Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Information Systems

School of Information Systems

12-2010

Sequence Alignment Based Analysis of Player Behavior in Massively Multiplayer Online Role-Playing Games (MMORPGs)

Kyong Jin SHIM
Singapore Management University, kjshim@smu.edu.sg

Jaideep SRIVASTAVA
University of Minnesota - Twin Cities

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Databases and Information Systems Commons, and the Numerical Analysis and Scientific Computing Commons

Sequence Alignment Based Analysis of Player Behavior in Massively Multiplayer Online Role-Playing Games (MMORPGs)

Kyong Jin Shim*, Jaideep Srivastava[†]

*Department of Computer Science & Engineering
University of Minnesota, Minneapolis, Minnesota 55455, USA
Email: kjshim@cs.umn.edu
[†]Department of Computer Science & Engineering
University of Minnesota, Minneapolis, Minnesota 55455, USA
Email: srivasta@cs.umn.edu

Abstract—This study proposes a sequence alignment-based behavior analysis framework (SABAF) developed for predicting inactive game players that either leave the game permanently or stop playing the game for a long period of time. Sequence similarity scores and derived statistics form profile databases of inactive players and active players from the past. SABAF uses global and local sequence alignment algorithms and a unique scoring scheme to measure similarity between activity sequences. SABAF is tested on the game player activity data of EverQuest II, a popular massively multiplayer online role-playing game developed by Sony Online Entertainment. SABAF consists of the following key components: 1) sequence alignment-based player profile databases, 2) feature selection schemes and prediction model building, and 3) decision support model for determining inactive players.

Keywords-User behavior, player behavior, sequence alignment, games, inactivity

I. INTRODUCTION

In recent years, many customer relationship management tools have been developed to understand, predict, and prevent customers from leaving permanently. This can be a costly process for the company as it leads to negative financial consequences. Numerous data analysis approaches such as Support Vector Machines as well as a variety of behavior profiling methods have been developed in the past. This study proposes a behavior analysis framework based on activity sequence alignments, namely Sequence Alignment-based Behavior Analysis Framework (SABAF). SABAF is designed to predict inactive game players that either leave the game permanently or stop playing the game for a long period of time. Early detection of potentially inactive players allows companies to pro-actively plan out the strategy to retain the customer basis by providing special customized or personalized offers and services.

SABAF consists of the following key components: 1) sequence alignment-based customer profile databases, 2) feature selection schemes and prediction model building, and 3) decision support model for determining inactive players. This paper discusses methods and experiments that test the effectiveness of SABAF on the game logs from EverQuest II (spanning over eight months), a popular massively multiplayer online role-playing game developed by Sony Online Entertainment. The results show that by choosing appropriate feature selection schemes and classification algorithms and experimentally adjusting the parameters, inactive players can be readily detected. This study provides comparisons between SABAF and the baseline method. This study aims to show that the data analysis methods based on sequence alignment methods can be successfully applied in inactivity prediction of game players. Additionally, we show that the novel feature selection schemes based on activity sequence alignments combined with selection of proper classification algorithms lead to inactivity prediction coverage higher than that achieved by only using aggregated activity information (i.e. total number of monsters killed, total number of quests completed, total instances of mentoring apprentices, and so forth).

II. BACKGROUND

A. Customer Relationship Management and Inactivity Analysis

The Customer Relationship Management domain has seen numerous tools developed to improve customer acquisition and retention and increase sales. Saturated markets and intensive competition have led companies in virtually all industries to pay special attention to retaining existing customer basis as numerous studies have shown that acquisition of new customers can be costly, a process which is many-fold more expensive than retaining the existing customers [3], [6]. One of the key components in the CRM tooling is that of predictive modeling and classification for prediction of inactive customers. Inactivity prediction models often deal with a large amount of customer data where one type of customer information is customer's activity over time. Various statistical and data analysis methods have been developed to enable timely detection of potentially inactive customers followed by strategic and effective customer retention efforts [17], [18], [19].

B. Games and Virtual Worlds

Massively Multiplayer Online Role-Playing Games (MMORPGs) are personal computer or console-based digital games where thousands of players can simultaneously sign on to the same online, persistent virtual world to interact and collaborate with each other through their in-game characters. Recent years have seen an explosive growth in digital game sales including MMORPGs. MMORPG game players spend an average of 21 hours per week playing games, which is close to the national average 28 hours per week spent watching television. MMORPG game players only watch an average of 7.7 hour of television per week [13]. As people spend more time in virtual environments, researchers have recently taken notice that virtual environments such as EverQuest II serve as a major mechanism for socialization [9]. In particular, educational research has found virtual environments to be a sound venue for studying learning, collaboration, social participation, literacy in online space, and learning trajectory at the individual level as well as at the group level.

III. SEQUENCE ALIGNMENT

Sequence alignment is a well-studied method for quantifying and visualizing similarity between sequences. One of the most prominent uses of sequence alignment has been in biological sciences where the technique has been used to compare genetic materials such as DNA, RNA, and protein sequences [5]. One well-known application of sequence alignment is searching against biological databases to find specific genes or motifs [4] as well as studying phylogenetic relationships via multiple sequence alignment [2]. Sequence alignment techniques operate in a global, semi-global, or local context. In DNA and RNA sequences, nucleotides in one sequence are aligned against those of the other sequence. In protein sequences, amino acids of one sequence are aligned against those of the other sequence. Depending on how similar a pair of nucleotides or amino acids are, a score is assigned for each nucleotide-to-nucleotide or amino acid-to-amino acid pair (or gaps if gaps are inserted during alignment). The sum of the individual scores amounts to a sequence similarity measure, indicating how similar the two sequences are.

This study uses this same concept to align sequences of player activities. Given a pair of players, SABAF aligns their activity sequences, assigns scores to each activity-to-activity pair using match, mismatch and gap penalty scores, and then computes the similarity score. By aligning an activity sequence of a player (whose future inactivity is unknown) with activity sequences of known active players and inactive players, SABAF predicts whether the given activity sequence is indicative of inactivity behavior in the future. A number of factors predispose sequence alignment algorithms for use in inactivity prediction, i.e. capability to identify high level patterns embedded within the alignment and a manageable

number of parameters to tweak in order to suit different types of data. Recent studies have applied sequence analysis method to time-series human behavior [14], [15], [16].

IV. EVERQUEST II GAME MECHANICS

Previous studies [9], [10], [11] describe the game mechanics of EverQuest II in details.

A. Game Subscriptions and Inactivity

There is no one unique definition of inactivity. EverQuest II requires a monthly fee to play the game (as of May 2010, the monthly fee is $14.99/month in US currency). A subscribing user creates an account in order to play the game. The subscription fee is per account. Using the account, the player can create one or more in-game characters. The purchase price of the game includes a free play period of 30 days. Optionally, extended services such as an online item database or guild hosting websites can be purchased by subscribers. As part of a free trial, players can download and play the game for free. In this study, we define inactivity as an event in which a subscribing user explicitly requests to discontinue the service or a subscribing user stops playing the game for a long period of time.

V. SEQUENCE ALIGNMENT-BASED BEHAVIOR ANALYSIS FRAMEWORK

A. Dataset

The study uses over eight months worth of player activity data on 'Guk' server (Player-versus-Environment) from January 1, 2006 to September 9, 2006. The dataset contains over 283 million (67% solo plays, 33% team plays) player-to-task records where over 135 million (35% solo plays, 65% team plays) of them are monster kills and quest related tasks. The dataset contains 63,707 distinct players across player levels 1 through 70 (the max level for the players at that time). In a more recent release, Sentinel's Fate, the game maker raised the level cap to 90. The dataset contains at the minimum the following information about game players and their characters: character id, character class, race, task, timestamp of task completion, group size (whether a given character grouped with one or more other characters in completing a task), average group level (if a given character played with one or more other characters, this value represents the average of player levels of all member characters of that group), experience (XP) points, and location (location in which the task was completed). Figure 1 shows the number of players leaving the game permanently at each player level. 70% of such players occur in the first 15 levels, 80% occur in the first 23 levels, and 90% occur in the first 42 levels.
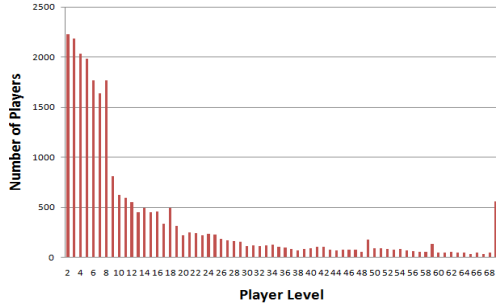
Figure 1.　Number of Inactive Players

```
Data: Player Activity Sequences (P) at levels (L)
Result: Sequence Alignment-Based Profile Databases
begin
    for l ∈ L do
        N ⟵ ∅
        S ⟵ ∅
        for P[i] ∈ P do
            for P[j] ∈ P do
                if i ≠ j then
                    N ⟵ Global(P[i], P[j])
                    S ⟵ Local(P[i], P[j])
                end
            end
            ComputeStatisticsInsertDB(N)
            ComputeStatisticsInsertDB(S)
        end
    end
end
```

Algorithm 1: SABAF Profile Database Construction

B. System Overview

The objective of SABAF is to predict whether a given player at a certain level i will 1) leave the game permanently in the future, for instance, at level $i + 1$ or 2) stop playing the game (and thereby producing no new activities) for more than 30 days. Figure 2 shows the workflow of SABAF. The next three sub-sections discuss in details the three key components which comprise the SABAF system.

C. Part 1 - Profile Database Construction and Sequence Alignment

The first component of SABAF builds sequence alignment profile databases. It builds two types of profile database, one containing global sequence alignments and the other containing local sequence alignments. Subsequently, in each type, it builds two profile databases, one for inactive players and the other for active players. Each profile database is segmented by player level. At each level i, it takes the players who becomes inactive at level $i + 1$ and creates for each player an activity sequence based on their activities at level i. Likewise, at each level i, it takes the players who stay active at level $i + 1$ and creates for each player an activity sequence based on their activities at level i. Hence, it ends up with N activity sequences belonging to the Active group and M activity sequences belong to the Inactive group. Algorithm 1 outlines the steps taken to create two profile databases each segmented by player level.

In Algorithm 1, Global function executes the Needleman-Wunsch global sequence alignment algorithm [8] and Local function executes the Smith-Waterman local sequence alignment algorithm [12]. At the core of all sequence alignment methods is an idea of assigning a score to an alignment. In this study, SABAF assigns +1 to a match, -1 to a mismatch, and -1 to a gap. Using these distance and similarity measure functions, a high score indicates a high level of similarity between two given player activity sequences.

A player statistic consists of the mean, median, and standard deviation of the alignment scores obtained from aligning his activity sequence at a particular level against all other players at the same level. SABAF executes Algo-

```
X: MMDMDDDMMDMQDDQ

Y: TMTMMTMTTQDDMD
```

Figure 3.　Player Activity Sequences

rithm 1 to construct 1) active players vs. active players, 2) active players vs. inactive players (same as inactive players vs. active players), and 3) inactive players vs. inactive players profile databases.

In order to evaluate the SABAF system, a baseline dataset is created. Two baseline profile databases are created; one for active players and one for inactive players. Each profile database is segmented by player level. At each level i, SABAF takes the players who becomes inactive at level $i + 1$ and creates for each player an activity vector of size 25, where each column represents one of the 25 different activities in the game. The element represents the frequency of each activity. Likewise, at each level i, SABAF takes the players who stays active at level $i + 1$ and creates for each player an activity vector. Figure 3 shows two sample player activity sequences. Each letter represents a certain task completed in the game; 'M' denotes a monster kill, 'D' denotes a death event, 'Q' denotes a quest, 'T' denotes a mentoring event. There are well over 20 different activities kept track of in the game logs. From left to right, it shows that player X completed two monster kills and died next to come back and complete another monster kill.

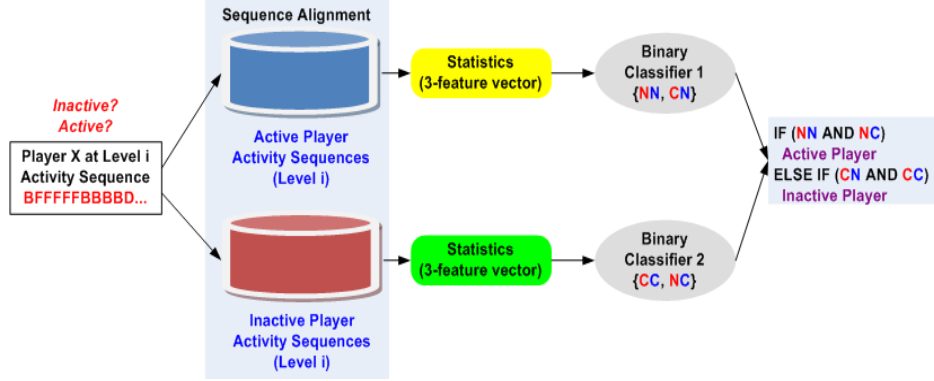Algorithm 2 outlines the steps of creating the two baseline profile databases.

Figure 2. SABAF Workflow

```
Data: Player Activity Sequences (P) for all levels (L)
Result: Baseline Profile Databases
begin
    for l ∈ L do
        for P[i] ∈ P do
            ComputeFrequencyInsertDB(P[i])
        end
    end
end
```

Algorithm 2: Baseline Profile Database Construction

D. Part 2 - Feature Selection and Predictive Model Construction

SABAF uses the open source project Weka [7] to build classifiers. The Experiments and Results sections show comparative performances of different schemes.

1) Baseline Classifiers: In building the baseline classifiers, SABAF uses the 25-feature vector as the main feature set with two labels, one representing active players and the other representing inactive players. A binary classifier is built for each player level. Each binary classifier classifies a given input (25-feature vector of a given player) into either the active players bucket or the inactive players bucket.

2) SABAF Classifiers: With respect to feature representation, SABAF uses three different schemes; one using only the global sequence alignment-based profiles, one using only the local sequence alignment-based profiles, and one using both the global sequence alignment-based profiles and the local sequence alignment-based profiles.

Next, SABAF builds two classifiers. Each SABAF classifier receives as the input a three-feature vector computed by aligning the activity sequence of a given player at level i (whose inactivity at level $i+1$ is unknown) against 1) all known active players' activity sequences at level i and 2) all known inactive players' activity sequences at level i.

The final decision on deciding whether the given input is of an active player or of an inactive player is made based on the outputs from both of these classifiers. The first classifier (Classifier 1) is built using the NN (active players vs. active players) profile database and the CN (inactive players vs. active players) profile database. And the second classifier (Classifier 2) is built using the CC (inactive players vs. inactive players) profile database and the NC (active players vs. inactive players) profile database.

E. Part 3 - Inactivity Prediction

Figure 2 shows the final decision making process of the SABAF system with respect to inactivity prediction. When Classifier 1 outputs NN and Classifier 2 outputs NC, the system outputs "active player" as the final decision. Likewise, when Classifier 1 outputs CN and Classifier 2 outputs CC, the system outputs "inactive player" as the final decision. The correctly identified inactive players are considered True Positives, the correctly identified active players True Negatives, the active players incorrectly identified as inactive players False Positives, and the inactive players incorrectly identified as active players False Negatives.

VI. EXPERIMENTS AND RESULTS

A. Evaluation

Evaluation focuses on the effects of 1) baseline versus SABAF feature selection schemes and 2) different classification algorithms on the True Positive and True Negative percentages. True Positives are those cases where SABAF correctly labels inactive players as inactive players. True Negatives are those cases where SABAF correctly labels active players as active players. False Positives are those cases SABAF labels active players as inactive players. False Negatives are those cases SABAF labels inactive players as active players.

Often times, companies attempt to provide some type of incentive or free goods in an effort to prevent potentially inactive customers from leaving their service completely. In other words, False Positives could mean to the company more staff and personnel time as well as revenue loss due to

Input	Classifier 1 Output	Classifier 2 Output	Decision	Type
Inactive Player	CN	CC	Inactive Player	TP
Inactive Player	NN	NC	Active Player	FN
Active Player	NN	NC	Active Player	TN
Active Player	CN	CC	Inactive Player	FP

Table I

LOGICS FOR DETERMINING TP, TN, FP, AND FN CASES

Algorithm	Best Feature Scheme	TP Percentage (Levels 2 to 69)
JRip	Baseline	0.547
J48	Baseline	0.501
SVM (RBF kernel)	Global & Local (SABAF)	0.903
SVM (Linear kernel)	Global & Local (SABAF)	0.915
Decision Table	Baseline	0.505
AdaBoost	Baseline	0.622
Logistic Regression	Baseline	0.539
Naive Bayes	Baseline	0.500
Neural Network	Baseline	0.594

Table II

OVERALL INACTIVITY PREDICTION COVERAGE (TRUE POSITIVE PERCENTAGE) ACROSS LEVELS 2 THROUGH 69

having to provide incentives and free goods to the customers that really are not potentially inactive customers. False Negatives could mean to the company future revenue loss due to permanent leaving of potentially inactive customers in the future. In either case, it leads to revenue loss for the company and therefore, this study focuses on maximizing both True Positive and True Negative percentages and thereby minimizing both False Positives and False Negatives. This study performs ten-fold cross validation on the game dataset and reports findings below.

Table I shows the logics for determining True Positive, True Negative, False Positive, and False Negative cases in use with SABAF binary classifiers.

Furthermore, as Figure 1 shows, a majority of the inactive players in the game belong to player levels 2 through 20 or so. In addition to reporting the overall coverage across all the player levels, this study reports prediction coverages for those player levels where the majority of the inactive players occur.

B. Comparison of Feature Selection Schemes and Classification Algorithms

This study evaluates the different feature selection schemes (baseline, global & local, global-only, and local-only) and different classification algorithms. First, the overall inactivity prediction coverage (measured as True Positive percentage) across players levels 2 through 69 are reported.

Table II shows that the Support Vector Machine (using Linear kernel) classifier combined with one of the SABAF feature selection schemes (Global and Local sequence alignments) produces the highest overall inactivity prediction coverage (measured as True Positive percentage).
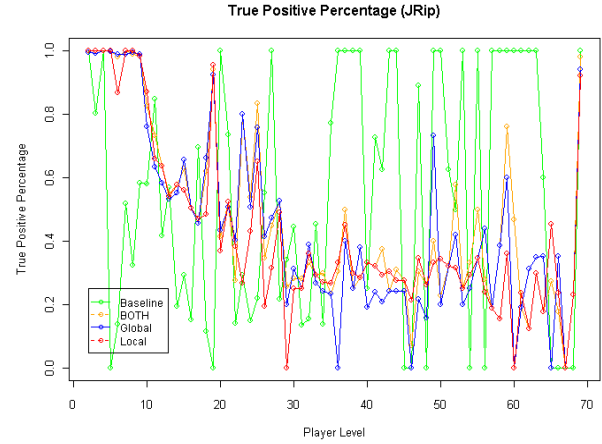

Figure 4. Inactivity Prediction Coverage (True Positive Percentage) - Comparison of Algorithms and Feature Selection Schemes. "BOTH" indicates the feature selection scheme of combining Global and Local sequence alignment information.
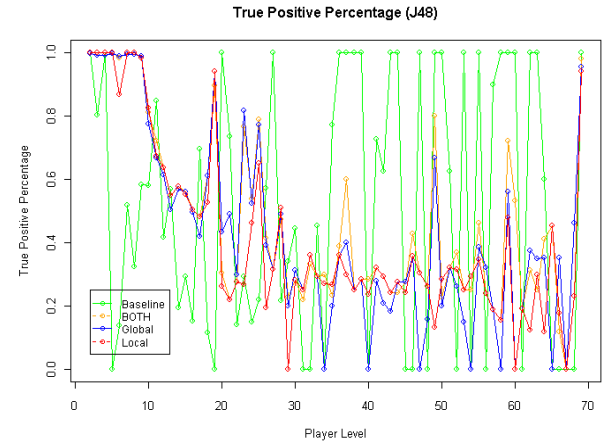

Figure 5. Inactivity Prediction Coverage (True Positive Percentage) - Comparison of Algorithms and Feature Selection Schemes. "BOTH" indicates the feature selection scheme of combining Global and Local sequence alignment information.

The overall coverage of this classifier using the Global and Local sequence alignment information, is 0.915 which is significantly larger than that produced by the same algorithm using the baseline feature selection scheme (0.527). In player levels beyond 20, the baseline scheme performs better than the SABAF feature representation schemes.

In most of the classification algorithms, all of the three SABAF feature selection schemes produce comparative prediction coverage with the exception of SVM (RBF kernel) and SVM (Linear kernel). Additionally, no one feature extraction scheme or no one classification algorithm works best across all the player levels. For instance, the SVM (Linear kernel) classifier produces the highest overall True Positive percentage, however, in player levels 11, 20, 28, 43 and 44, the baseline feature selection scheme produces a
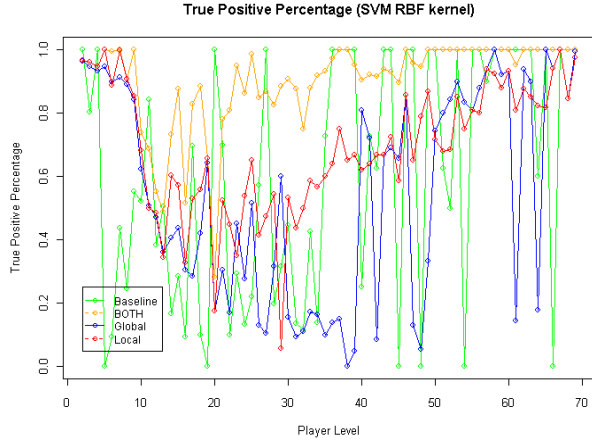
Figure 6. Inactivity Prediction Coverage (True Positive Percentage) - Comparison of Algorithms and Feature Selection Schemes. "BOTH" indicates the feature selection scheme of combining Global and Local sequence alignment information.
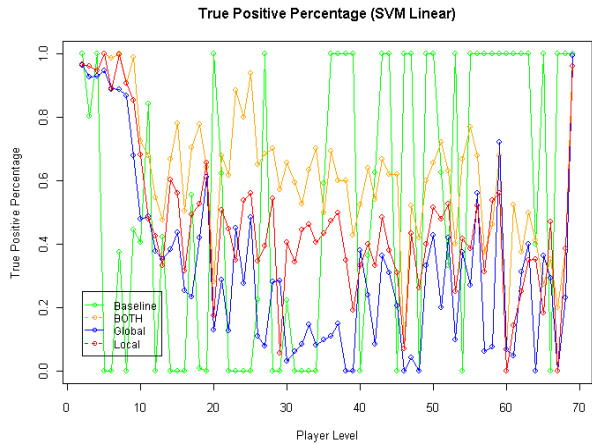

Figure 7. Inactivity Prediction Coverage (True Positive Percentage) - Comparison of Algorithms and Feature Selection Schemes. "BOTH" indicates the feature selection scheme of combining Global and Local sequence alignment information.

considerably higher coverage.

Nearly 80% of the inactive players occur in player levels 2 through 23. Table 3 shows the comparative performances of different classification algorithms and feature selection schemes on this 80% of the inactive player population.

With respect to the first 23 player levels (comprising 80% of inactive player population), in most of the classification algorithms, the SABAF feature selection schemes lead to inactivity prediction coverage higher than that produced by using the baseline feature selection scheme.

Overall, the SABAF system performs better than the baseline in lower levels (23 player levels), covering some 80% of the inactive player population. However, the low True Positive percentage in higher levels mean potentially inactive players missed out. In order to achieve an even

Algorithm	Best Feature Scheme	TP Percentage (Levels 2 to 23)
JRip	Global & Local (SABAF)	0.747
J48	Global (SABAF)	0.712
SVM (RBF kernel)	Global (SABAF)	0.798
SVM (Linear kernel)	Global & Local (SABAF)	0.972
Decision Table	Global & Local (SABAF)	0.730
AdaBoost	Global & Local (SABAF)	0.533
Logistic Regression	Global & Local (SABAF)	0.509
Naive Bayes	Global & Local (SABAF)	0.506
Neural Network	Global & Local (SABAF)	0.600

Table III
INACTIVITY PREDICTION COVERAGE (TRUE POSITIVE PERCENTAGE) ACROSS LEVELS 2 THROUGH 23 (80% OF INACTIVE PLAYER POPULATION)
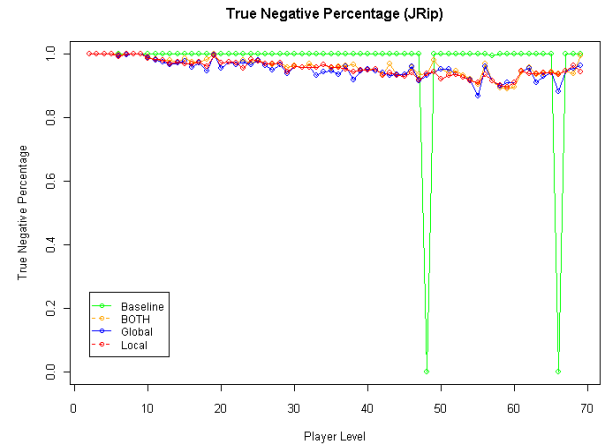

Figure 8. Inactivity Prediction Coverage (True Negative Percentage) - Comparison of Algorithms and Feature Selection Schemes. "BOTH" indicates the feature selection scheme of combining Global and Local sequence alignment information.

higher overall coverage with an improved True Positive percentage in higher levels, it is best to use the baseline methods in higher levels while using the SABAF methods in lower levels.

Next, we evaluate True Negative percentages produced by different feature and algorithm combinations.

While the baseline methods consistently produce near 100% True Negative percentages as the player level goes up, the True Negative percentages produced by the SABAF methods decrease, indicating that the SABAF methods in higher levels generate more False Positives than the baseline method.

VII. CONCLUSION AND FUTURE DIRECTIONS

This paper introduces a sequence alignment-based behavior analysis framework (SABAF) developed for inactivity prediction. Sequence similarity scores and derived statistics form profile databases of known inactive players and active players from the past. The proposed system uses global and local sequence alignment algorithms and a unique scoring scheme to measure similarity between activity sequences
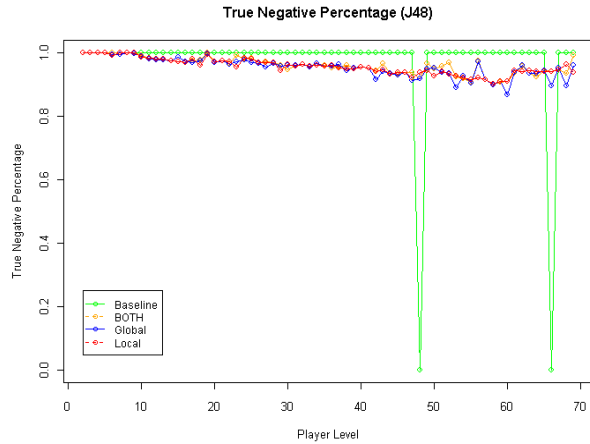
Figure 9. Inactivity Prediction Coverage (True Negative Percentage) - Comparison of Algorithms and Feature Selection Schemes. "BOTH" indicates the feature selection scheme of combining Global and Local sequence alignment information.
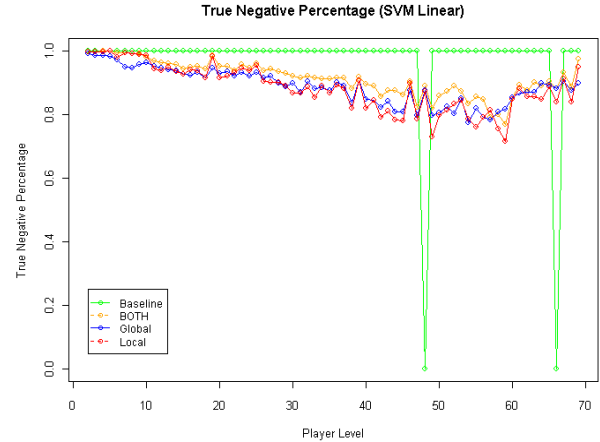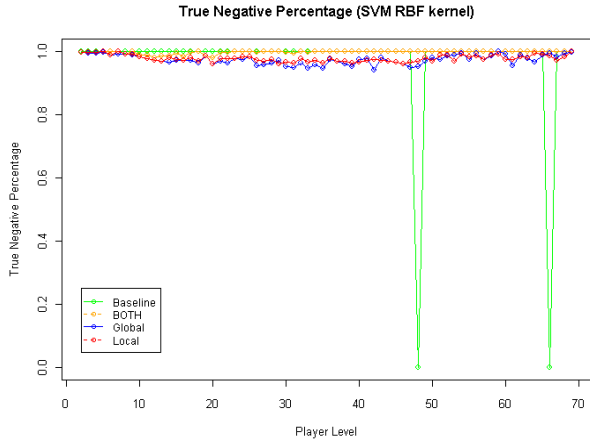

Figure 11. Inactivity Prediction Coverage (True Negative Percentage) - Comparison of Algorithms and Feature Selection Schemes. "BOTH" indicates the feature selection scheme of combining Global and Local sequence alignment information.

Figure 10. Inactivity Prediction Coverage (True Negative Percentage) - Comparison of Algorithms and Feature Selection Schemes. "BOTH" indicates the feature selection scheme of combining Global and Local sequence alignment information.

of game players. SABAF is tested on the game player activity data of EverQuest II. The system consists of the following key components: 1) sequence alignment-based customer profile databases, 2) feature selection schemes and prediction model building, and 3) decision support model for determining inactive players. The results show that by choosing appropriate feature selection schemes and classification algorithms, inactive players can be readily detected. This study provides comparisons between the SABAF system and the baseline method.

Our findings indicate that the Support Vector Machine (using Linear kernel) classifier combined with the Global & Local feature selection scheme produces the highest overall True Positive percentage. In most of the classification algorithms, all of the three SABAF feature selection schemes

produce comparative prediction coverages with the exception of Logistic Regression, SVM (RBF kernel), and SVM (Linear kernel). We report that no one feature extraction scheme or no one classification algorithm works best across all the player levels. For instance, the SVM (Linear kernel) classifier produces the highest overall True Positive percentage, however, in player levels 11, 20, 28, 43 and 44, the baseline feature selection scheme produces a considerably higher coverage. With respect to the first 23 player levels (comprising 80% of inactive player population), in most of the classification algorithms, the SABAF feature selection schemes lead to inactivity prediction coverage higher than that produced by using the baseline feature selection scheme. Overall, the SABAF system performs better than the baseline in lower levels (23 player levels), covering some 80% of the inactive player population. However, the low True Positive percentage in higher levels mean potentially inactive players missed out. In order to achieve an even higher overall coverage with an improved True Positive percentage in higher levels, it is best to use the baseline methods in higher levels while using the SABAF methods in lower levels. In terms of True Negatives, while the baseline methods consistently produce near 100% True Negative percentages as the player level goes up, the True Negative percentages produced by the SABAF methods decrease, indicating that the SABAF methods in higher levels generate more False Positives than the baseline method.

An extension to the current work involves segmenting the players by character class. A previous study [11] reports that the selection of class at character creation limits the character to certain activities as reflected in player-to-task interaction records in the game logs, hence, the activity signatures are different from one class to another. The current implementation of the SABAF system lumps all

classes into one bucket due to the fact that segmentation by class leads to buckets too small for algorithms such as Support Vector Machines to train on. Yet another addition to this study is to leverage a variety of social networks in EverQuest II (i.e. housing trust network, raid group network, and guild network) to further segment the player population based on social interactions over time and perform sequence alignment within each segment.

REFERENCES

[1] Burez, J., Van den Poel, D. (2008) "Handling class imbalance in customer churn prediction", Expert Systems with Applications, In Press, available online 16 May 2008.

[2] Carrillo, H., and Lipman, D. The multiple sequence alignment problem in biology. SIAM Journal of Applied Math 48(5): 1073-1082, 1988.

[3] Coussement, K., Van den Poel, D. (2006) "Churn Prediction in Subscription Services: an Application of Support Vector Machines While Comparing Two Parameter-Selection Techniques", Working Paper 2006/412, Ghent University.

[4] Gelfand, M., Mironov, A., and Pevzner, P. Gene recognition via splices sequence alignment. Proc. Natl. Acad. Sci. USA, 93:9061-9066, 1996.

[5] Goad, W. and Kanehisa, M. Pattern recognition in nucleic acid sequences: a general method for finding local homologies and symmetries. Nucleic Acids Research, 10: 247-263, 1982.

[6] Hadden, J., Tiwari, A., Roy, R., Ruta, D. (2005) "Computer assisted customer churn management: State-of-the-art and future trends", Computers & Operations Research, 34, pp. 2902-2917.

[7] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. (2009), "The WEKA Data Mining Software: An Update", SIGKDD Explorations, Volume 11, Issue 1.

[8] Needleman, S.B. and Wunsch, C.D. (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins". Journal of Molecular Biology 48 (3): 4433.

[9] Shim, K.J., Ahmad, M., Pathak, N., Srivastava, J., "Inferring Player Rating from Performance Data in Massively Multiplayer Online Role-Playing Games (MMORPGs)," cse, vol. 4, pp.1199-1204, 2009 International Conference on Computational Science and Engineering, 2009.

[10] Shim, K.J., Sharan, R., Srivastava, J., "Player Performance Prediction in Massively Multiplayer Online Role-Playing Games (MMORPGs)," 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD 2010. Hyderabad, India, June 21 - June 24, 2010, Proceedings.

[11] Shim, K.J., Srivastava, J., "Behavioral Profiles of Character Types in EverQuest II," IEEE Conference on Computational Intelligence and Games. Copenhagen, Denmark, August 18 - August 21, 2010, Proceedings.

[12] Smith, T.F. and Waterman, M.S. (1981). "Identification of Common Molecular Subsequences". Journal of Molecular Biology 147: 195197.

[13] Yee, N. MMORPG Hours vs. TV Hours. http://www.nickyee.com/daedalus/archives/000891.php.

[14] Prinzie, A., Van den Poel, D., "Incorporating sequential information into traditional classification models by using an element/position-sensitive SAM," DECISION SUPPORT SYSTEMS, Volume 42, Issue 2, pages 508-526, Nov 2006.

[15] Hamuro, Y., Kawata, H., Katoh, N. and Yada, K., A Machine Learning Algorithm for Analyzing String Patterns Helps to Discover Simple and Interpretable Business Rules from Purchase History, Progress in Discovery Science, LNAI 2281, Springer, pp.565-575, 2002.

[16] Hamuro, Y., Katoh, N., Ip, E. H., Cheung, S. L., Yada, K., Combining Information Fusion with String Pattern Analysis: A New Method for Predicting Future Purchase Behavior, V. Torra(ed.), Information Fusion in Data Mining, Studies in Fuzziness and Soft Computing, Vol.123, Springer, pp.161-187, 2003.

[17] Datta, P., Masand, B., Mani, D.R. and Li, B., Automated cellular modeling and prediction on a large scale, Artif. Intell. Rev., vol. 14, no. 6, pp. 485502, 2000.

[18] Hadden, J., Tiwari, A., Roy, R., and Ruta, D., Churn prediction using complaints data, in Proceedings Of World Academy Of Science, Engineering and Technology.

[19] Morik, K. and Kopcke, H., Analysing customer churn in insurance data: a case study, in PKDD 04: Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, (New York, NY, USA), pp. 325336, Springer-Verlag New York, Inc., 2004.