Anaphora Resolution in Bengali

A Rule Based Approach

Related Work

- Using an off-the-shelf AR system : GuiTAR-based Pronominal Anaphora Resolution in Bengali - 2013, ISICAL

  [https://www.aclweb.org/anthology/P13-2023]

- CRF, 2nd stage uses BART : Anaphora Resolution for Bengali - An Experiment with Domain Adaptation - 2013, IIT-Patna & University of Trento

  [http://www.scielo.org.mx/pdf//cys/v17n2/v17n2a4.pdf]

- CRF, Decision Tree Algorithm : Anaphora Resolution for Bengali, Hindi, and Tamil Using *RandomTreeAlgorithm* in Weka

  [https://www.academia.edu/3275626/Anaphora_Resolution_for_Bengali_Hindi_and_Tamil_Using_RandomTree_Algorithm_in_Weka]

Key

Anaphora Resolution in Bangla Language, IJCA 2016, NY, USA

Tazbeea Tazakka, Md. Asifuzzaman, Sabir Ismail - Shahjalal University of Science and Technology

- pronouns and verbs are considered as anaphors with nouns as antecedents
- first attempt, check if ∃ any relation among the classifications of pronouns and verbs with those of nouns OR which types of pronouns refer to which type of nouns
- findings: anaphor and antecedent in Bangla matches with each other in some factors - number, gender, person
- if it is a personal pronoun, then status of person can be differentiated such as honorable, normal and negotiable
- reports 80% accuracy - take it with a pinch of salt

[https://pdfs.semanticscholar.org/086c/b74ff9f2dcb915ec84e3be69fb567d859ff6.pdf]

Dependencies

Rule based anaphora resolution in Hindi, Singla et. al - Thapar University, ICCIDS 2017 [https://ieeexplore.ieee.org/document/8272666]

- Entity Resolution (ER) and pronominal forms
- Use Karaka Relations : Reflexive, Spatial, Relative, First, Second, etc.

What about Bengali? - Availability of Dependency Parsers

1. ISICAL - Modified MALT parser - Issues: format & encoding
2. LTRC - Work in progress

Rule-based Pronominal Anaphora

identify pronouns

identify all possible antecedents for the particular pronoun

rule out antecedents not satisfying certain constraints

in an ideal scenario, “and then there were ~~n~~one”

features used -

1. *Number*
2. *Person*
3. *Status*
4. *Gender*
5. *POS*
6. *Morph* ←

Number

singular_pronouns = ["তুমি", "তুই", "সে", "আপনি", "তিনি", "তার", "তোমার", "তোর", "আপনার", "আমার", "ওর"]

plural_pronouns = ["তোমরা", "তোরা", "তারা", "আপনারা", "তোমাদের", "তোদের", "আপনাদের", "আমাদের", "ওদের"]

Examples

Singular : রাম বই পড়ছে । সে খুব ভালো । [Ram book read. He very good.]

Plural : অভিজ্ঞানরা দার্জিলিং বেড়াতে গেছে । তারা পরশু ফিরবে । [(Abhigyan & co.) Darjeeling tour gone. They (day after tomorrow) return.]

Person

first_person_pronouns = ["আমি", "আমার"]

second_person_pronouns = ["তুমি", "তুই", "আপনি", "তোমার", "তোর", "আপনার"]

third_person_pronouns = ["সে", "তিনি", "তার", "তারা", "ওদের", "ওর"]

Example

Third Person : শেলি ছবি আঁকছে । সেটি রঙিন । [Shelly picture drawing. It colourful.]

First Person : আমার নাম যদু। আমি কলকাতায় থাকি। [My name Jodu. Me (in Kolkata) stay.]

Status

status_formal = ["আপনি", "তিনি", "আপনার", "আপনারা", "আপনাদের"]

status_informal = ["তুমি", "সে", "তার", "তোমার", "তোমরা", "তোমাদের"]

status_close = ["তুই", "তোর", "তোরা", "তোদের", "ওদের", "ওর"]

Honorific features

Examples

Formal : রামবাবু সমাজের মাথা। ওনাকে সকলে ভয় পায়। [Ram-babu{honorific marker} society's head. He everyone (feared by) feel.]

Informal : রাম বই পড়ছে। সে খুব ভালো। [Ram book read. He very good.]

System

a feature extraction and a disambiguation module

accept a set of sentences

identifies anaphors - identifies their possible antecedents

rules out antecedents - picks the best possible option

Note : NO *cataphoras* detected

testing :

1. manually tagged Bengali data created using Google’s transliteration tools
2. large corpus of pre-tagged Bengali newspaper articles

DEMO

Issues

- gender feature
- free word order
- constituency parsers and Hobb’s algorithm
- strong assumptions about nature of antecedents
- what about cataphora*?*
- lack of semantic information
- entity recognition
- real-world data
- abstract anaphora

Done.