Postdoctoral Research Proposal: High Performance Data Mining for Increased Player Personalization and Integration of E-Commerce in Massively Multiplayer Online Games

June 16, 2003

Abstract

Massively Multiplayer Online Games (MMOGs) server-systems are currently becoming some of the *largest parallel computer clusters on earth* with thousands of processing units serving hundreds of thousands paying human players. *Norwegian game companies*, e.g. Funcom AS, are currently among the pioneers of MMOG services. In order to cater an increasingly large and heterogeneus player community (gender-, age- and interest-wise), there is a need to develop and adapt automated techniques based on data mining and recommender systems in order to provide increased personalization. *These methods are likely to transferable to e-learning based on interactive entertainment*, and due to their scalability requirements (handling TeraBytes of data), they can *possibly also be applied in scientific data mining, e.g. the tremendous amount of space data gathered by NASA or experimental data at CERN.*

1 Context

Massively Multiplayer Online Games (MMOGs) server-systems are currently becoming some of the **largest parallel computer clusters on earth** with thousands of processing units serving a very large number of simultaneous human players, typically in the five- to seven-figure realm. One example is the Korean company NCSoft that provides the game Lineage with about 1.2 million paying players. The characteristic property of MMOGs is that players take part in a persistent virtual world where they communicate, cooperate, fight, build virtual characters, express feelings and emotions, and undertake game-related missions [5, 4].

1.1 The Economy of the Electronic Game Market

The global electronic game market size is currently approximately USD 31.7 billion, which *is larger than the entire movie industry.* This includes the emerging MMOG market of approximately USD 300-400 million (some estimate as high as USD 600 million [13]). With an expected annual revenue growth of 70% the MMOG market size is anticipated to reach USD 2.7 billion (per year) in year 2006 [24].

The business models of MMOGs vary, but currently most are billed subscription services, where players not necessarily pay for the MMOG client, but instead pay a monthly fee (typically USD 10-20) for being connected to the MMOG world. For mobile games

(a) Androids mingling

(b) Profession selection

Figure 1: Screenshots from Funcom's Anarchy Online

the granularity of payments are usually smaller, typically per minute or even per message (SMS) [10].

1.2 The Norwegian Game Industry

Norway has a substantial and fast growing electronic game industry, which throughout the 1990s focused on developing games for PCs and game consoles (e.g. Nintendo 64, Sega Dreamcas and Sony Playstation), in recent years the focus has shifted into several specialized directions: 1) **Massively Multiplayer Online Games** - e.g. Funcom AS, Fifth Season AS [1] and RazorWax[2] are all among the MMOG pioneers, 2) **Games for Mobile Phones and Devices** - e.g. Eyeone AS, Plutonium Software and Triggerduck, 3) **Custom-made games** for TV and advertisement - e.g. Innerloop, and 4) **Traditional games** for PCs and game consoles - e.g. Art Plant AS and Caprino Video Games

Possibly future outlook are mobile phone clients for MMOGs, and custom-made mobile games for TV and advertisement purposes. Due to problems with software piracy and player expectations about communicating with other players, the traditional games which can't connect to the mobile or traditional Internet is likely to become a small niche in the near future.

2 Emerging Research Problems for MMOGs

The MMOG is rapidly maturing in terms of facing the issues of *robustness*, *scalability* and *security*, a sign of this is that there are now several companies that provide MMOG hosting services, including engines (components) that makes it easier for game companies to create, launch and provide a MMOG service. With the basic MMOG infrastructures in place, the research focus is gradually moving towards meeting intra-game challenges such as:

[1] They sold the Planetarion MMOG to Simulation Technologies Ltd. in December 2002

[2] They got Greek investors and moved the entire company to Athens, Greece

2.1 How to meet individual expectations with an increasingly heterogeneus player community?

A *common stereotype* of the game player as an antisocial adolescent male is rapidly becoming outdated, e.g. for MMOGs the majority of players is over 20 years old, and for Electronic Arts' SIMS Online more than 60% of the more than 100 thousand players are female. Another change is that so-called casual players is growing rapidly, and they have other needs and expectations than "hardcore" players who typically spend 20+ hours per week on the game.

The challenges of MMOGs is to cater the needs of all of the player types in a best possible way, but due to their scale (hundreds of thousands of players) this needs to be (partially) automated possibly using personalization techniques and algorithms. Such techniques exists (typically recommender systems using collaborative filtering and data mining) in e-commerce settings, e.g. Amazon.com's automatic book recommendation [9]. However, *such techniques are not suited for the open, much less rigid and very large virtual world environments provided by MMOGs*. Two essential basic questions are:

1. **How to detect that a player is bored?**
2. **How to determine which stimuli a (bored) player needs in order to be satisfied with the MMOG?**

Automated detection of *boredom* requires some kind of data mining, but as there are no standards or standard ways of logging player actions in MMOGs, so this has to be researched. The concrete research problems to enable :

1. **Which player data should be selected to be logged and how should they be preprocessed?** (feature extraction problem [14], *related research areas include Multimedia Data Mining and Web Usage Mining*). Examples of data for logging are: time, intra-game location, action, surrounding descriptions (graphics and sound), interacting players and non-personal characters [17]
2. **How to store and represent the selected data?** In a MMOG world with hundreds of thousands simultaneous players, the data to be mined grow extremely fast, quickly generating *TeraBytes* of data. *related research areas include storage and data mining of scientific data*, e.g. data gathered from space by NASA and experimental data at CERN.
3. **How to data mine those data?** Due to the continously growing data volumes, there is likely to be no time for batch-based data mining, so incremental data mining approaches must be used in order to provide updated patterns. Distributed and parallel data mining must be used to handle the large volumes of data logged [6].
4. **How to easily test approaches such as logging and storage mechanisms, and data mining algorithms in a non-production MMOG-like environment?** Introducing new and untested data mining and personalization methods in a production environment is *not an option*, it has to be tested in less vulnerable environment, e.g. in a simulation of a MMOG.

A related area is Artificial in Computer Games [8, 1], where the focus is on creating increasingly intelligent non-personal characters (i.e. agents), in particular warrior-style, e.g. monsters in Quake [7].

2.2 How to increase the revenue of the MMOG service by introducing intra-game e-commerce?

Cummins proposed online games as a potential environment for e-commerce services [2], this makes sense both from a MMOG provider and a player perspective.

For the MMOG provider, introducing intra-game e-commerce gives an opportunity for additional revenue-streams of significant size, e.g. by letting online stores have a virtual presence in the game and collect fees per sale or per lease-period. E.g. if Amazon.com was present with one or several virtual stores in NCSoft's Lineage or Funcom's Anarchy Online, it could potentially generate new revenue both for Amazon.com and the MMOG providers. A related opportunity for increased MMOG revenue is intra-game *product placement*, this is increasingly common for single-player games, *but little research has been done on creating methods to measure and test effects of product placement in MMOGs.*

MMOGs provides a much richer, flexible and in most eyes nicer graphical 3D environment than the HTML-based pages usually provided by e-commerce stores. So there is an opportunity that players, if it was possible to chose, would select a virtual store in their favorite MMOG over the web-based store counterpart. *By extrapolating on this thought, the future web might be likely to become more similar to todays MMOGs.*

Both intra-game e-commerce and product placement requires research related to data mining for finding patterns and optimizing the benefits of them for all 3 parties: players, MMOG provider and external e-commerce service or product vendor.

3 Methods

In order to perform research on data mining for increased player personalization (section 2.1) and intra-game e-commerce (section 2.2) in MMOGs we focus on methods we have been been working with over time.

1. **Feature selection and data logging** in MMOGs. Our related previous work include [17, 19, 3, 23].
2. **Application and Improvements of High Performance Data Mining and Recommender Algorithms** in MMOGs. Our related previous work include [22, 21, 20, 16].
3. **Agent-based user modelling and simulation of players** in MMOGs. Our related previous work include [23, 19, 3, 12, 11, 18, 15]
4. **Empirical and statistical methods, in particular factorial design**. Related previous work include [19].
5. **Analysis and Data Mining of real player data from the game industry**

Common for all of the above approaches is a focus on continued implementation of related software tools **Zereal** (MMOG simulator) and **Incridge** (parallel data mining tool).

4 Goals

The research goals are as follows (based on problems presented in the sections 2.1 and 2.2)

1. **Develop a flexible and useful logging format for behavioral data mining in MMOGs**, preferably based on the Predictive Model Markup Language (PMML) defined by the Data Mining group
2. **Find useful data mining approaches for detecting patterns of bored players in MMOGs**
3. **Find useful data mining approaches for detecting patterns that measures and explains the effect of introducing intra-game e-commerce, e.g. a bookstore or internet banking**

The goals is obtain a minimum of 2 journal or book chapter publications and 6 refereed conference and workshop publications.

Another goal is to arrange both a national (Norwegian, for the national game industry) and an international workshop (e.g. at ACM SIGKDD) on the topic of data mining and personalization in MMOGs together with other researchers in the area, e.g. dr. Ruck Thawonmas at Ritsumeikan University in Japan.

5 Schedule

The proposed research plan is planned to start March 1st, 2004. (This is continued work based on the post.doc candidate's PhD research):
Apr.2004 - Workshop with Funcom, May.2004 - 1st conference submission, Aug.2004 - 1st journal submission, Oct.2004 - 2nd conference submission, Dec.2004 - 3rd conference submission, Mar.2004 - 2nd journal submission, May.2004 - 4th conference submission, Jul.2004 - 5th conference submission, Nov.2004 - 6th conference submission, Feb.2006 - Finish Post.doc evaluation report

6 Research Partners

Collaborating international research parters during the post.doc period include: 1) Associate Professor Dr. Ruck Thawonmas, Intelligent Computer Entertainment Lab., Ritsumeikan University, Japan, 2) Dr. Steven Willmott, LSI, Universitat Politecnica de Catalunya, Barcelona, Spain, 3) Dr. Staffan Bjork, Studio Manager, PLAY, Interactive Institute and Guest Lecturer, Dep. of Computing Science, Chalmers Technical University, Gothenburg, Sweden. In addition the Digital Games Research Association (`http://www.digra.org`), where the post.doc candidate is a board member, will be a part of the research network.

Industry partner is Product Manager Jørgen Tharaldsen and V.P. of sales & marketing Nicolai Mikkelsen at Funcom AS in Oslo, Norway.

Local research partners at NTNU include prof. Svein E. Bratsberg (Autonomic Computing and Scalable Databases), assoc. prof. Dr. Anne C. Elster (High Performance Parallel Computing and Algorithms) and Magnus L. Hetland (High-Performance Data Mining of time-series) and MSc project students.

References

[1] Robert St. Amant and R. Michael Young. Artificial Intelligence and Interactive Entertainment. *Intelligence Magazine*, 12(2):17–19, 2001.

[2] Nizami Cummins. Integrating E-Commerce and Games. *Journal of Personal and Ubiquitous Computing*, 6(5-6):362–370, December 2002.

[3] Håvard Engum, Jørgen V. Iversen, and Øyvind Rein. Zereal: A Semi-realistic Simulator of Massively Multiplayer Games. Technical report, IDI, NTNU, December 2002.

[4] Robert Gehorsam. The Coming Revolution in Massively Multiuser Persistent Worlds. *IEEE Computer*, 36(4):93–95, April 2003.

[5] Moon Ihlwan. The champs in online games. *Business Week*, (30):27–28, July 2001.

[6] Hillol Kargupta and Philip Chan, editors. *Advances in Distributed and Parallel Knowledge Discovery*. AAAI/MIT Press, 2000.

[7] John Laird. It Knows What You're going To Do: Adding Anticipation to a Quakebot. In *Proceedings of the 5th International Conference on Autonomous Agents*. ACM, ACM Press, 2001.

[8] John Laird and Michael Van Kent. Human-Level AI's Killer Application: Interactive Computer Games. *AI Magazine*, 22(2):15–26, 2001.

[9] Greg Linden, Brent Smith, and Jeremy York. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing*, 7(1):76–80, January/February 2003.

[10] Ian Maccines, Janusz Moneta, Julio Caraballo, and Dominic Sarni. Business Models for Mobile Content: The Case of M-Games. *The International Journal of Electronic Commerce & Business Media: Special Section on Electronic Commerce in Entertainment and Media*, 12(4):218–227, 2002.

[11] Mihhail Matskin. Collaborative Advertising over Internet with Agents. In *Proceedings of the 12th International Conference on Database and Expert Systems, 1st Workshop on Web Based Collaboration (WBC'01)*, pages 509–513. IEEE Computer Society, September 2001.

[12] Mihhail Matskin and Amund Tveit. Mobile Commerce Agents in WAP-based Services. *Journal of Database Management - Special Issue on Mobile Commerce*, pages 27–35, July-September 2001.

[13] PCGameWorld.com. Online Gaming Booms. Published online, `http://www.pcgameworld.com/story.php/id/234/`, June 2002.

[14] Dorian Pyle. *Data Preparation for Data Mining*. Morgan Kaufmann Publishers and Academic Press, 1st edition, 1999.

[15] Amund Tveit. A Survey of Agent-Oriented Software Engineering. Published online, `http://www.jfipa.org/publications/AgentOrientedSoftwareEngineering/`, May 2001.

[16] Amund Tveit. Peer-to-peer based Recommendations for Mobile Commerce. In *Proceedings of the 1st International Workshop on Mobile Commerce*, pages 26–29. ACM Press, July 2001.

[17] Amund Tveit. Game Usage Mining: Information Gathering for Knowledge Discovery in Massively Multiplayer Games. In Hamid R. Arabnia and Youngson Mun, editors, *Proceedings of the International Conference on Internet Computing (IC'2002), session on Web Mining*, volume III, pages 636–642. CSREA Press, June 2002.

[18] Amund Tveit. jfipa: An Architecture for Agent-based Grid Computing. In *Proceedings of AISB'02 Convention, Symposium on AI and Grid Computing*, April 2002.

[19] Amund Tveit. Scalability Analysis of the Zereal Massively Multiplayer Game Simulator. Technical Report ISSN: 0802-6394, 12/02, IDI, NTNU, December 2002.

[20] Amund Tveit and Håvard Engum. Parallelization of the Incremental Proximal Support Vector Machine Classifier using a Heap-based Tree Topology. Submitted to ECML/PKDD Workshop on Parallel and Distributed Computing for Machine Learning, May 2003.

[21] Amund Tveit and Magnus Lie Hetland. Multicategory Incremental Proximal Support Vector Classifiers. In *Proceedings of the 7th International Conference on Knowledge-Based Information & Engineering Systems (KES'03, forthcoming)*, Lecture Notes in Artificial Intelligence. Springer-Verlag, September 2003.

[22] Amund Tveit, Magnus Lie Hetland, and Håvard Engum. Incremental and Decremental Proximal Support Vector Classification using Decay Coefficients. In *Proceedings of the 5th International Conference on Data Warehousing and Knowledge Discovery (DaWaK'03, forthcoming)*, Lecture Notes in Artificial Intelligence. Springer-Verlag, September 2003.

[23] Amund Tveit, Øyvind Rein, Jørgen V. Iversen, and Mihhail Matskin. Zereal: A Mobile Agent-based Simulator of Massively Multiplayer Online Games. Submitted to Intl. Conf. on Intelligent Virtual Agents, May 2003.

[24] Zona Inc. and Executive Summary Consulting Inc. State of Massive Multiplayer Online Games 2002: A New World in Electronic Gaming. Technical report, Zona Inc., Redwood City, California, USA, October 2002.