

PROC IML vs. PROC REG: Erfolgsprognostik in Online Rollenspielen (MMORPG)

Winfried Hering
FernUniversität in Hagen
58097 Hagen
Winfried.Hering@FernUni-Hagen.de

Rolf T. Wigand
University of Arkansas at Little Rock
Little Rock, AR, USA
rtwigand@ualr.edu

Nitin Agarwal
University of Arkansas at Little Rock
Little Rock, AR, USA
nxagarwal@ualr.edu

Zusammenfassung

Das Prognostizieren von Potenzialen ist ein wichtiger Bestandteil von Organisationsstudien, Verhaltensforschungen, CRM-Anwendungen und vielen anderen Gebieten. Dieses relative Potential, welches auch als Erfolg interpretiert werden kann, wird von verschiedenen Faktoren wie Erfahrung, Ressource-Management, Kooperation, Kommunikation usw. beeinflusst. Ziel dieser Arbeit ist die Erfolgsprognostik in Massively Multiplayer Online Role-Playing Games (MMORPG). Dazu werden zwei Data Mining Methoden miteinander verglichen, ein stochastischer Ansatz mit der Markov Kette und eine klassische Regressionsanalyse. Die Grundlage dafür bilden die Daten des MMORPG, bei dem bis zu 20.000 Spieler auf einem Server spielen. Das Spielfeld auf diesem Server ist begrenzt an Raum und Ressourcen, wodurch sich die Mitspieler bald in einem sozialen Dilemma befinden. Das Ziel des Spiels, der Bau eines *Weltwunders*, ist von einem einzelnen Spieler nicht zu erreichen, und es benötigt die Zusammenarbeit von mehreren Spielern. Dazu ist Kooperation und Koordination zwischen den Spielern unerlässlich. Zusammengefasst ist das Spiel damit komplex, endlich, an Ressourcen begrenzt und kooperationsbedingt dynamisch. Damit weist das Spiel Travian sehr viele Parallelen zu Projekten und Abteilungen in der Berufswelt auf und ist prädestiniert für Untersuchungen von virtuellen Organisationen. Nach der Entscheidung welche Data Mining Methode die geeignetere ist, wird anhand dieser ein finales Prognosemodell erstellt. Anhand dieses Modells zur Potenzialprognostik wird nach Verhaltensmustern gesucht, um ggf. Handlungsempfehlungen für erfolgreiche Spieler ableiten zu können. Ebenso wird das Spiel nach unterschiedlichen Spielphasen durchsucht, in denen die Spieler ihr Verhalten und ihre Strategie wechseln.

Schlüsselwörter: Data Mining, Erfolgsprognostik, MMORPG, Online-Rollenspiele, Regression, Markov Kette

1 Einleitung und Zielsetzung

Ziel dieser Arbeit ist die Erfolgsprognostik in virtuellen Organisationen. Ebenso wird untersucht, welches Data Mining Verfahren dafür am besten geeignet ist. Studien mit virtuellen Organisationen bergen einige spezielle Probleme. Größere Feldstudien mit virtuellen Gruppen sind eher selten und es gibt oft wenige quantitative oder qualitative Daten [1] [2]. Laboruntersuchungen hingegen haben oft sehr kleine Gruppen und sind kurzlebige Simulationen, in denen die Teilnehmer ein geringes Identifikationsgefühl haben. In diesem Projekt arbeiten wir mit Daten eines browserbasierenden Massively Multiplayer Online Role-Playing Game (MMORPG). In diesen Online-Spielen sind die Personen in realistischer Art und Weise über eine längere Zeitspanne miteinander in einer virtuellen Welt integriert. Durch diese Rahmenbedingungen haben die Daten ein großes Potenzial für Studien in Sozial-, Verhaltens- und Wirtschaftswissenschaften [1] [2].

Die University of Arkansas in Little Rock hat Daten eines MMORPG mit dem Namen Travian zur Verfügung gestellt bekommen. Travian ist ein Echtzeit-Strategiespiel (engl. real-time strategy, abgekürzt RTS), in dem viele tausend Spieler auf einem Server, begrenzt an Raum und Ressourcen, spielen. Dadurch befinden sich die Mitspieler bald in einem sozialen Dilemma, indem sie um Ressourcen kämpfen, kooperieren oder handeln müssen. Das Ziel des Spieles ist ein „Weltwunder“ zu errichten und dieses bis zu einer maximalen Stufe gemeinsam auszubauen. Dieses Ziel ist von einem einzelnen Spieler nicht zu erreichen und deshalb können Spieler Allianzen mit bis zu 60 Mitgliedern gründen. Diese Allianzen sind durch Hierarchien und einem Berechtigungskonzept organisiert und die Spieler können durch spielinterne Foren und einem Nachrichtensystem miteinander kommunizieren. Da das Spiel als Einzelspieler nicht gewonnen werden kann, ist Kommunikation unerlässlich in diesem Spiel. Kurz beschrieben, das Spiel ist komplex, endlich, an Ressourcen begrenzt und benötigt Kooperationen zwischen den Spielern. Die Beziehung zwischen Spielern kann mit der Beziehung von Arbeitskollegen in ihrem wirklichen Beruf durchaus verglichen werden. Auch viele Aktivitäten in Spielen sind ähnlich zu denen in der täglichen Berufswelt [3]. Deshalb eignen sich die Daten dieses Spiels auch um ggf. generelle Rückschlüsse auf virtuelle Organisationen zu schließen.

Um zu entscheiden, welches Data Mining Verfahren am besten für diese Art von Erfolgsprognostik geeignet ist, wird als stochastischer Ansatz eine Markov Kette einer Regressionsanalyse gegenübergestellt. Als Realisierungswerkzeug wurden die SAS Module Base SAS mit der Makrosprache, SAS/GRAPH, SAS/IML und SAS/STAT genutzt.

2 Daten

Um die Zusammenhänge besser zu verstehen, werden die Daten des Spiels Travian beschrieben und wie diese für die Datenanalyse aufbereitet wurden. Die Daten für dieses Projekt sind von dem Echtzeit-MMOPRG Travian, in dem bis zu 20.000 Spieler auf einem Server zusammen spielen können. Wir haben die Daten für eine Spielzeit von 3 ½ Monaten in einer überwachten Umgebung aufgezeichnet. In diesem Zeitraum wurde das Spiel einmal vom Anfang bis Ende durchgespielt. Am Anfang des Spieles bekommt jeder Spieler ein Dorf auf dem Spielfeld zugeordnet. In diesem Dorf gibt es verschiedene Ressourcen wie Holz, Eisen, Korn und Lehm. Diese kann der Spieler nutzen, um Gebäude oder Einheiten wie Soldaten etc. zu bauen. Während des Spielverlaufs können sich die Spieler entweder kaufmännisch, defensiv oder räuberisch und aggressiv ausrichten. Bei der defensiven Ausrichtung werden sie neue Dörfer gründen, und durch Handel mit anderen Mitspielern bekommen sie die benötigten Ressourcen. Wenn ein Spieler eine aggressive Spielweise wählt, überfällt er mit seinen Soldaten andere Dörfer und erobert diese. Da es sich bei Travian um ein 24 Stunden Onlinespiel handelt, können die Spieler jederzeit angegriffen werden, selbst wenn diese gerade offline sind und nicht aktiv am Spielgeschehen teilnehmen. Dies ist einer von vielen Gründen, warum es sinnvoll ist, mit anderen Spielern zu kooperieren, die ihre Dörfer im Notfall gegenseitig verteidigen. Diese Zusammenarbeit wird durch das Bilden von Allianzen erleichtert. Ein weiterer und schwerwiegenderer Grund für Allianzen ist, dass das finale Ziel des Spiels nicht von einem einzelnen Spieler erreicht werden kann. Das Erzeugen von Ressourcen zum Aufbau des Weltwunders, sowie das Verteidigen sind als einzelne Person unmöglich. Die Allianzen entwickeln sich von einfachen Zusammenschlüssen schnell zu gut organisierten und autoritär hierarchisch aufgebauten Einheiten. Ein erfolgreiches Zusammenspiel in dieser Einheit kann eine Schlüsselrolle zum Erfolg in diesem Spiel sein.

Als Daten für diese Arbeit nutzten wir 478 Variablen in 30 Tabellen für insgesamt 7.406 Spieler. Die Tabellen sind fachlich getrennt, z.B. gibt es eine Tabelle für Gebäude. Mit einer eindeutigen Gebäude-ID werden dort alle Gebäude beschrieben, die die Spieler haben. Zusätzlich gibt es Informationen zu Ausbaustufen und den benötigten Ressourcen, um diese Gebäude zu bauen. Eine andere Tabelle beschreibt die Angriffe zwischen den Spielern, d.h. jede Bewegung, Entscheidung usw. ist erfasst. Neben dem Angreifer und Verteidiger sind dort Felder für die Anzahl und Art der Truppen sowie den Verlusten auf beiden Seiten. Anhand dieser Daten und weiterer Tabellen kann man sich ein detailliertes Bild über die Spieler und den Spielverlauf machen. Zusätzlich gibt es 2,4 Millionen Nachrichten aus dem spielinternen Kommunikationssystem. Für externe Kommunikationen wie z.B. Skype-Gespräche liegen leider keine Daten vor. Die Spieler haben im Vorfeld zugestimmt, dass die Daten des Spiels für spätere Forschungszwecke genutzt werden dürfen.

Das Ziel dieses Projektes ist den *Erfolg* der Spieler zu erkennen. Daher muss der Erfolg für unseren Sachverhalt definiert werden. Als Erfolg wird der Status des einzelnen Spielers gewertet, den er am Ende des Spiels erreicht hat. Der ultimative Erfolg in Travian ist es, das Weltwunder vollständig auszubauen. Dabei zählen für uns alle Spieler als „Winner“, die den Bau des Weltwunders unterstützt bzw. das Weltwunder militä-

risch verteidigt haben. Eine reine Aufteilung zwischen Gewinnern und Verlierern wäre aus unserer Sicht nicht ausreichend. Vergleichbar zu den Olympischen Spielen ist es nicht nur ein Erfolg, die Goldmedaille zu gewinnen, sondern auch überhaupt auf die Siegertreppe oder in den Finalkampf zu kommen. Analog zu unserem Spiel sind das die Spieler, die bereits angefangen haben das Weltwunder zu errichten, es aber nicht vollständig ausgebaut haben. Diese Spieler werden als „Finalist“ bezeichnet. Auch sich bereits für die Olympischen Spiele zu qualifizieren, ist für viele Personen bereits ein erstrebenswerter Erfolg. Diesen Vergleich kann man auf die Spieler anwenden, die es geschafft haben, bis zum Ende des Spiels zu überleben. Sie haben sich 3 ½ Monate gegen die Angriffe der Gegner gewehrt und damit auch einen Teilerfolg erreicht. Diese Spielergruppe wird hier „Survivor“ genannt. Schlussendlich gibt es noch die Spieler, die im Laufe des Spiels ausgeschieden sind und nicht mehr am Spielgeschehen teilnehmen; diese letzte Gruppe sind die „non-Survivor“. Diese vier Gruppen, Winner, Finalist, Survivor und non-Survivor, sind verschiedene Zustände oder Stufen von Erfolg, die die Spieler während der Spielzeit in Travian haben.

Eine gute Datenqualität und passend aufbereitete Daten sind für das Data Mining unerlässlich. Wie schon der wohlbekannte Ausdruck "garbage in, garbage out" verdeutlicht, ist es nicht möglich, nutzbare Ergebnisse zu erzielen, wenn die Daten noch voller Verunreinigungen sind [4]. Nachdem die Daten aufbereitet sind, werden sie in drei Datentöpfe aufgeteilt. Dabei werden 50% der Daten als Trainingsdaten genutzt, um das Modell zu erstellen. Zur Optimierung der Daten nimmt man 25% als Validierungsdaten und schlussendlich die letzten 25% als Testdaten, um die Qualität zu überprüfen (siehe Abbildung 1).

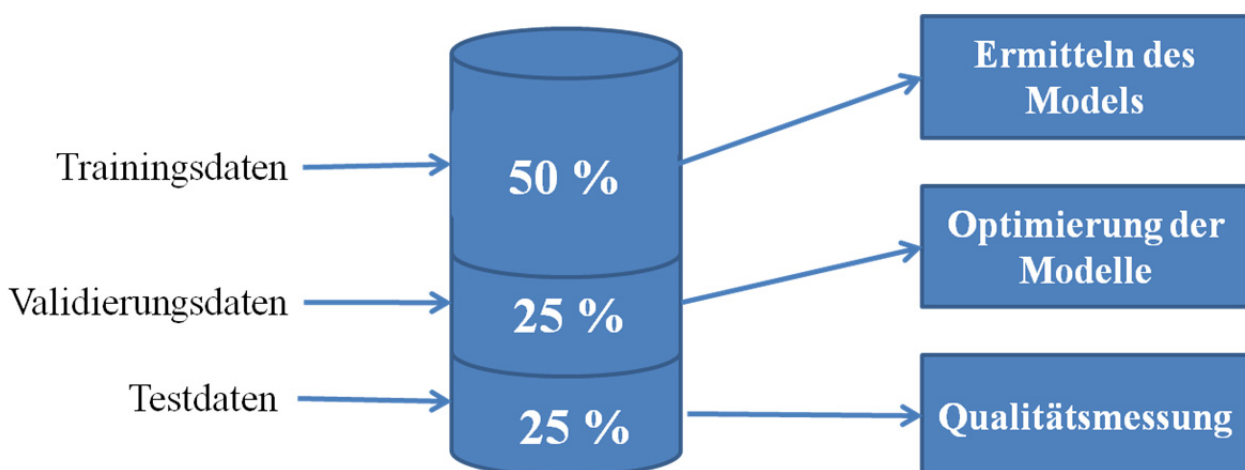


Abbildung 1: Trainings-, Validierungs-, und Testdaten

3 Markov Kette

Die erste der beiden Data Mining Methoden ist die Markov-Kette. Das ist ein stochastischer Ansatz, um unbekannte Zustände anhand von Wahrscheinlichkeitsverteilungen zu ermitteln. Die Grundvoraussetzung dafür ist, dass es eine feste Anzahl von Zuständen gibt, und man die Wahrscheinlichkeiten, dass ein Zustand in einen anderen übergeht, kennt [5]. Die Übergangswahrscheinlichkeit p wird in einer stochastischen Matrix und die Zustände werden in einem Vektor z dargestellt. Durch Multiplikation beider kann man die Zustandswahrscheinlichkeit für die nächste Zeitperiode berechnen: $z_1 = z_0 * p$. Will man den Zustand für einen beliebigen Zeitpunkt n bestimmen, sieht die Formel wie folgt aus: $z_n = z_0 * p^n$.

Wie in Teil 2 erwähnt, gibt es vier Zustände: non-Survivor, Survivor, Finalist, Winner. Der Vektor A , der die Zustände beinhaltet, sieht wie folgt aus $A = \{non - Survivor, Survivor, Finalist, Winner\}$ und wird zukünftig mit den Abkürzungen $A = \{N, S, F, W\}$ beschrieben. Die Wahrscheinlichkeit, dass ein Spieler von einem in den anderen Zustand wechseln wird, ist beispielhaft in Form einer 4x4 stochastischen Matrix in Abbildung 2 dargestellt. Die für dieses Beispiel verwendeten Werte sind zufällig gewählt.

Definition:

Eine stochastische Matrix definiert sich wie folgt: Ist A eine endliche Menge und $\mathbf{p} = (p_{ij})$ mit $i, j \in A$ eine $A \times A$ Matrix, die folgende Eigenschaften besitzt,

$$\sum_{j \in A} p_{ij} = 1$$

$$p_{ij} \geq 0$$

dann heißt \mathbf{p} stochastische Matrix. [5]

	N	S	F	W
N	1	0	0	0
S	0,2	0,7	0,09	0,01
F	0,1	0,1	0,6	0,2
W	0	0	0	1

Abbildung 2: Wechselmöglichkeiten zwischen den Zuständen (stochastische Matrix)

Die Matrix in Abbildung 2 drückt aus, dass ein Spieler mit dem Zustand, dass er einmal ein non-Survivor oder Winner ist, seinen Zustand nicht mehr ändern kann. Grafisch ist der Sachverhalt in Abbildung 3 dargestellt. Non-Survivor nehmen nicht mehr an dem aktuellen Spiel teil und können nicht mehr erneut am Spielgeschehen mitwirken. Nachdem es geschafft wurde, ein Weltwunder komplett auszubauen, wird das Spiel beendet

und es werden die Winner ermittelt. Die Wahrscheinlichkeit, dass ein Survivor zum non-Survivor wird, p_{21} , liegt bei 20%, dass er Survivor bleibt, p_{22} , ist 70%. Dass ein Survivor eine Zustandsveränderung zum Finalist hat, p_{23} , ist 9% und eine Wahrscheinlichkeit von einem Prozent liegt vor für eine Veränderung von Survivor zu Winner. Die Wahrscheinlichkeiten eines Finalisten sehen wie folgt aus:

- Finalist \rightarrow non-Survivor = $p_{31} = 10\%$
- Finalist \rightarrow Survivor = $p_{32} = 10\%$
- Finalist \rightarrow Finalist = $p_{33} = 60\%$
- Finalist \rightarrow Winner = $p_{34} = 20\%$

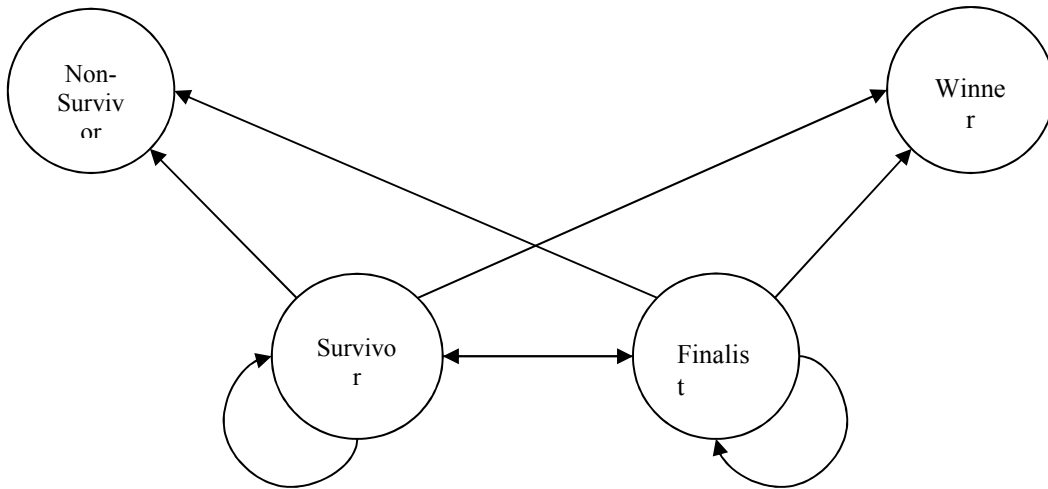


Abbildung 3: Grafische Darstellung der Wechsellmöglichkeiten zwischen den Zuständen

SAS hat für Matrix Operationen eine eigens entwickelte Lösung, SAS/IML - Interactive Matrix Language. Mit einem selbst entwickelten IML-Makro wird nun der Vektor mit der Initialwahrscheinlichkeit so lange mit den stochastischen Matrizen multipliziert, bis der Endzeitpunkt des Spiels erreicht wurde (siehe Abbildung 4). Zu diesem Zeitpunkt hat der Spieler nun seine endgültigen Zustandswahrscheinlichkeiten. Mit diesen Wahrscheinlichkeiten wird ein eindeutiger Zustand bestimmt. Die Ergebnisse der Methode werden im Teil 5 vorgestellt.

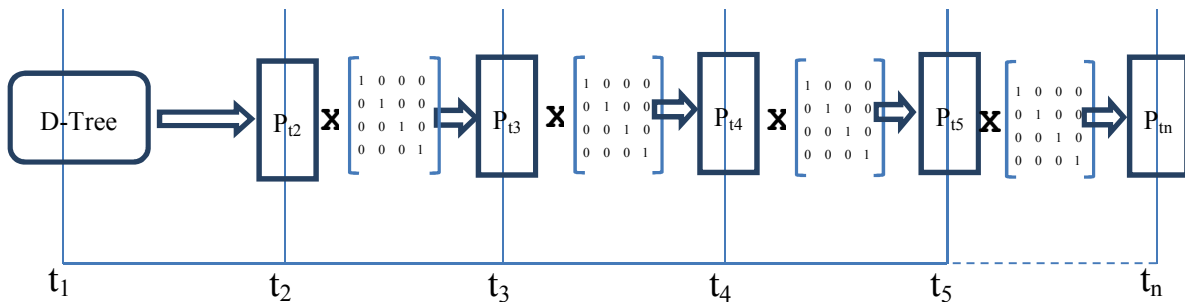


Abbildung 4: Ermittlung der Zustandswahrscheinlichkeit bei der Markov Kette

4 Regressionsanalyse

Die Regressionsanalyse ist ein statistisches Analyseverfahren, um die Beziehung zwischen einer abhängigen und einer oder mehreren unabhängigen Variablen zu modellieren [6]. Für eindimensionale Fälle, in denen es nur eine unabhängige Variable gibt, sieht die mathematische Formulierung wie folgt aus:

$$y = f(x) + e$$

Die Variable y stellt die abhängige, zu prognostizierende Variable und x die unabhängigen Variablen dar. Das Residuum e beschreibt den Fehler, der je nach Komplexität des Modells variiert. Bei mehreren unabhängigen Variablen, die ein mehrdimensionales Modell darstellen, das auch multiples Regressionsmodell genannt werden kann, lautet die mathematische Formulierung wie folgt:

$$y = f(x_1, x_2, \dots, x_n) + e$$

Wenn nun auch die abhängige Variable mehrdimensional ist, spricht man von einem multivariaten Regressionsmodell. Dieser Sachverhalt trifft auch auf diese Arbeit zu. Die unabhängige Variable y_t ist der Zustand des Spielers zu einem gewissen Zeitpunkt t . Damit kann dann der Zustand des Spielers von verschiedenen Zeitpunkten aus dem Spiel heraus vorausgesagt werden. Die unabhängigen Variablen $x_1 \dots x_n$ werden aus den zur Verfügung stehenden Daten ermittelt, die im Teil 2 genauer beschrieben sind. Das Ergebnis der multivariaten Regressionsanalyse ist ein Score-Modell. Dieses Score-Modell enthält einen Anfangskoeffizienten β_0 , die unabhängigen Variablen $x_1 \dots x_n$ mit den gewichteten Koeffizienten $\beta_1 \dots \beta_n$ und beschreibt den Zustand des Spielers zu einem Zeitpunkt t des Spiels.

Formeln des Score-Modells:

Zustandsermittlung für: zum Zeitpunkt t mit n verschiedenen unabhängigen Variablen:

$$y_t = \beta_{t0} + x_{t1} * \beta_{t1} + x_{t2} * \beta_{t2} + \dots x_{tn} * \beta_{tn}$$

Wie bei der Realisierung des Markov-Modells wurde auch für die Regressionsanalyse SAS als Umsetzungswerkzeug genutzt. Gewählt wurde die Prozedur REG für multivariate lineare Regressionen. Die Prozedur erstellt für die abhängige Variable (den Zustand) zum Zeitpunkt t ein Score-Modell. Das Score-Modell, angewendet auf die Daten, gibt die Wahrscheinlichkeit an, dass sich ein Spieler, zum Zeitpunkt t in dem Zustand befindet (siehe Abbildung 5).

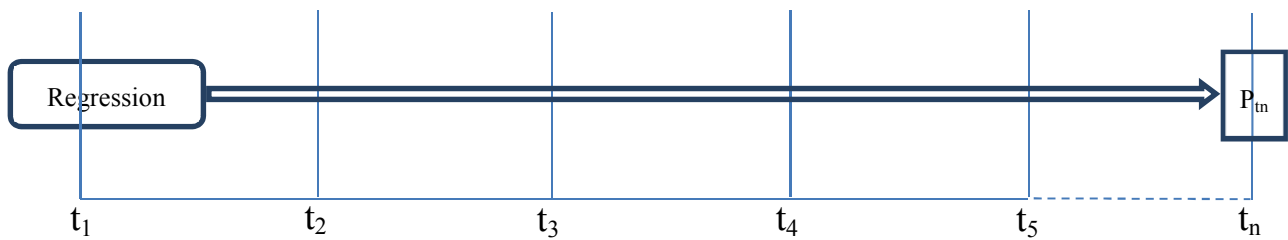


Abbildung 5: Ermittlung der Zustandswahrscheinlichkeit bei der Regressionsanalyse

5 Modellvergleich

Anhand der Testdaten kann mittels einer einfachen Fehlerklassifizierung die Qualität der einzelnen Modelle ermittelt werden. Die Anzahl der richtig erkannten Zustände in Prozent ist der Messwert; dabei wurden die Ergebnisse der einzelnen Zustände jeweils für sich bewertet. Ein besonderes Augenmerk verdienen dabei die Zustände Winner und Finalist. Damit eine grundsätzliche Aussage über die Qualität der Methoden gemacht werden kann, gibt es zu Kontrollzwecken eine Zufallsgruppe.

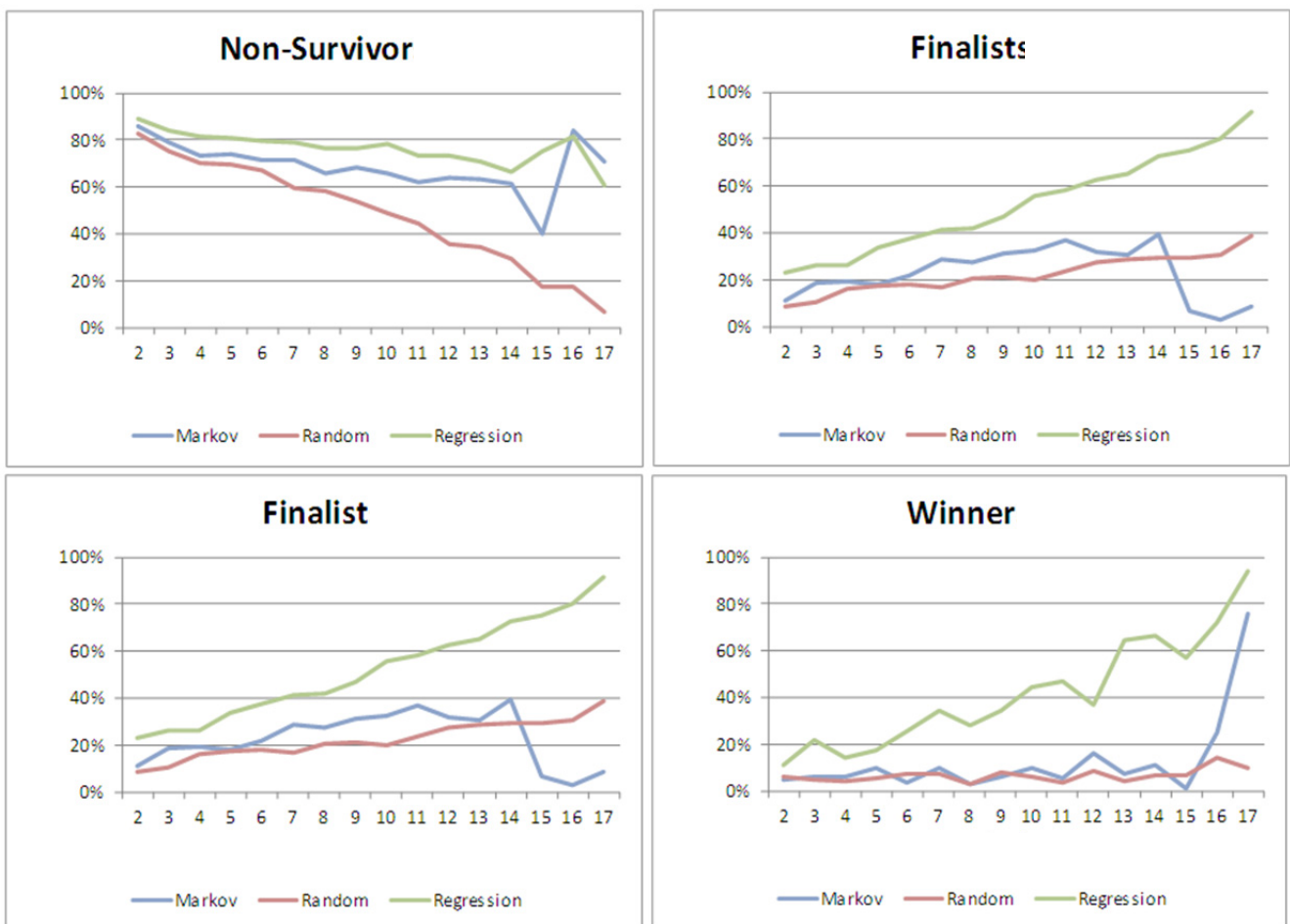


Abbildung 6: Qualitative Methodenbewertung

Es ist deutlich zu erkennen, dass sich Regressionsanalysen besser eignen, um den Erfolg von Spielern vorauszusagen, als Markov-Ketten. Besonders für die entscheidenden Gruppen Finalist und Winner ist das Regressionsmodell über die gesamte Spielzeit hinweg um ein vielfaches besser. Ein entscheidender Grund ist, dass Markov-Ketten hauptsächlich auf den Verteilungen der Zustandsveränderungen basieren und weitere Informationen über die Spieler während des Spiels nicht berücksichtigt werden. Bei Regressionsanalysen werden auch diese Informationen genutzt, um die Modelle für die verschiedenen Zustände und Zeiträume zu erstellen. Für die weiteren Untersuchungen wurden die Ergebnisse der Regressionsanalyse angewendet.

6 Mustererkennung

Die besser geeignete Methode, die Regressionsanalyse, wurde genauer untersucht und genutzt, um die Eigenschaften für erfolgreiche Spieler zu erkennen. Durch die Regressionsanalyse wurde je Woche und Zustand ein Score-Modell erstellt. Dieses Score-Modell errechnet für jeden Spieler eine Affinität für den jeweiligen Zustand. Die errechnete Affinität für den einzelnen Zustand ist relativ und liegt zwischen 0 für eine sehr geringe, bis 1 für eine sehr hohe Affinität und kann auch jeden beliebigen Wert zwischen 0 und 1 annehmen. Dieses Modell wird von verschiedenen Faktoren, wie Ressourcen, Angriffstärke, Anzahl von Spielern in der Allianz oder Rang der Allianz, beeinflusst. Jeder Ausdruck im Modell bekommt einen Faktor zugeordnet, der einer Gewichtung gleichkommt.

Beispiel eines Score-Modells für Winner:

```
score=0,02035
+AVG_NUM_RAUB_ALLY * 0,0000000008
+( AVG_NUM_RAUB_ALLY>438.779.627,571) * -0,50254
+( 45.758,429 < NUM_KILLS_OFF_ALLY<= 122.060,429 ) * 0,1371489116
+( NUM_KILLS_OFF_ALLY>2.934.822) * -0,95585
+( NUM_KILLS_DEF_ALLY>1.308.473) * 0,7693
+( NUM_PLAYER_ALLY>54) * 0.3655849755
+( 53 <NUM_PLAYER_ALLY<= 54) * -0,9117624115
+( 0 <NUM_RANG_ALLY<= 4) * 0,7634821806
+( 25 <NUM_RANG_ALLY<= 31 ) * -0,1028500351
```

Diese Score-Modelle beschreiben, welche Eigenschaften ein Spieler haben muss, um in dem Spiel erfolgreich zu sein. Zur Verdeutlichung wird kurz auf das oben beispielhaft gezeigte Score-Modell eingegangen. Der Ausdruck „(NUM_KILLS_OFF_ALLY > 2.934.822) * -0,95585“ bedeutet, dass ein Spieler, der in der Allianz Angriffspunkte von über 2,9 Millionen hat, -0,95585 Affinitätspunkte bekommt und weniger erfolgreich ist. Ein anderes Beispiel ist „(NUM_KILLS_DEF_ALLY > 1.308.473) * 0,7693“ der ausdrückt, dass Spieler, die Mitglied in einer Allianz sind und die sehr viel Defensivpunkte haben, 0,7693 Punkte bekommen und erfolgreicher sind.

Für das angegebene Beispiel lautet die Beschreibung eines erfolgreichen Spielers wie folgt:

- Die Allianz hat eine hohe positive Raubbilanz, d.h. sie raubt mehr als sie beraubt wird.
- Die Allianz hat keine maximale Raubbilanz, d.h. sie konzentriert sich nicht ausschließlich auf das Ausrauben anderer Allianzen.
- Die Allianz hat durchschnittliche Angriffspunkte, aber beschränkt sich nicht nur auf den Angriff.
- Die Allianz hat sehr hohe Verteidigungspunkte.
- Die Allianz hat zwischen 54 und 60 Mitgliedern
- Die Allianz hat einen sehr hohen Rang, besser als 5.
- Individuelle Attribute des Spielers sind nicht mehr bedeutend. Es ist in diesem Score-Modell nur noch entscheidend, wie gut die Allianz ist, in der sich der Spieler befindet.

Diese Beschreibung wurde für jede einzelne Woche vorgenommen, um zu erkennen, ob sich die Eigenschaften eines erfolgreichen Spielers über die Spielzeitpunkte hinweg verändern. Dabei wurden drei verschiedene Phasen entdeckt (siehe Abbildung 7):

- Start: Die einzelnen Spieler spielen für sich alleine. Sie verschaffen sich eine gute Position in Ihrer Umgebung. Erste Allianzen haben noch keine starke Bedeutung.
- Aufbau: Diese Phase ist die längste Phase. Die Spieler bilden Allianzen, koordinieren mit anderen Spielern und die individuellen Eigenschaften werden weniger wichtig.
- Finale: Nur starke Allianzen erreichen das finale Ziel, individuelle Attribute sind nicht mehr relevant.

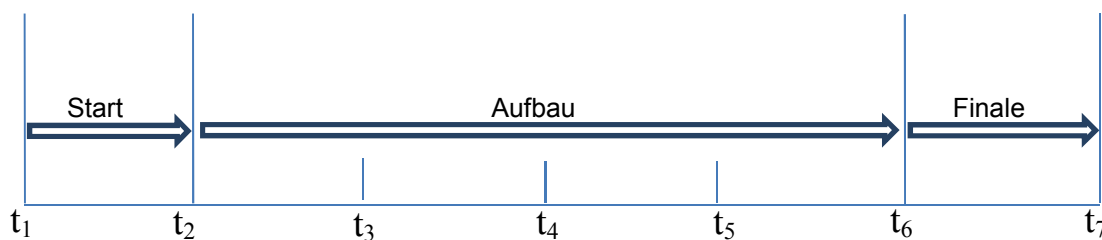


Abbildung 7: Spielphasen

Start:

- Spieler spielen aggressiv, um schnell zu wachsen.
- Spieler nutzen Gold und sind sehr aktiv.
- Die Allianzen haben eine durchschnittliche Größe.
- Die Allianzen spielen aggressiv und müssen sich nicht verteidigen.

Aufbau:

- Spieler nutzen Gold.
- Die Spieler sind stark und haben einen guten Rang.
- Die Allianz spielt aggressiv.
- Die Allianz hat durchschnittliche Defensivpunkte.
- Die Allianz hat viele Mitglieder.

Finale:

- Individuelle Attribute des Spielers sind nicht mehr bedeutend.
- Die Allianz spielt weniger aggressiv.
- Die Allianz ist sehr groß und hat viele Mitglieder.
- Die Allianz hat einen sehr guten Rang.

7 Fazit

Aus dem Vergleich der beiden Data Mining Methoden ergibt sich, dass Regressionsanalysen besser geeignet sind als Markov Ketten, um den Erfolg in MMORPG vorzusagen. Wie man erkennen kann, ist die Frage, wie ein Spieler vorgehen muss, um erfolgreich zu sein, nicht mit einer einzelnen Formel oder Strategie zu beantworten. Was man aber aus den Ergebnissen dieses Projektes schließen kann, ist dass es wichtig ist zum richtigen Zeitpunkt im Spiel die richtige Ausrichtung zu haben. Am Anfang ist es noch wichtig, dass der Spieler als Individuum schnell wächst, aber im Spielverlauf ist es immer bedeutender zu kooperieren und in einer starken Allianz zu sein.

Zukünftig kann man die Datenbasis erweitern, um noch genauere und stabilere Ergebnisse zu bekommen. Zur Erweiterung kann man Daten aus mehreren Spielen nutzen oder die Datenbasis erweitern, indem man weitere Datentransformationen vornimmt. Eine weitere Verbesserung kann durch eine Veränderung der Bewertung erreicht werden. Anstatt einer einfachen Fehlerklassifizierung, kann eine kostensensitive Klassifizierung genutzt werden. Mit dieser kann berücksichtigt werden, dass es „mehr falsch“ ist, wenn man einen Winner fehlerhafterweise als non-Survivor klassifiziert hat, anstatt diesen als Finalisten zu kennzeichnen.

Schlussendlich lässt sich sagen, dass es möglich ist, den Erfolg in virtuellen Organisationen, am Beispiel von MMORPG, vorauszusagen. Man kann annehmen, dass man auch in anderen Themenbereichen, die ähnliche Eigenschaften wie das MMOPRG Travian haben, die Methodik der Regressionsanalyse nutzen kann, um Erfolge zu prognostizieren. Diese Themengebiete sollten analog zu MMORPG komplex, endlich, an Ressourcen begrenzt und kooperationsbedingt dynamisch sein. Denkbare wären z.B. der Erfolg von Studenten im Studium, oder das Gelingen von Projekten in der Wirtschaft. Eine unerlässliche Voraussetzung dafür ist natürlich eine ausreichende und aussagekräftige Datenbasis.

Literatur

- [1] W. S. Bainbridge, "The Scientific Research Potential of Virtual Worlds," *Science*, pp. 472-476, 27 Juli 2007.
- [2] R. T. Wigand, N. Agarwal, I. Osesina, W. Hering, M. A. Korsgaard, A. Picot and M. Drescher, "Social Network Indices as Performance Predictors in a Virtual Organization.," in *Proceedings of the Fourth International Conference on Computational Aspects of Social Networks*, Sao Carlos, Brazil, 2012.
- [3] D. Williams, N. Ducheneaut, L. Xiong, Y. Zhang, N. Yee and E. Nickell, "From Tree House to Barracks; The Social Life of Guilds in World of Warcraft," *Games and Culture*, pp. 338-361, Oktober 2006.
- [4] J. Han, M. Kamber and J. Pei, *Data Mining: Concepts and Techniques*, Elsevier Verlag, 2011.
- [5] D. Meintrup and S. Schäffler, *Stochastik: Theorie und Anwendungen*, Springer Verlag, 2004.
- [6] N. R. Draper and H. Smith, *Applied Regression Analysis*, New York: Wiley, 1998.