A Survey on Anaphora Resolution

Ishani Sharma
Department of Computer Science Engineering
Jaypee University of Information Technology
Himachal Pradesh, India

Pradeep Kumar Singh
Department of Computer Science Engineering
Jaypee University of Information Technology
Himachal Pradesh, India

ABSTRACT

Anaphora occurs very frequently in written texts and spoken dialogues. Almost all NLP applications such as machine translation, information extraction, automatic summarization, question answering system, natural language generation, etc., require successful identification and resolution of anaphora. Though the significant amount of work has been done in English and other European languages, the computational work, in reference to Hindi, is lagging far behind. In this paper, we present a review of work done in the field of anaphora resolution in Hindi and other languages. There are plenty of techniques which have been developed on anaphora resolution but very less effort has been taken on Hindi language. Here we are discussing various features that includes pronoun resolution, noun resolution, and various methods such as centering, hobbs, gazetter and lappin and leass methods.

Keywords

Anaphora, Anaphora Resolution, Centering, Hobbs, Gazetteer method, Lappin and Leass, Natural Language Processing.

1. INTRODUCTION

Natural language processing (NLP) is basically used to fill the communication gap between the computer and the human being. Artificial intelligence is used in NLP. It has many challenges but anaphora is one of its main challenges. Anaphora is a phenomena is which one expression depends on another expression. Anaphora is also defined as the linguistic expression in which act as reference to other linguistic form that precedes it. The word which is pointing back is called anaphor and to which it refers to is called antecedent. Consider a simple example:

Director of JUIT ordered an investigation against the accounts manager. He claims that the manager is involved in the fees scam.

In the above example, ‘He’ is refer to the ‘director of JUIT’.

The process of identifying the referent is known as ‘Anaphora Resolution’.

The significant amount of research has been done on English and other European languages. But the efficient research on anaphora resolution in Hindi is still pending. Since Hindi is free word language so it leads to many ambiguities.

2. RELATED WORK

A literature survey of existing anaphora resolution approaches is presented in this section. The earlier work done in the anaphora resolution is for English, Hindi, and other European languages. Since very less effort is given on hindi languages

Lakhmani, et. al.[1], present anaphora resolution for Hindi language. Anaphora resolution is a key problem in natural language processing. The researching weight of work has been done in English and other European language, and in Hindi language efficient work is pending. And that work completed in different section. That first section of the present paper review of anaphora resolution work in Hindi language. And the second part they solve problem of syntactic and semantic structure of Hindi pronoun. Then after thread part is define source constraint which will form to the task of anaphora resolution. And in last they perform different experiment on different kind of data set. And also they get results is obtain 71%.

Feng, et. al.[2], discussed about the Chinese Natural Language Interface for Navigation in Mobile GIS. In that “Mobile GIS” and “Voice technology” both combination has improved the intelligent degree of mobile GIS. Natural language quick convert to GIS commands they can possible because of field of scientific research. This paper study of natural language sentences and GIS command understanding method of mobile voice GIS. Using a machine learning method conversion between natural language and GIS commands.

Poesio, et. al.[3], discussed the system’s design using two algorithms implemented in the current version of the system for descriptions and pronoun resolution. Develop one tool GUITAR (General Tool for Anaphora Resolution) In this paper they briefly discuss the architecture and implementation of the system, as well as some preliminary evaluation results.

Khan, et. al.[4], discuss about the natural language processing or define all the language using different country. There are many NLP applications such as machine translation, question answering system, automatic summarization etc. In this paper there are three techniques are use first is statistical machine translation, second parallel corpus, and third one is phrase-based translation. and also we discuss about Indian language to translate a many language like Gujarati, Hindi, Bengali, Telugu, Tamil, Urdu that all language are translating in Indian language by using the Machine translation technique.

Pal, et. al.[5], proposed natural language processing(NLP) like as machine translator, automatic summarization, etc. its required empathy and resolution of anaphora. Almost work is done in English and other European languages in hindi language existent a review of work done in the ground of anaphora resolution in Hindi. This paper is not work committed to pronominal anaphora.

Steinberger,et. al.[6], proposed the use anaphoric information in latent semantic analysis LSA) and discuss its application. Anaphoric information is automatically mind a new release of our own anaphora resolution system, GUITAR, is includes proper noun resolution. Anaphoric information is used to check consistency of summary produced our summarizer.

Sontakke, et. al.[7], proposed data are stored in the database and the database is the source of all the information which are used by the human or any all other information. Information is required for do the some work in every human life day-to-day.

And this database has an impotent role playing in computer and internet uses. Database management system are used for to accessing, storing and retrieving the data or information which are stored in the system. Whereas all the people are not getting to how to access database because of they have no knowledge of database language. That is why we need to Find new method or technique with the use of NLP (Natural Language Processing)to access the database. The new development of method is called Natural Language Interface to Database (NLIDB).In this method no need to learn any database language those who not accessing the data .So they can give query in their native language such as ,Hindi, Gujrati, English ,etc. And it also give the resonance or result in the same language.

Lakhmani,et. al.[8], proposed challenges is to determine the noun .how they related to with pronoun and how they refer to each other. This is known as Anaphora resolution. Generally there are three main algorithms works for anaphora resolution. That is Hobbs , Centering and Leppin leass Algorithm. This three algorithm are works for hindi language. As hindi language is more difficult to other European language.so many terms and method need to resolving anaphora.

Sinha,et. al.[9], presents a system overview of an English to Hindi Machine-Aided Translation System named AnglaHindi. Anglabharti is a pseudo-interlingual rule-based translation methodology. The system generates approximately 90% acceptable translation in case of simple, compound and complex sentences upto a length of 20 words.

Singh,et. al.[10], presents a computational model for anaphora resolution in Hindi that is based on Gazetteer method. The experiment conducted on different data set - data set 1 - children story - 65%, data set 2 - news article - 63%, data set 3 bioghraphy from wikipidea- 83%.

Chopra, et. al.[11], presents about how Anaphora Resolution is useful in performing computation linguistic task in various Natural languages including the Indian languages. It also tell about the how the Anaphora

Resolution is conducive in handling unknown words in Named Entity Recognition.

Dutta, et. al.[12], proposed the application of Hobbs algorithm for pronominal resolution in Hindi which is used to solve reflexive and possessive pronouns.

Uppalapu, et. al.[13], presents an algorithm which is in line with S-List (Prasad and Strube, 2000) for resolving the Hindi third person pronouns. It also show that there is an improvement in the performance of the S-List algorithm by taking two lists one is present and second is past instead of one. It has given 61.11%,77.45% of result on different data sets.

Devi, et. al.[14], present a generic anaphora engine for Indian languages, which are mostly resource -poor languages. It has analysed the similarities and variations between pronouns and their agreement with antecedents in Indian languages.

Dakwale, et. al.[15], present a hybrid approach to resolve Entity-pronoun references in Hindi. The results show that, use of dependency structures provides syntactic knowledge which helps to resolve some specific types of references. Semantic information such as animacy and Named Entity categories further helps to improve the resolution accuracy.

Lakhmani, et. al.[16], proposed the pronominal anaphora resolution for Hindi Language using Gazetteer method. This model uses Recency factor as the baseline ,Animistic knowledge is introduced to the model which forms the criteria of classification of different nouns and pronouns. It has given approx 60 to 70% of result.

Mehla, et. al.[17], presents a comprehensive study about the anaphora resolution ,Event Anaphora resolution and Entity resolution.

Duttaa, et. al.[18], present machine learning approach for the classification indirect anaphora in Hindi corpus. Based on the semantic structure provided by the collocation patterns following the pronoun is also carried out.It has given 12.44% result on indirect anaphora.

Tetreault, et. al.[19], presents a modification of the CT-based approach called the Left-Right Centering approach (LRC). Psycholinguistic research claims that listeners try to resolve references as soon as they hear an anaphor. If new information appears that contradicts this choice of antecedent, they reanalyze and find another antecedent. This psycholinguistic fact is modeled in the LRC. In a further modification (LRC-F), information about the subject of the utterance is also encoded.

Mitkov, et. al.[20], proposed the inaccuracies in the preprocessing stage in anaphora resolution lead to a significant overall reduction in the performance of the system, for systems that use automated preprocessing. Due to this, it is not entirely fair to compare machine learning approaches that use automated preprocessing with knowledge-based techniques that had the advantage of having manually preprocessed input available.

Cardie, et. al.[21], proposed the three extra-linguistic modifications to the Soon Algorithm and got statistically significant improvement in performance. This supports the Mitkov (1997) results, in the domain of machine learning approaches, showing that co-reference systems can be improved by “the proper interaction of classification, training and clustering techniques”

Bates, et. al.[22], presents only contrasts progressive and simple past tense. However, it is likely that significant differences in information status are seen for other tenses and aspects.

Soon, et. al.[23], presented the results that were comparable to non-learning techniques for the first time. They resolved not just pronouns but all definite descriptions. They used a small annotated corpus to obtain training data to create feature vectors.

Hobbs, et. al.[24], present the first results which obtained an impressive accuracy in pronoun resolution. The evaluation was done manually, as also all preprocessing. This makes it difficult to compare performance with other approaches develop later. However, Hobbs algorithm remains the main algorithm that many syntactically based approaches still use, even though they augment it with other knowledge sources.

Lappin, et. al.[25], proposed a model that calculates the discourse salience of a candidate based on different factors that are calculated dynamically and use this salience measure to rank potential candidates. They do not use costly semantic or real world knowledge in evaluating antecedents, other than gender and number agreement.

Saha, et. al.[27], proposed models using Multi-objective Optimization (MOO) techniques based on Genetic Algorithm. It does not focus on other factors like number agreement.

3. CONCLUSION

Anaphora resolution is an important aspect in Natural Language Processing. After going through the literature review and research papers on Natural Language Processing and related areas, we conclude that all the work is done in the anaphora resolution in different languages which includes English, and various other European languages but less work is done in Hindi language. Many techniques like Lappin and Leass, Centering, Hobbs, Gazetteer method are used to generate wordlist and traverse the sentence and then compare sentence to the wordlist, after comparing the word, it will store in index and print the result. For Anaphora Resolution on Hindi language we find Gazetteer method more appropriate.

4. REFERENCES

[1] Lakhmani P, Singh S. Anaphora Resolution in Hindi Language.

[2] Feng J, Xu N. Using Chinese Natural Language Interfaces for Navigation in Mobile GIS.

[3] Poesio M, Kabadjov MA. A General-Purpose, Off-the-shelf Anaphora Resolution Module: Implementation and Preliminary Evaluation. InLREC 2004 May.

[4] N.A.J. khan. Statistical machine translation of Indian language: A survey”

[5] Pal TL, Dutta K, Singh P. Anaphora Resolution in Hindi: Issues and Challenges. International Journal of Computer Applications. 2012 Mar;42(18).

[6] Steinberger J, Poesio M, Kabadjov MA, Ježek K. Two uses of anaphora resolution in summarization. Information Processing & Management. 2007 Nov 30;43(6):1663-80.

[7] Sontakke AR, Pimpalkar A. A Review Paper on Hindi Language Graphical User Interface to Relational Database using NLP.

[8] Lakhmani P, Singh S, Morwal S. Performance Analysis of two Anaphora Resolution System for Hindi Language.

[9] Sinha RM, Jain A. AnglaHindi: an English to Hindi machine-aided translation system. MT Summit IX, New Orleans, USA. 2003 Sep 23:494-7.

[10] S. Singh, P. Lakhmani, P. Mathur and S. Morwal. ANAPHORA RESOLUTION IN HINDI LANGUAGE USING GAZETTEER METHOD.

[11] D. Chopra, G.N. Purohit. HANDLING AMBIGUITIES AND UNKNOWN WORDS IN NAMED ENTITY RECOGNITION USING ANAPHORA RESOLUTION.

[12] Dutta K, Prakash N, Kaushik S. Resolving pronominal anaphora in hindi using hobbs algorithm. Web Journal of Formal Computation and Cognitive Linguistics. 2008 Jan;1(10):5607-5607.

[13] B. Uppalapu, D. M. Sharma. Pronoun Resolution For Hindi.

[14] Devi SL, Ram RV, Rao PR. A Generic Anaphora Resolution Engine for Indian Languages. InCOLING 2014 (pp. 1824-1833).

[15] Dakwale P, Mujadia V, Sharma DM. A Hybrid Approach for Anaphora Resolution in Hindi. InIJCNLP 2013 (pp. 977-981).

[16] P. Lakhmani, S.mita Singh2, Dr. P. Mathur. Gazetteer Method for Resolving Pronominal Anaphora in Hindi Language.

[17] K. Mehla, Karambir and A. Jangra. Event Anaphora Resolution in Natural Language Processing for Hindi text.

[18] Dutta K, Kaushik S, Prakash N. Machine Learning Approach for the Classification of Demonstrative Pronouns for Indirect Anaphora in Hindi News Items. The Prague Bulletin of Mathematical Linguistics. 2011 Apr 1;95:33-50.

[19] Tetreault JR. A corpus-based evaluation of centering and pronoun resolution. Computational Linguistics. 2001 Dec 1;27(4):507-20.

[20] Mitkov, Ruslan; Boguraev, Branimir and Lappin, Shalom, 2001, Introduction to the Special Issue on Computational Anaphora Resolution in Computational Linguistics, Volume 27, Number 4, 2001

[21] Ng V, Cardie C. Improving machine learning approaches to coreference resolution. InProceedings of the 40th Annual Meeting on Association for Computational Linguistics 2002 Jul 6 (pp. 104-111). Association for Computational Linguistics.

[22] Harris CL, Bates EA. Clausal backgrounding and pronominal reference: A functionalist approach to c-command. Language and cognitive processes. 2002 Jun 1;17(3):237-69.

[23] Soon WM, Ng HT, Lim DC. A machine learning approach to coreference resolution of noun phrases. Computational linguistics. 2001 Dec;27(4):521-44.

[24] Hobbs JR. Resolving pronoun references. Lingua. 1978 Apr 30;44(4):311-38..

[25] Lappin S, Leass HJ. An algorithm for pronominal anaphora resolution. Computational linguistics. 1994 Dec 1;20(4):535-61.

[26] Mitkov R. Factors in anaphora resolution: they are not the only things that matter: a case study based on two different approaches. InProceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts 1997 Jul 11 (pp. 14-21). Association for Computational Linguistics.

[27] Saha S, Ekbal A, Uryupina O, Poesio M. Single and multi-objective optimization for feature selection in anaphora resolution. InIJCNLP 2011 (pp. 93-101).