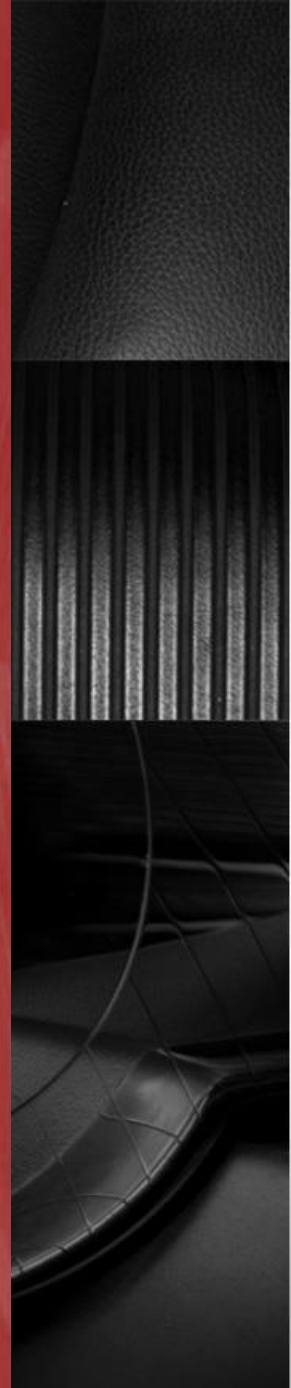# Action Discovery and Recognition in Psychological Videos
&
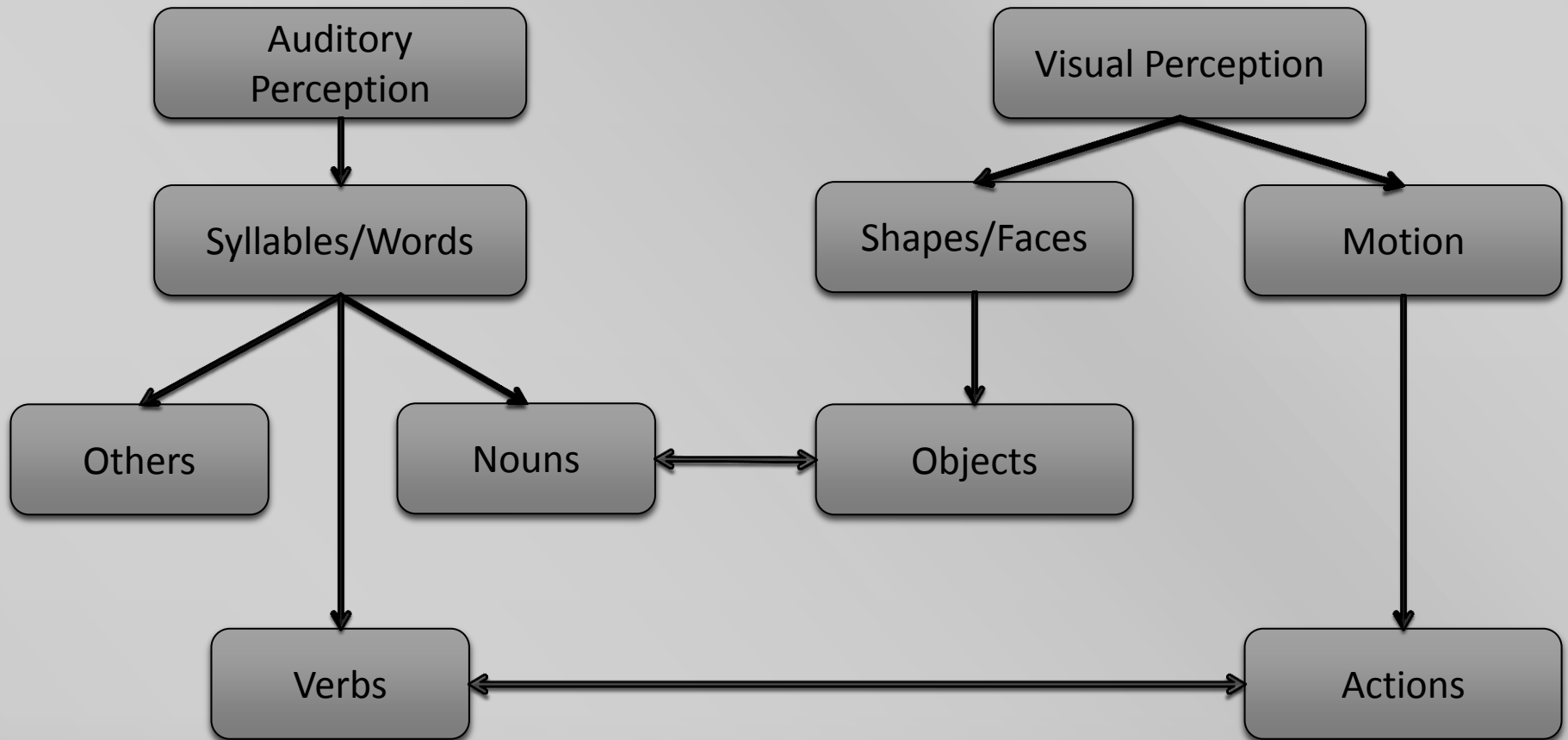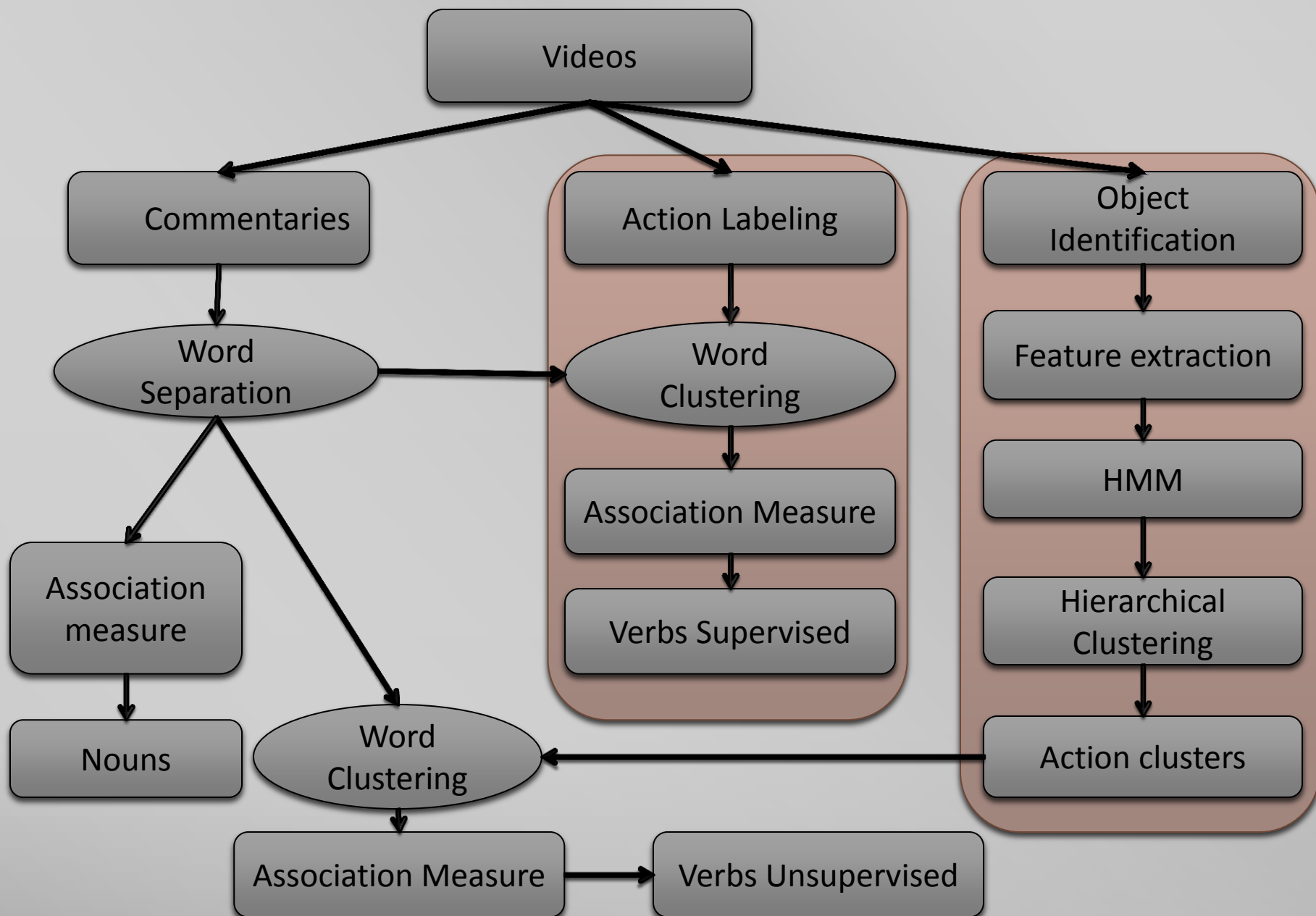# Noun and Verb Modeling

By: Diwakar Chauhan

Guide : Amit Mukerjee

email : {diwakarc, amit}@cse.iitk.ac.in
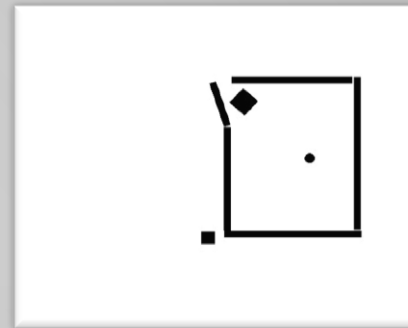
Language learning by Babies

Psychological Videos

- Manually created goal driven videos

- Difficult to generate automatically

- Available Videos

  - Heider Simmel Video

  - Frith Happes Animations
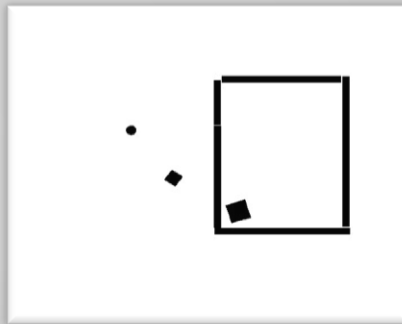
Stills from Heider Simmel Video

Actions in Psychological videos

- Limited actions(chase, go away, come closer) in Heider-Simmel Video

- Need of generalization

- Actions in Frith-Happe Animations

  - Large Variety of actions

  - Chase, push, pull, rotate, play, dance
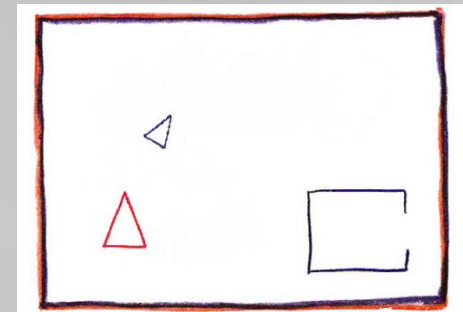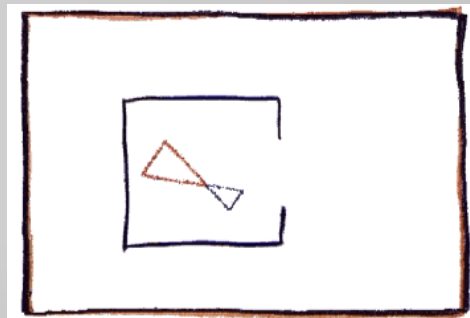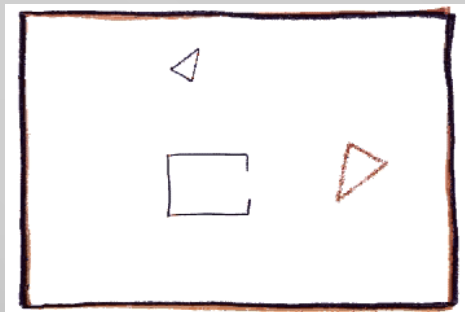
Stills from Frith Happe Animations

Preprocessing of Data and Feature Extraction

- Identification of triangles and Rectangle

  - Line detection : Hough Transform

  - Vertices : Intersection of sides of triangle
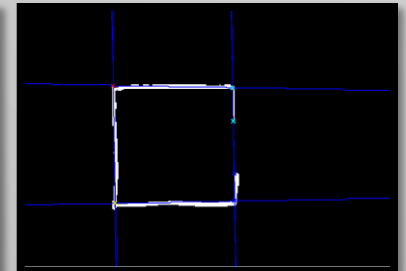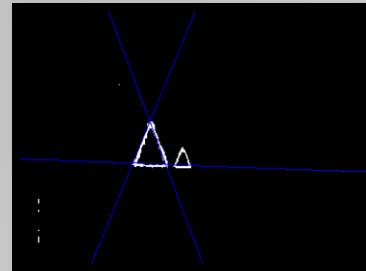
- Feature Extraction

  - Features relevant to motion and interaction of triangles

  - **Centroids of triangles**

    - Captures relative positions of triangles

  - **Orientation of triangles**

    - Capture direction and relative orientations of triangles

Preprocessing of Data and Feature Extraction

- Feature Extraction

  - **Visibility of triangles**

    - Measure of how much one triangle can see other triangle
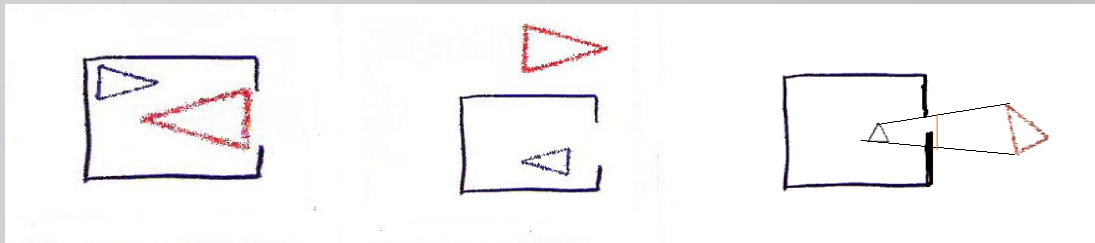
- Feature Vector

$$\begin{bmatrix} x_1 & y_1 & \theta_1 \\ x_2 & y_2 & \theta_2 \\ & v_{12} & \end{bmatrix} \text{ or } \quad \begin{matrix} x_1 & y_1 & \theta_1 \\ x_2 & y_2 & \theta_2 \end{matrix} \quad or \quad \begin{matrix} x_1 & y_1 & sin\theta_1 & cos\theta_1 \\ x_2 & y_2 & sin\theta_2 & cos\theta_2 \end{matrix}$$

Hidden Markov Models

- Dynamic Bayesian Networks

- Modeled as :-

$$\gamma = \{A, B, \pi\}$$

Where A, B and π are state transition,

observation symbol and initial

probability distribution

- Segment the video into image sequences of some length and train HMM on them
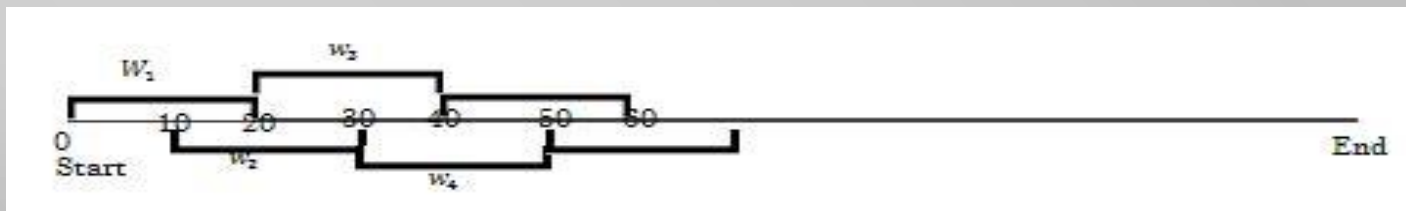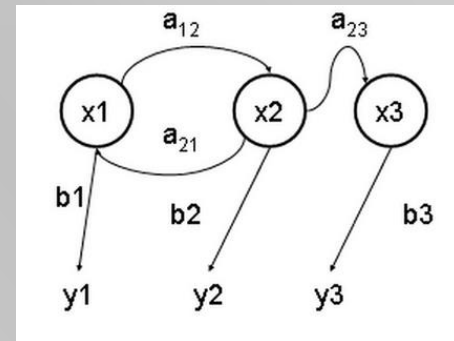
Distance in HMMs and Hierarchical Clustering

- Trained HMMs on all image sequences

- Calculated the Mutual Acceptance measure between HMMs

- Created cluster tree of image segments based on this measure

- Wards method used for cluster distances

- Cut the tree at appropriate height to get 6-8 clusters

- Clusters used for unsupervised verb learning
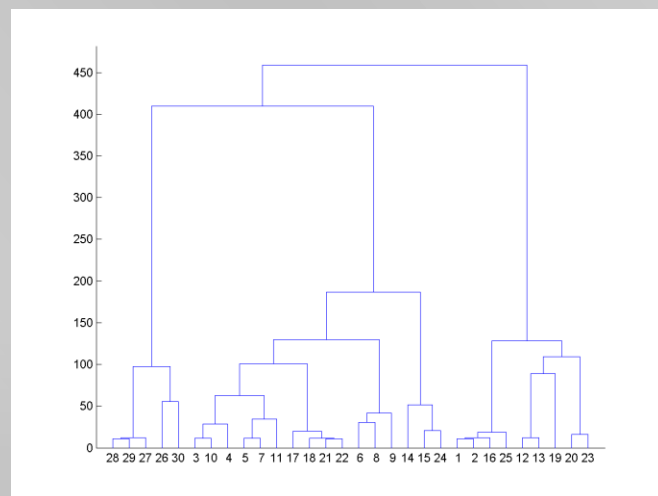
Language

- Lexicons

    - Words or basic units consisting a language (e.g. pitifully)

- Morphology

    - Smallest meaningful elements

    - Word forming from the morpheme (e.g. pity + ful + ly)

- Phonology

    - Syllables (pi + ty + ful + ly)

- Syntax

    - Organization of words to form sentence

- Lexical and phonological properties of a language are considered here

Hindi Corpus

- Obtained Hindi Corpus from CFILT, IIT Bombay

- Corpus has more than 2.8 Million words

- For each unique word in corpus, calculated its frequency and fraction of frequency

- Used it in filtering the equally occurring words in the commentary

- Also Used to remove very common words in results

Commentary Collection and Movement Record

- Collected commentaries in two languages

    - 22 Commentaries in Hindi for Coaxing video

    - 13 Commentaries in English for Coaxing video

- Instructions to subjects before they gives commentary

- Converted audio to texts manually

- Merged commentaries of all subjects

- Manually recorded the motion intervals

    of both triangles in video

```
1   48   एक बक्से के अंदर एक छोटा त्रिकोण एक बड़ा त्रिकोण है
49  76   दोनो आपस में लड़ रहे हैं
77  86
87  102  और दोनो घूम रहे हैं
103 120
121 139  छोटा त्रिकोण
140 170  बड़ा त्रिकोण बाहर आ गया बक्से के
171 197  छोटा त्रिकोण अभी भी अंदर है
198 207
208 238  बड़ा त्रिकोण भी अंदर आ गया अब
239 296  छोटा त्रिकोण बड़े त्रिकोण को बाहर लाने की कोशिश कर रहा है
297 327
328 374  बड़ा त्रिकोण अब छोटे त्रिकोण को बक्से के बाहर धकेल रहा है
375 400
401 428  और अब छोटा त्रिकोण बाहर आ गया है
429 460  और बड़ा त्रिकोण अभी भी बक्से के अंदर है
```

Words and K-Gram Calculation

- Determined all the unique words in the commentaries and their frequency

- Removed common words from the commentary by relative frequency in corpus and in commentary for all words

- Found 1, 2, 3, 4 grams from filtered words

- Determined the frequency and relative frequency of k-grams separately

Noun Learning and Association Measures

- **Definitions**

    - **Concept :**

        - Perception of object or action

        - Concept denoted by c_i and word denoted by w_j

        - Determined by motion records of triangles and simultaneous utterances by subjects

    - Joint Probability

        $$jp\ (c_{i,}\ w_j) =\ Prob\ (w_j\ |c_i)^*\ prob(c_i) = \frac{time\ when\ concept\ is\ attended\ and\ w_j\ is\ spoken}{total\ Time}$$

    - Mutual Information

        $$mi\ (c_{i,}\ w_j) = jp\ (c_{i,}\ w_j)^* \log(\frac{jp(c_i,w_j)}{fraction\ w_j\ is\ uttered\ in\ concept\ *\ fraction\ concept\ is\ attended})$$

    - Relative Frequency

        $$rf(c_{i,}\ w_j) = \frac{Frequency\ of\ w_j\ when\ c_i\ is\ attended}{Total\ frequency\ of\ w_j}$$

Noun Learning

- Concepts used :

  - Red triangle is c_1 and blue triangle is c_2

  - When c_1 is moving

  - When c_2 is moving

  - When c_1 is moving but c_2 is not moving

  - When c_1 is not moving but c_2 is moving

  - When both c_1 and c_2 are moving

  - When both c_1 and c_2 are not moving

- Calculated Mutual Information and relative Frequency measures of all k-grams for each of these concepts

Supervised Action Recognition

- Manually labeled the video with actions

- Total 9 segments of coaxing video labeled

- Learned HMMs on all of these segments
  - Feature vectors were same as that in unsupervised learning

- Calculated Mutual Acceptance measure on all these HMMs

- Merged the clusters with very low HMM distance

- Left with 7 clusters in the end

45	93	घूम रहे, खेल रहे
134	150	गया, बाहर चला गया
151	175	घूम रहा
176	195	अंदर आ रहा
210	230	टक्कर मार रहा, रोक रहा
235	249	ले जा रहा, खींच रहा
310	410	निकाल रहा, ले जा रहा, धकेल रहा, फेक रहा
415	487	बंद कर दिया, खड़ा हो गया, रोक लिया
557	598	घूम रहे, गोल घूमना

Supervised Verb Learning

- For the intervals labeled, obtained overlapping commentaries

- Selected commentary based on overlapping threshold

- Calculated the relative frequency words all words in each cluster/interval

- Filtered the words based on threshold

- This contains very less nouns because nouns are uniform in all clusters/intervals

- Removed the most common words by matching with 1000 most common words in corpus

Unsupervised Verb Learning

- Merged all the overlapping clusters from the results of cluster tree cutting

- Removed all very small intervals

    - Very less chance that, they will have actions

- For all the remaining intervals, obtained fairly overlapping commentaries

- Calculated the Relative Frequency measure for all the words in each cluster

- Filtered words based on threshold

- Removed the most common words by comparing them with top 1000 most common words in corpus

Results : Noun Learning

A. Concept 1(Red Triangle)

- Monograms
  - Joint Probability

त्रिभुज	0.24467923381
बाहर	0.236622073579
लाल	0.143508665248
बड़ा	0.123517786561

  - Mutual Information

त्रिकोण	0.509903357248
बाहर	0.404395887243
बड़ा	0.335477967263
त्रिभुज	0.206563466488

- Bigrams
  - Joint Probability

लाल त्रिभुज	0.0997263605959
नीला त्रिभुज	0.0540437823047
छोटे त्रिभुज	0.0446944359988
बाहर धकेल	0.0410459106111

  - Mutual Information

लाल त्रिभुज	0.120043032228
बाहर धकेल	0.117456453802
छोटे त्रिकोण	0.101200299431
छोटे त्रिभुज	0.0886936045256

Results : Noun Learning

B. Concept 2(Blue Triangle)

- Monograms
  - Joint Probability

बाहर	0.255092733354
त्रिभुज	0.243083003953
लाल	0.135679537853
त्रिकोण	0.0938735177866

  - Mutual Information

बाहर	0.448007998822
त्रिकोण	0.43001677713
छोटे	0.21587483377
त्रिभुज	0.188039314433

- Bigrams
  - Joint Probability

लाल त्रिभुज	0.103636363636
नीला त्रिभुज	0.060428701733
नीले त्रिभुज	0.0485709942232
त्रिभुज बाहर	0.0460626330192

  - Mutual Information

लाल त्रिभुज	0.151236728231
बाहर निकल	0.0997442010428
छोटे त्रिभुज	0.0903335555828
नीले त्रिभुज	0.070829260188

Results : Noun Learning

C. Concept 1 but Not Concept 2(Strong Red Triangle)

- Monograms
  - Joint Probability

त्रिभुज	0.033900881727
बड़ा	0.0294922468836
त्रिकोण	0.0288081483734
छोटा	0.0207509881423

  - Mutual Information

त्रिकोण	0.158270612882
बड़ा	0.0945487135243
छोटा	0.0646137109017
टकरा	0.0230077459893

- Bigrams
  - Joint Probability

बड़ा त्रिभुज	0.0101094557616
लाल त्रिभुज	0.00980541197933
छोटा त्रिभुज	0.00896929157799
नीला त्रिभुज	0.00790513833992

  - Mutual Information

छोटा त्रिभुज	0.0251906359027
बड़ा त्रिभुज	0.0230570612744
त्रिकोण छोटे	0.0206180446272
छोटे त्रिकोण	0.0206180446272

Results : Noun Learning

D. Not Concept 1 but Concept 2(Strong Blue Triangle)

- Monograms
  - Joint Probability

त्रिभुज	0.0332167832168
बाहर	0.0296442687747
नीला	0.0239434478565
लाल	0.0170264518091

  - Mutual Information

नीला	0.049575448637
बाहर	0.0367927268328
कोशिश	0.031638426055
अंदर	0.0281128780209

- Bigrams
  - Joint Probability

नीला त्रिभुज	0.0239434478565
लाल त्रिभुज	0.0142900577683
त्रिभुज बाहर	0.011477652782
त्रिभुज लाल	0.00592885375494

  - Mutual Information

नीला त्रिभुज	0.0548524131978
अंदर ना	0.0181769486602
लाल त्रिभुज	0.013284393001
वो बाहर	0.0132640733406

Results : Supervised Verb Learning

- Ground Truth

- Results

Interval	Verbs
45-93, 151-175,557-598	घूम रहे, खेल रहे, घूम रहा, गोल घूमना
134 - 150	गया, बाहर चला गया
176 - 195	अंदर आ रहा
210 - 230	टक्कर मार रहा, रोक रहा
235 - 249	ले जा रहा, खींच रहा
310 - 410	निकाल रहा, ले जा रहा, धकेल रहा, फेक रहा
415 - 487	बंद कर दिया, खड़ा हो गया, रोक लिया

Verbs
खुश, पकड़, गोल, घूम, कोण, दोनो, दोस्त
घूमे
वापस
निकाल, धकेल, फेक, तरीक़ो, धक्का, सिरे, जुड़, धक्के, मारकर, छोटा, अधूरे, धकेलने, चतुर्भुज, नीले
रास्ता, बड़े, खड़ा, युध, दरवाजा, ना, भगा, विरूध, पाए, पड़ा, वो, रोक

Results : Unsupervised Verb Learning

- Ground Truth

Interval	Expected
190-255	
150-195	
520-595	
370-405	
400-455, 460-515	

- Results

Verbs Discovered
भिड़ा, हिचक, निकालना, चिढ़ा, शरमा, खींचने, टक्करमारने
डरता, खेलना, घूमे, घूमते
वापस
युध,

Conclusion and Further work

- The nouns are successfully discovered

- Verbs obtained in Supervised Learning are vey much same as the ground truth

- In the unsupervised learning, the verbs are not mapped very well to action

  - Clustering by HMM

  - Changing time lag in commentary by each subject

- Syntax of the language can also be learned after verbs are discovered

Questions?