Fuzzy conditional random fields for temporal data mining

Intan Nurma Yulita and Atje Setiawan Abdullah
Department of Computer Science, Faculty of Mathematics and Natural Sciences
Universitas Padjadjaran, Indonesia
E-mail: intan.nurma@unpad.ac.id, atjesetiawan@gmail.com

Abstract. Temporal data mining is one of the interesting problems in computer science and its application has been performed in a wide variety of fields. The difference between the temporal data mining and data mining is the use of variable time. Therefore, the method used must be capable of processing variables of time. Compared with other methods, conditional random field has advantages in the processing variables of time. The method is a directed graph models that has been widely applied for segmenting and labelling sequence data that appears in various domains. In this study, we proposed use of Fuzzy Logic to be applied in Conditional Random Fields to overcome the problems of uncertainty. The experiment is compared Fuzzy Conditional Random Fields, Conditional Random Fields, and Hidden Markov Models. The result showed that accuracy of Fuzzy Conditional Random Fields is the best.

1. Introduction
Several studies have introduced the method of for temporal data mining for example use artificial neural networks [1], [2], decision tree, and support vector machine. The other hand, the widely used method for sequence data is Hidden Markov Models (HMM). Research conducted by Rabiner is a preliminary study using HMM for dividing and labeling sequence data that appears in various domains such as voice recognition [3]. Since then, HMM is widely used in other pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, and bioinformatics. Other studies also suggest the use of HMM for the classification of motion [4], [5], hand gestures [6], and speech [7]. However, HMM has some limitations such as distribution of the conditional probability of hidden variable x only at time t, not along duration since HMM observation is a scalar (1 dimension). In order to achieve better performance than HMM, Laferty et al proposes the use of Conditional Random Fields (CRF) to divide and label the sequence data but it can overcome the limitations of the HMM [8]. CRF is a directed graph models that has been widely applied for segmenting and labeling sequence data that appears in various domains. CRF is able to combine the features of a complex sequence of observations of X which does not require the assumption of non-independence between observations in variable X. Unlike HMM, CRF accommodate observation data vectors and capable of solving the existing label is inherently biased in HMM.

2. Fuzzy Inference System
Fuzzy Inference System (FIS) is selected to implement fuzzy logic for this study. FIS have two advantages over other fuzzy logic techniques. It is better than the other in handling the linguistic
concepts and also mapping the concept of non linear between input and output [9]. FIS consists of three main steps, namely:

(i) **Fuzzification**
In this step, every crisp input is converted into fuzzy input. To change it, then built a diagram of the membership function for each input.

(ii) **Inference**
In this step, every fuzzy input is converted into fuzzy output. It can be done by building a number of rules that map fuzzy input into output targets.

(iii) **Defuzzification**
In this step, conversion is done fuzzy output into a crisp output. The process is also built a diagram of membership function but it has a different purpose than fuzzification.

3. **Conditional Random Fields**
Input CRF is a segmentation of gesture phase with the overall features are used, while the output is a label of gesture phase. To build CRF in labeling the gesture phase, the necessary set of feature functions $f_i$. In the CRF, each of the features has several input variables, namely:

(i) Segment of gesture phase ($s$)
(ii) The position $i$ of the phase segment
(iii) Label the current phase $l_i$
(iv) Label $l_{i-j}$ from the previous phases or can add $l_{i+j}$ of the phases thereafter.

And the output is real-valued probability [10]. If the feature functions only observe a single phase with its previous phase, the function is $f_i(s, i, l_i, l_{i-1})$. As an example of the features are $f_i(s, i, l_i, l_{i-1}) = 1$ if $l_i$ is rest and $l_{i-1}$ is hold so that if the weights 1 associated with this $f_i$ large and positive, the labeling of the phases being observed labeled rest. To build the CRF, the necessary number of feature functionality. The function of this feature is obtained through a combination of the observation of the observation before or afterwards. Large observations will be used depending on the size of the window to be used. If the window size ($w$) is 0 then the observation phase took only one segment without having a dependency with another segment of the mobile phase. But if the window size is 1, then shift observations will overlap with one segment before / after. In this study, experiment will test change the window size on the performance of the CRF so that it can be seen the influence of the dependence of a segment of the mobile phase to the other segments. Furthermore, the features that have been obtained are given weight. Both of these play an important role in modeling the label of each mobile phase. Modelling is indicated by a score of a label which is calculated using the following formula:

$$score(l|s) = \sum_{j=1}^{m} \sum_{i=1}^{n} \lambda_j f_j(s, i, l_i, l_{i-1})$$ (1)

The first summation calculating the features $j$ and summing the two for each position $i$ of the segmentation of the mobile phase. Weight of features affect CRF in discovering the proper labeling of test data is provided so that the invention proper weight for each feature needs to be done. The weight of this feature is obtained through a learning process of the data used. But if the initial weights are assigned randomly then execution time to obtain the optimal weight value will be longer. The easiest way to gain weight is done by calculating all combinations or sequences for possible configurations labels and count all scores one by one through brute force. But this mechanism will require enormous computing process, therefore some ways that can be done through optimization, one of which is the algorithm Broyden-Fletcher-Goldfarb-Shanno (BFGS) is applied in this study. The final step of the process of labeling the gesture phase was
### Table 1: Dataset

| No | Attributes                                      |
|----|------------------------------------------------|
| 1  | lhx - Position of left hand (x coordinate)     |
| 2  | lhy - Position of left hand (y coordinate)     |
| 3  | lhz - Position of left hand (z coordinate)     |
| 4  | rhx - Position of right hand (x coordinate)    |
| 5  | rhy - Position of right hand (y coordinate)    |
| 6  | rhz - Position of right hand (z coordinate)    |
| 7  | hx - Position of head (x coordinate)           |
| 8  | hy - Position of head (y coordinate)           |
| 9  | hz - Position of head (z coordinate)           |
| 10 | sx - Position of spine (x coordinate)          |
| 11 | sy - Position of spine (y coordinate)          |
| 12 | sz - Position of spine (z coordinate)          |
| 13 | lwx - Position of left wrist (x coordinate)    |
| 14 | lwy - Position of left wrist (y coordinate)    |
| 15 | lwz - Position of left wrist (z coordinate)    |
| 16 | rwx - Position of right wrist (x coordinate)   |
| 17 | rwy - Position of right wrist (y coordinate)   |
| 18 | rwz - Position of right wrist (z coordinate)   |

...continued...

4. **Experiment**

4.1. **Data set**

The data set used in this study came from School of Art, Sciences and Humanities University of Sao Paulo [11]. The data set consists of temporal segmentation of gesture. It is composed by three data recorded using Microsoft Kinect sensor. For each data has 18 features, namely: For the classification, it uses only five phases: Rest (1), Preparation (2), Stroke (3), Hold (4), and Retraction (5).

4.2. **Flowchart**

Three scenarios of temporal data mining in this study are described in the Figure 1.

The mechanism of this research has three scenarios, namely the implementation of HMM, CRF and Fuzzy CRF. To be able to process the data, then HMM requires vector quantization (VQ). VQ in this study is built using K-Means Clustering. While on Fuzzy CRF, it needs additional preprocessing of feature extraction to minimize number of input on the Fuzzy Inference System. The greater the input, the more rules to be built and it will improve the execution time. Some step is done before modeling is started, namely:

...
(i) **Preprocessing**
First scenario implements vector quantization to make all input to conform with Hidden Markov Models, and scenario 3 implements feature extraction and fuzzy inference system.

(ii) **Vector Quantization**
In order to process HMM, an observation sequence is needed. By using K-means clustering algorithm, Vector Quantization is used to construct observation sequence. The process has two steps: the formation of codebook and the determination of codebook index.

(iii) **Feature Extraction**
The process is done by replacing some feature vector with one represented feature vector by Euclidean Distance. The output of this study is six represented features.

(iv) **Fuzzy Inference System**
The system is developed by six inputs, 64 rules, and an output. The system is done by trial and error in this study. The process in this study is shown in Figure 2.

In this study, it will implements Hidden Markov Models, Conditional Random Fields, and also Fuzzy Conditional Random Fields. Hidden Markov Models is a directed graph that represents the joint probability between class and observation. But the observation of Hidden Markov Models confined to one dimension of data between two time segments. Therefore, Conditional
Random Fields implemented to overcome it. The other hand, to handle uncertainties in gesture phase segmentation, Condition Random Fields will be combined with Fuzzy Inference Systems.

4.3. Testing
The experimental process conducted through three data to test some models produced by HMM, CRF, and Fuzzy CRF. For each data, each part of the data divided into test data and the rest into training data by cross-validation scheme. This process is illustrated in Figure 3.

5. Results
The results for three scenarios of temporal data mining are presented in Table 2 and Table 3 respectively

5.1. Accuracy
Accuracy of three methods to test data shown in Table 2. From the experiments that have been done, the percentage of accuracy with HMM, the value only ranged from 20-38, whereas using CRF ranged between 50-65, Fuzzy CRF has a range of between 57-77. It can be concluded is superior to the CRF HMM modeling based on observations due to several segments of time and data features. Accuracy is increased when combined with Fuzzy CRF because this technique is able to overcome the problem of loss of information as it happens if without using Fuzzy.

5.2. Execution time
Time execution of three methods to test data shown in Table 3. The execution time in able 3 are calculated based on only for the classification. From experiments, it is known that CRF has been combined with Fuzzy experiencing a decrease in execution time for process modeling. CRF
Table 3: Execution Time (seconds)

| Fold | CRF   | Fuzzy CRF |
|------|-------|-----------|
| 1    | 115.06| 33.50     |
| 2    | 49.00 | 48.24     |
| 3    | 52.71 | 50.10     |
| Average | 72.26 | 43.95     |

execution time ranged from 48-116 seconds but when combined with Fuzzy Logic, the execution time becomes 33-51 seconds for each fold. This is because the number of features is reduced when it was done Fuzzy Inference System.

6. Conclusion and Future Work
From the analysis of the performance of Fuzzy Conditional Random Fields by using the data in this study, it can be concluded Fuzzy Conditional Random Fields has the highest accuracy. Also, it has the smallest execution time due to the reduction in the number of variables used for classification Since our method is an effective way for temporal Data Mining, it is highly recommended for tuning the parameter of Fuzzy Inference System by using Genetic Algorithms or Artificial Neural Networks.

Acknowledgment
The Authors thank to Rector Universitas Padjadjaran and Direktorat Riset dan Pengabdian Pada Masyarakat (DRPM) Universitas Padjadjaran who gave funding for dissemination of this paper, and also Korean Mathematical Society (KMS) who gave funding for dissemination of this paper through travel grant.

References
[1] Priyanka Gaur 2012 Artificial Neural Networks in Data Mining International Journal of Electronics and Computer Science Engineering 1(3)
[2] Sigillito V G, Wing S P, Hutton L V, Baker K B 1989 Classification of radar returns from the ionosphere using neural networks Johns Hopkins APL Technical Digest 10 pp 262–266
[3] Lawrence Rabiner, Bing-Hwang Juang 1993 Fundamentals Of Speech Recognition Prentice Hall
[4] Bobick a F, Ivanov Y a 1998 Action recognition using probabilistic parsing In Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No.98CB36231). IEEE Comput. Soc pp 196–202
[5] Kim T, Shakhnarovich G, Urtasun R 2010 Sparse Coding for Learning Interpretable Spatio-Temporal Primitives NIPS pp 1–9
[6] Elmezain M, Al-hamadi A 2008 A Hidden Markov Model-Based Isolated and Meaningful Hand Gesture Recognition In Proceeding of World Academy of Science, Technology, Engineering and Technology pp 394–401
[7] Yulita I N, The H L 2012 Fuzzy Hidden Markov Models for Indonesian Speech Classification Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII) 16(3)
[8] Sutton C and McCallum A 2010 An Introduction to Conditional Random Fields Foundations and Trends pp 1–87
[9] Serge Guillaume 2001 Designing fuzzy inference systems from data: an interpretability-oriented review IEEE Transactions on Fuzzy Systems 9(3) pp 426–443
[10] Lafferty J, McCallum A 2001 Conditional Random Fields?: Probabilistic Models for Segmenting and Labeling Sequence Data Conditional Random Fields?: Probabilistic Models for Segmenting and (Icml) pp 282–289
[11] Madeo R C B, Wagner P K, PERES S M 2013 A Review of Temporal Aspects of Hand Gesture Analysis Applied to Discourse Analysis and Natural Conversation International Journal of Computer Science and Information Technology 5 pp 1–20