Coreference Resolution for Better Information Retrieval from Indian Classical Music Forums.

Joe Cheri Ross, Sachin Pawar, Pushpak Bhattacharyya
Indian Institute of Technology Bombay
{joe,sachinpawar,pb}@cse.iitb.ac.in

Abstract

Information retrieval from music related text is an integral part of Music Information Retrieval(MIR) augmenting content based MIR. Discussion forums on music are rich sources of information gathered from a wider audience. There are a few music forums related to Indian classical music having notable information pertaining to entities including artiste, music concepts including raga, location etc. The forum posts generally contain anaphoric references to the main topic of discussion or to an intermediate mention. Coreference resolution assists resolving anaphoric references thus improving the yield of relation extraction from the posts. In this paper we explore features for supervised approach to coreference resolution felicitous to discourses of aforementioned nature. Along with prevalent features for coreference resolution, we experimented with grammatical features obtained from dependency parsing. Results with Naive Bayes and SVM classifiers are compared along with analysis of different relevant features. The dependency role of the mentions, specifically the mentions to be checked for coreference and other mentions in the vicinity, are found to be relevant for coreference resolution, especially in short discourse of text.

1 Introduction

Serving the queries to obtain the appropriate music content is challenging when the query contains meta information about the content. Having ample meta information on the music content and associated concepts, complements the knowledge base contributing to better content based MIR. This includes information about artistes, performances, music concepts etc. Even though most of the knowledge sources on Indian music tradition confines to ancient scriptures and few recent literature, quite a good number of websites including discussion forums and blogs have been introduced providing relevant information. Rasikas.org (Rasikas.org,) is one among the prominent discussion forums where the users engage in discussions related to Carnatic music topics comprising ragas, talas, artistes etc.

Extraction of relevant information from vast text resources spread across the web is challenging due to the nature of the text content in these websites. The major share of information on Indian classical music is seen as unstructured form in forums, blogs and other websites which are partially monitored. Identification of entities is a critical step in information extraction followed by identification of relations between them. In forums like rasikas.org posts are written in informal language with pronominal and alias mentions referring to the main topic of discussion or to another related entity mentioned in the discourse. Though most of the pronominal and nominal expressions present contributing to relations refers to some other entity, the exact antecedent of each coreferential mention has to be resolved for extraction of exact relation. It is commonly observed that the main topic of a post is referred by pronominal or alias mention. Presence of a large number of such sentences having potential relations present, make coreference resolution an inevitable part of information extraction from forums. The process of checking whether two expressions are coreferent to each other is termed as coreference resolution

(Soon et al., 2001)

The well-known discussion forum of Carnatic music Rasikas.org, is taken for our study. Enrolled with a good number of music loving users, the forum discusses many relevant topics on Carnatic music providing valued information. Sordo et al. evaluated information extraction from the same forum using contextual information (Sordo et al., 2012). Integration of natural language processing methods yields better coverage for the extracted relations.

Largely the entities are mentioned using pronominal and nominal mentions in this forum. Resolution of these coreferences is crucial in increasing recall of relation extraction from forums. Coreference resolution identifies the real world entity, an expression is referring to (Cherry and Bergsma, 2005). Though a widely researched area, coreference resolution will have to be applied differently considering the characteristics of the text in these forums. Supervised approach has been widely used in coreference resolution (Rahman and Vincent Ng, 2009; Soon et al., 2001; Aone and Bennett, 1995; McCarthy, 1996). We examine the commonly used conventional features and its variants that suits this domain of text. Soon et al. and Vincent et al. have investigated an extensive list of features for coreference resolution. Most of these methods model this problem as classification of mention-pair as coreferent or non-coreferent.

Research on coreference resolution for similar domains of text are reported. Ding et al. has discussed features for supervised approach to coreference resolution for opinion mining where the discourse of text is short as in forum posts (Ding and Liu, 2010). Hendrickx et al. experimented their coreference resolution with unstructured text in news paper articles, user comments and blog data targeting opinion mining (Hendrickx and Hoste, 2009).

Coreference resolution in this domain is restricted to resolve coreferential relations between entities within a discourse of a post. We follow a supervised approach with mention-pair model, learning to identify two mentions are coreferent or not. Mention pairs are constructed from the annotated mentions from the posts. Along with standard set of proven features, dependency parse based features and its proposed variants are found to contribute to increase in accuracy. Dependency parse features are intended to capture the characteristics of the human process of coreference resolution, getting the role of a mention in the corresponding sentence and thus obtaining the relation between the mentions in the pair. Kong et al. proposed anaphora resolution employing dependency driven tree kernel based method (Kong et al., 2010). Vincent utilized dependency parsing to identify the subject or object relation of a mention with a verb (Vincent Ng, 2007). Uryupina, Recasens et al. and McCarthy et al. have investigated similar features based on grammatical roles. (Uryupina, 2006; Recasens and Hovy, 2009; McCarthy and Lehnert, 1995).

This work delves into analyzing the relevance of dependency parse based features and other features with the limited annotated music forum (rasikas.org) data available. The common characteristics of the text mitigates the problem due to data insufficiency, helping to make sufficient observations for improving coreference resolution in web forums of similar nature.

2 Features for coreference resolution

Coreference resolution task classifies a given pair of mentions as coreferent or not through the features capturing the coreferent characteristics. Supervised approach takes positive and negative pairs from the mention pairs formed from the annotation. The model is trained to identify whether a pair of mentions is coreferent or not. When a mention pair (m_i,m_j) is considered to be coreferent m_i is the antecedent to which m_j refers to. The system is trained with mention pairs formed from a set of mentions in training data. The features are extracted for each mention pair selected for training the system. The annotated mentions in the forum posts serve as the ground truth for the experiments. The coreferent mentions are annotated with the same id. These mentions carrying the same id belongs to the same cluster. The positive mention pairs are formed from mentions which are marked coreferent such that for a pair (m_i, m_j), m_i occurs before m_j and m_i is not a pronoun, definite phrase or a demonstrative phrase. Positive instances are formed between the first mention in a cluster and rest of the mentions in the same cluster. Each negative mention pair instance have mentions from different clusters.

The commonly accepted features for coreference resolution discussed in (Soon et al., 2001)

and features suitable for text content in forum posts are evaluated . These features include

String match(STR_MH): This is true when the string of both the mentions in the pair are similar. Fuzzy matching is employed to discard minor changes in the strings. This is based on the assumption that identical strings refer to the same entity.

Alias(AL): This feature is true when the second mention is the part of the first mention or the second mention is the acronym of the first.

Same sentence(SAME_SENT): This feature is true when both the mentions are from the same sentence.

Second mention pronoun(PRN): This feature is true when the second mention of the pair is a pronoun.

Second mention definite noun phrase(DEF_NP): This feature is true when the second mention of the pair is a definite noun phrase.

Second mention demonstrative noun phrase(DEM_NP): This feature is true when the second mention of the pair is a demonstrative noun phrase. Demonstrative noun phrase starts with the word *this*, *that*, *these* or *those*. ex. This person.

First mention proper noun(PRPN1): This feature is true when the first mention is a proper noun referring to a person, place or a concept name. In Indian music domain it can be a name of an aritste, instrument, raga name etc.

Second mention proper noun(PRPN2): This feature is true when the second mention of the pair is a proper noun.

Though aforementioned features are significant for showing the coreferent characteristics of a mention pair, the role of a mention in a discourse and its relation with other mentions are prime features in coreference identification. Apart from analysing whether a mention in the pair is a subject or object of a sentence, we also analyze the role of other mentions coming in between the mentions of the pair under consideration. This helps to figure out the existence of any other potential antecedent for the anaphora in the mention pair. The existence of a potential antecedent should decrease the probability of the mention pair considered to be coreferent. The role of a mention is determined with the help of dependency parse of a sentence.

First mention subject(SUBJ1): This feature is true when first mention is a subject of any verb in the sentence

Second mention subject(SUBJ2): Similar to the previous feature, but for second mention in the pair.

First mention object(OBJ1): This feature is true when first mention is an object of any verb in the sentence.

Second mention object(OBJ2): Similar to the previous feature, but for second mention in the pair.

For coreferent pairs it is generally observed that the first mention forms the subject of the sentence in which the mention occurs.

Subject mention between(SUBJ_BET): This feature is true if there exists another mention in between the mentions under consideration, taking the subject role in the occurring sentence. Such a mention have higher chances for being coreferent with the anaphoric mention in the pair. This helps in reducing the chances of a mention pair getting classified as coreferent when there is a potential candidate mention

Subject mention associated with root verb between(ROOT_SUBJ_BET): This is similar to the previous feature except for this is true only when the mention is associated with the root verb of the sentence. Having another mention associated with the root verb occurring after the candidate antecedent in the pair, increases the probability of this mention being the actual antecedent. This feature is experimented to analyse the false positives created by SUBJ_BET feature.

We have tried both Naive Bayes and SVM approaches for training and classification. A positive mention pair (m_i, m_j) is selected from the annotation in such a way that m_i occurs before m_j in the discourse and they are coreferents. All such pairs are selected provided m_i is a proper noun such as artiste name, music concept, instrument name etc. A negative mention pair is formed from mentions belonging to different coreferent clusters in the same discourse. In order to maximize the utilization of the available annotated data, negative pairs are formed between a mention and all preceding proper noun mentions belonging to dif-

ferent coreferent clusters.

3 Experimental Setup

Forum	#Posts	#Sent.	#M	#P	#N
Raga & Alapana	42	228	104	124	189
Vidwans & Vidushis	29	182	62	194	181

Table 1: Details of annotated posts. (#Posts= No. of posts #Sent= No. of sentences in the forum. #M= No. of annotated mentions #P= positive mention pairs formed #N= negative mention pairs formed)

Coreference annotation is done majorly on two forums; discussing carnatic ragas and carnatic singers. Mention pairs for training and testing are formed from the limited annotations done on 71 posts from the forums. Table 1 provides information on the number of posts, sentences and annotated mentions and number of positive and negative mention pairs available in each forum. Mention pairs are extracted from the annotated mentions in the posts. Table 2 shows the statistics of the type of mentions present in the posts, showing the count and percentage contributed by each type to the total mentions count. Compared to proper nouns and pronouns present in the posts, number of definite and demonstrative phrases is not high. Dependency parse output of sentences in the posts is obtained from Stanford dependency parser for procuring dependency relation based features (Marneffe et al., 2008).

Experiments carried out with Naive Bayes and SVM classifiers. K-fold cross validation was performed taking k=2. Considering the size of the dataset k is taken as 2. SVM classifier is tried with linear, polynomial and RBF kernels. Evaluation is done computing precision, recall and F. Experiments with various kernels taking the best performing parameters for each of the kernel. This is identified through experimentation with a range of values for the parameters and finding the parameter configuration giving the highest F measure. The best among the methods in terms of F measure is used for subsequent experiments.

4 Results

Table 3 compares Naive Bayes and SVM with different kernels. Precision, recall and F measures are computed for each of these configuration. In this coreference resolution problem, True positives(TP) is defined as the number of coreferent pairs identified as coreferents; false positives(FP) as number of non coreferent pairs identified as coreferents; true negatives(TN) as number of non coreferent pairs identified as non coreferents; false negatives(FN) as actual coreferent pairs not identified as coreferents by the classifier. Precision is computed as Prec= TP/(TP+FP), recall as Recall= TP/(TP+FN) and F measure as
F= $2 \times$prec$\times$recall/(prec+recall).

The reported result in table 3 is based on the best performing parameter for each kernel. The c parameter is varied in the range 2^{-5} to 2^{5} for linear kernel and γ is varied in the range 2^{-5} to 2^{5} for RBF. Polynomial kernel is experimented with degrees ranging from 1 to 5. These above experiments are performed with a certain set of features giving acceptable results. Comparing F values, Naive Bayes performs better on the given data and successive experiments are done with Naive Bayes.

Classifier	Precision	Recall	F
Naive Bayes	0.720	0.827	**0.77**
SVM- Linear (c=0.125, c=0.25)	0.691	0.846	0.754
SVM- Polynomial (degree=2)	0.676	0.862	0.746
SVM-RBF (γ=0.125)	0.688	0.846	0.751

Table 2: Performance of Naive Bayes and SVM classifiers(with different kernels).

Table 4 shows improvement or decline caused by different class of features. Experiment A shows the results with some of the basic features commonly used in coreference resolution tasks. Experiment B adds definite and demonstrative noun phrase features. Though there is a decrease in precision, it produces improvement in recall and F. The presence of reasonable number of definite and demonstrative noun phrases makes this improvement. The significance of dependency parse based features is very clear from the increase in F on introduction of dependency parse based features in experiment A. The features SUBJ1, SUBJ2, OBJ1 and OBJ2 are critical in classification of mention pair, as these expose the role of a mention in the corresponding sentence and the relation between

Exp Id	Features	Precision	Recall	F
A	STR_MH, AL, SAME_SENT,PRN, PRPN1, PRPN2	0.687	0.207	0.319
B	STR_MH, AL, SAME_SENT,PRN,DEF_NP, DEM_NP, PRPN1, PRPN2	0.568	0.355	0.437
C	STR_MH, AL, SAME_SENT,PRN, PRPN1, PRPN2, SUBJ1, OBJ1, SUBJ2, OBJ2	0.726	0.818	**0.769**
D	STR_MH, AL, SAME_SENT,PRN, PRPN1, PRPN2, SUBJ1, OBJ1, SUBJ2, OBJ2, SUBJ_BET	0.731	0.802	0.765
E	STR_MH, AL, SAME_SENT,PRN, PRPN1, PRPN2, SUBJ1, OBJ1, SUBJ2, OBJ2, ROOT_SUBJ_BET	0.725	0.805	0.763

Table 3: Results with different feature combinations

the mentions.

Introduction of feature SUBJ_BET helped to get rid of false positives which are having another mention which forms subject in the occurring sentence. In most of these cases the third mention forms the actual antecedent of the anaphoric mention in the pair under consideration. At the same time this feature led to certain false negatives because of the existence of other subject mentions in between the actual coreferent mention pairs. Experiment C shows improvement in results on introduction of SUBJ_BET feature.

Feature ROOT_SUBJ_BET is a refinement of SUBJ_BET, introduced to consider only the potential mentions coming in between the mentions of the pair under consideration expecting to reduce false negatives. Certain false negatives which got introduced by the previous feature were removed. This feature thus contributes to increase in recall, but shows a fall in recall since certain valid subject mentions in between having no association with root verb are ignored. However these mentions are coreferent with the anaphoric mention in the pair.

The observations show that gender identification and semantic class identification can rectify the misclassification of many instances. (`Narayanaswamy, She`) is obviously a non-coreferent mention pair both are of different gender. Mention pair (`Hamsabrahmari, him`) with mentions from different semantic classes is not coreferent. The existing gender identification systems are inappropriate for Indian names. In Indian classical music domain, semantic class identification is not limited to identification of an entity as person,location etc, but it has to be extended to identification of an entity as raga, other Indian music concepts etc. The dictionary based approach we have tried is yet to be matured to enable semantic class feature.

5 Discussion

5.1 Comparison with existing systems

Existing coreference resolution systems are experimented with the forum posts in rasikas.org to understand the general behaviour of state of the art systems on this kind of text. Stanford coreference system (Raghunathan et al., 2010) and Cherrypicker (Rahman and Vincent Ng, 2011) are taken for analysis. Results are not compared in terms of precision, recall and accuracy, since the results of our system are analysed based on classification of a mention pair. Both the systems have problem with extraction of mentions from the forum posts, especially with the noun phrases involving Indian names. That itself bring down the performance of these systems to a large extent. Since mention extraction step is not automated in our system, we do not intend to compare the performance based on mention extraction of the above mentioned systems.

Both Cherrypicker and Stanford system does not identify many important mentions in the text. Cherrypicker identifies many long noun phrases having 3-4 words as 2 different mentions and misses most of the definite and demonstrative noun phrases. Stanford system is better in identifying demonstrative and definite noun phrases as mentions. Stanford system takes long words as mentions, failing to resolve Indian name mentions in the sentences. Even from the correctly identified mentions, the precision is observed very low. The results are not quantified here as our metrics are based on mention pairs.

These observations show that existing systems fail in resolving the coreferences from text of this nature with entities from Indian music domain.

The longer terms and names present in the text creates confusion to the existing coreference resolution systems.

6 Conclusion and Future work

This paper focused on coreference resolution for music related forum posts motivated by its importance in information extraction. Different set of features are analysed, after picking the better performing classifier. The experiments asserted the importance of features incorporating dependence relations of the mentions. The grammatical errors in the forum text make it challenging to extract dependency relations accurately for all mentions. Even though the available annotated dataset is limited, the common characteristics of reference mentions in this corpus helped in making sufficient observations.

The observations clearly reflects the need to have features containing gender and semantic class information. Taking into account the distinctions of the forums post in comparison with other forms of texts, better features are to be identified to make supervised approach comparable to human approach to coreference resolution.

References

Cherry, Colin, and Shane Bergsma. 2005. An expectation maximization approach to pronoun resolution. In *Proceedings of the Ninth Conference on Computational Natural Language Learning.*

Sordo, Mohamed, Joan Serr, Gopala K. Koduri, and Xavier Serra. 2012. Extracting Semantic Information from an Online Carnatic Music Forum. In *ISMIR.*

Rahman, Altaf, and Vincent Ng. 2009. supervised models for coreference resolution. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing.*

Soon, Wee Meng, Hwee Tou Ng, and Daniel Chung Yong Lim. 2001. A machine learning approach to coreference resolution of noun phrases. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics.*

Aone, Chinatsu, and Scott William Bennett. 1995. Evaluating Automated and Manual Acquisition of Anaphora Resolution Strategies. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics.*

McCarthy, Joseph F. 1996. A Trainable Approach to Coreference Resolution for Information Extraction. *PhD diss., University of Massachusetts Amherst.*

Xiaowen Ding, and Bing Liu. 2010. Resolving object and attribute coreference in opinion mining. In *Proceedings of the 23rd International Conference on Computational Linguistics.*

Iris Hendrickx, and Veronique Hoste. 2009. Coreference resolution on blogs and commented news. In *Anaphora Processing and Applications.*

Fang Kong, Guodong Zhou, Longhua Qian, and Qiaoming Zhu. 2010. Dependency-driven anaphoricity determination for coreference resolution. In *Proceedings of the 23rd International Conference on Computational Linguistics.*

Vincent Ng. 2007. Semantic class induction and coreference resolution. In *ACL.*

Uryupina, Olga. 2006. Coreference resolution with and without linguistic knowledge. In *Proceedings of the Language Resources and Evaluation Conference.*

Marta Recasens, and Eduard Hovy. 2009. A deeper look into features for coreference resolution. *Springer Berlin Heidelberg.*

McCarthy, Joseph F., and Wendy G. Lehnert. 1995. Using decision trees for conference resolution. In *Proceedings of the 14th international joint conference on Artificial intelligence.*

De Marneffe, Marie-Catherine, and Christopher D. Manning. 2008. The Stanford typed dependencies representation. In *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation.*

Altaf Rahman, and Vincent Ng. 2011. Narrowing the modeling gap: a cluster-ranking approach to coreference resolution. *Journal of Artificial Intelligence Research 40.1.*

Karthik Raghunathan, Heeyoung Lee, Sudarshan Rangarajan, Nathanael Chambers, Mihai Surdeanu, Dan Jurafsky, and Christopher Manning. 2010. A multi-pass sieve for coreference resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics.*

Rasikas.org, [online], Available at: http://www.rasikas.org/ [Accessed 10 May. 2014].