

Automatic Action Discovery and Hindi Word Mapping in Psychological videos

Diwakar Chauhan
Guide : Amitabha Mukerjee

{diwakarc, amit}@cse.iitk.ac.in

April 8, 2013

Abstract :

Human learning process involves learning the identity of objects(nouns) the relationship between the object(preposition) and the interaction of objects. An infant learns the identities of objects in first place than verbs. The reason behind this may be that verbs therefore activities, require more information. Naigles(1990)[?] demonstrated that while learning meaning of verbs, infants use the syntactic information. She proved that given a illegal verbs in a sentence the children use the syntax to guess the meaning of the verb.

However difficult the learning process of infants may be, they do not analyse large amount of data in order to learn the meanings of verbs or recognize the activities. They learn from the competent speakers who know relate the words to the event's objects in the environment.[?] And the children can extend their learning from one event to another. Here we are trying to understand the leaning process of infant and based on that learning process, try to discover actions in psychological videos and map Hindi verbs to them.

Introduction :

Actions can be based on the the variations in the shape and size change , poses[?] or orientation of the object. For example a person walking can be identified by motion of his hands and legs relative to his body. Similar can be said about the actions involving multiple agents. But what if the action which either don't need any shape changes to be expressed or the agent is an abstract object. These type of actions are represented either by specific motion of single object or particular type of motions and interaction between the agents. In this project we analyse the second type of actions. We use psychological videos[?].

Most of these animations contain only two agents which are triangles. The videos represent a particular transitive action. The videos, each of them represents certain action either complex or simple. By complex action we mean that the action constitutes smaller actions, e.g. the "Coxing" in the video consists of many instances of "Pushing" and "Rotating Jointly". We take some of the videos and apply HMM to on the feature vectors extracted from the frames of video. We take frames in small consecutive groups. We create HMM for each of this groups and evaluate the mutual acceptance measure for each of the groups. Based on this measure, hierarchical clustering is created. This cluster tree is later cut at some point to produce certain number of clusters. At the same time commentary is taken for the video. This commentary is processed to remove less important words and association measures are calculated for each of the remaining words. Based on this association measure, the nouns are identified.

Related Work :

Activity recognition has been a very much worked upon topic by researchers after 1980. Most of the work has been done on human activity recognition. E. Tapia [?] have used different sensors to collect data from home and based on that, they recognise activities. Vail [?] have formulated activity recognition problem as temporal classification. They use CRF for recognising activity of robots. After that they compare the results with HMM classification.

Considering of psychological videos, Heider-Simmel Videos [?] have been major attention for noun, verb recognition and linguistic mapping. Mukerjee and Satish(2008)[?] have used unsupervised approach to cluster visual events into action classes. According to them, in some visual, the objects in the focus are more likely to be the part of the action

happening in the videos. They have used merge neural gas algorithm to cluster the events. But there has been very less work done on activity recognition on psychological videos.

Database :

In this project we have used the animations[?] created at University of cognitive Neuroscience, UK. These animations were created to study autism in human. Autism creates Theory of Mind deficiency in children and in adults. People suffering from this are not able to predict the actions and thoughts of the other people or objects or in other way they are not able to make a theory about the reaction or interpretation of their surroundings. Traditionally autism was tested by false-belief test. False belief is an attribute in human, which makes them able to believe that other people in world can have different belief than him about same phenomena. This quality is developed in the childhood. But adults with autism were able to pass the false-belief test.

These animations named Frith-Happe animations had better success in capturing autism in adults. There are three major classes of animations. "Random" , "Goal-Desired", and "Theory of Mind". Each class is made to test different psychological aspect of mind. Each of these class have 4 videos in it.

We used these videos because these videos represent intentionality in them. Generate such video automatically with those is a very difficult problem.

Commentaries:

Commentaries are one of the very crucial part of the project because the accuracy of the results depends on how consistent the words are in the commentary with the video.

Taking Commentaries : We recorded Hindi commentaries primarily on two videos Chasing and Coaxing from various speakers. The process of recording the commentaries was as follows:-

These two videos were shown to the user 2-3 times and then each user was given certain instruction about the content type of the commentaries. The instruction was :

"Describe the objects in the video and their interaction"

Then the commentaries were taken. Further these commentaries are time-stamped according to the videos.

Results :