Annotation of Human Motion Capture Data using Conditional Random Fields

Mert Değirmenci
Department of Computer Engineering,
Middle East Technical University, Turkey
mert.degirmenci@ceng.metu.edu.tr

Anıl Sevim
Department of Computer Engineering,
Middle East Technical University, Turkey
anil.sevim@ceng.metu.edu.tr
Outline

- Introduction
- Literature Survey
- Problem Definition & Algorithm
- Experimental Evaluation
- Future Work
- Conclusion
Introduction

• Problem is classifying and labelling motion capture data

• Tagging motion sequence is a multi-label supervised learning problem in general

• A linear-chain Conditional Random Field (CRF) used to tackle the problem

• Each example supplied to CRF can be represented as (x,y) pairs, where x denote a motion data sequence and y denote a sequence of corresponding tags.
**Literature Survey**


*Features such as 3D point positions, joint angles, and PCA reducted information have been used to compare and identify a new motion sequence.[1-3]*
Literature Survey (cont.)

  *Arikan et al. have applied SVM classifiers to allow users annotate large collections of human motion data quickly.* [4].

  *Manabe et al. propose a stochastic model of human motion for showing the differences between each observed motion.* [5].
Problem Definition & Algorithm

Task Definition:

• In this study, we investigate annotation of human motion sequence by set of tags.

• It is fairly challenging task to classify a human motion data since it lacks categorical structure.

• During transitions of activities, there can be patterns of different motion classes.

• Human motion also exhibits ambiguous time-scales.

• Persons walking speeds and patterns may vary greatly.
Problem Definition & Algorithm (cont.)

Task Definition (cont.):

- Human perception system may select different frames of the sequence of a walking human to identify the activity.
- For more complex activities, the problem exhibits larger temporal dependency and longer contexts of information may need to be examined.
- Varying time slices for different motion categories makes identifying key frames even more complicated.
- A human may perform sequence of simple activities through the process of much complicated motion activity.
- Even seemingly basic categories to humans can often be decomposed into more primitive activities.
Problem Definition & Algorithm (cont.)

Algorithm Definition:

- Conditional Random Fields (CRF) is a discriminative probabilistic model which is a special case of log-linear models used for labeling and parsing sequential data.

- CRF is can be represented as an undirected graphical model where $X$ denote the observation sequence to be labeled and $Y$ denote the label sequence.
**Problem Definition & Algorithm (cont.)**

**Algorithm Definition (cont.):**

- Log linear models express the probability of a possible output tags $y$, with corresponding input motion sequence $x$. $F_j$'s are feature functions and $w_j$'s are weights of the feature functions to be learned from training examples. Denominator $Z(x, w)$ is a normalization factor for the probability to fall into $[0, 1]$.

$$p(y|x; w) = \frac{\exp(\sum_j w_j F_j(x, y))}{Z(x, w)}$$

- $F_j$'s represent high level feature functions composed by $f_j$'s, which are low level feature function that can depend only on current label, previous label, current position $i$, and motion sequence $x$.

$$F_j(x, y) = \sum_i f_j(y_{i-1}, y_i, x, i)$$
Problem Definition & Algorithm (cont.)

Algorithm Definition (cont.):

- For a trained CRF with weights $w$, following rigorous mathematical derivations show us that inference on $C$ is possible in reasonable time with dynamic programming. Our aim of inferring a label sequence for the given and unobserved motion sequence can be mathematically expressed as computing $y'$.

$$
\tilde{y}' = \arg\max_{y} p(y|x; w) = \arg\max_{y} \sum_j w_j F_j(x, \tilde{y})
$$

- If length of $x$ is $n$, and cardinality of set of labels is $m$, $y'$ can be computed in $O(m^2.n)$ time. This dynamic programming algorithm turns out to belong Viterbi class of algorithms which are also utilized in HMMs.

$$
U(k, v) = \max_{y_{k-1}} [U(k - 1, y_{k-1}) + g_k(y_{k-1}, v)]
$$
Problem Definition & Algorithm (cont.)

Algorithm Definition (cont.):

- Log linear models express the probability of a possible output tags \( y \), with corresponding input motion sequence \( x \). \( F_j \)'s are feature functions and \( w_j \)'s are weights of the feature functions to be learned from training examples. Denominator \( Z(x, w) \) is a normalization factor for the probability to fall into \([0,1]\).

\[
p(y|x; w) = \frac{\exp(\sum_j w_j F_j(x, y))}{Z(x, w)}
\]

- \( F_j \)'s represent high level feature functions composed by \( f_j \)'s, which are low level feature function that can depend only on current label, previous label, current position \( i \), and motion sequence \( x \).

\[
F_j(x, y) = \sum_i f_j(y_{i-1}, y_i, x, i)
\]
Experimental Evaluation

- Experiments conducted on CMU database.
- In order to supply the structured input that the CRF requires, downloaded files are preprocessed.
- The training examples in CMU database are variable length mocap files associated with one label. The training examples are composed of randomly sampled labels with fixed durations.
- The training set is composed of 621 examples whereas the size of test set is 123. Each training example can be regarded as a set of frame chunks where each chunk is 10 frame. The frame consists of sensor outputs at a specific time instant.
Experimental Evaluation (cont.)

- The overall accuracy of the algorithm is measured by the metrics below:

\[
Detection = \frac{100TP}{TP + FN} \\
Branching = \frac{FP}{TP} \\
Quality = \frac{100TP}{TP + FP + FN}
\]

- TP, FP, and FN are abbreviations for True Positives, False Positives, and False Negatives respectively.

- The Detection metric measures the performance of the annotator. The Branching metric measures the delineation performance. And Quality can be thought as measure of overall performance of the algorithm.
Experimental Evaluation (cont.)

• When sufficiently large number of training examples are randomly sampled from the entire distribution, two types of feature functions dominates the final decision on labeling.

• These are angle & magnitude statistics feature functions.

• The information gain obtained by using these feature functions are given in figures below:
Experimental Evaluation (cont.)

- The Detection metric for our algorithm with the same test environment is 80%. The branching metric has been found as 0.24, and the overall quality is approximately 77%.

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>B</th>
<th>Q</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manabe</td>
<td>70%</td>
<td>0.25</td>
<td>69%</td>
<td>[7]</td>
</tr>
<tr>
<td>Yamato</td>
<td>75%</td>
<td>0.32</td>
<td>72%</td>
<td>[8]</td>
</tr>
<tr>
<td>Muller</td>
<td>73%</td>
<td>1.65</td>
<td>56%</td>
<td>[9]</td>
</tr>
</tbody>
</table>

- Figure below elaborates the results on specific labels.
Experimental Evaluation (cont.)

• The result of a single experiment where the training example consists of 3 chunks.

• The outputs are "walk", "run", and "leap" respectively.

• In this study, we have suggested an algorithm to classify a human motion capture data into one of the previously trained categories. The CRF is known to be the best performing algorithm on sequential data labeling where there are dependencies between adjacent labels.

• Our proposed feature functions capture the intrinsic information in human motion data such as angular variations, label transition probabilities, and the magnitude changes in sensor outputs.
Future Work

- Mocap databases such as CMU can be used in other works.
- A reliable classifier can be useful in anomaly detection, gait disease diagnosis, and content-based video querying.
- Suggested stochastic CRF based framework can be used for 2D images, videos etc.
- Labels can be increased and specialized according to desired work.
Conclusion

- Conditional random fields offer a unique combination of properties: discriminatively trained models for sequence segmentation and labeling.
- Advantages of CRFs are especially important in human motion recognition.
- The sequential data obtained by motion capture systems exhibit high dependencies between several key frames.
- HMM based algorithms fail to explain these dependencies since they have strict assumptions over the input distributions.
- The experimental results confirmed the inherent superiorities of CRF based models to HMM based approaches for human motion annotation.