

Semantic Annotation with RescoredESA: Rescoring Concept Features Generated From Explicit Semantic Analysis

Zhuoren Jiang
Dalian Maritime University

Miao Chen*
Indiana University Bloomington

Xiaozhong Liu
Indiana University Bloomington

INTRODUCTION

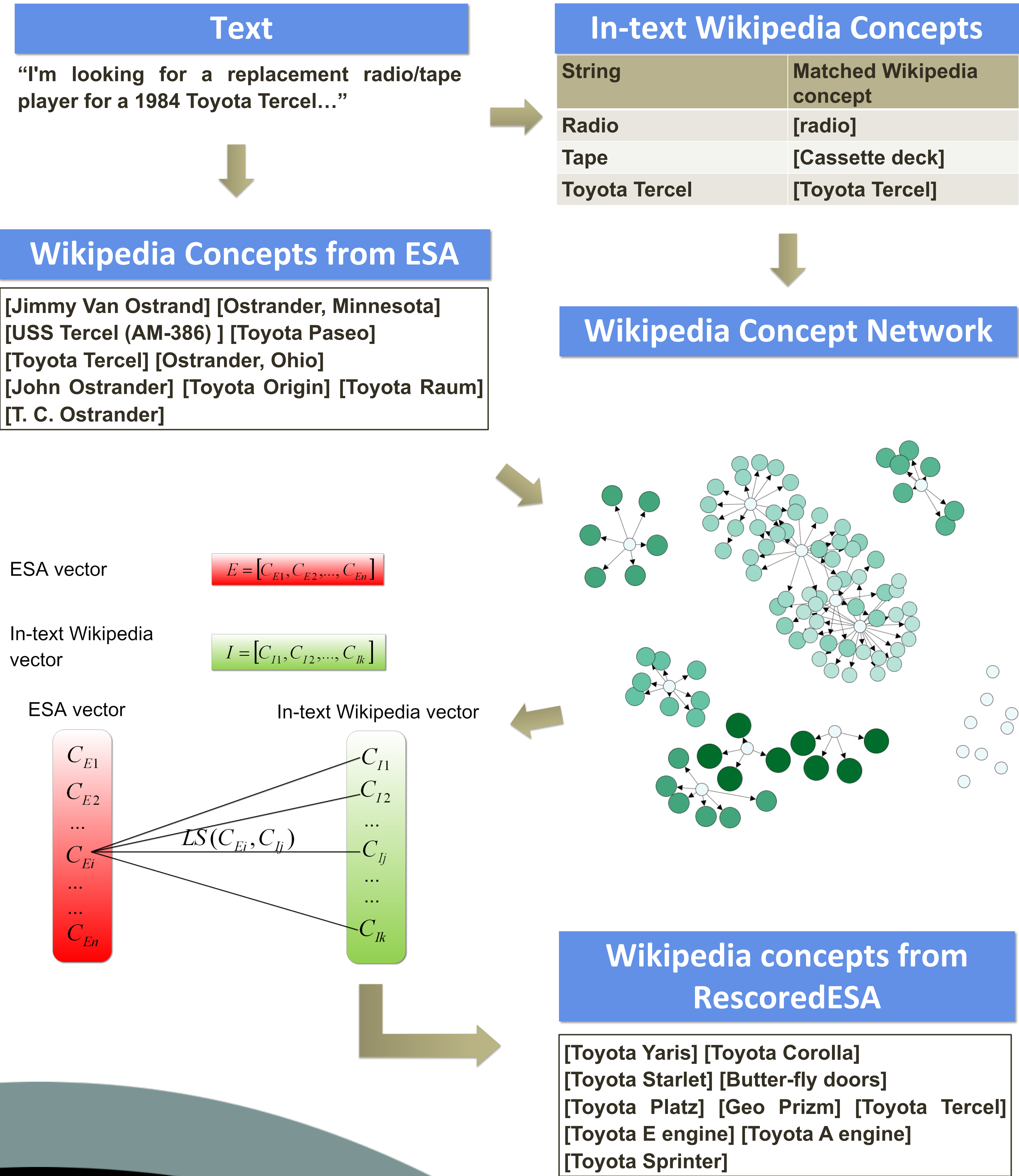
Annotating text with a set of concepts reveals semantically important entities and thus is an important step in semantic annotation. This study we try to extend a semantic annotation approach, Explicit Semantic Analysis (ESA), by enriching it with context and knowledge base.

ESA formula: $v_E = v_d \bullet M^T$

RescoredESA:

$$r_i = \frac{e_i + \alpha \left(\frac{\sum_{j=1}^K LS(C_{E_i}, C_{I_j})}{k} \right)}{1 + \alpha}$$

THE RESCORED ESA APPROACH



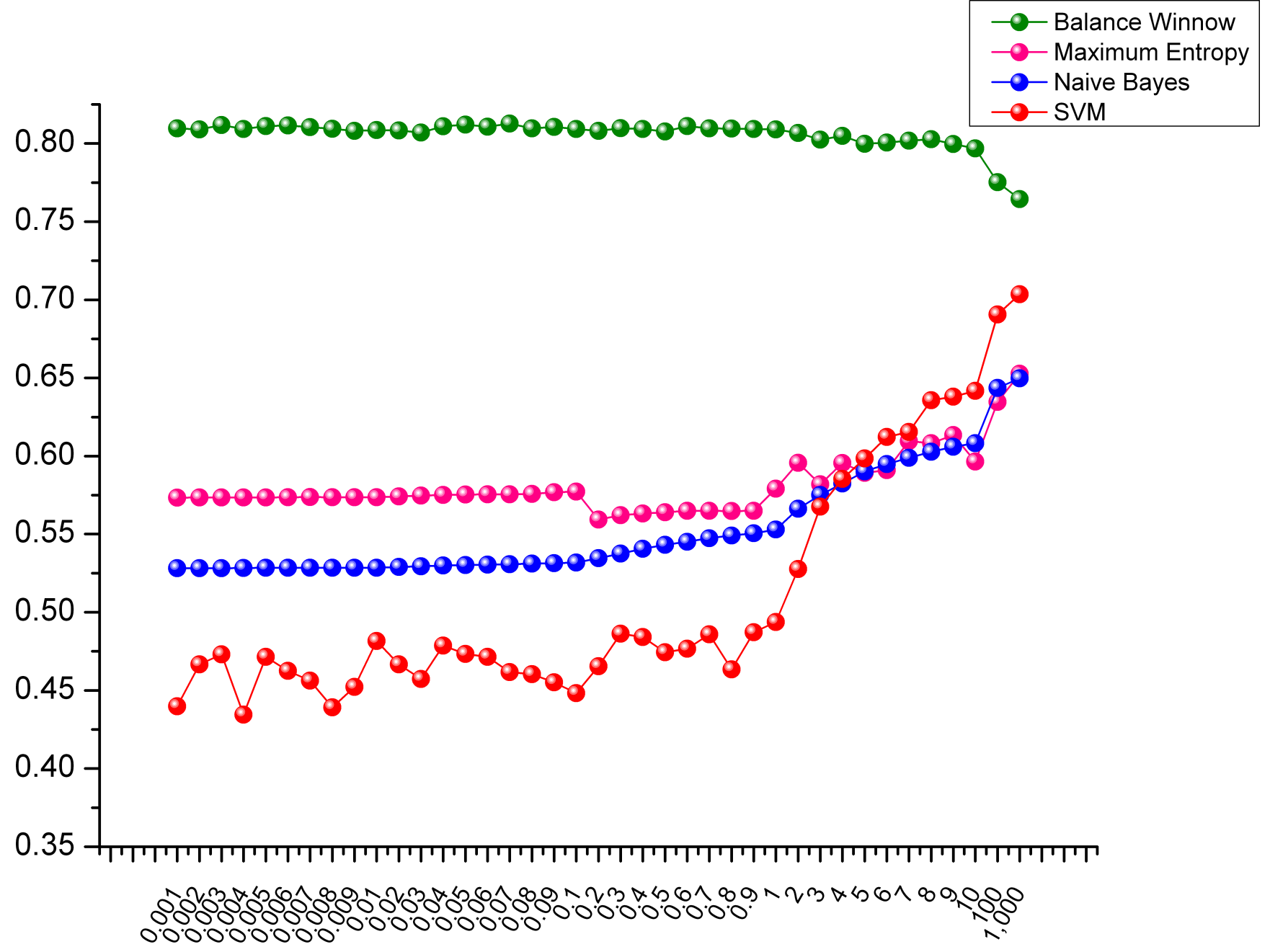
EXPERIMENT

In order to examine performance of RescoredESA, we use the 20 newsgroups data, a widely used news data set for text classification. RescoredESA, along with other comparable approaches, are evaluated using their performance on classifying 20 newsgroup - that is, first representing text with these approaches and then perform text classification based on the representation results.

RESULTS

The results show that for the three text categorization methods other than BalanceWinnow, the general trend is that: Concat (BOW + RescoredESA) > BOW > InText > RescoredESA > ESA. For BalancedWinnow, the order is similar but with ESA and RescoredESA reversed: Concat(BOW + RescoredESA) > BOW > InText > ESA > RescoredESA.

Parameter α



Parameter γ

