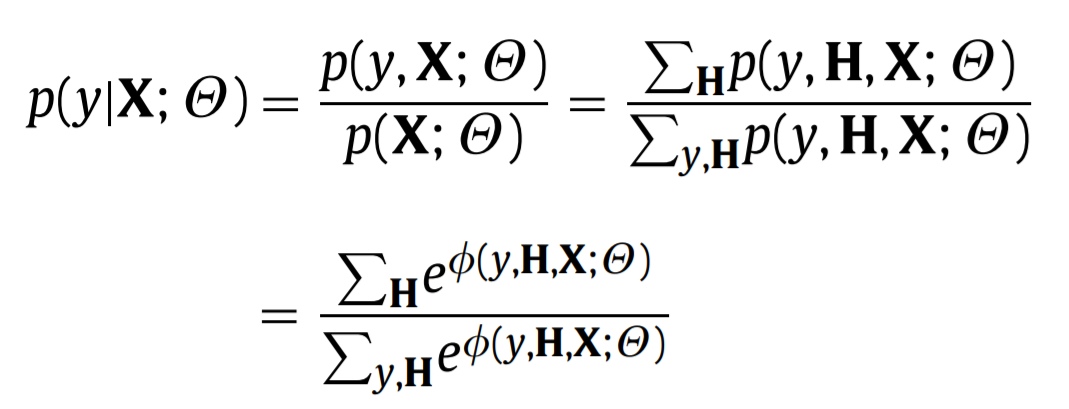
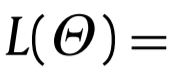
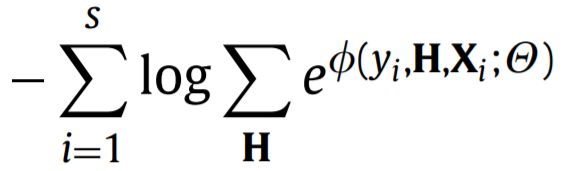
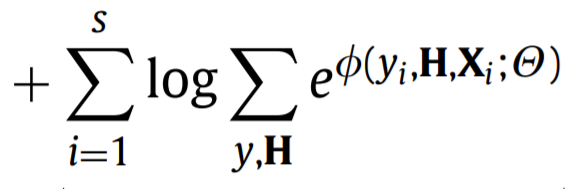
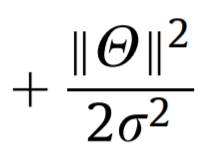
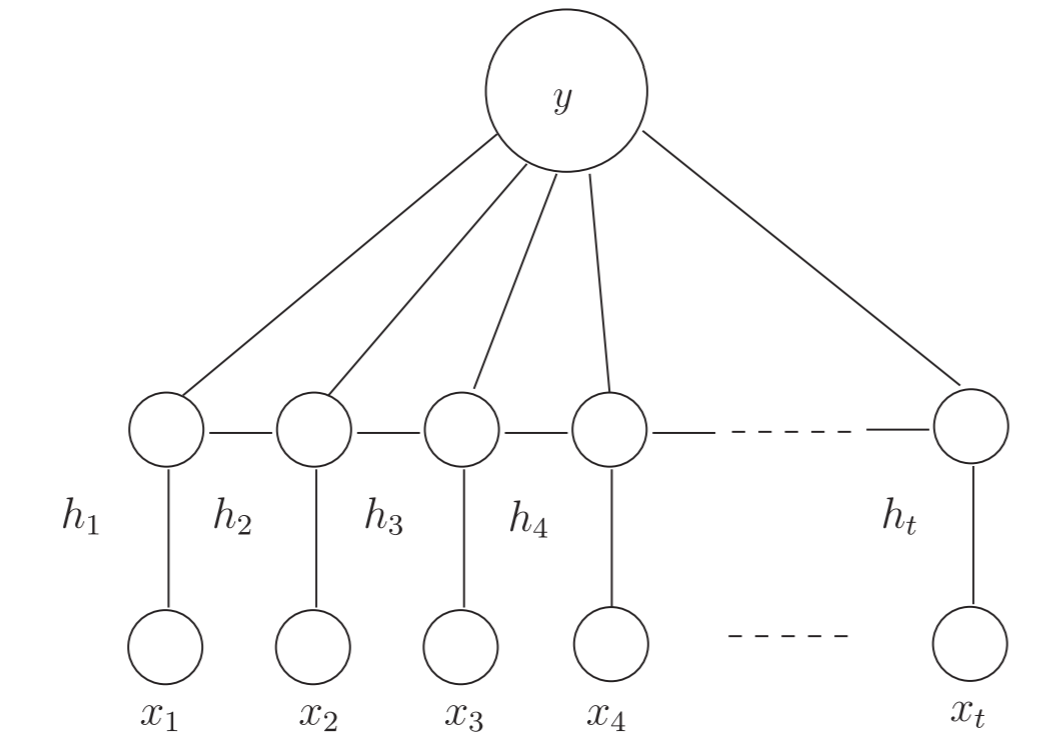
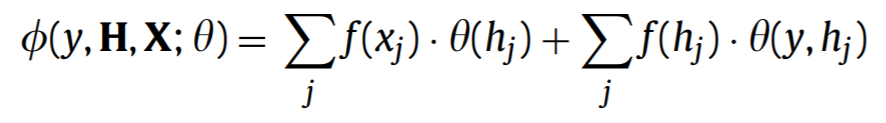
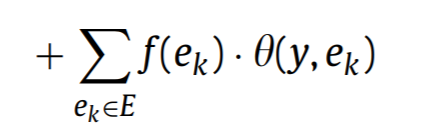
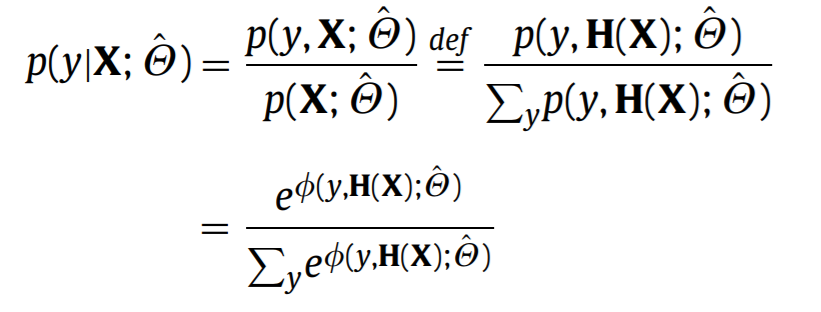
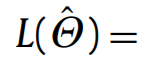
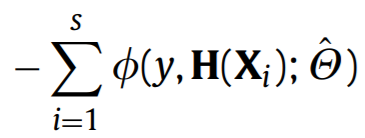
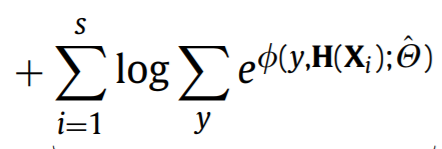
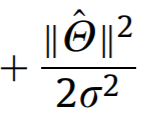
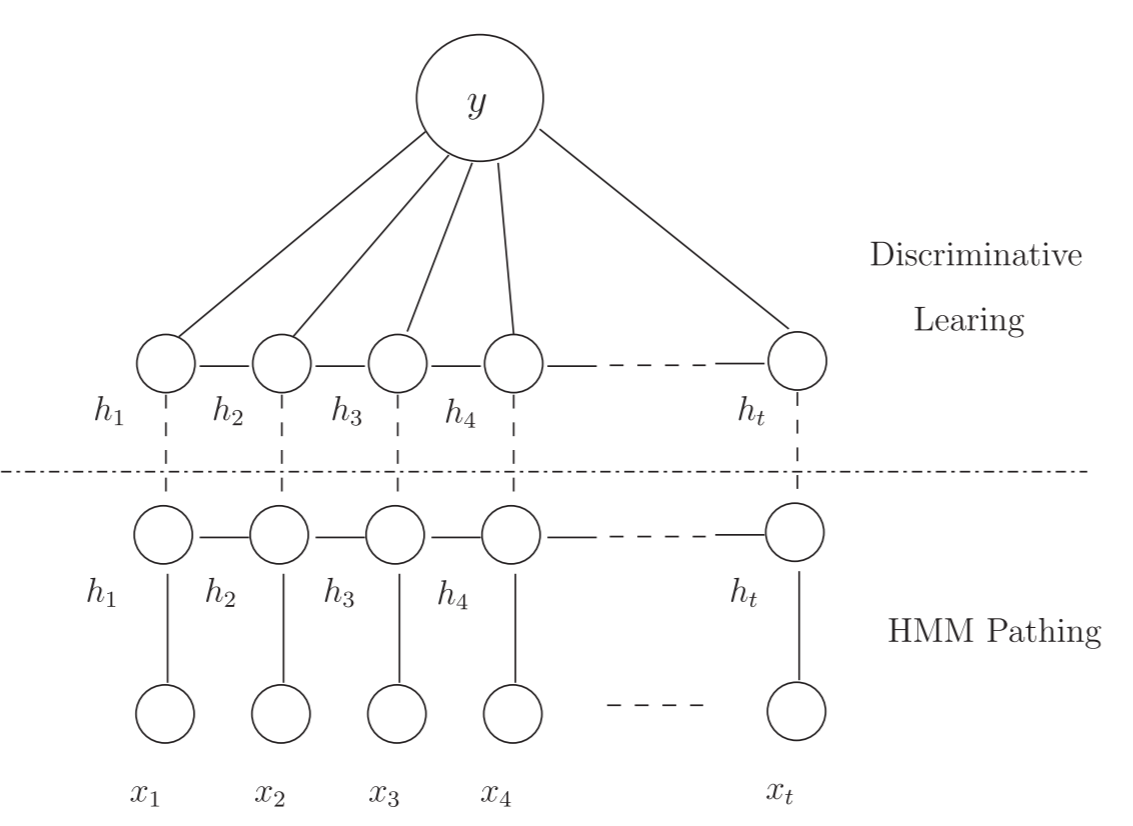
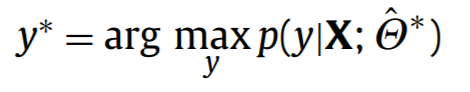
**Paper Review: “Action Categorization with Modified Hidden Conditional Random Field”**

* Paper presents a method for action categorization with a modified hidden conditional random field (HCRF)
  + Effective silhouette-based action features are extracted using motion moments and spectrum of chain code
* Formulate modified HCRF (mHCRF) to have a guaranteed global optimum in the modelling of the temporal action dependencies after the HMM pathing stage
* Experiment results on action categorization using this model are compared favourable against several existing model-based methods, including GMM, SVM, log reg, HMM, CRF, and HCRF
* Two key elements in modelling human actions: local appearance and temporal dependences
* Silhouette-based action recognition has been popular
  + I.e. an action is represented by a series of body shapes
* Silhouette usually extracted by estimation of background or given a known fixed background
* Other feature representations of action include:
  + Space-time interest points
  + Optical flow
  + Motion template
  + Space-time volumes
  + Shape context from still images
* Numerous methods have been proposed (mostly graphical models) for learning temporal dependencies between consecutive frames
* HMM is a baseline approach for modelling temporal dependencies
  + Model parameters are estimated based on the optimization of the joint prob between observations and separate labels (marginalized over hidden variables)
  + Hence, generative model sare not optimized based on conditional Bayesian info
* Though HMM is good in many applications, for pattern discrimination a common consensus is that ideal model should be derived and optimized based on maximising a discriminative function
  + Hence, HMM not optimal
* Hence use of CRFs,
  + However, CRFs cannot incorporate the need for labelling a whole sequence as an action, and also cannot capture the intermediate structures using hidden state variables
  + HCRFs proposed to overcome these issues
* Compared to CRF, HCRF is capable of incorporating a sequence label into the optimization of observation conditional probabilities
* However, due to non-convex nature of objective function of HCRF, performance depends heavily on its parameter initialization
  + Not guaranteeing to give good results in a real application
* To address this, the paper formulates a mHCRF based on HMM pathing and prove objective function of mHCRF is convex which a global optimum after the hidden variables becomes observable
* Further develop effective approach to silhouette-based action recognition using mHCRF
* Extract both a set of spectrum features using Fourier transform applied to chain code of silhouettes and a set of motion moment features
* Relationship between whole sequence label and temporal dependencies is then learned using mHCRF
* Finally, compare this approach to other techniques, e.g. HMM, GMM, log reg, SVM, CRF< and HCRF for action categorization
* Describing silhouette of objects by their chain code has been widely adopted for shape retrieval/recognition or matching
* Chain code itself not invariant to shape orientation change caused by possible 3D pose changes from human-body actions
  + If human body rotated by ‘a’, corresponding chain code ‘C(p)’ will be shifted by offset, e.g.
  + Let ‘C(p+)’ be resulting chain code after rotation
  + To obtain rotation invariance, perform Fourier transform on ‘C(p)’, resulting in abs(F(C(p))) = abs(F(C(p+))
  + First ‘n’ components of Fourier spectrum selected as our action features (i.e. spectrum features)
* Spectrum features good for capturing actions that cause body shape change
  + E.g. in a bending/walking sequence
* However, human body actions not always necessarily associated with significant body shape change
  + E.g. in jumping sequence, human silhouette does not change a great deal over time
  + Results in its spectrum features being less discriminative
* To overcome this problem, utilize motion moment features computed based on human silhouettes
* Those silhouettes are extracted from already-known background using binary sequences
* Found that for binary images sequences, simple inter-frame differencing method performs well
  + For simplicity and computational proficiency, used to detect motion changes resulting in binary image
  + I.e. at time ‘t’, the binary image ‘Bt’ obtained by difference between 2 consecutive frames,
  + Bt = Bt+1 - Bt
  + Set of moment features are then extracted from each binary image Bt
* To make moment features more robust to noise, reject those features extracted from regions with small ‘At’:
  + At = t[x,y]
* Moment features shown to be more distinctive than the spectrum features in a jumping sequence and in a waving sequence
* Indicates that these two types of features are complementary
* HCRF first used for speech classification
  + Then applied to gesture and object recognition
* Given sequence composed of a set of ‘n’ local observations {x1, x2, ….} denoted by ‘X’ and class labels y Y, we want to find a mapping ‘P(y|x)’ between them (y conditioned on X)
* HCRF defined as:

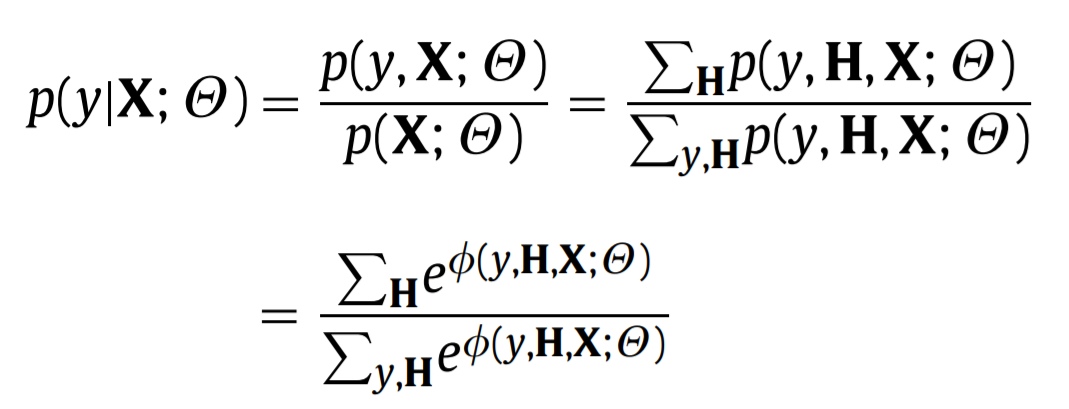


Where ‘’ = set of parameters of the model and ‘H’ = {h1, h2, …., hn}

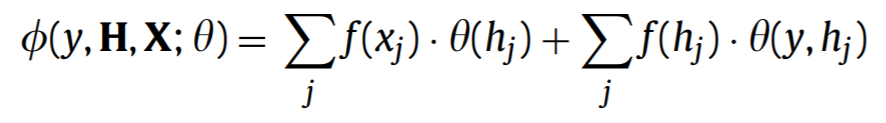
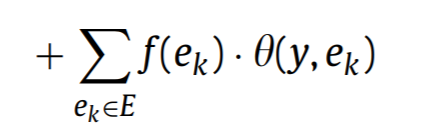
* Each hi H captures certain underlying structure of each class and H is the set of hidden states of the model
* is the potential function which measures compatibility between a label ‘y’, a set of observations ‘X’, and a configuration of hidden variables ‘H’
* Based on ML estimation, regularized version of objective function of HCRF is:
*    
* Where ‘s’ = total # of train sequences with known class labels
* Best params \*=argminϴ L() can be found by gradient descent using Quasi-Newton optimisation
* Objective function and gradient can be written in terms of marginal distributions over hidden variables
* Can see that HCRF is not convex (terms 2 and 3 in above equation don’t guarantee convexity)
  + Global converge heavily depends on initialization
* Important to normalize each data first since sum of potential could result in infinite values in inference process for gradient calculation (could cause numerical instability)
* Potential function (in context of action categorization) can be defined in terms of the following forms where observations interact with hidden states and the sequence class labels interact with both the individual hidden node and the edges between hidden nodes
* 
* Where ‘y’ = class label of entire sequence, ‘h­­i’ = hidden state at node ‘i’, and ‘X’ = observation sequence
*  
  + Where ‘e­k’ is an edge between nodes ‘j’ and ‘j’’
* In action recognition, HCRF graph model is defined as a chain where each node corresponds to hidden state variable at time ‘t’
  + ‘f(x­j)’ is feature vector of node ‘j’
  + ‘f(hj)’ is feature vector corresponding to hidden node ‘j’
  + ‘f(ek)’ is feature vector corresponding to edge between node ‘j’ and ‘j’’
* Requirement to carefully select initial params of HCRF limits its usefulness
* Idea is to make hidden variables observable under the condition of learning the HMM
  + Once hidden variables become ‘observable’ to HCRF, objective function can be shown to be convex
* Approach done in two steps: automatic HMM pathing, and discriminative learning w/ global optimum
* First, learn an HMM for each action class
  + # of hidden states automatically selected by a GMM using minimum description length (MDL)
* Then, compute Viterbi path for each training sequence
  + ‘HMM pathing’
* Thus, node of each training sequence is labelled by the learnt class specific HMM and this procedure makes the hidden states ‘observable’
* The Viterbi path is inferred, which makes the hidden states of each training sequence observable so we can use them directly in the next step
* After HMM pathing stage, model function of HCRF becomes:
* 
* Objective function becomes:
*    
* Can see that the effect of HMM pathing stage is to change the relation between H, x, into function form
  + That is, the hidden variable becomes direct function of observation via ‘H(x)’
  + Evident that L() now convex as 1st part is linear function (viewed as convex), 2nd part is a log-sum-exp term (convex), and 3rd is a quad function (convex), hence overall is convex
  + 
* Proposed approach retains advantage of the general HCRF (i.e. discriminative learning the parameters of the edge potentials, the temporal dependencies interacted w/ action sequence label)
  + Complexity of the approach is reduced by making the hidden variables observable
* Note that objective function of CRF is ‘log (H = observed node label instead of the whole sequence class label), which is different to the one outlined here
* Once estimated the model parameters, the test of new sequence is straightforward by maximizing posterior probability of learnt model with parameters
  + Thus, the final decision rule is: 
* To test the proposal method for action categorization, tested it against other models including GMM, HMM, log reg, SVM, and CRF
* Dataset contains 10 action classes with 93 low res video sequences of 9 people performing: running, walking, jumping-jack, jumping-forward-on-two-legs, jumping-in-place-on-two-legs, galloping-sideways, waving-two-hands, waving-one-hand, bending, and skipping
* Silhouette of each frame is extracted based on subtraction of the median background from each sequence and threshold in colour space
  + Shape chain code is extracted and Fourier transform performed
  + First 10 magnitudes of Fourier response used as spectrum features
* For HMM pathing stage, use EM algorithm to optimize model parameters
* In training a normal HMM (for comparison with mHCRF), the number of hidden states and transition matrix are automatically initialized using MDL criteria over the whole training set
  + Then learn HMM for each class respectively
  + For another comparison with mHCRF, learn for each class a GMM
  + Compared to CRF, where a single CRF is learned for all action classes and then infer the Viterbi path for each test sequence (label of whole sequence is computed as the most frequence frame labels in the Viterbi path)
* HCRF outperforms CRF due to its ability of discriminative learning of hidden states structures, while mHCRF performs the best of all models
* Regarding the features, moment features usually better or comparable than spectrum features
  + Indicates that moment features have greater discriminative potential than spectrum features
  + Combination of both gives better results than using alone
* HMM pathing stage in mHCRF approach only outputs the optimal path (i.e. optimal hidden states sequence) of observation sequence based on Viterbi
  + Doesn’t output action label of the whole sequence
* ‘Run’ and ‘skip’ action sequences misclassified by CRF approach, but correctly classified by ‘mHCRF’ approach despite similar visual appearance
* In this paper, focused on classification of action images as a whole, rather than identifying the detailed body configurations
* Worth noting that another line of research in motion activity recognition is based on human parts
  + In those, task of motion recognition can also be performed by firstly identifying the body configurations and then inferring relationships between candidate body parts
* Some factors needed to be taken into account in the silhouette extraction process when background in unknown (e.g. shadow, illumination changes, camouflage)
* Other possible features besides moment and spectrum (e.g. shape, context features)

**Significant Points and Takeaways from Paper**

* Formulate modified HCRF (mHCRF) to have a guaranteed global optimum in the modelling of the temporal action dependencies after the HMM pathing stage
* Through HMM is good in many applications, for pattern discrimination a common consensus is that ideal model should be derived and optimized based on maximising a discrim function, hence HMM not optimal
* However, CRFs cannot incorporate the need for labelling a whole sequence as an action, and also cannot capture the intermediate structures using hidden state variables; HCRFs proposed to overcome these issues
* Compared to CRF, HCRF is capable of incorporating a sequence label into the optimization of observation conditional probabilities; however, due to non-convex nature of objective function of HCRF, performance depends heavily on its parameter initialization; not guaranteeing to give good results in a real application
* To address this, the paper formulates a mHCRF based on HMM pathing and prove objective function of mHCRF is convex which a global optimum after the hidden variables becomes observable
* Spectrum features good for capturing actions that cause body shape change (e.g. in a bending/walking sequence); however, human body actions not always necessarily associated with significant body shape change (e.g. in jumping sequence), which results in its spectrum features being less discriminative
* To overcome this problem, utilize motion moment features computed based on human silhouettes
* Moment features shown to be more distinct than spectrum features in a jumping and in a waving sequence
* HCRF defined as:



Where ‘’ = set of parameters of the model and ‘H’ = {h1, h2, …., hn}

* Potential function defined as:  
  + Where ‘e­k’ is an edge between nodes ‘j’ and ‘j’
* Requirement to carefully select initial params of HCRF limits its usefulness
* Idea is to make hidden variables observable under the condition of learning the HMM
  + Once hidden variables become ‘observable’ to HCRF, objective function can be shown to be convex
* First, learn an HMM for each class, then compute the Viterbi path for each training sequence (‘HMM pathing’); thus the node of each training sequence is labelled by the learnt class specific HMM and procedure makes the hidden states ‘observable’
* Proposed approach retains advantage of the general HCRF (i.e. discriminative learning the parameters of the edge potentials, the temporal dependencies interacted w/ action sequence label)
  + Complexity of the approach is reduced by making the hidden variables observable
* HCRF outperforms CRF due to its ability of discriminative learning of hidden states structures, while mHCRF performs the best of all models
* Regarding the features, moment features usually better or comparable than spectrum features
  + Indicates that moment features have greater discriminative potential than spectrum features
  + Combination of both gives better results than using alone