On the Empirical Evaluation of Author Identification Hybrid Method

Seifeddine Mechti[1], Maher Jaoua[2],Rim Faiz[1, 3], Lamia HadrichBelguith[2] and Bassem Bsir[2]
[1]LARODEC Laboratory, ISG of Tunis B.P.1088, 2000 Le Bardo, Tunisia
[2]ANLP Group, MIRACL Laboratory, University of Sfax, 3018,Sfax Tunsia
[3]IHEC of Carthage, 2016 Carthage Présidence , Tunisia

1. Preliminaries

Introduction

- "Author attribution consists in identifying the author, one of a list, who wrote a particular anonymous text [Stamatatos et al., 2014].

Objective

- In this task : “was a particular text written by a well-defined author?”.
- Build HyTAI system (Hybrid Tool for author Identification).

2. Hybrid Method

- **L’hybridation** : Stylistic analysis + Statistical analysis

→ Statistical Analysis : Delta Rules

→ Stylistic Analysis: Lexical features, syntactic features, Caracters and n-grams

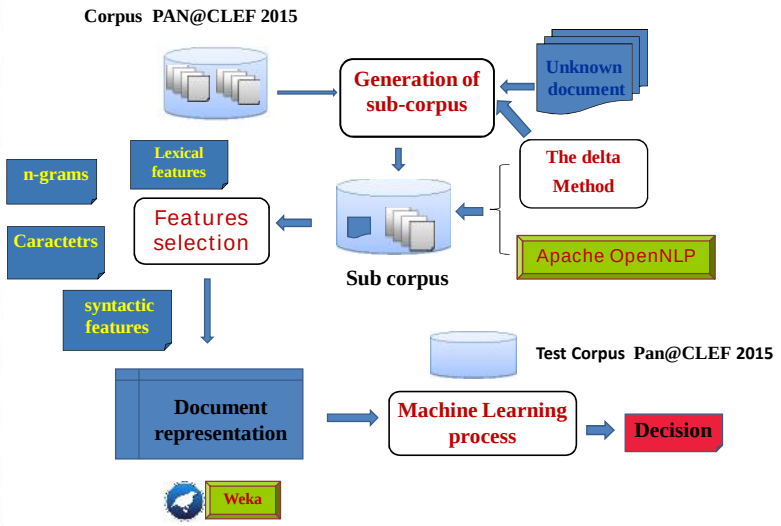
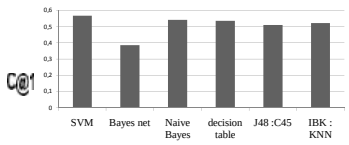
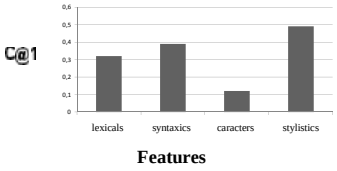


Fig.1. Architecture of HyTAI System

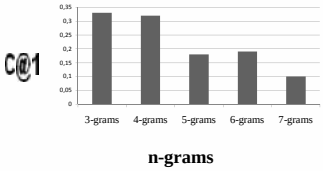
3. Experiments



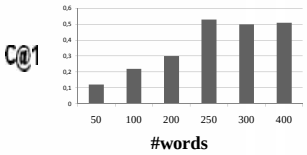
a) The accuracy of different classifiers



b) The c@1 performance stylistic features



c) The c@1 performance according to the n-grams



d) The c@1 performance according to the number of words

Fig.2. Author Identification Histograms

4. Results

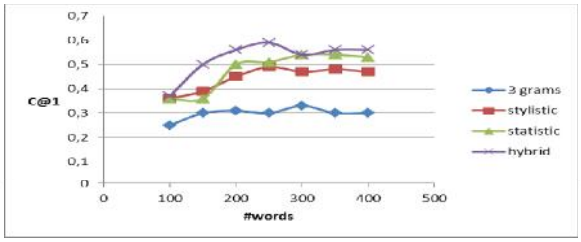


Fig.3. The C@1 Performance of different features according to words number

- SVM classifier : the best results
- The best c@1 obtained : number of terms m=250
- The HyTAI system : 0.59 c@1

5. Conclusion

Results : the interest of hybridization and the importance of statistical features

Future work

To build a framework for intrinsic plagiarism detection based on author identification