Toward Verb frame extraction: Clustering Verb Arguments

Vidyadhar Rao
Chandrakanth M
Mentor: Abhilash I

Over view

- Introduction
- Problem of study
- Our Approach
  - Clustering of arguments
- Experiments
- Conclusion

Introduction

- Verbs
  - Actions, events and other complex semantic processes
- Verb frames?
  - Building blocks of computational models
  - Useful in parsing
  - Useful NLP resource

Hindi Verb Frames

- Hindi verb frames
  - Specify # of Mandatory arguments
  - Syntactico-Semantic Relations
  - Vibhakti
- Eg: Verbframe for xEnA

arc-label	vibhakti	lextype
k1	0	n
k2	ko	n

Problem

- Clustering Verb Arguments
  - # mandatory arguments of the verb
  - Vibhakti of these arguments
  - Syntactico-semantic relation (Future work)
- Why Clustering?
  - Underlying premise: For a given v*erb, arguments share similar properties.*
  - *Example: All arguments that stand in k1 relation take post position ‘ne’*

Approach

1. Extraction of simple sentences
   - Consider only ***monosemous*** verb (Hindi WordNet)
2. Extraction of potential argument list
   - All the NP chunks from the shallow parser output.
3. Clustering of potential arguments

Experimental Setup

- Corpus used
  - amar ujala: 324MB
- Monosemous verbs
  - Extracted from Hindi word net.

Extracting Simple Sentences

- Shallow parser used, but very slow
  - Took ~20 hrs to run 1200 files
  - Avg # of sentences in a file = 25
  - Took 40 days to run on 32, 000 files
- 8800 sentences simple sentences extracted

Experiments .. contd

- Top 7 monosemous verbs extracted ..
  - Beja 1429
  - KarIxa 253
  - mara 245
  - harA 205
  - liKA 172
  - beca 159
  - Gera 105

- We have considered verbs with at least freq of 100 simple sentences

Clustering arguments

- Used weka tool
  - to cluster the obtained instances of arguments (NPs).
- Results through figures

Features used

- Distance from verb
  - Mandatory arguments tend to appear closer to verb
- Frequency of the argument
  - Mandatory arguments appear more often with the verb
- Post position, Gender, Number, Person
- Combination of these.

Guided clustering using Linguistic cues

- Eliminating non arguments
  - Genitives and others
- Directly identifying arguments
  - Derivational features

Derivational cues

- Using a richer set of semantic features
  - Derivational morphological features
- Example (agentive derivations):

Suffix	Example
-ar	कुम्हार सुनार लुहार
-ek	लेकक, मारेक

- Easy to compute semantic features
- Ref: Chapter #5, Hindi, Yamuna Kachru

Conlcusion

- Clustering (using rationale features) makes explicit the regularities in usage of verbs with its arguments
- These methods also give frequency of usage.
- A through evaluation yet to be done