

# Towards Role Detection in Virtual Worlds

CARSTEN EICKHOFF

Delft University of Technology

c.eickhoff@tudelft.nl

and

VICTOR P. LAVRENKO

The University of Edinburgh

vlavrenk@inf.ed.ac.uk

---

Virtual worlds are a topic of steadily growing relevance. Some of the providers report user numbers that exceed the population of entire nations in the real world. Virtual worlds typically provide a high degree of complexity, which in some areas approaches the real world's richness of detail. Without "living" in any given virtual world it is hard to get insights about that world and its inhabitants. Knowledge about users' roles within a virtual world can be of socio-economical and scientific interest. This work describes an automatic means of inferring such roles based on textual communication. We give an introduction into virtual worlds and formalize the task of virtual world role detection as well as evaluating its performance against a manually annotated large-scale corpus. We present and discuss various approaches towards virtual world role detection, showing that performance close to that of human judges can be achieved. We close by demonstrating a successful application of role detection in a free to play MMORPG.

Categories and Subject Descriptors: H.1.2 [Models and Principles]: User/Machine Systems—*Human Factors*; H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing—*Linguistic Processing*; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*Retrieval Models*

General Terms: Experimentation, Human factors

Additional Key Words and Phrases: Accessibility, Online games, MMORPG, User modelling, Virtual world

---

## 1. INTRODUCTION

Virtual worlds (VWs) are an invention of the last 20 years. They had their beginnings in the early 1990s and have since then been attracting a steadily growing number of users. Virtual worlds are persistent environments within which users can interact with each other. Most virtual worlds have a very high degree of complexity which in certain aspects even approximates the real world's richness of detail. Made

---

Authors' addresses:

C. Eickhoff, Delft University of Technology, Mekelweg 4, 2628 CD Delft, Netherlands

V.P. Lavrenko, The University of Edinburgh, 10 Crichton Street, Edinburgh, UK

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.

© 20YY ACM 0000-0000/20YY/0000-100001 \$5.00

possible by the spread of affordable broadband Internet connections, the number of virtual world users has been steadily rising in recent years. MMOData, an on-line project on virtual world statistics [MMO 2009], collected information about the total number of active virtual world subscriptions. Even with occasional declines, this currently means active communities of a size in the region of 17 million people. This equals the populations of Chile or the Netherlands. Such vast user numbers have been attracting significant interest in very different areas of the research community. Rizzo et al. [Rizzo et al. 2006] used virtual worlds for the treatment of Post Traumatic Stress Disorder caused by extreme experiences such as wars, hostage situations or terrorist attacks. The authors' explicit application was the reintegration of Iraq war veterans into normal civilian life.

Another research idea which was introduced very early and since then has been pursued with great effort is the educational use of virtual worlds. Although most serious researchers agree that we are far from declaring real world schools and universities obsolete, projects such as NICE [Johnson et al. 1998] show the idea's practical applicability.

A very interesting application of virtual worlds was discovered merely by chance through a bug in the popular on-line role playing game World of Warcraft. On September 13th, 2005, the game was expanded by an encounter in which one of the players' opponents could cause an infectious disease that would harm and eventually kill the avatars. This plague, known as "Corrupted Blood", was supposed to be limited to one particular zone of the virtual world. Players, however, managed to exploit a glitch in the game to extract the disease from that zone and to spread it in the virtual world's capitals. Thousands of avatars died and playing became almost impossible because of malicious disease spreaders who would try to infect as many people as possible before eventually dying themselves. The whole incident could only be ended by Blizzard, the game's provider, resetting the entire virtual world. Due to the strong parallels to real world diseases as for example AIDS, epidemiologists and behavioural scientists became aware of the situation and expressed their strong interest in using virtual worlds as a means of simulating disease spread and patient behaviour [Balicer 2007].

Leuski and Lavrenko in 2006 published their results on event detection in virtual worlds [Leuski and Lavrenko 2006]. They followed the intuition that in any environment, regardless if it is the real world or a virtual one, there is a dependency between its inhabitants' communication and their actions. Using a language modelling approach, they predicted events in an on-line role playing game based on the players' chat messages.

The desire for role detection originates from the high degree of complexity commonly observed in virtual worlds. Many of the social and economical structures found in these worlds are results of real people's interactions. As a consequence, they become as complex and hard to predict as their real world equivalents. We as inhabitants of the real world have, through education and studies, assembled a great amount of knowledge about the role distribution in our environment. We know who is president and who is a famous actor. Especially in today's globalised world, this set of important individuals of public interest becomes more and more universal for people from very different cultural and ethnic backgrounds. This knowledge,

however, has to be acquired over a long period of being an inhabitant of the world. For young children or members of a tribal society who spend their lives in remote areas it might not at all be obvious, or even important, who is the current president of the United States of America.

We are talking about the exact same situation when observing virtual worlds from the outside. At a glance, we can not easily decide who is a prominent member of society and who is not. Without spending time in the virtual world, talking to people and having our own experiences, we will not be able to figure out the role distribution in this alien society. Role detection tries to analyse an environment's communication, thus, figuring out who takes which role(s) in it without forcing human observers to actually enter the world themselves.

### 1.1 Practical Application

Apart from the theoretical fascination of coming closer to understanding complex environments in an automated way, there are several practical applications for role detection. We will briefly discuss three very different approaches of using and benefiting from virtual world role detection.

#### **Virtual world Yellow Pages**

Virtual worlds with their high degree of complexity, their changeability and expandability, represent environments which demand a great amount of time spent within them in order to fully understand all the underlying concepts, before allowing the user to benefit from all the world's possible features. Especially users who are new to the virtual world or who went on hiatus are easily overwhelmed and intimidated by the richness of features they offer. A role detection approach to ease the initial barrier would follow the idea of the real world's Yellow Pages. Such a directory of the role distribution within the virtual world would help new inhabitants finding their way around by easily being able to find out about prominent members of society who are known to pursue certain professions or activities which the new user is interested in. Due to the strong parallels between this scenario and several fields of Information Retrieval such as expert finding or recommendation we will discuss it in more detail in Section 7 where we describe an example setting of automatically generated yellow pages for virtual worlds.

#### **Bot detection**

Most virtual worlds require their users to harvest resources, collect goods or simply earn virtual currency. Since these tasks are often perceived as time-consuming or tedious, players have been trying to employ automated agents, so-called bots, for these jobs. The bot logs into the virtual world and takes control of the avatar to perform a given task. However, most VW providers ban the usage of bots. They see the virtual world's economy, the non-botting players' gaming experience and, in the long run, their profit threatened. Over the past years, there has been an arms race between bot developers and virtual world providers for better and better stealth and bot detection techniques, which resulted in a succession of lawsuits with significant values in litigation.

The currently strongest element of camouflage for the bots is the huge number of users who are logged into the virtual world at any time. With several thousand users

logged in even at night time, virtual world providers cannot realistically afford to have their employees inspect a reasonable percentage of users for bot usage. Role detection might bring a change by modelling “regular” user behaviour and then spotting cases which differ dramatically from this standard. The number of such reported suspected bot users could be made small enough to allow actual human employees to only check on those few which the automatic classification was very sure about. This might gain virtual world providers significant savings in their operational costs.

An important fact to note about this particular application is the chat-based nature of our approach. We purely consider communication-related features in this work. Some years ago this might not have been possible as back then the clear opinion was that MMOG bots do not chat. Nowadays, however, this notion becomes more and more blurred by spamming bots and conversational agents employed to further hide the fact that the avatar is automated. With respect to these developments, a bot detection scheme that, among other aspects, analyses player communication would be timely and desirable.

### Enhancing event detection

After having introduced two possible applications which are directed at providers and users of virtual worlds, this third one aims to enhance virtual world research. Earlier, we described Leuski and Lavrenko’s method [Leuski and Lavrenko 2006] of using language modelling to predict virtual world events. Their approach treats all chat messages equally in order to come to a judgement. An effective role detection system would probably be able to further increase event detection performance by determining individuals who fulfil certain roles of importance and then giving more weight to these individuals’ messages. This would for example enable event detection to treat a blacksmith’s utterances differently from those of a seasoned warlord. Depending on the context, one or the other might be the more valuable source of information.

## 1.2 MMORPGs

The recently popular and highly frequented virtual worlds can be grouped into several categories, the most important of which we will introduce briefly:

**Games.** By far the most popular commercial virtual world application are on-line games. Players either band together to fight strong monsters or opposing groups of players.

**Social VW.** The main feature of this kind of virtual world is social interaction. Probably the most famous specimen of this category is Second Life [Sec 2010], a VW that is very closely aligned to the real world. While virtual worlds belonging to this category will seldom project their users into outer space or the middle ages, they rather offer a communication platform for people with similar interests.

**(Combat) Simulation.** Virtual worlds with emphasis on the simulation aspect let users take the role of air plane pilots, air control tower staff, army generals or medieval lords. Connected through the Internet the users emulate a real world topic to a very high degree of detail. Often even the protocol of communication is similar or identical to the one used by their real world examples.

As Table I shows, the virtual world market is clearly dominated by games. Within

Table I. Virtual world market shares by VW type

VW Type	Subscription market share
(Combat) Simulation	0.2%
Social VW	1.9%
Games	97.9%

the class of on-line games there are several different genres. With a share of 94%, Massively Multiplayer Online Roleplaying Games (MMORPGs) are more and more seen as a synonym for the whole on-line gaming sector. This project will narrow its focus further down on MMORPGs, to be able to treat their intricacies appropriately. Specializing on MMORPGs will offer us a more homogeneous field of study while their great market share means that our research results will still remain representative. The main idea of MMORPGs is to allow the user to create an avatar with a basic set of features. They often also offer a range of different virtual races and professions to choose from. The player typically enters the game with his avatar having low attribute values (e.g., strength or intelligence) and with hardly any equipment. At this stage, all but the weakest opponents will impose great dangers to him. Therefore, one of the central ideas of MMORPGs is to advance one's avatar by defeating monsters, thus, gaining experience points and better equipment. This kind of game is called player versus environment (PVE) since the player's avatar only fights system-controlled opponents. The alternative principle is player versus player combat (PVP) where rivalling players fight against each other. The final standard feature all MMORPGs share is an in-game chat system which allows players to communicate in the virtual world. We will base our role detection on the messages sent through this chat system.

The contributions of this work are fourfold: (1) We formalize the novel task of virtual world role detection and identify the essential steps to achieve it. (2) We evaluate a wide range of state of the art means of information retrieval in terms of their applicability for this setting. (3) A role detection approach is introduced that approximates human performance. (4) We demonstrate the applicability of our method by providing an automatically generated directory of proficient players for an MMORPG. The remainder of this article is structured as follows. In Section 2, we formalize the task of role detection in Virtual Worlds. Section 3 gives an overview of related work followed by a definition of the evaluation plan along which detection performance is measured in Section 4. In Section 5, we introduce our role detection approaches which are evaluated in Section 6. In Section 7, we describe a real world application of our role detection method as well as the user community's feedback on it. Finally, Section 8 discusses the results and gives future perspectives on role detection.

## 2. STATE OF THE ART & RELATED WORK

Currently, **virtual worlds** (VWs), regardless whether applied for games or other purposes, do not conduct automatic analyses of user behaviour. There are several examples of qualitative academic studies of user behaviour in virtual worlds (e.g., [Ducheneaut et al. 2006; Williams 2006; Chen and Duh 2007]) but none of them attempted automatic categorization of users according to their activities. Although

more and more aspects of social networks are being integrated, hardly any of the well-studied data mining and information retrieval techniques have been applied by platform providers. The closest match to the applications and methods discussed in this work is the grouping aid functionality introduced by several Massively Multi-player Online Games (MMOGs) such as for example World of Warcraft. Such tools allow players to declare their desired operative role (e.g., healer) and find prospective group mates with complementary roles. However, there is no automatic role detection and the roles to choose from are functional rather than behavioural. The activity-based roles employed in this work are defined in a much more fine-grained way and are expected to hold greater benefit for VW users if supported appropriately.

Although there has not been any prior research on virtual world role detection known to us, advances in related disciplines are highly relevant to the topic of this work and deserve discussion. Previous work on automatic bot detection in virtual worlds has mainly focused on server-side features as for example avatar trajectories [Chen et al. 2008] or traffic analysis [Chen et al. 2006]. Thawonmas et al. analysed avatar behaviour to identify bots [Thawonmas et al. 2008]. While the intuition behind their approach and ours is similar, the underlying notions of behaviour are different. The authors concentrated exclusively on the usage of VW spells and attacks whereas our classification scheme will try to distinguish more abstract concepts such as activities and roles based on user communication.

Interestingly, we can note an interesting change in the research community's attitude towards this task. Not long ago, in 1994, Mauldin concluded his survey of state of the art conversational agents [Mauldin 1994] by stating that while being an interesting problem, there was no practical use for machines passing the Turing Test. This view has changed radically in recent years. Nowadays, we face MMOG bots aspiring to imitate human chat behaviour in order to conceal their artificial nature, and machines, such as IBM's Watson, winning game shows by answering natural language questions. In the field of crowdsourcing, direct financial gains can be made from disguising machines as workers and letting them perform paid tasks.

Another related discipline from the field of information retrieval is expert finding [Balog et al. 2006]. Identifying proficient individuals within the boundaries of a given organization or network can be beneficial for our aim of gaining knowledge about the social structures of a virtual world. Especially with respect to the "Virtual World Yellow Pages" application discussed previously, the parallels become obvious. With the growing popularity of social networks such as Facebook or its local equivalents, a significant amount of recent research has been dedicated to finding experts in such networks [Maybury et al. 2001; Zhang et al. 2010; McDonald 2003]. Considering VW features such as friend lists or scheduled group events, VWs can be viewed as a form of social network augmented by a virtual environment. While expert finding approaches largely consider corpora of (professional) documents, our approach will exploit communication data to find knowledgeable people. Assuming a role affiliation to be the equivalent of expertise in that role's core capabilities, we employ similar techniques tailored towards the domain-specific information available in virtual worlds (e.g., spatial and directional knowledge). Leuski and Lavrenko's work on virtual world event detection [Leuski and Lavrenko

2006] is the first and currently only application of means of IR for a virtual world. Events were defined as groups of players defeating monsters in the game. Based on the success of the 2006 project, which yielded an error rate of only 10% at predicting raids at a given time and location, this work will take an additional step and also predict the precise kind of activity performed as well as the participants' roles.

### 3. ROLE DETECTION

We will begin the formal description of virtual world (VW) role detection by introducing the relevant activities and roles that are commonly observed in Massively Multiplayer Online Roleplaying Games (MMORPGs). The research presented in this work is based on the Blade Mistress Corpus [Eickhoff 2009], a manually annotated collection of 11,227 MMORPG chat messages originating from a non-commercial MMORPG. The messages are organized in 849 conversations, each annotated with activity and role labels.

#### 3.1 Activities

The Blade Mistress corpus distinguishes between 6 different types of functional activities in which avatars can be involved. Table II shows the distribution of activity types in the corpus.

**Adventuring.** Adventuring conversations are related to one or more players' efforts to defeat a VW monster and in turn gain its treasure. Typical adventuring roles are *leaders* who coordinate the action and *adventurers* who carry out the plan.

**Trading.** Trading conversations are concerned with the buying, selling or exchanging of virtual goods. They commonly involve the roles of *buyers* and *sellers*.

**Crafting.** Many MMORPGs offer the possibility to create virtual goods (e.g., a sword). Crafting conversations address the made-to-measure creation of goods. The common roles observed in crafting are *craftsmen* and *customers*. The difference to trading conversations lies in the product customisation aspect which mere trading lacks.

**Social Structures.** A common feature of virtual worlds is the possibility to form social networks and user groups. Conversations about this activity are related to the creation or administration of such structures. The related roles are *organizers* who guide and manage a social structure and *members* who are part of that structure.

**Knowledge Transfer.** Due to most VW's high degree of complexity there is a constant need for exchange about VW mechanisms. Knowledge transfers commonly identify *teachers* and *learners*.

**Story telling.** The activities above are to be used when annotating conversations by users who are actually involved in an action. In order to differentiate this from just telling someone about a recent VW event, "story telling" was introduced. It involves *narrators* and *listeners*.

**Other.** All previously described activities are closely VW-related. There is, however, a substantial amount of non-VW chatter present in the messages. In order to represent everything which could not be clearly associated to any of the above activities, we introduced the "other" category.

Table II. Global activity distribution

Activity	Share
Knowledge Transfer	23.91%
Adventuring	22.73%
Other	22.5%
Trading	14.49%
Social Structures	6.83%
Story Telling	6.12%
Crafting	4.12%

Table III. Global role distribution

Role	Share
Adventurer	15.71%
Learner	12.76%
Teacher	11.15%
Seller	7.56%
Leader	7.02%
Buyer	6.93%
Member	4.03%
Listener	3.35%
Organizer	2.8%
Narrator	2.77%
Craftsman	2.33%
Customer	1.79%
Other	21.8%

### 3.2 Roles

The previous activity descriptions mentioned two typical roles for each of the activities. These are the most commonly observed roles for those activities. We will further refer to them as *essential* roles since the nature of the activity is defined by their presence. Most roles can be taken by multiple individuals in a conversation. It is for example not unusual to have several sellers barter the price of a certain (shared) good with one or more buyers. The opposite case in which an essential role is not incorporated by at least one avatar is rather unlikely.

Analogously to the “other” activity there is an “other” role, as well, to account for unrelated chatter within a conversation. Table III shows the distribution of roles in the Blade Mistress corpus.

### 3.3 Role Detection as a task

Since role detection in a complex environment is not a trivial task, we will split it up into three sub tasks which separately are easier to accomplish. To illustrate the scope and complexity of these sub tasks, we will inspect a small portion of the Blade Mistress Corpus for each of the steps. This corpus will also be used for training and evaluation of our classifiers. Each entry contains a time stamp, the sender, the message and the sender’s coordinates in the VW.

```
01:07:49 2,1 oldy_chick says is this for keeps?
01:07:54 2,1 Joe2 says yeah free
01:08:05 2,1 oldy_chick says thank u very much
01:08:10 2,1 Joe2 says welp off to do some lving
01:08:22 3,1 oldy_chick says ill help
01:08:26 17,8 Belanya says why you following me?
01:08:34 5,0 Joe2 says no need i dont kill monsters
01:08:37 18,10 Belanya tells everyone under skills unless you havent purchase it yet
01:08:40 5,0 Joe2 says just hit with fists
01:08:42 5,0 oldy_chick says ?
```



**01:08:54 5,0 Joe2 says they dont die that way but i still get xp**  
**01:09:02 5,0 oldy\_chick says oh**  
 01:09:04 16,8 Belanya says owy!  
**01:09:06 5,0 oldy\_chick says cool**  
 01:13:57 14,29 Belanya tells everyone 1100 exp to lvl 22 :D  
 01:14:16 72,8 Presideon tells everyone how do you find out your level?  
 01:14:21 14,30 Belanya tells everyone your dodge lvl  
 01:14:24 34,11 Joe2 tells everyone dodge  
 01:14:34 72,4 Presideon tells everyone where's that  
 01:14:34 34,11 Lorian says 102 till lvl 18  
 01:14:37 34,11 oldy\_chick says 3  
 01:14:51 34,11 Joe2 says 600 till 17  
 01:15:14 34,11 Lorian says were's are lvl 27 you wont get xp

The 8 minutes of communication shown above contain various conversations between players. It contains examples of knowledge transfer, the end phase of a trading activity and some chatter. The part which we will mainly focus on in this example is the knowledge transfer between oldy\_chick and Joe2 from 01:08:10 until 01:09:06.

The first observation to be made is that the conversation in which Joe2 explains a certain exploit of the game mechanics to oldy\_chick is preceded and trailed by unrelated messages from other players as well as the two involved ones. To enhance readability, the related messages in this knowledge transfer were highlighted in bold type face. Apart from unrelated messages surrounding the conversation, we also have to detect unrelated messages being sent within it. An example would be the clearly irrelevant utterances of Belanya, who is not even at the same location as the other two.

### Message grouping

Precisely as in the real world, virtual world roles depend strongly on their context. A person might take several very different roles depending on what he or she does and who else is present. A successful manager who is in charge during most of the day might for example fulfil an entirely different role in the Judo club which he just joined two weeks ago and where he still has to learn a lot about the group's structure and the sport itself. In order to take this fact into account, roles are not assigned globally, but within the context of each conversation. Global trends can be inferred from the roles which a person regularly takes.

The first step towards role detection is therefore grouping those messages from the chat log that belong to a single coherent conversation. With regard to our example conversation, grouping would fulfil the purpose of clustering the bold messages together, separating them from the others. Although humans can handle this task of deciding which information belongs to which topic without difficulties, it is probably the hardest of the three sub tasks in terms of automation.

### Activity labelling

Similar to their real world counterparts, VW roles tend to depend on the type of activity that is carried out. Unless the roles are defined very abstractly (e.g., just distinguishing dominant from subordinate behaviour), information about the conversation's topic is essential for role detection. Since we defined groups as sets of messages belonging to a single activity, this means assigning a class label for each system-generated message group.

We again inspect our sample conversation which has now been grouped ideally, all unrelated messages before, between, and after the important ones have been stripped away. The aim of this sub task is now to find a single class label for the kind of activity taking place. In our example we would like the system to classify this group of messages as a *knowledge transfer*.

```
01:08:10 2,1 Joe2 says welp off to do some lvlng
01:08:22 3,1 oldy_chick says ill help
01:08:34 5,0 Joe2 says no need i dont kill monsters
01:08:40 5,0 Joe2 says just hit with fists
01:08:42 5,0 oldy_chick says ?
01:08:54 5,0 Joe2 says they dont die that way but i still get xp
01:09:02 5,0 oldy_chick says oh
01:09:06 5,0 oldy_chick says cool
```

### Role labelling

With the messages grouped into conversations and their activities annotated, the third sub task is to finally assign an activity-dependent role to every participating person in the conversation.

In the knowledge transfer example, we would like Joe2 to be classified as a *teacher* and oldy\_chick, who gets information about the exploit, as a *learner*.

## 4. EVALUATION PLAN

This section will briefly introduce the evaluation metrics which we will use to determine the quality of our role detection system. The evaluation will be done separately for each sub task. The Blade Mistress Corpus consists of virtual world (VW) conversations grouped and annotated with activity and role labels by human judges to form a reliable evaluation corpus. It was split into distinct training (90%) and test sets (10%). Model training and parameter tuning will be carried out on the training set. The test set will be exclusively used to report the final performance. In the following, we will call the corpus' judgements *reference* groups/labels as opposed to automatically created *hypothesis* groups/labels.

### 4.1 Message grouping

The evaluation of the first task will be done in terms of misses and false alarms according to Topic Detection and Tracking (TDT) Phase 2 [Doddington 1998]. We decided to follow this well-known evaluation scheme as the message grouping task shows the same work flow and identifies the same quality criteria as the established topic detection/tracking task from the literature. The miss probability, in this context, expresses the likelihood of a message related to a certain virtual world

proceeding not being put into the relevant group. The false alarm probability, on the other hand, stands for the likelihood of an unrelated message being put into the group. Reducing misses and false alarms will create better groupings since related messages from the VW are grouped together (low misses) and unrelated utterances should not end up in the same group (low false alarms). The first step is to find the most similar hypothesis group  $H_j$  for each reference group  $R_i$ . This mapping  $H(R_i) = \operatorname{argmin}_{j=1}^J \{D(R_i, H_j)\}$  will be done by determining the distance between  $R_i$  and all system generated groups. The group with the smallest distance will be selected, whereas distance is defined as:

$$D(R_i, H_j) = P_{Miss}(R_i, H_j) * c + P_{FalseAlarm}(R_i, H_j) * (1 - c)$$

with

$$P_{Miss}(R_i, H_j) = \frac{N_{Miss}(R_i, H_j)}{|R_i|}$$

$$P_{FalseAlarm}(R_i, H_j) = \frac{N_{FalseAlarm}(R_i, H_j)}{|M - R_i|}$$

$c$  is the relative cost of a miss in relation to a false alarm.

$N_{Miss}(R_i, H_j)$  is the number of messages in  $R_i$  which are not in  $H_j$ .

$N_{FalseAlarm}(R_i, H_j)$  is the number of messages in  $H_j$  which are not in  $R_i$ .

$|G_i|$  is the number of messages in group  $G_i$ .

$M$  is the set of all messages  $m_1..m_{|M|}$  in the corpus.

Once the mapping between hypothesis groups and reference groups is defined, we can compute the miss and false alarm probabilities over the whole corpus as:

$$P_{Miss} = \sum_i P(i) P_{Miss}(R_i, H(R_i))$$

$$P_{FalseAlarm} = \sum_i P(i) P_{FalseAlarm}(R_i, H(R_i))$$

Where  $P(i) = \frac{|G_i|}{|M|}$  is the group size divided by the collection size.

## 4.2 Activity detection

The activity detection task will be evaluated in terms of classification accuracy over all reference groups.

$$Accuracy = \frac{1}{|R|} \sum_{i=1}^{|R|} \delta(R_i, H\{R_i\})$$

with

$$\delta(R_i, H_j) = \begin{cases} 1 & \text{if } \alpha(R_i) = \alpha(H_j) \\ 0 & \text{otherwise} \end{cases}$$

Where  $\alpha(R_i)$  is the activity label assigned to message group  $R_i$ .

The intuition behind this approach is the simple notion of visiting all reference clusters and checking whether they carry the same activity label as the most similar system-generated cluster. For each match, we add 1 to our total sum. In the end, we normalize by the size of the reference set, ending up with the share of clusters which were labelled correctly.

### 4.3 Role detection

In analogy to activity detection, the role detection task will be evaluated through classification accuracy over all reference groups  $R_i$  and all players  $P_{R_i,j}$  within each group.

$$Accuracy = \frac{1}{|R|} \sum_{i=1}^{|R|} \frac{1}{|P_{R_i}|} \sum_{j=1}^{|P_{R_i}|} \Delta(P_{R_i,j}, P_{H(R_i),j})$$

with

$$\Delta(P_r, P_h) = \begin{cases} 1 & \text{if } \rho(P_r) = \rho(P_h) \\ 0 & \text{otherwise} \end{cases}$$

Where  $\rho(P_r)$  is the role label assigned to player  $P_r$ .

This evaluation scheme is based on the same notion of iterating over all reference groups that we used for activity detection. Instead of inspecting activity labels, at this point, we compare the role labels per person in each group. In the end, we obtain the share of role labels which has been correctly assigned.

One of the key limitations of the proposed evaluation scheme is the central role of the mapping  $H(R_i)$  between hypothesis and reference clusters. Low overlap between clusters will propagate to activity and role labelling performance. As a consequence, the resulting performance should be interpreted as conservative estimates of the actual performance. To limit this effect, we will use gold standard groups for activity and role detection.

## 5. ROLE DETECTION APPROACHES

After having introduced the relevant sub tasks of virtual world role detection and the means of evaluating detection performance, this section will describe our method in detail.

### 5.1 Message grouping

Our message grouping strategy is a rule-based system that takes into account the senders' positions, their means of communication, as well as the time elapsed between messages  $m$ . We considered a combination of two types of rules: (1) Connectivity rules determine whether several users are able to communicate. Some of the VW's chat channels are position-dependent (e.g., *say* or *shout*), others are directly sent to one or more players (e.g., *whisper* or *guild*), and some are indefinite broadcasts to the entire VW. Thus, we consider only messages from authors who were able to read a given message group's previous stream of communication.

(2) Freshness rules are concerned with the delay between messages. If two messages are issued with a significant communication gap between them, it is unlikely that they belong to the same conversation. Once connectivity and freshness are ensured, we expand the message group by  $m$ .

## 5.2 Activity/ Role labelling

In this section, we will introduce different n-gram language models (LM) and vector space models (VSM) to represent the likelihood of a group of messages belonging to a given activity or role.

### Uni-gram Language Models

Message groups are split into white space-separated lists of words, punctuation is removed and the words are mapped to lower case. The (smoothed) probability of observing each single word  $w$  given the model's activity/ role is multiplied into the final LM score.

$$p(G|LM_A) * p(A) = p(A) \prod_{m \in G} \prod_{w \in m} LM_A(w)$$

It consists of the prior class probability

$$p(A) = \frac{\#Groups \text{ labeled } A}{\#Groups \in C}$$

and

$$LM_A(w) = \lambda \frac{\#w, A}{|A|} + (1 - \lambda) \frac{\#w}{|C|}$$

a uni-gram-based language model. This model interpolates the relative frequency of  $w$  in the training corpus for activity  $A$  and its relative frequency over all models according to the Jelinek-Mercer smoothing principle [Jelinek and Mercer 1980].

### Character n-gram LM

Message groups are split into white space-separated lists of words, punctuation is removed and the words are mapped to lower case. The resulting words are split into word-internal character sequences of length  $n$ . Those are then used as terms to compute the LM score as before.

The character n-gram representation of “*Did you find the mystery sword?*” for  $n = 2$  is:

$$v = \{di, id, yo, ou, fi, in, nd, th, he, my, \\ ys, st, te, er, ry, sw, wo, or, rd\}$$

As an alternative to using language models, we propose applying various vector space models. Each of them represents message groups as vectors in high dimensional spaces.

### Binary Vector Space Model

This simplest version of a vector space model as a message group representation contains one vector component for each unique term in the group of messages. We chose this model to evaluate the impact of general term presence or absence on activity/role detection performance.

### Frequency-based Vector Space Models

This kind of VSM was used to measure the effect of word frequency information on detection quality. Each vector component contains the absolute frequency of a unique word.

### Character n-gram Vector Space Models

In difference to the previous two modelling approaches that used whole words for the vectors' components, this model employs word-internal character n-grams and their frequencies to represent message groups. While the first two models were non-parametric, the choice of n-gram length  $n$  will be an important step towards obtaining good results. This model is interesting for its tendency to reveal characteristic patterns on the sub-word level.

### Non-verbal features

The models introduced above exclusively used words or character n-grams as features. In order to not only consider *what* people say but also *how* (where, when, etc.) they say it, the previous models can be expanded by a range of non-verbal features generated from the message groups. These features are:

- Mean message size in words/ size variance
- Total number of messages/ words
- Most frequently used communication channel's ID
- Mean x and y coordinates of senders
- Mean/ variance time interval between messages
- Total number of involved players
- Travelled geographical distance during conversation

These features bear information about shallow message properties, means of communication and movement patterns. Each of which might contain information about the role the sender takes. They are injected as additional dimensions in the vector space. In order to do this, we have to take into account the very different scales of the values. Previously, we only had to process term frequencies which are on the same scale. Now, however, we introduce new dimensions that might exceed those counts by far and that would therefore give a too high overall importance to the feature. We prevent this by normalizing the features' scales and mapping them onto word frequency scale in the following way:

$$c_i = \frac{x_i - \min_i}{\max_i - \min_i} \max_{freq} \lambda_i$$

Where

- $c_i$  is the rescaled value of the  $i$ -th feature,
- $x_i$  is the feature's original value,
- $\min_i$  is the smallest observed value for feature  $i$ ,
- $\max_i$  is the largest observed value for feature  $i$ ,
- $\max_{freq}$  is the largest observed term frequency.

The minima and maxima needed for this mapping were calculated over the whole corpus of messages. To tune the features' individual contributions to the overall classification, the model is finally expanded by a vector  $\Lambda$  of weights  $\lambda_i$  for each feature component  $c_i$ .

### VSM strategies

The main task for vector space classification is finding the decision boundaries in the high dimensional space. The Centroid method approaches this task by introducing mean vectors for each class. They represent the center of mass of their class. The intuition behind this is checking how similar a given conversation is to the “average knowledge transfer”, the “average monster hunt”, etc. Instead of considering one centroid per activity or role, we can employ a k-nearest-neighbour (KNN) approach that keeps all known training instances and assigns the locally dominant label to each new message group.

### VSM similarity measures

Both, KNN and Centroid methods employ notions of distance between entities in the vector space. We evaluated the following similarity measures to determine nearest neighbours/ closest centroids. They were chosen to inspect the suitability of various approaches of distance measurement for the domain of role detection.

- Inner Product
- Euclidean Distance
- Dice Coefficient
- Cosine Coefficient
- Jaccard Coefficient
- TF/IDF

While most of the above are widely known distance metrics, the TF/IDF formula is a more sophisticated measure from the field of information retrieval that deserves further elaboration. The central idea of this approach is to compare the relative frequency of a term  $w$  between two message groups. The main element of the equation is the term frequency  $tf_{w,\vec{g}}$  which represents the frequency of word  $w$  in message group  $g$ . Since larger message groups contain more words, the group size  $|G|$  multiplied by a squashing factor  $K$  is used for normalization. Finally, words which are generally frequent tend to contain little topical meaning (e.g., “the” or “and”) whereas very infrequent words can contain a lot of meaning (e.g., “hibernation”). To account for this, the notion of *inverse document frequency*,  $df_w$ , was introduced. This represents the number of documents containing term  $w$ . TF/IDF gives only very little weight to frequent words (i.e., words with  $df_w$  close to the collection size  $|C|$ ).

$$sim(\vec{r}, \vec{u}) = \sum_w \frac{tf_{w,\vec{r}}}{tf_{w,\vec{r}} + K * |R|} \frac{tf_{w,\vec{u}}}{tf_{w,\vec{u}} + K * |U|} (\log \frac{|C|+1}{df_w+0.5})^2$$

## 6. PERFORMANCE EVALUATION

In this section, we will evaluate different means of role detection along the formulae described in Section 4.

In order to obtain unbiased evaluation results, the corpus was split into distinct training and test sets. Training and evaluation were done exclusively on the training set using stratified 10-fold cross validation. When the preferred model for a sub task had been chosen and its parameters had been tuned, this model was run once on the test set. This final performance is reported without further tuning or training.

Table IV. Grouping performance on the training set

$\delta_t$	$P_{Miss}$	$P_{False Alarm}$	Mean
10	93.32%	2.69%	48.01%
20	85.04%	3.42%	44.32%
40	61.17%	4.24%	32.71%
60	40.77%	4.52%	22.65%
80	32.69%	7.98%	20.34%
100	28.31%	10.44%	19.38%
120	26.01%	12.81%	19.41%
<b>140</b>	<b>21.78%</b>	<b>14.17%</b>	<b>17.98%</b>
160	21.51%	16.4%	18.96%
180	21.66%	17.5%	19.58%

A part of the test set was redundantly annotated<sup>1</sup> in order to measure human agreement for the three tasks.

### 6.1 Message Grouping

As outlined in Section 4, the grouping task is evaluated in terms of misses and false alarms. For a practical application of message grouping techniques, one will have to decide on a weighting scheme that reflects the relative cost of a miss with respect to a false alarm. For our evaluation, we chose uniform weights, making misses and false alarms equally important.

The main variable in message grouping is the time  $\delta_t$  after which we close a potential conversation cluster. Lower values of  $\delta_t$  mean less false alarms at the cost of more misses.

Table IV shows the influence of  $\delta_t$  on the grouping performance on the training set. Figure 1 depicts the error function for different settings of  $\delta_t$ . In order to get a better understanding of these results, we computed Kappa agreement scores for the corpus' redundantly annotated share of the groupings. Human annotators were able to achieve  $K = 0.82$  whereas our system reached  $K = 0.51$ .

Message grouping appears to be a domain in which human annotators outperform our grouping system by a significant percentage. The lowest overall combination of misses and false alarms could be achieved by a  $\delta_t$  of 140 seconds. This setting was used on the test set and achieved a  $P_{Miss}$  of 35.91% and a  $P_{FalseAlarm}$  of 11.04% for unseen data.

### 6.2 Activity and Role Detection

Because of the high number of possible combinations of models and similarity metrics proposed in Section 5, we will evaluate them separately for each model. The best model will be selected and its parameters be tuned before reporting its performance on the unseen test data.

### SVM baseline

In order to compare our results to state of the art text classification techniques, we

<sup>1</sup>For 20% of the test set there are second judgements. An additional third judgement was collected for 6% of the messages.



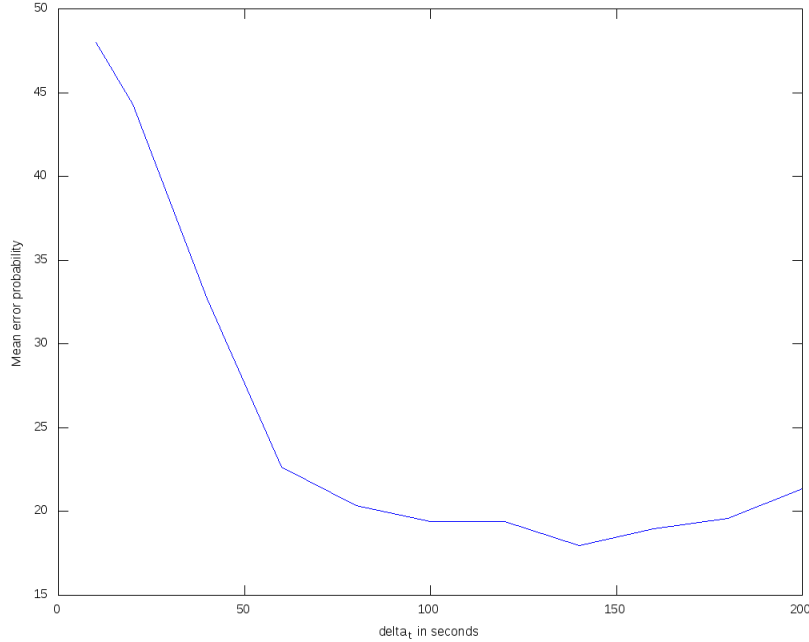


Fig. 1. Grouping performance on the training set

will include a support vector machine (SVM) classifier using uni-gram frequencies and our set of non-verbal features for activity/role classification. SVMs have been trained and tested using the LibSVM library [Chang and Lin 2001] (C-SVM, radial basis kernel, cost = 1,  $\epsilon = 0.001$ ,  $\gamma = 0.01$ ). We used tf/idf-weighted terms and the rescaled non-verbal features described in 5.2 for our SVMs. The above parameter settings are the best-performing settings determined through a series of iterative experiments.

### Activity detection

Table V shows a comparison of the various model-measure combinations and our SVM baseline. The general tendency we could observe was a superior performance of the Centroid method for this application. The classification results of vector space models expanded by 12 non-verbal features was tested for term frequency-based VSMs. The high number of weights creates a vast space of potential settings. For reasons of time, we could not exhaustively inspect these settings. Although prior testing on smaller hand-labelled data sets suggested a potential benefit from these non-verbal features, we unfortunately could not achieve gains in detection performance on the final corpus. Further fine-tuning and experimentation with the feature weights might, however, produce better results.

For the final run on the unseen test set we chose a 4th-order character n-gram

Table V. Activity detection results on training data

Classifier	Accuracy
Random Baseline	23.91%
Character 3-gram LM	29.09%
<b>Euclidean/Char 4-gram VSM/Centroid</b>	<b>38.34%</b>
TF/IDF/Binary VSM/KNN	21.95%
Euclidean/Frequency VSM/KNN	20.91%
Euclidean/Binary VSM/Centroid	28.7%
Euclidean/Character 4-gram/KNN	23.12%
SVM (uni-gram freq + shallow features)	34.03%

Table VI. Role detection results on training data

Classifier	Accuracy
Random Baseline	21.28%
Euclidean/Binary VSM/Centroid	39.23%
Cosine/Character 3-gram VSM/KNN	35.48%
Uni-gram LM	38.10%
Euclidean/Character 4-gram VSM/Centroid	36.29%
Jaccard/Frequency VSM/KNN	38.79%
<b>Jaccard/Binary VSM/KNN</b>	<b>39.84%</b>
SVM (uni-gram freq + shallow features)	38.12%

vector space model using Euclidean Distance as similarity function with the Centroid method. This combination achieved a very reliable overall performance on the training data. With this set-up we could obtain an accuracy of 36.47% on unseen test data. Human annotators reached a score of 47.29% ( $K = 0.55$ ).

### Role detection

Evaluation procedure and set-up for this task are very similar to the previous ones. We determined the best model and its parameter settings on the training data and finally evaluated the system with a single run on the test set. Table VI shows the best-performing model-measure combinations for role classification. Similar though the tasks may be, we could clearly recognize that the most successful activity detection approaches are by no means bound to be the best ones for role detection. The Centroid method for example, that performed very well on activity detection, does not yield convincing results for role detection. Another difference to activity classification results is the performance of the language modelling approach. Although they are amongst the best models, the former tendency of character n-gram language models outperforming uni-gram language models was inverted. For role classification, the accuracy rises with the model order, thus approaching uni-gram word model accuracy.

For the final classifier we chose a binary vector space model using KNN and the Jaccard Coefficient as this has proven to be a reliable combination. Run on the test set's unseen data, the system could achieve a final classification accuracy of 36.08% in comparison to the human annotators' 44.96% ( $K = 0.49$ ).

Table VII. Detection results on unseen test data

Method	Activity labels	Role Labels
SVM Baseline	31.51%	33.04%
VSM approach	36.47%	36.08%
Human performance	47.29%	44.96%

### Performance on unseen test collection

Table VII shows our vector space approach’s performance on the unseen test collection in comparison to the SVM baseline for activity and role labelling. While not being able to match human accuracy, our vector space methods outperformed the baseline for both labelling tasks at  $\alpha < 0.05$  significance level (Determined using Wilcoxon Signed Rank Test).

## 7. APPLICATION EXAMPLE

In Section 1.1, we motivated several scenarios for which virtual world (VW) role detection can be beneficial. After having shown that automatic role detection is achievable with results close to human performance, we will now revisit the virtual world yellow pages scenario in more detail. According to a survey on players’ reasons for quitting their VW subscriptions [Yee 2003], one of the most frequently mentioned causes was the player’s circle of friends stopping to play the game. Massively Multiplayer Online Roleplaying Games (MMORPGs) focus heavily on the grouping aspect, offering manifold challenges for the players that can not be overcome by a single individual. Players quitting the game, therefore, often cause series of subscription cancellations as the users lose their accustomed groups. It appears crucial for preserving long-term player satisfaction to provide a lively social environment in the VW. At the same time, meeting new people in virtual worlds seems to be a hard task. World of Warcraft, the market leader in MMORPGs, recently introduced various means of easing grouping and companion finding in order to limit user frustration and the resulting cancellation chains. Their system currently allows to search for players that want to fight the same monsters. This intention has to be manually indicated when entering the grouping queue. This rather straightforward approach is not suitable for providing recommendations of complex concepts such as the roles described in this work. We are confident that offering more fine-grained avatar recommendations based on VW role detection will be widely appreciated by the user community.

### 7.1 Set-up

We conducted our applicability study at the example of the free MMORPG *The Mana World* [tmw 2010]. The game offers all common MMORPG features with the only exception of the crafting system being limited to just one trade (alchemy) at that point of development. We decided for this particular VW as it is one of the few sophisticated games with Open Source licenses available for both, the client and server code. This enabled us to offer our own game server and conveniently record and evaluate player communication while at the same point avoiding privacy issues by being able to inform the players of our research interest before they started playing in the VW. The server was available for 6 weeks during which we collected



Fig. 2. Open MMORPG: The Mana World

data from 284 distinct players.

## 7.2 Features offered

For this first applicability study we offered only a basic range of yellow page features to the inhabitants of our virtual world. The services described below were offered on the web site on which we also advertised our VW server. The two types of services available were (1) a browsable directory of players and (2) a friend recommendation system.

### Player directory

To build up a browsable directory of players, we represent each user as a vector with one dimension per role type<sup>2</sup>. The values in these dimensions are the shares of their overall play time that the users spend in the relevant role. Players for whom the value in a certain dimension lies significantly above the global average are listed for that role(s). The web interface finally allows players to browse the list of players by their preferred roles in the game in order to ease grouping or to find the knowledgeable person on a certain topic.

### Friend recommendation

To show the possible impact of role detection on friend recommendation, we employ two different measures, mutual similarity and complementarity of activity and role preferences. A high score in one of the measures will lead to the relevant avatar being recommended as a friend. Players with similar interests (expressed through a similar role distribution) are valuable sources for experience exchange or joint activities. To measure similarity, we make use of the role distribution vectors

<sup>2</sup>Activities and roles were determined according to the best-performing methods evaluated in Section 6. We exclude the “other” role at this stage as it does not have a well-defined functional use.

from the player directory. The actual similarity score is determined by computing Euclidean Distance between the vectors. In Section 3.2, we described roles as complementary pairs belonging to a common activity. We will exploit this design by measuring the degree of complementarity between two players' interests. A player with a high complementarity score often takes the roles that are needed for interaction. Following complementarity, an impassioned leader could find people who frequently take the role of adventurers. Complementarity scores are, again, obtained by measuring Euclidean Distance between role distribution vectors. This time, however, the player's vector is transformed in such a way that the positions of the two complementary roles are swapped before comparing it to a prospective friend's vector. This for example results in the player's preference for crafting being compared to each potential friend's previous preference for having things crafted. The friend recommendation at the current stage is exclusively based on the findings of our role detection. For further use in large-scale environments, state of the art predictors of friend relations such as the number of mutual friends or the guild affiliation should be incorporated into the model.

### 7.3 Conclusion & Community response

In order to get an impression of how useful our yellow pages system is, we asked each user to give us a brief feedback on how much she or he liked the recommendation and player directory features on a scale from 1 "Did not like it at all" to 5 "Liked it a lot". They additionally had the possibility to provide more detailed textual comments. During the time of the experiment we collected 179 feedback responses on our system. The mean popularity score of the player directory was a very high  $\mu = 3.95$  with a standard deviation  $\sigma = 0.28$ . The friend recommendation feature was rated  $\mu = 3.1$  and  $\sigma = 0.505$ . The detailed responses were largely positive. Especially the player directory was mentioned frequently. Several players explicitly expressed their desire for similar functionalities being implemented into their favourite commercial on-line games. The friend recommendation was not as popular and should, for real use, be expanded by features like the number of mutual friends. Players were hesitant to approach (possibly interesting) recommendations whom they did not know at all. With our applicability study, we showcased an example of how to exploit the findings of this work in a real world environment. The generic fashion in which our models are phrased, making no static assumptions about the concrete type of VW or its purpose, allow for a flexible use of our method. In this section, we showed how the insights gained from our analysis of one MMORPG (Blade Mistress) could be successfully ported to another platform (The Mana World) and was able to provide an enhanced game experience there. Due to the huge quantities of archived textual player communication, the creation of role detection functionality following the scheme outlined in this work, can be considered a feasible task for providers of commercial VWs. Potential benefits include: (1) Greater enjoyment of players due to sophisticated VW recommendation functions as described in this section. (2) Better understanding of player behaviour that may be exploited when designing new content following the needs and preferences of the user base. (3) The possibility of targeted advertisement services affiliated to many current VWs. Knowledge of a player's preferred VW activities and roles may allow for more accurate personalised ad placement.

## 8. CONCLUSION

In this work, we introduced the novel task of virtual world (VW) role detection and formalized the necessary steps to achieve it. Due to strong parallels to the real world, virtual worlds can be used as a controlled test environment for our approaches. We employed techniques of information retrieval in the VW domain and evaluated their performance based on a manually annotated corpus of VW chat messages. With the resulting classification method, we were able to outperform the baseline method, a state of the art text classification technique, at significance level. Message grouping turned out to be an unambiguous task for human annotators who showed good agreement rates. As we already anticipated earlier, this task is not trivial for automatic systems and will remain a challenging problem to dedicate further research to. Identifying activities and user roles based on chat communication has been shown to be an area in which our automatic methods perform reliably. We closely approximated human performance while our approaches consistently outperformed the text classification baseline.

Based on the positive recent development of the virtual world market and the success of our applicability experiment we are convinced that role detection is a valuable research field of growing importance. Recent developments on the Massively Multiplayer Online Games (MMOG) market show increasingly strong tendencies towards the integration of social networking features into games. As we have seen in other domains, the load of available additional information (e.g., friend relationships, notions of preference, content sharing, event organization, etc.) can be exploited in multiple ways to achieve greater financial gains for VW providers but also greater immersion and enjoyment for users. Although virtual worlds are currently lacking behind on-line social networks in terms of automatic analysis and data mining features, they could aspire to drive the entire field because of their close connection between (virtual) reality and ubiquitous means of communication that, despite all advances in mobile telecommunication, are not yet present in the real world. Further practical applications to be studied and evaluated include more fine-grained means of automatic discourse understanding guided by user roles, automatic analyses of MMOG activity distributions as a means of evaluating the community's acceptance of novel game features, as well as confirming our findings for alternative position-aware chat domains such as geo-tagged Twitter messages. Future work will follow three major directions: (1) An in-depth exploration of VW activities and roles will be conducted to reduce the still significant share of conversations that could not be assigned a functional activity/role and therefore currently ended up in the "other" category. While we acknowledge the presence of non-VW chatter in the data, we believe that a more atomic definition of activities and roles may serve to more clearly separate functional categories from noise. (2) Text-based classification schemes in highly collaborative social media domains often suffer from the multitude of languages used for communication. The same holds for MMOGs. At this point, we disregarded any non-English message in order to not obstruct the performance of term frequency-based methods. Future extensions on this work will employ cross-language models to also benefit from non-English traces of information. (3) Currently, the role detection scheme is exclusively based on communication features. However, it is to be assumed that role affiliation is subject

to the influence of several additional factors. For example, we can assume certain role types to be observed together more frequently, thus, motivating an inspection of the interrelationship between players in order to propagate prior probabilities of role affiliation across networks built along friendships, guild memberships, or other VW connections.

## REFERENCES

2009. MMOData. [www.MMOData.net](http://www.MMOData.net).
2010. Second life. <http://secondlife.com/>.
2010. The Mana World - A free open source 2D MMORPG in development. <http://themanaworld.org/>.
- BALICER, R. 2007. Modeling infectious diseases dissemination through online role-playing games. *Epidemiology* 18, 2, 260.
- BALOG, K., AZZOPARDI, L., AND DE RIJKE, M. 2006. Formal models for expert finding in enterprise corpora. In *ACM SIGIR*.
- CHANG, C.-C. AND LIN, C.-J. 2001. *LIBSVM: a library for support vector machines*. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- CHEN, K., JIANG, J., HUANG, P., CHU, H., LEI, C., AND CHEN, W. 2006. Identifying MMORPG bots: A traffic analysis approach. In *ACM SIGCHI*.
- CHEN, K., LIAO, A., PAO, H., AND CHU, H. 2008. Game bot detection based on avatar trajectory. *Entertainment Computing-ICEC*, 94–105.
- CHEN, V. AND DUH, H. 2007. Understanding social interaction in world of warcraft. In *Proceedings of the international conference on Advances in computer entertainment technology*. ACM.
- DODDINGTON, G. 1998. The topic detection and tracking phase 2 (TDT2) evaluation plan. In *Proceedings DARPA Broadcast News Transcription and Understanding Workshop*.
- DUCHENEAUT, N., YEE, N., NICKELL, E., AND MOORE, R. 2006. Alone together?: exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 407–416.
- EICKHOFF, C. 2009. Role Detection in Virtual Worlds, M.Sc Thesis.
- JELINEK, F. AND MERCER, R. 1980. Interpolated estimation of Markov source parameters from sparse data. *Pattern recognition in practice* 23, 381.
- JOHNSON, A., ROUSSOS, M., LEIGH, J., VASILAKIS, C., BARNES, C., AND MOHER, T. 1998. The NICE project: Learning together in a virtual world. In *IEEE 1998 Virtual Reality Annual International Symposium, 1998. Proceedings*.
- LEUSKI, A. AND LAVRENKO, V. 2006. Tracking dragon-hunters with language models. In *CIKM*. ACM Press, New York, NY, USA.
- MAULDIN, M. 1994. Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition. In *Proceedings of the National Conference on Artificial Intelligence*. JOHN WILEY & SONS LTD, 16–16.
- MAYBURY, M., D'AMORE, R., AND HOUSE, D. 2001. Expert finding for collaborative virtual environments. *Communications of the ACM* 44, 12, 55–56.
- MCDONALD, D. 2003. Recommending collaboration with social networks: a comparative evaluation. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 600.
- RIZZO, A., PAIR, J., GRAAP, K., MANSON, B., MCNERNEY, P., WIEDERHOLD, B., WIEDERHOLD, M., AND SPIRA, J. 2006. A virtual reality exposure therapy application for Iraq War military personnel with post traumatic stress disorder: From training to toy to treatment. *NATO Security through Science Series E Human and Societal Dynamics* 6, 235.
- THAWONMAS, R., KASHIFUJI, Y., AND CHEN, K. 2008. Detection of MMORPG bots based on behavior analysis. In *Proceedings International Conference on Advances in Computer Entertainment Technology*. ACM, 91–94.
- WILLIAMS, D. 2006. From tree house to barracks: The social life of guilds in world of warcraft. *Urbana* 51, 61801.
- YEE, N. 2003. The Daedalus Project - The psychology of MMORPGs. <http://www.nickyee.com/daedalus/archives/000342.php>.
- ZHANG, J., TANG, J., AND LI, J. 2010. Expert finding in a social network. *Advances in Databases: Concepts, Systems and Applications*, 1066–1069.