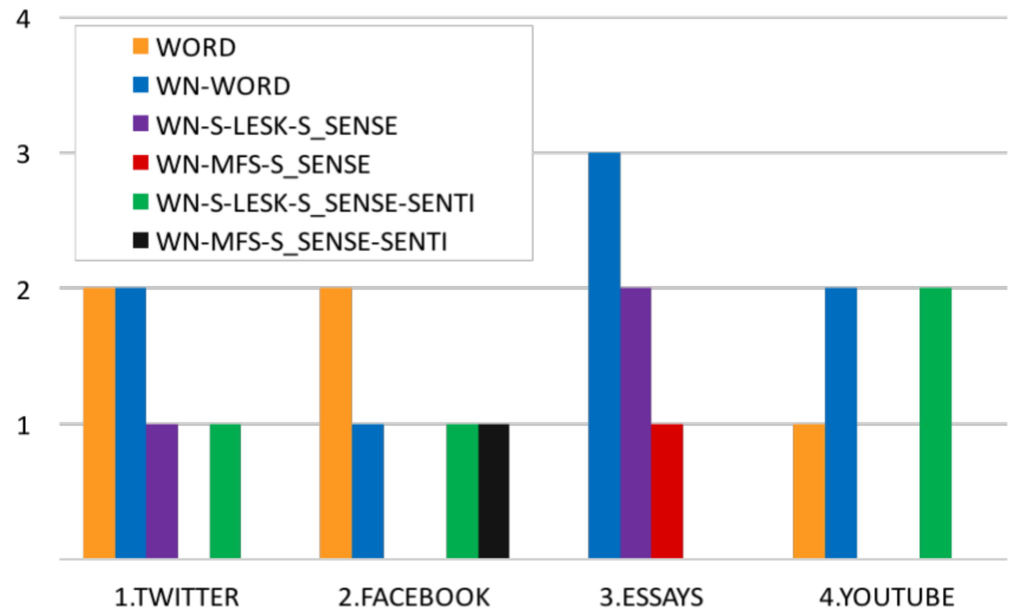



Figure 3: The overall number of times that each feature setting achieves the best performance in the four datasets.

# Selective.WSD vs All.WSD

- The ***Selective.WSD*** method works better than the normal WSD method
- We increase the number of topK features, the performance will drop.

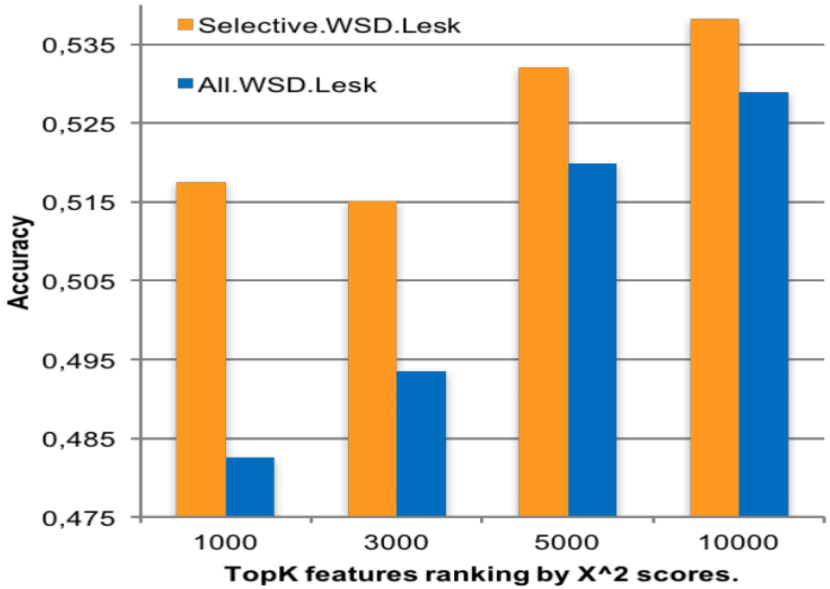



Figure 4: A test on cEXT personal trait of ESSAYS dataset to compare between Selective.WSD and All.WSD.

# Comparison with the SOTA results

Table 3: Performance in comparison with the state-of-the-art results on the FACEBOOK dataset.

| Trait | Majumder et al. (2017) | Ours (Majority.Acc) |
|---|---|---|
| cOPN | 62.68 | **72.10** (70.40) |
| cCON | 57.30 | 56.80 (52.00) |
| cEXT | 58.09 | **62.10** (38.40) |
| cAGR | 56.71 | 55.80 (53.60) |
| cNEU | 59.38 | **61.70** (39.60) |
| Avg | 58.83 | 58.64 (50.80) |

Deep Learning

# Impact of WSD on APC

- WSD does not generally lead to an improvement in classification results
- However
  - In contrary to previous beliefs (Sanderson, 1994; Gonzalo et al., 1998), the performance of the WSD algorithms is not the major issue
  - Rather, it is the reduction of the representative scope of bag-of-words

# Impact of WSD on APC

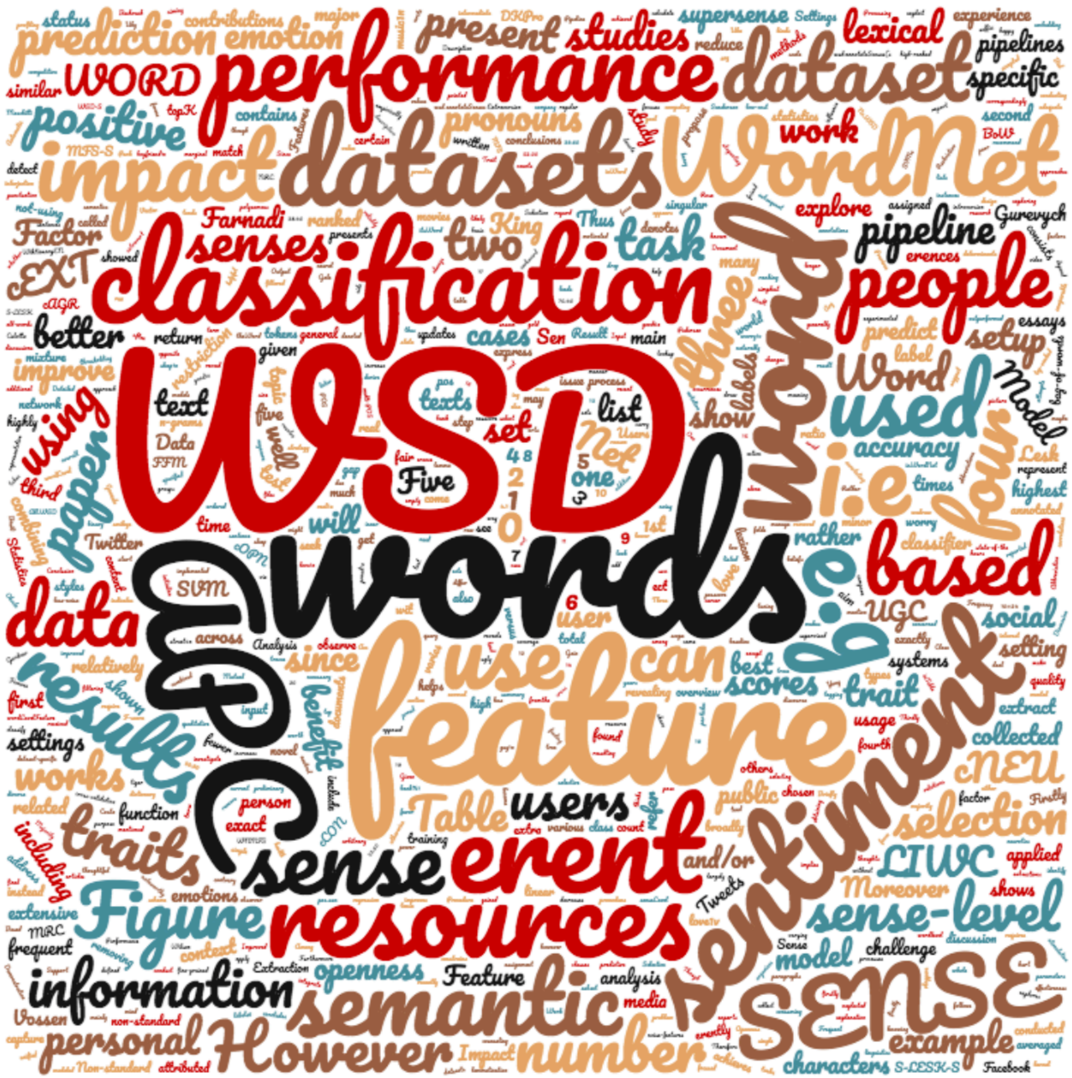
- Rather, it is the reduction of the representative scope of bag-of-words:
  - In the WN-WORD setup, the word **worry** is ranked to predict ***extraversion*** with chi = .007,
  - While the sense **worry%1v** is ranked to predict ***introversion*** with chi = -.004.
- While the effect of WSD itself in a BoW setup is marginal, we observe that the WSD quality is rather high

| WORD | $\chi^2$ | WN-WORD | $\chi^2$ |
|---|---|---|---|
| love | .012 | love | .026 |
| boyfriend | .008 | music | .010 |
| ’d | .008 | sleep | .009 |
| me | .007 | assignment | .009 |
| so | .006 | proud | .008 |
| people | .006 | boyfriend | .007 |
| much | .005 | worry | .007 |
| we | .005 | people | .007 |
| thinks | .005 | awkward | .007 |

| WN-MFS | $\chi^2$ | WN-S-LESK | $\chi^2$ |
|---|---|---|---|
| $\text{love}_1 v$ | .016 | $\text{love}_1 v$ | .017 |
| $\text{music}_1 n$ | .009 | $\text{assignment}_1 n$ | .009 |
| $\text{guy}_1 n$ | .009 | $\text{sleep}_1 v$ | .008 |
| $\text{good}_1 a$ | .009 | $\text{street}_4 n$ | .007 |
| $\text{proud}_1 a$ | .008 | $\text{love}_1 n$ | .006 |
| $\text{assignment}_1 n$ | .008 | $\text{sleep}_1 n$ | .006 |
| $\text{boyfriend}_1 n$ | .008 | $\text{music}_1 n$ | .005 |
| $\text{real}_1 a$ | .006 | $\text{good}_6 a$ | .005 |
| $\text{sleep}_1 v$ | .006 | $\text{proud}_3 a$ | .004 |

Table 4: The highest ranked features for Extraversion on the ESSAYS dataset, averaged across the 10 cross-validation folds, using the $\chi^2$ feature selection.

# Why GWC2018?




UMEÅ UNIVERSITET

# Conclusion

- Main contributions:
  - **WSD** and **semantic** and **sentiment** information to pose an improved performance in APC
  - Using a dictionary (e.g., WordNet, WiktionaryEN) to remove noise-features often works well in most datasets.
  - Applying WSD alone, in general, does not work in APC
    - Especially on not-well-written UGC data.
    - Our proposed Selective.WSD works better than the basic WSD
  - Through away the previous beliefs on performance of WSD
    - The performance of the WSD algorithms is the major issue for stagnating performance (Sanderson, 1994; Gonzalo et al., 1998)
    - Rather:
      - (1) the reduction of the representative scope of bag-of-words
      - (2) the reduction of the impact of multi-POS words (since those are assigned different senses)

Thank you