Hans Pongratz, Reinhard Keil (Hrsg.): Die 13. E-Learning Fachtagung Informatik,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015 307

Analyse erfolgreicher Studenten in Massive Open Online Courses

Marcus Klüsener[1] und Albrecht Fortenbacher[2]

Abstract: Massive Open Online Courses (MOOCs) besitzen das Potenzial, die Hochschulbildung zu skalieren und für viele Teilnehmer zugänglich zu machen. Plattformen wie Coursera, edX oder auch Iversity sind in diesem Bereich sehr erfolgreich. Trotz der unbestreitbaren Erfolge bleibt die niedrige Completion Rate in vielen MOOCs ein Problem. Ziel dieser Arbeit ist es erfolgreiche Studenten zu identifizieren und durch Bewertung ihrer Merkmale, den Dozenten handlungsweisende Informationen zur Verfügung zu stellen. Dazu wird untersucht, wie solche Informationen aus den Lernaktivitäten erfolgreicher Studenten abgeleitet werden können. Die Merkmale erfolgreicher Studenten werden zu einem Profil verbunden und können als Grundlage für Empfehlungen an ‚Risikostudenten' verwendet werden, um deren Chancen zu erhöhen, einen MOOC erfolgreich abzuschließen. Dazu wurde ein Analyse-Tool entwickelt, das Merkmale von Studenten aus großen MOOC-Foren von Iversity bestimmt, mit Methoden des maschinellen Lernens analysiert und auf eine intuitive Weise visualisiert.

Keywords: Completion Rate, Lernprofile, Predictive Analytics, Klassifikation, Forenanalyse.

1 Einleitung

In den letzten Jahren hat das Interesse an Massive Open Online Courses (MOOC) stark zugenommen. Mit Videolektionen, begleitenden Texten und Prüfungen bieten sie viele interessante Lernerfahrungen für Studenten. Aktuelle Studien konnten belegen, dass das Vorwissen der entscheidende Faktor für erfolgreiches Abschneiden in MOOCs ist [Ke15]. In dieser Arbeit wurden bewusst nur Indikatoren für den Studienerfolg aus den Diskussionsforen abgeleitet. Diese ermöglichen den Dozenten noch während des laufenden Kurses Einfluss auf die Studenten zu nehmen mit dem Ziel die Completion Rate zu verbessern.

Bei den in MOOCs beobachteten, sehr geringen Completion Rates – von zum Teil unter 2% – entsteht eine starke Asymmetrie zwischen erfolgreichen und nicht erfolgreichen Studenten. Eine Möglichkeit, diese Asymmetrie auszugleichen, ist es, einen Trade-off zwischen Accuracy und Recall mit der Receiver Operator Characteristic (ROC) zu finden. Eine weitere Möglichkeit besteht darin, nur diejenigen Studenten zu betrachten, die eine für die Fragestellung relevante Zielgruppe darstellen [He15], im konkreten Fall also nur die im Kurs aktiven Studenten. Dadurch können bis zu 50% aller Studenten von der Betrachtung ausgeschlossen werden, was wiederum die Completion Rate erhöht.

[1] HTW-Berlin, Treskowallee 8, 10318 Berlin, m.kluesener@htw-berlin.de
[2] HTW-Berlin, Treskowallee 8, 10318 Berlin, forte@htw-berlin.de

2 Analysen der Studienleistung

2.1 Datenbasis

Als Datenbasis dieser Analyse dienen drei MOOCs, die in den Jahren 2013 und 2014 auf Iversity.org durchgeführt wurden. Insgesamt waren in diesen MOOCs 106327 Studenten eingeschrieben und es liegen 7826320 Aktivitätsdaten und 21825 Forenbeiträge vor. Aus der Datenbasis wurden 14 Merkmale direkt extrahiert, z. B. die Anzahl der Upvotes oder der Downvotes, und durch Analysen berechnet, z. B. die Anzahl der Bilder, und zu einem Lernprofil zusammengefügt. Die Studienleistung wird in diesen drei MOOCs anhand des Fortschritts im Kurs bewertet. Zum erfolgreichen Bestehen des Kurses müssen mindestens 80% der Videolektionen angesehen worden sein.

2.2 Klassifikation

Es wurden die im Bereich Educational Data Mining erfolgreich eingesetzten Algorithmen wie Entscheidungsbäume, Random Forest, Decision Rules, Step Regression und Logistische Regression angewendet und mit 10-fach stratifizierter Kreuzvalidierung validiert. Als relevante Zielgruppe im Sinne von [HE15] wurden diejenigen Studierenden betrachtet, die mindestens einmal mit dem Kurs interagiert haben, z. B. durch Betrachtung eines Videos. Es wurde ein Algorithmus ausgewählt, der hinsichtlich Interpretierbarkeit, Accuracy und Recall der erfolgreichen Studenten die besten Ergebnisse aufweist.

2.3 Explorative Analyse

Die explorative Analyse wurde mit dem Open-Source-Tool LEMO durchgeführt. Im Rahmen einer Masterarbeit [Kl15] wurden Scatterplots implementiert. Sie gestatten Annahmen über die Ursachen der beobachteten Daten zu bilden und eine Basis für weitere Analysen bereitzustellen. Ferner ermöglichen sie eine interaktive Untersuchung der Daten sowie den Vergleich mit den Klassifikationsergebnissen. Den Achsen des Scatterplots können dazu die Merkmale der Studenten frei wählbar zugewiesen werden.

3 Ergebnisse

Als bester Klassifikator konnte die Logistische Regression alle drei MOOCs klassifizieren. Die Abb. 1 zeigt die maximale Accuracy der Klassifikation und den dazugehörigen Recall der beiden Klassen. Um den Recall-Wert der erfolgreichen Studenten zu optimieren wurde der Schwellwert des Klassifikators mit der ROC angepasst. So konnte der Recall der erfolgreichen Studenten in allen untersuchten MOOCs auf mindestens 50% erhöht werden, während die Accuracy über 80% blieb.

	Maximale Accuracy			Trade-off zwischen Accuracy und Recall	
	Accuracy	Recall nicht erfolgr. Studenten	Recall erfolgr. Studenten	Accuracy	Recall erfolgr. Studenten
Mathe-MOOC1	97,90	0,999	0,080	84,54	0,66
Storytelling-MOOC	91,39	0,995	0,132	83,38	0,50
Mathe-MOOC2	97,68	0,998	0,083	80,12	0,52

Abb. 1: Ergebnis der Logistischen Regression mit der Accuracy und Recall der erfolgreichen Studenten in den untersuchten MOOCs

Die Odds-Ratios zeigen, dass die Anzahl der gegebenen Upvotes besonders stark mit erfolgreichem Abschneiden korreliert. Weiterhin auffällig ist die Relevanz der Anzahl der gegebenen Antworten. Im LEMO-Tool kann visuell nachvollzogen werden, dass z. B. im Mathe-MOOC 2 die Anzahl der Antworten bei den Studenten, die nur bis zur Mitte des Kurses aktiv waren, am höchsten ist. Die Anzahl der geschriebenen Wörter hingegen steigt kontinuierlich bis zum Ende des Kurses an.

4 Diskussion und Schlussfolgerung

Die Anpassung der Logistischen Regression mit der ROC konnte die asymmetrische Verteilung der Studenten ausgleichen und mindestens 50% der erfolgreichen Studenten mit über 80% Accuracy in jedem der drei MOOCs finden. Als entscheidendes und kursübergreifendes Merkmal der erfolgreichen Studenten ergibt sich die Anzahl der gegeben Upvotes. Die große Relevanz der Upvotes deutet darauf hin, dass das Lesen und Bewerten von Beiträgen anderer Studenten entscheidend für den Studienerfolg ist.

Die hier vorgelegten Ergebnisse wurden ohne Kenntnis des jeweiligen didaktischen Konzepts erzielt. Die Scatterplot-Darstellungen der Klassifikationsergebnisse sind besonders für Dozenten und Tutoren ein mächtiges Werkzeug, da sie auf eingängige Art Paare von Merkmalen mit den Klassifikationsergebnissen korrelieren und eine explorative Analyse des Lernverhaltens unterstützen.

Literaturverzeichnis

[He15] He, Jiazhen; Bailey, James; Rubinstein, Benjamin IP; Zhang, Rui: Identifying At-Risk Students in Massive Open Online Courses. 2015.

[Ke15] Kennedy, Gregor; Coffrin ,Carleton; de Barba, Paula; Corrin, Linda: Predicting success: how learners' prior knowledge, skills and activities predict MOOC performance. In Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (LAK '15). ACM, New York, NY, USA, 136-140, 2015.

[Kl15] Klüsener, Marcus: Vorhersage der Studienleistung durch Forenanalyse und Klassifikationsverfahren im Learning-Analytics-Tool LEMO, Masterarbeit Hochschule für Technik und Wirtschaft Berlin, 2015.