ARTICLE IN PRESS

JID: KNOSYS [m5G;July 20, 2016;6:45]

Knowledge-Based Systems 000 (2016) 1–13

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A generalized framework for anaphora resolution in Indian languages

Utpal Kumar Sikdar, Asif Ekbal*, Sriparna Saha

Department of Computer Science and Engineering Indian Institute of Technology Patna, India

ARTICLE INFO

Article history:
Received 16 May 2015
Revised 27 June 2016
Accepted 28 June 2016
Available online xxx

Keywords:
Multiobjective optimization (MOO)
Single objective optimization (SOO)
Conditional random field (CRF)
Support vector machine (SVM)

ABSTRACT

In this paper, we propose a joint model of feature selection and ensemble learning for anaphora resolution in the resource-poor environment like the Indian languages. The proposed approach is based on multi-objective differential evolution (DE) that optimises five coreference resolution scorers, namely MUC, BCUB, CEAFM, CEAFE and BLANC. The main goal is to determine the best combination of different mention classifiers and the most relevant set of features for anaphora resolution. The proposed method is evaluated for three leading Indian languages, namely Hindi, Bengali and Tamil. Experiments on the benchmark datasets of ICON-2011 Shared Task on Anaphora Resolution in Indian Languages show that our proposed approach attains good level of accuracies, which are often better with respect to the state-of-the-art systems. It achieves the F-measure values of 71.89%, 59.61%, 52.55% 34.45% and 72.52% for MUC, BCUB, CEAFM, CEAFE and BLANC, respectively, for Bengali language. For Hindi we obtain the F-measure values of 33.27%, 63.06%, 49.59%, 49.06% and 55.45% for MUC, BCUB, CEAFM, CEAFE and BLANC metrics, respectively. In order to further show the efficacy of our proposed algorithm, we evaluate with Tamil, a language that belongs to a different family. This shows the F-measure values of 31.79%, 64.67%, 46.81%, 45.29% and 52.80% for MUC, BCUB, CEAFM, CEAFE and BLANC metrics, respectively. Experiments on Dutch show the F-measure values of 17.67%, 74.43%, 58.08%, 59.21% and 55.58% for MUC, BCUB, CEAFM, CEAFE and BLANC metrics, respectively.

© 2016 Published by Elsevier B.V.

1. Introduction

The anaphora or coreference resolution is the process of identifying noun phrases that denote the same real world entities [1,21,30,46]. Many crucial application areas of Natural Language Processing (NLP), for example, Information Extraction [11], Text Summarization [36], Question Answering [16], Text Retrieval [30], Machine Translation [14,18,23,24] etc. require the task of coreference resolution to be performed. There have been significant amount of works in this area, but most of these focus mainly for the languages such as English [5,26,27], due to the availability of different lexical resources and large corpora such as ACE [44] and OntoNotes [45]. We use a state-of-the-art English coreference resolution system, BART [42] for our task, and so proper adaptation had to be carried out for the resource-scare Indian languages.

India is a multilingual country with great cultural and linguistic diversities. There are 22 officially spoken languages in India. However, there has not been significant number of works on anaphora resolution involving Indian languages due to its resource-constrained nature, i.e., annotated corpora and other lexical resources are not readily available in the required measure. Literature shows that the existing works on anaphora resolution involving Indian languages are a few in number, and they cover only few of the languages like Bengali, Hindi and Tamil [2,33,34,39]. However, based on these works it is difficult to get a comprehensive view of the research on anaphora resolution related to Indian languages because each of these was developed using the self-generated datasets, and evaluation setups are not the same. Therefore, it is not fair to compare between the algorithms reported in these works.

The first benchmark setup for anaphora resolution involving Indian languages was established in ICON-2011 NLP Tools Contest on Anaphora Resolution[1]. Out of the six participating teams, four addressed the issues of anaphora resolution in Bengali, and one each for Hindi and Tamil. Apart from these some other works have been reported in [9,31–33] for anaphora resolution, especially for Bengali and Hindi. A system for anaphora resolution in Bengali is reported in [32], where various models for mention detection were developed, and their impacts on anaphora resolution were reported. In another work, Senapati and Garain [31] have shown how an off-the-shelf anaphora resolution system can be effectively

* Corresponding author. fax.: +91 612 2277383.
E-mail addresses: utpal.sikdar@iitp.ac.in (U. Kumar Sikdar), asif@iitp.ac.in, asif.ekbal@gmail.com (A. Ekbal), sriparna@iitp.ac.in (S. Saha).

[1] http://ltrc.iiit.ac.in/icon2011/contests.html

http://dx.doi.org/10.1016/j.knosys.2016.06.033
0950-7051/© 2016 Published by Elsevier B.V.

Please cite this article as: U. Kumar Sikdar et al., A generalized framework for anaphora resolution in Indian languages, Knowledge-Based Systems (2016), http://dx.doi.org/10.1016/j.knosys.2016.06.033

used for Bengali. A more recent study on anaphora resolution in Bengali can be found in [33]. In recent times a generic framework for anaphora resolution in Indian languages has been reported in [9].

Literature survey shows that a large collection of multiobjective evolutionary approaches are developed to solve several real-life problems in recent years. In [38], a multiobjective symbiotic organism search based approach is developed for solving time-cost-labor utilization trade-off problem. Similarly in [48], a performance comparison is conducted between generational and steady-state asynchronous multi-objective evolutionary algorithms for computationally-intensive problems. In [10], a multiobjective evolutionary approach is developed for named entity recognition in chemical domain. The paper proposes a joint model of feature selection and parameter optimization of classifier. The technique is based on multobjective genetic algorithms. The algorithm that we propose here is a joint model, but it performs ensemble learning for mention detection and feature selection for coreference resolution. Unlike [10], we do not make use of genetic algorithm, and use differential evolution to build our models. Genetic algorithm and differential evolution are two different optimization strategies. Two problems, NER solved in [10] and coreference resolution solved here are completely different.

In anaphora resolution there does not exist any globally accepted metric for measuring the performance, and each of Muc, Bcub, ceaf, Blanc metrics exhibits significantly different behavior. The most popular coreference resolution scorer is Muc [43] which is defined based on the number of common links between reference (true) and system prediction (response). The Muc scorer ignores single mention entity as this does not contain any link in it. An error which is encountered because of linking two large groups could do more damage compared to one link connecting two small groups. Bagga and Baldwin [4] present their B-cubed evaluation to deal with these issues. While the B-cubed metric fixes some of the shortcomings of the Muc scorer problem, it has also its own problems. The Bcub scorer is based on entities containing the mentions. But when precision or recall is calculated by comparing entities, it may be the case that an entity may be considered more than once. For this problem, Luo introduced one new metric called Constrained Entity-Alignment F-Measure (CEAF) [20] based on the aligned entities in key (true) and response (system output). In [28], authors have proposed another scorer called BiLateral Assessment of Noun-phrase Coreference (Blanc). This scorer takes into account the number of coreference and non-coreference links. Because of these different views, systems optimized according to one metric often tend to perform poorly with respect to the others, and therefore comparing performance among different systems introduces difficulties. Hence, we decide to optimize all the well-known metrics simultaneously.

In anaphora resolution an early attempt for explicit optimization was proposed in [22], where no significant performance improvement was observed over the baseline that was constructed with all the available features. A systematic effort of manual feature selection on the benchmark datasets was carried out in [40], where total 600 features were evaluated. The very first attempt for automatic optimization of anaphora resolution was carried out in [15], where the usability of evolutionary genetic algorithms (GAs) is investigated. The authors have suggested that such a technique may yield significant performance improvements over the Muc-6/7 datasets. The concept of multi-objective optimization (MOO) for feature selection in anaphora resolution has been addressed in [29] for the English language, where GA was used as an optimization technique. A recent study for feature selection in anaphora resolution for Bengali can be found in [33]. In contrast to [29], we have not made use of GA as an optimization technique, and performed feature selection for a non-English language. The algorithm proposed here is based on differential evolution (DE), which is not similar to GA. The method proposed in [33] is based on single objective optimization (SOO). While SOO concentrates in optimizing only one function, MOO simultaneously optimizes more than one objective functions. The output of MOO produces a set of solutions on the Pareto optimal front. Each of these solutions is equally important from the algorithmic point of view. Hence, one interesting aspect of our algorithm is that depending upon the need user can pick up any solution. In contrast to all these previous works, we propose here a method that performs feature selection and ensemble learning jointly. The algorithm can simultaneously yield the best ensemble model for mention detection and the most relevant features for anaphora resolution. In none of the previous approaches these two problems were modeled in this way. Other existing works such as [31,32] do not address the problems of feature selection and/or ensemble learning.

The key contributions of this work are four-fold, *viz.* (i) building a generic framework for anaphora resolution in a less-resourced environment such as for Indian languages; (ii) adapting an existing state-of-the-art English co-reference resolution system for one European language (i.e., Dutch) which has completely different orthography and characteristics; (iii). joint model for ensemble learning and feature selection, especially for this kind of application; and (iv) the use of evolutionary algorithm such as DE to simultaneously optimize the coreference metrics such as Muc, Bcub, Ceafm, Ceafe and Blanc using the concepts of multiobjective optimization.

The rest of the paper is structured as follows. Section 2 elaborately discusses our proposed methods that include technique for developing mention detection, technique for feature selection, and the features used for anaphora resolution. In Section 3 we report detailed evaluation results along with the necessary analysis. Finally, in Section 4, we conclude the paper.

2. Proposed method

In this section, we present a joint model for feature selection and ensemble construction. Performance of anaphora resolution greatly depends on the mentions (also called markables). Mentions are basically the noun phrases that form the core parts in anaphoric relations. A good performing mention detector might be useful to develop an accurate model for anaphora resolution. Here we build several models for mention detection using heuristics and machine learning (c.f Section 2.1). As machine learner we use Conditional Random Field (CRF) [17] and Support Vector Machine (SVM)[41]. All these models of mention detection are combined together to further increase the performance. The noun phrases extracted from this combined model are used as the candidate markables to anaphora resolution. Anaphora resolver is trained with a set of features which were implemented without using much domain-dependent resources and/or tools. We propose a framework based on multiobjective optimization (MOO) that determines the best ensemble of mention detectors and the best feature combination for anaphora resolution that optimizes several evaluation scorers. In particular we optimize Muc, Bcub, Ceafm, Ceafe and Blanc. We generate N mention detection models ($N = 10$) for mention detection and implement M features ($M = 18$) for anaphora resolution. The proposed approach is general in nature, and therefore applicable to many less-resourced languages. An overall architecture of the proposed model is shown in Fig. 1.

2.1. Models for mention detection

Mention or markable denotes the terms or phrases that participate in anaphoric relations. Accuracy of mention detection plays an important role in the overall performance of anaphora resolution. We develop the following 10 models for mention detection.

Please cite this article as: U. Kumar Sikdar et al., A generalized framework for anaphora resolution in Indian languages, Knowledge-Based Systems (2016), http://dx.doi.org/10.1016/j.knosys.2016.06.033

Download English Version:

https://daneshyari.com/en/article/4946543

Download Persian Version:

https://daneshyari.com/article/4946543

Daneshyari.com