

A RULE-BASED APPROACH FOR ANAPHORA RESOLUTION IN BENGALI

Tathagata Raha
2018114017

WHAT IS ANAPHORA RESOLUTION?

- ▶ **Anaphora Resolution == the problem of resolving what a pronoun, or a noun phrase refers to.**
- ▶ 1) John helped Mary.
2) He was kind.

ABSTRACT

- ▶ This project is a rule-based approach to resolve pronominal anaphors in Bengali
- ▶ Takes a tagged sentence as input
- ▶ Output the pronominal anaphors, possible antecedents and the antecedents to the corresponding anaphors.

RELATED WORK

- GuiTAR-based Pronominal Anaphora Resolution in Bengali by Apurabalal Senapati and Utpal Garain - 2013, ISICAL

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

- Anaphora Resolution for Bengali using BART - An Experiment with Domain Adaptation - 2013, IIT-Patna & University of Trento

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

RULE BASED APPROACHES IN INDIAN LANGUAGE

A rule-based approach for Hindi has been developed by LTRC, IIIT Hyderabad which uses dependency parsing for disambiguating the pronominal anaphors and the referents.

KEY IDEA BEHIND MY APPROACH

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

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

IMPLEMENTATION

INPUT

- ▶ Here at first, we take a Bengali sentence as input where each token will contain POS, number, person and honorific information.

INPUT

- ▶ For example:
- ▶ রাম will be tagged as রাম/NP/S/T/I
- ▶ The first tag contains the word in UTF-8 format
- ▶ The second tag specifies the POStag of the word
- ▶ The third tag gives the number information.
 - S-for singular
 - P-for plural
 - NA-for not applicable
- ▶ The fourth tag gives the person information
 - F-for first person
 - S-for second person
 - T-for third person
 - NA-for not applicable
- ▶ The fifth tag gives the honorific information
 - F-Formal
 - I-Informal
 - C-Close

STEPS

- ▶ Identify all the pronouns i.e. all the tokens with PRP or PRP\$ tag
- ▶ Identify all the antecedents.
- ▶ Rule out all the possible antecedents that doesn't satisfy certain constraints
- ▶ Features used-
 - ▶ POStag
 - ▶ Number
 - ▶ Person
 - ▶ Status
 - ▶ Morph features

PRONOUN CLASSIFICATION

- ▶ Singular pronouns: ["তুমি", "তুই", "সে", "আপনি", "তিনি", "তার", "তোমার", "তোর", "আপনার", "আমার", "ওর"]
- ▶ Plural pronouns: ["তোমরা", "তোরা", "তারা", "আপনারা", "তোমাদের", "তোদের", "আপনাদের", "আমাদের", "ওদের"]

PRONOUN CLASSIFICATION

- ▶ First person pronouns: ["আমি", "আমার"]
- ▶ Second person pronouns: ["তুমি", "তুই", "আপনি", "তোমার", "তোর", "আপনার"]
- ▶ Third person pronouns: ["সে", "তিনি", "তার", "তারা", "ওদের", "ওর"]

PRONOUN CLASSIFICATION

- ▶ Formal status: ["আপনি", "তিনি", "আপনার", "আপনার", "আপনাদের"]
- ▶ Informal status: ["তুমি", "সে", "তার", "তোমার", "তোমরা", "তোমাদের"]
- ▶ Close status: ["তুই", "তোর", "তোরা", "তোদের", "ওদের", "ওর"]

PRONOUN CLASSIFICATION

- ▶ Pronoun ending with 'টা', 'টি', 'টির', 'টিকে', 'টাকে' refer to non-human entities.

DEMOS

► INPUT:

► রাম/NP/S/T/I বই/NN/S/T/I পড়ছে/VB/S/T/I I সে/PRP/S/T/I খব/JJ/S/T/I ভালো/JJ/S/T/I I

► OUTPUT:

► Pronominal anaphores:

► [{'word': 'সে', 'id': 5, 'POS': 'PRP', 'number': 'S', 'person': 'T', 'honor': 'I'}]

► Possible antecedents:

► [{'word': 'রাম', 'id': 1, 'POS': 'NP', 'number': 'S', 'person': 'T', 'honor': 'I'}, {'word': 'বই', 'id': 2, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}]

► Set of antecedents that matched the constraints for সে :

► [{'word': 'রাম', 'id': 1, 'POS': 'NP', 'number': 'S', 'person': 'T', 'honor': 'I'}]

► VERDICT: Success

- ▶ INPUT:
- ▶ রামের/NP/S/T/I বল/NN/S/T/I আছে/VB/NA/NA/NAI সে/PRP/S/T/I ওটাকে/PRP/S/T/I খব/ADJ/NA/NA/NA ভালোবাসে/VB/NA/NA/NAI সেটি/PRP/S/T/I গোলাকার/VB/NA/NA/NAI
- ▶ OUTPUT:
- ▶ Pronominal anaphores:
- ▶ [{ 'word': 'সে', 'id': 4, 'POS': 'PRP', 'number': 'S', 'person': 'T', 'honor': 'I' }, { 'word': 'ওটাকে', 'id': 5, 'POS': 'PRP', 'number': 'S', 'person': 'T', 'honor': 'I' }, { 'word': 'সেটি', 'id': 8, 'POS': 'PRP', 'number': 'S', 'person': 'T', 'honor': 'I' }]
- ▶ Possible antecedents:
- ▶ [{ 'word': 'রামের', 'id': 1, 'POS': 'NP', 'number': 'S', 'person': 'T', 'honor': 'I' }, { 'word': 'বল', 'id': 2, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I' }]
- ▶ Set of antecedents that matched the constraints for সে :
- ▶ [{ 'word': 'রামের', 'id': 1, 'POS': 'NP', 'number': 'S', 'person': 'T', 'honor': 'I' }]
- ▶ Set of antecedents that matched the constraints for ওটাকে :
- ▶ [{ 'word': 'বল', 'id': 2, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I' }]
- ▶ Set of antecedents that matched the constraints for সেটি :
- ▶ [{ 'word': 'বল', 'id': 2, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I' }]
- ▶ VERDICT: Success

► INPUT:

► রামবাবু/NP/S/T/F সমাজের/NN/S/T/I মাথা/NN/S/T/I | ওনাকে/PRP/S/T/F সকলে/QT/P/T/I ভয়/JJ/S/T/I
পায়/VB/S/T/I |

► OUTPUT:

► Pronominal anaphores:

► [{'word': 'ওনাকে', 'id': 5, 'POS': 'PRP', 'number': 'S', 'person': 'T', 'honor': 'F'}]

► Possible antecedents:

► [{'word': 'রামবাবু', 'id': 1, 'POS': 'NP', 'number': 'S', 'person': 'T', 'honor': 'F'}, {'word': 'সমাজের', 'id': 2, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}, {'word': 'মাথা', 'id': 3, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}]

► Set of antecedents that matched the constraints for ওনাকে :

► [{'word': 'রামবাবু', 'id': 1, 'POS': 'NP', 'number': 'S', 'person': 'T', 'honor': 'F'}]

► VERDICT: Success

► INPUT:

► রাম/NP/S/T/I আর/CONJ/S/T/I যদ/NP/S/T/I গান/NN/S/T/I করছে/VP/S/T/I I তারা/PRP/P/T/I
ভালো/JJ/S/T/I গান/NN/S/T/I গায়/VB/S/T/I I

► OUTPUT:

► Pronominal anaphores:

► [{'word': 'তারা', 'id': 7, 'POS': 'PRP', 'number': 'P', 'person': 'T', 'honor': 'I'}]

► Possible antecedents:

► [{'word': 'রাম', 'id': 1, 'POS': 'NP', 'number': 'S', 'person': 'T', 'honor': 'I'}, {'word': 'যদ', 'id': 3, 'POS': 'NP', 'number': 'S', 'person': 'T', 'honor': 'I'}, {'word': 'গান', 'id': 4, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}, {'word': 'গান', 'id': 9, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}]

► None of the antecedents matched the constraints for তারা

► VERDICT: Failure

► INPUT:

► পাখির/NN/S/T/I বাসা/NN/S/T/I গাছে/NN/S/T/I I সেটির/PRP/S/T/I রং/NN/S/T/I নীল/NN/S/T/I I

► OUTPUT:

► Pronominal anaphores:

► [{'word': 'সেটির', 'id': 5, 'POS': 'PRP', 'number': 'S', 'person': 'T', 'honor': 'I'}]

► Possible antecedents:


► [{'word': 'পাখির', 'id': 1, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}, {'word': 'বাসা', 'id': 2, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}, {'word': 'গাছে', 'id': 3, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}, {'word': 'রং', 'id': 6, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}, {'word': 'নীল', 'id': 7, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}]

► Set of antecedents that matched the constraints for সেটির :

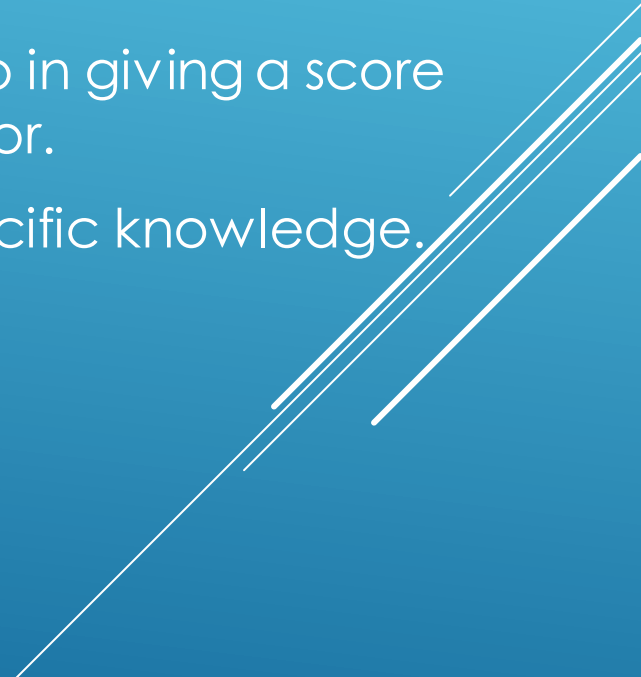
► [{'word': 'পাখির', 'id': 1, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}, {'word': 'বাসা', 'id': 2, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}, {'word': 'গাছে', 'id': 3, 'POS': 'NN', 'number': 'S', 'person': 'T', 'honor': 'I'}]

► VERDICT: Debatable

PROBLEMS

- ▶ Gender
 - ▶ Lack of semantic information
 - ▶ Can only recognize anaphors properly in short amount of sentences.
 - ▶ Abstract anaphors
 - ▶ Real-world data
 - ▶ Anaphors relating to verbs
- 
- A series of white lines of varying lengths and slopes are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

FUTURE WORK

- ▶ Developing a dependency parser for Bengali and by using the depending information would enhance the performance to a great extent
 - ▶ Using neural network models like LSTMs and Siamese nets can help in giving a score to the possible antecedents and mark its relevance with a anaphor.
 - ▶ By capturing the semantic information for a global or domain specific knowledge.
- 
- A series of white lines of varying lengths and slopes are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.