Generating Concise Rules for Retrieving Human Motions from Large Datasets

Tomohiko Mukai and Ken-ichi Wakisaka and Shigeru Kuriyama
Toyohashi University of Technology
1-1 Hibarigaoka, Tenpaku-cho, Toyohashi, 441-8580 Aichi, Japan
{mukai@|k_wakisaka@val.|kuriyama@}ics.tut.ac.jp

Abstract

This paper proposes a method for retrieving human motion data with concise retrieval rules based on the spatio-temporal features of motion appearance. Our method first converts motion clip into a form of clausal language that represents geometrical relations between body parts and their temporal relationship. A retrieval rule is then learned from the set of manually classified examples using inductive logic programming (ILP). ILP automatically discovers the essential rule in the same clausal form with a user-defined hypothesis-testing procedure. All motions are indexed using this clausal language, and the desired clips are retrieved by subsequence matching using the rule. Such rule-based retrieval offers reasonable performance and the rule can be intuitively edited in the same language form.

Keywords: motion capture, motion indexing, motion appearance feature, inductive logic

1 Introduction

Automated retrieval methods of human motion data have been proposed using some feature analysis techniques. However, existing methods have one major problem with query formulation. The numerical methods (Chiu et al., 2004) use a short motion clip as a query, but the similarity measure between motions should be manually defined with fine parameter tunings. The learning-based methods (Arikan et al., 2003) implicitly obtain a motion classifier which can not be modified after the learning. The template-based methods (Müller and Röder, 2006) use many binary symbols so as to enable them to represent a wide variety of human movement, making the notation often redundant for retrieval problem. Although the heuristic method (Müller et al., 2008) eliminates the redundancy of the template, the conventional method does not guarantee the minimality of the resulting template and requires high computational cost.

In the fields of artificial intelligence, a general induction technique has been developed to discover an effective solution to multiple classification problems. The induction method often uses a logical language such as symbolic and clausal language to represent the training data, and discovers a concise classification rule using logical programming, called inductive logic programming (ILP) (Muggleton, 1995). ILP analyzes an essential rule presented in the explicit logical language, and provides a programmable framework based on the same logical language to control its learning procedure.

We propose a rule generation technique for human motion retrieval using ILP. Our method first computes a set of spatio-temporal features of motion appearance in the form of a multivalued logical expression. An ILP framework then discovers an essential classification rule, which is composed of a few logical expressions, by analyzing an intrinsic difference among the set of training motion clips. The desirable segments are retrieved from a database using the discovered rule by specifying the name of the motion class. Moreover, such a retrieval rule can be easily edited in the form of logical language to improve the retrieval accuracy. Consequently, our system provides flexible motion retrieval with semi-automated rule generation.

2 Algorithm

We here explain how to discover the retrieval rule. Given training data, they are manually segmented into clips of unit movement and classified into multiple semantic classes. One class is then chosen as a positive class and others are used as a negative class. Each training clip is represented by a set of clauses corresponding to the spatio-temporal motion features. The inductive learning discovers a retrieval rule consisted of as few clauses as possible so that the resulting rule explains common features of the positive examples and no features of the negative ones.

2.1 Spatio-temporal features of motion appearance

Given a training motion clip, several key-poses are first extracted from the training data to reduce the computational cost of the learning. After selecting the first key-pose at the first frame of the motion sequence, the next key-pose is sequentially searched until the pose distance to the previous key-pose exceeds a given threshold. Next, the multivalued spatial features are computed at each key-frame and then represented in the clausal form like $has\_sf(f_i[l_i])$, where f_i and l_i denote a name and quantization index of a spatial feature, respectively. This clause means that the training data has a pose indicating a spatial feature f_i with the quantization index l_i, where we omit $[l_i]$ for binary features for simplicity. Our definition of spatial feature includes 31 geometrical features proposed in (Müller and Röder, 2006) and 4 additional customized features.

We also define two types of temporal features that explain the duration of a spatial feature and a temporal relation between different spatial features. The duration is represented by two clauses: $long(f_i[l_i])$ and $short(f_i[l_i])$, which indicate longer and shorter duration than 0.5 sec, respectively. The temporal relation is represented by a clause: $after(f_i[l_i], f_j[l_j], l_t)$, for the spatial feature $f_j[l_j]$ appearing after $f_i[l_i]$ with a quantized delay l_t. The time delay index l_t is represented by three symbols: *short* ($[0, 0.25)$ sec), *middle* ($[0.25, 0.5)$ sec), and *long* ($[0.5, 1.0)$ sec), where these time ranges are experimentally optimized.

2.2 Simplification of retrieval rules

Given the clauses of spatio-temporal feature of a training data, the ILP framework discovers the retrieval rule for each motion class. We use a public ILP system, called Progol (Muggleton, 1995), which uses a programmable hypothesis-testing procedure to discover an essential rule. It uses both positive and negative examples to discover a rule that is obeyed by the positive examples and is excluded by the negative examples. This learning model often results in too strict a retrieval rule, which can be reduced by relaxing the tolerance of the quantization error of multivalued feature.

The learning criterion of ILP is the minimality of the clauses used in retrieval rules. Multiple clauses can often be substituted with a simpler clause based on a syllogism and other reasoning. ILP introduces the substitution procedure with user-defined logical expressions represented in the clausal form for discovering the retrieval rule that consists of as few clauses as possible. We define the subsumption relation of multivalued feature, which is modeled by a combinational structure. The basic component of the structure is the quantization index of multivalued feature. The ILP system then selects the most appropriate subset to best describe the feature of training data.

2.3 Subsequence search with space windows

By specifying the name of motion class, motion segments are retrieved by a subsequence search using the retrieval rule associated with the specified class. Our system sequentially searches the subsequence that includes all constituent clauses of the retrieval rule from the motion sequence. The discrete representation of spatial feature often decreases retrieval accuracy because its quantization index is computed by regularly discretizing the geometrical distance between body parts. The space window is therefore introduced for tolerating the small variation of simple quantization of multivalued feature. If a quantization index l_f is assigned to the interval $[d_i, d_{i+1})$, where d_i and d_{i+1} are the geometrical distance between body parts, the retrieval process uses a wider range $[d_i - \alpha, d_{i+1} + \alpha)$ for discriminating the region of the quantization index l_f. The margin α is experimentally optimized by the quantization interval $\alpha = 0.5|d_{i+1} - d_i|$.

Table 1: Retrieval rules discovered from training dataset, where the number in [] represents a quantization index of multivalued features.

Class	Retrieval rule
Cartwheel	long(lhand_up[2]) & long(gradient) & long(move_upward)
ElbowToKnee	long(larm_bent[2])
	long(larm_bent[1]) & after(larm_bent[0], lfoot_up[0], middle)
	short(rarm_bent[1]) & has_sf(lleg_bent[0])
JumpingJack	long(move_upward) & short(lhand_up[1]) & after(rhand_up[2], lhand_up[0], short)
Lie	long(lying)
Sit	long(move_upward) & long(gradient) & long(body_bent[1])
Squat	long(rhand_up[0]) & has_sf(body_bent[0])
	long(lhand_up[1]) & long(rhand_up[1]) & has_sf(body_bent[0])
Toss	long(move_upward) & short(larm_bent[0]) & after(larm_bent[1], rhand_up[1], middle)

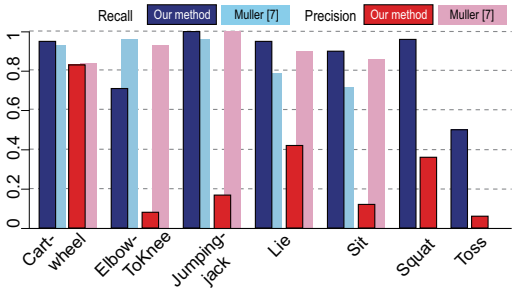

Figure 1: Retrieval results of seven motion classes. The dark-colored bars and light-colored ones represents the performance of our method and existing method (Müller et al., 2008), respectively.

3 Experimental result

The retrieval performance of our method is compared with the existing heuristic method (Müller et al., 2008) under almost the same experimental condition. We experimentally retrieved motion segments from a large public collection of motion capture data (http://www.mpi-inf.mpg.de/resources/HDM05). We manually segment a whole motion sequence of 120 minutes into 5481 clips of unit movement and arranged them into 99 motion classes. The training dataset consists of 7 motion classes; *Cartwheel*(6/21), *ElbowToKnee*(17/58), *JumpingJack*(15/52), *Lie*(6/20), *Sit*(6/20), *Squat*(16/56), and *Toss*(4/14)), where the two numbers in () denote the number of training data of each class and the total number of motion clips, respectively.

3.1 Discovery of retrieval rules

Table 1 shows the retrieval rules for the seven motion classes which are discovered using the training dataset. It shows that a motion clip is classified as a cartwheel motion if the actor bend his/her body and raises his/her left hand, and moves upward for a long period. The number of constituent clauses is determined according to the uniqueness of movements in comparison with other motion classes. For example, the rule of *Lie* motions only has one clause because the spatial feature of *Lying* appears only in the *Lie* motion class. On the other hand, the ILP framework discovers multiple retrieval rules for *ElbowToKnee* and *Squat*, and the desired segment is retrieved using all rules. This indicates that these motion classes can be respectively divided into subclasses. In fact, the training dataset of *ElbowToKnee* includes symmetric motions.

3.2 Retrieval by discovered rule

The statistics of the retrieval performance is shown in Figure 1. Average computational time of the retrieval is about 10 milliseconds, which is fast enough for practical usage. High recall indicates that the subsequence matching with a retrieval rule successfully retrieves almost all relevant motions. The low recall of *Toss* is probably caused by the overfitting problem; the ILP generalizes the small common part of the example motions which do not appear in other *Toss* motions that are not used in the learning. On the other hand, the precision values are remarkably lower than those of the existing method except for *Cartwheel*. This means that the discovered rules can not exclude the non-relevant motions because the number of training data is too small to generalize the retrieval rule for the size of the entire database. However, a large number of examples often lead to a failure in learning because of the limited memory capacity. We consider that the accuracy of our retrieval method becomes acceptable for the practical motion database because the retrieval performance could be improved by a manual editing.

3.3 Extensions of rule-based retrieval

The manual editing can make the retrieval rule more distinctive and often improves the retrieval accuracy. For example, we modify the second clause of the rule of *Cartwheel* from *long(gradient) & long(lhand_up[2])* to *long(somersault)*, based on our knowledge that cartwheel motion includes a handstanding pose. This modification increases the precision and recall from 0.83 to 1.0 and 0.95 to 1.0, respectively. This improvement is attributed to the constraint of *somersault* stricter than that of the *gradient* where both features appear in most cartwheel motions. Such an artificial decision can be integrated into the rule by simple text editing.

Our rule-based method can also use a motion clip for a retrieval key. Given a query clip, every retrieval rule is checked if it categorizes the query motion into one of the given class, and the validated rules are then used for retrieving the similar motion segments. We use the short motion clip composed of several types of gymnastic movements for the retrieval query. Our system validates that the query motion clip consists of subsequences categorized as *ElbowToKnee* and *Squat*. The related motion clips are then retrieved from the database using the two corresponding retrieval rules. This approach enables the retrieval of a semantically similar motion with a large difference in appearance. This property can overcome the limitations in existing numerical techniques.

4 Conclusions

This paper has proposed a rule generation technique for motion retrieval using ILP. The clausal formulation provides a meaningful representation of human motion and its retrieval rules. The retrieval rules are efficiently learned within the ILP framework from a set of manually classified training data. The discovered rule is directly edited in the clausal form. By specifying the name of a motion class, motion segments are efficiently retrieved from a large database using the rule assigned to the motion class with the space windows. Our system also retrieves the motions using a short motion clip for the retrieval query, which actually uses the retrieval rule associated with the query clip.

The major limitation of our method is that the retrieval rule can not be incrementally learned. Another limitation is that our method requires fine adjustment of many numerical parameters. Furthermore, the manual segmentation of the training motions often affects the retrieval accuracy, which is a general issue in example-based motion retrieval techniques. These problems could be alleviated by statistically optimizing the thresholds or using a fuzzy representation in the logical expression. Our future work also involves the investigation of the adaptive sampling method for selecting training data essential to rule generation.

Acknowledgement

This work was supported by the MEXT Grant-in-Aid for Young Scientists (B) 19700090 and Scientific Research (B) 18300068.

References

Arikan, O., Forsyth, D. A., and O'Brien, J. F. (2003). Motion synthesis from annotations. *ACM Transactions on Graphics*, 22(3):402–408.

Chiu, C.-Y., Chao, S.-P., Wu, M.-Y., Yang, S.-N., and Lin, H.-C. (2004). Content-based retrieval for human motion data. *Journal of Visual Communication and Image Representation (Special Issue on Multimedia Database Management Systems)*, 15(3):446–466.

Muggleton, S. (1995). Inverse entailment and progol. *New Generation Computing*, 13:245–286.

Müller, M., Demuth, B., and Rosenhahn, B. (2008). An evolutionary approach for learning motion class patterns. In *Symposium of the Germann Association for Pattern Recognition*.

Müller, M. and Röder, T. (2006). Motion templates for automatic classification and retrieval of motion capture data. In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation 2006*, pages 137–146.