Journal of Advanced Computational Intelligence and Intelligent Informatics, Special issue on Intelligence Techniques in Computer Games and Simulations, **vol. 12, no. 2, pp. 150-155, 2008**.

Paper: jc12-2-3265: 2008/2/25

Haar Wavelets for Online-Game Player Classification with Dynamic Time Warping

Ruck Thawonmas and Keita Iizuka

Intelligent Computer Entertainment Laboratory, Graduate School of Science and Engineering, Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan
E-mail: ruck@ci.ritsumei.ac.jp

[Received May 14, 2007; accepted September 12, 2007]

Online game players' action sequences, while important to understand their behavior, usually contain noise and/or redundancy, making them unnecessarily long. To acquire briefer sequences representative of players' features, we apply the Haar wavelet transform to action sequences and reconstruct them from selected wavelet coefficients. Results indicate that this approach is effective in classification when the *k*-nearest neighbor classifier is used to classify players based on dynamic time warping distances between reconstructed sequences.

Keywords: online game, player classification, wavelet transform, dynamic time warping, action sequences

1. Introduction

Keeping online game players playing requires finding and fulfilling players' demands and tailoring appropriately contents to each player or group of players. In virtual worlds such as online games, players are typically identified by performance type such as killer, achiever, explorer, and socializer [1]. Such types should be exploited to provide game contents that players favor. Examples of content include more hunting opportunities for killers, a wider variety of collectable items for achievers, longer missions for explorers, and a higher frequency of social events for socializers.

Online game players' action sequences are crucial for classifying player types, but noise or redundancy makes them needlessly long. Rather than directly using action sequences in classifying players, we previously used normalized action frequency vectors (NAFV) [2] and hidden markov models (HMM) [3]. The NAFV requires no parameter settings, costs relatively little, and effectively classifies players whose action frequency distinctly differs. It is less effective, however, when differences are less apparent despite dissimilar action sequences. The HMM [4] classifies sequence data well, but its performance depends on the initial structure and parameters.

We proposed a parameter-free approach [5] using action transition probability and considering action information in frequency and order. However, we found this approach unsuitable for classifying players with similar local behaviors but with different global structures, such as classification of Type-I players performing mission A before mission B and Type-II players vice versa. This is because the action transition probability represents only local changes in the sequence of interest.

In this paper, we apply Haar wavelet transform [6] to action sequences and reconstruct them from selected wavelet coefficients. By this, more compact sequences are obtained representing important player features covering both local and global information. We employ the k-nearest neighbor classifier [7] to classify unknown players and dynamic time warping [8] to calculate distances between the reconstructed sequence of an unknown player and those of known players. We evaluate this approach with action sequences from an online game, "The ICE" we are developing.

Related work in the literature focuses on players' trails or time series of locations visited, such as using trails to determine the distance over time among members of a social group [9] or a visualization tool for visualizing player flows in virtual environments such as virtual museums [10]. Other work focuses on visualizing important parts of trails [11] and on player clustering based on trails [12].

This paper is organized as follows: Section 2 discusses the Haar wavelet transform, Section 3 details dynamic time warping, Sections 4 and 5 propose selecting wavelet coefficients and evaluate performance, and Section 6 summarizes the paper and mentions future work.

2. Haar Wavelet Transform

In the Haar wavelet transform concept, decomposition involves obtaining Haar wavelet coefficients from an action sequence. Reconstruction involves recovering the original sequence from obtained coefficients. Our proposal for selecting coefficients to achieve more compact sequences representing important features is detailed in Section 4.

We assume that length L of a sequence is a power of 2 and $q = \log_2(L)$. The ith Haar wavelet coefficient at

Table 1. Example of Haar wavelet transform.

Resolution	Averages $x_{(k,i)}$	Coefficients $d_{(k,i)}$
$k=4$	(9, 7, 1, 2, 4, 8, 5, 2)	-
$k=3$	(8, 1.5, 6, 3.5)	(1, -0.5, -2, 1.5)
$k=2$	(4.75, 4.75)	(3.25, 1.25)
$k=1$	(4.75)	(0)

resolution order k, $d_{(k,i)}$, is derived by

$$d_{(k,i)} = \frac{x_{(k+1,2i-1)} - x_{(k+1,2i)}}{2} \quad \ldots\ldots\ldots \quad (1)$$

where $x_{(k,i)} = \frac{x_{(k+1,2i-1)}+x_{(k+1,2i)}}{2}$ is the ith average at order k between two corresponding adjacent values in order $k+1$. Note that with this representation, $k_{max} = q$, and the original sequence is represented by $x = x_{(q,1)}, x_{(q,2)}, \ldots, x_{(q,L)}$. An example of Haar wavelet decomposition of the sequence 9, 7, 1, 2, 4, 8, 5, 2 is shown in **Table 1**.

Reconstruction of a given sequence from its Haar wavelet coefficients and averages is done using the following formulas:

$$x_{(k,2i-1)} = x_{(k-1,i)} + d_{(k-1,i)} \quad \ldots\ldots\ldots \quad (2)$$

$$x_{(k,2i)} = x_{(k-1,i)} - d_{(k-1,i)}. \quad \ldots\ldots\ldots \quad (3)$$

3. Dynamic Time Warping for Action Sequences

3.1. Action Coding

Let O denote the set of action symbols of interest and $|O|$ their number. As in [8], action sequence $S = S(1), S(2), \ldots, S(L)$ is numerically coded into $|O| \times L$ time-series matrix $X = [X(1), X(2), \ldots, X(L)]$, where $X(i)$ is a column vector with the element indexing the action symbol of $S(i)$ being 1 and other elements 0.

Consider, for example, the set of action symbols $O = \{A, B, C\}$, and thus $|O| = 3$, where symbols A, B, and C are represented by column vectors $[100]^t$, $[010]^t$, and $[001]^t$. In an action sequence such as $S = A, B, C, C$, it is coded to $X = [[100]^t, [010]^t, [001]^t, [001]^t]$.

3.2. Dynamic Time Warping

Two time series of interest are considered similar if they have the same structures, i.e., rise and fall patterns, although they might have different scales on the time axis. A good measurement for deriving the distance between such series is the dynamic time warping (DTW) distance. The DTW distance between time-series matrices X and Y, $D(X,Y)$, having lengths L_X and L_Y, is defined as follows [8]:

$$D(X,Y) = g(L_X, L_Y) \quad \ldots\ldots\ldots\ldots \quad (4)$$

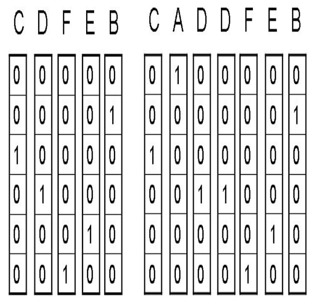

Fig. 1. Time-series matrices X and Y.

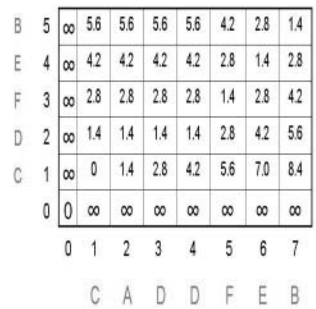

Fig. 2. Derivation of dynamic time warping distance between X and Y.

where

$$g(i,j) = min \begin{cases} g(i,j-1) + d(i,j) \\ g(i-1,j-1) + d(i,j) \\ g(i-1,j) + d(i,j) \end{cases} \quad \ldots \quad (5)$$

$$g(i,0) = \begin{cases} 0 & i=0 \\ \infty & i>0 \end{cases} \quad \ldots\ldots\ldots\ldots \quad (6)$$

$$g(0,j) = \begin{cases} 0 & j=0 \\ \infty & j>0 \end{cases} \quad \ldots\ldots\ldots\ldots \quad (7)$$

and $d(i,j)$ is the Euclidean distance between $X(i)$ and $Y(j)$.

Consider, for example, the set of symbols $O = \{A, B, C, D, E, F\}$ and two action sequences $x = C, D, F, E, B$ and $y = C, A, D, D, F, E, B$. The DTW distance between corresponding time-series matrices X and Y (**Fig. 1**), $D(X,Y)$, is 1.4, derived as shown in **Fig. 2**.

4. Reduction of Action Sequences

Given the set of action symbols O, below, we describe our procedure for reducing the length of an action sequence of interest. As an example, we use action sequence $x = A, B, C, C, A, B, C, C, A, A, A, B, B, B, A, A$, where $O = \{A, B, C\}$ and thus $|O| = 3$.

- Derive corresponding time-series matrix X for action

$$\begin{array}{cccccccccccccccc} A & B & C & C & A & B & C & C & A & A & A & B & B & B & A & A \end{array}$$
$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Fig. 3. Time-series matrix X.

$$\begin{array}{c|c|cc|cccc|cccccccc} & d(k=1) & \multicolumn{2}{c|}{d(k=2)} & \multicolumn{4}{c|}{d(k=3)} & \multicolumn{8}{c}{d(k=4)} \\ 0.4375 & -0.1875 & 0 & 0.125 & 0.25 & 0.25 & 0.25 & -0.5 & 0.5 & 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 \\ 0.3125 & -0.0625 & 0 & -0.125 & 0.25 & 0.25 & -0.25 & 0.5 & -0.5 & 0 & -0.5 & 0 & 0 & -0.5 & 0 & 0 \\ 0.25 & 0.25 & 0 & 0 & -0.5 & -0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}$$

Fig. 4. X after decomposition.

$$\begin{bmatrix} 0.4375 & -0.1875 & 0 & 0.125 & 0 & 0 & 0 & -0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3125 & -0.0625 & 0 & -0.125 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Fig. 5. X after coefficient selection.

$$\begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 & 0.125 & 0.125 & 0.125 & 0.125 & 0.625 & 0.625 & 0.625 & 0.625 & 0 & 0 & 1 & 1 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 1 & 1 & 0 & 0 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Fig. 6. Reconstructed X.

sequence of interest x having length L[1]. In our example, time-series matrix X is shown in **Fig. 3**.

- Decompose each row in X to obtain Haar wavelet coefficients. Resulting coefficients for X are given in **Fig. 4**.
- Reconstruct each row in X with selected Haar wavelet coefficients as follows:

  Following the same recipe in [6], the number of Haar wavelet coefficients used to evaluate performance is heuristically set to $\min(L-1, \lfloor \log_2 L \times 4 \rfloor)$. Reconstruction of each row in X is started from the coefficient at the lowest resolution order, i.e., $d(1,1)$, to those at the next higher order, etc. At a given resolution order, when the number of remaining coefficients that must be selected for reconstruction is less the number of coefficients in that order, we propose selecting remaining coefficients based on their total energy value in decreasing order, where total energy of $d_{(k,i)}$, $E_{(k,i)}$, is defined as

$$E_{(k,i)} = \sum_{n=1}^{|O|} d^2_{(n,k,i)} \quad \ldots\ldots\ldots\ldots \quad (8)$$

  where $d_{(n,k,i)}$ is $d_{(k,i)}$ decomposed at row n of X. Remaining coefficients are then reset to zero. Assume that four coefficients are selected in our example; X resulting after selection of coefficients is shown in **Fig. 5**.

- Reconstruct the action sequence with the above coefficients (**Fig. 6**).
- Reduce the size of X by sampling down a group of repetitive and consecutive elements at each reconstructed row to one element [6]. The reduced X is given in **Fig. 7**.

1. We assume that length L of each action sequence is adjusted so that L is a power of 2.

$$\begin{bmatrix} 0.25 & 0.125 & 0.625 & 0 & 1 \\ 0.25 & 0.25 & 0.25 & 1 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{bmatrix}$$

Fig. 7. Reduced X.

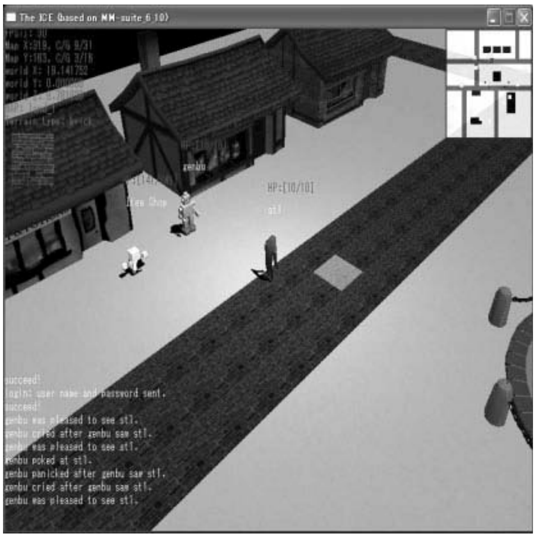

Fig. 8. A screen shot of The ICE.

5. Performance Evaluation

5.1. Players' Logs

We obtained players' logs from the online game The ICE, now being developed at our laboratory, a screen shot of which is shown in **Fig. 8**. The main game objects involved in the study were the mission master (MM), who assigned a mission to a player character (PC); three non-player characters (NPCs), Ceris, Rodth, and Gelec, statically located at different locations and with whom PC must interact -(chat, help, trade)- to complete the assigned mission; the item-shop NPC from which PC bought items; and monster ants, randomly located throughout the map, that PC must exterminate. Actions available in The ICE are summarized in **Table 2**, where for Talk the initial letter of a corresponding NPC is used.

A group of 30 male test-players, on average 20 years of age, participated. These players consisted of third-year and fourth-year computer science undergraduate students with no experience in playing The ICE. They were equally divided into three subgroups G1, G2, and G3. Subgroup members were then asked to complete three mission sessions in the following order:

G1: M1 session → M2 session → M3 session,

G2: M3 session → M1 session → M2 session,

G3: M2 session → M3 session → M1 session.

During a mission session, players were asked to repeat their missions but permitted to quit on their own will. They could move freely and attempt any command, but were not allowed to do any unassigned mission. A brief description of each mission is given as follows:

M1: Item Delivery. PC must deliver an item from MM to a specific NPC and then deliver an item from that NPC to another NPC, etc.

M2: Item Trade. PC must trade with NPCs to increase the amount of money initially provided by MM. PC uses the initial money to buy one of the three items from the item-shop NPC and sells the item at a higher price to one of the three NPCs, who only buy a particular item.

M3: Monster Ant Extermination. PC must help Rodth by exterminating five monster ants.

Table 2. Action list of The ICE.

Action	Symbol
Attack with a snow ball	a
Chat	c
Walk	w
Trade	t
Talk	(*)
Pick up potion	p
Use potion	u
Dead	d
Warp	r

Table 3. Action-sequence length statistics before and after reduction.

	MEAN	VAR
Before	885.8	286504.7
After	30.1	13.9

After checking players' logs, we found, not surprisingly, that most players soon dropped their game missions to play the game on their own. This made it difficult to correctly classify them based on their logs. To classify players, below, we selected four players from each subgroup who played the assigned missions. **Table 3** shows the mean and variance of lengths before and after action sequence are shorten.

5.2. Player Classification

As a classifier, we use the k nearest neighbor (k-nn) classifier. We determined classification performance in three cases, mutually differing in the calculation of the distance between a pair of players, i.e.,

Case 1: The sum of DTW distances between rows of the reduced time-series matrices was used.

Case 2: The sum of DTW distances between rows of original time-series matrices was used.

Table 4. Case 1 classification performance.

$k=1$

	G1	G2	G3
G1	3	1	0
G2	1	3	0
G3	0	0	4

$k=3$

	G1	G2	G3
G1	3	0	1
G2	1	3	0
G3	0	0	4

$k=5$

	G1	G2	G3
G1	2	1	1
G2	1	3	0
G3	1	0	3

Table 5. Case 2 classification performance.

$k=1$

	G1	G2	G3
G1	4	0	0
G2	0	4	0
G3	3	1	0

$k=3$

	G1	G2	G3
G1	4	0	0
G2	0	4	0
G3	2	2	0

$k=5$

	G1	G2	G3
G1	4	0	0
G2	1	3	0
G3	3	1	0

Table 6. Case 3 classification performance.

$k=1$

	G1	G2	G3
G1	2	1	1
G2	1	0	3
G3	0	2	2

$k=3$

	G1	G2	G3
G1	3	0	1
G2	1	0	3
G3	2	2	0

$k=5$

	G1	G2	G3
G1	3	0	1
G2	2	0	2
G3	1	3	0

Case 3: The sum of Euclidean distances between rows of action-transition-probability matrices [5] was used.

Tables 4-6 show the classification performance for each case over three variations of k, in which individual results were obtained by leave-one-out cross-validation [7]. The i, jth element indicates the number of times the k-nn classifier labels an unknown player of subgroup Gi to subgroup Gj. The k-nn classifier performance in case 1 outperforms those of other cases.

Analyzing results, we found that the k-nn in case 2 had difficulty in identifying G3 because partial sequences of multiple attacks typically seen in M3 were also seen in M2. The k-nn in case 1 presented no such difficulty indicating that the proposed distance measure is more reliable. The k-nn in case 3 had the worst performance because the distance measure in [5] could not cope with global structures, i.e., mission sequences. However, it could well classify the players in G1 because they walked a lot during play, making their transition probability from walk to walk relative high and easily classifiable.

6. Conclusions and Future Work

We have described how to calculate the distance between players' action sequences for classifying online-game players. The described distance measure is the dynamic time warping distance between the reduced time-series matrices of a pair of players of interest. We applied the Haar wavelet transform to decompose time-series matrices derived from action sequences and to reconstruct them based on a set of Haar wavelet coefficients selected by our energy-based proposal. We evaluated performance using the online game The ICE, indicating that the k-nn classifier using the aforementioned distance measure outperforms those using conventional distance measures. We plan to apply the distance measure to online-game player clustering and visualization.

References:

[1] R. Bartle, "Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs," The Journal of Virtual Environments, Vol.1, No.1, 1996.

[2] R. Thawonmas, J. Y. Ho, and Y. Matsumoto, "User Type Identification in Virtual Worlds," K. Arai, H. Deguchi, H. Matsui (Eds.), Agent-Based Modeling Meets Gaming Simulation (Post-Proceedings of the Session Conf. of the ISAGA, Int. Simulation and Gaming Association, 2003), Series: Springer Series on Agent Based Social Systems, Vol.2, Springer, pp. 79-88, 2006.

[3] Y. Matsumoto and R. Thawonmas, "MMOG Player Classification Using Hidden Markov Models," Matthias Rauterberg (Ed.), Lecture Notes in Computer Science, Proc. ICEC 2004, Vol.3166, pp. 429-434, 2004.

[4] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proc. IEEE, Vol.77, No.2, pp. 257-285, 1989.

[5] R. Thawonmas and J. Y. Ho, "Classification of Online Game Players Using Action Transition Probabilities and Kullback Leibler Entropies," The Journal of Advanced Computational Intelligence and Intelligent Informatics, Special issue on Advances in Intelligent Data Processing, Vol.11, No.3, pp. 319-326, 2007.

[6] K. Chan, A. Fu, and C. Yu, "Haar Wavelets for Efficient Similarity Search of Time-Series: With and Without Time Warping," IEEE Trans. Knowl. Data Eng., Vol.15, No.3, pp. 686-705, 2003.

[7] S. M. Weiss and C. A. Kulikowski, "Computer Systems That Learn," Morgan Kaufmann Publishers, San Mateo, CA, 1991.

[8] P. Somervuo, "Online Algorithm for the Self-Organizing Map of Symbol Strings," Neural Networks, Vol.17, No.8-9, pp. 1231-1239, 2004.

[9] K. Borner and S. Penumarthy, "Social Diffusion Patterns in Three-Dimensional Virtual Worlds," Information Visualization Journal, Vol.2, No.3, pp. 182-198, 2003.

[10] L. Chittaro, R. Ranon, and L. Ieronutti, "VU-Flow: A Visualization Tool for Analyzing Navigation in Virtual Environments," IEEE Transactions on Visualization and Computer Graphics, Vol.12, No.6, pp. 1475-1485, 2006.

[11] R. Thawonmas, M. Hirano, and M. Kurashige, "Cellular Automata and Hilditch Thinning for Extraction of User Paths in Online Games," CD-ROM, Proc. NETGAMES 2006 & 5th Workshop on Network & System Support for Games, 2006.

[12] R. Thawonmas, M. Kurashige, and K. T. Chen, "Detection of Landmarks for Clustering of Online-Game Players," The Int. Journal of Virtual Reality, Vol.6, No.3, pp. 11-16, 2007.

Name:
Ruck Thawonmas

Affiliation:
Intelligent Computer Entertainment Laboratory, Department of Human and Computer Intelligence, Ritsumeikan University

Address:
1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

Brief Biographical History:
April, 2002 Joined Ritsumeikan University as an Associate Professor
April, 2004 Promoted to a Full Professor

Main Works:
• R. Thawonmas and S. Abe, "A Novel Approach to Feature Selection Based on Analysis of Class Regions," IEEE Trans. Systems, Man, and Cybernetics, Vol.27, Part B, No.2, pp. 196-207, 1997.
• R. Thawonmas, A. Cichocki, and S. Amari, "A Cascade Neural Network for Blind Signal Extraction without Spurious Equilibria," IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E81-A, No.9, pp. 1833-1846, 1998.

Membership in Academic Societies:
• Senior Member, Institute of Electrical and Electronics Engineers (IEEE)
• Association for Computing Machinery (ACM)
• The Institute of Electronics Information and Communication Engineers (IEICE)
• Information Processing Society of Japan (IPSJ)
• International Game Developers Association (IGDA)
• Virtual Reality Society of Japan (VRSJ)
• Game Amusement Society (GAS)
• Digital Games Research Association Japan (DiGRA Japan)

Name:
Keita Iizuka

Affiliation:
Solution Development Team, Solution Development Dept., Bandai Networks Co., Ltd.

Address:
21F, 1-6-1 Higashi-shimbashi, Minato-ku, Tokyo 105-7421, Japan

Brief Biographical History:
April, 2001-March, 2005 Undergraduate student at the Department of Computer Science, Ritsumeikan University (Graduated in March, 2005)
April, 2005-March, 2007 Master's student at the Graduate School of Science and Engineering, Ritsumeikan University (Graduated in March, 2007)

Main Works:
• R. Thawonmas and K. Iizuka, "Haar Wavelets for Online-Game Player Classification with Time Warping," Proc. Int. Symposium on Intelligence Techniques in Computer Games and Simulations, Mar., 2007.