

Modulempfehlungen über Vorgänger- und Nachfolgemodule

Patrick Bittner¹

Abstract: Das moderne Studium stellt die Studierenden vor inhaltliche, organisatorische und technische Herausforderungen. Nicht jeder Studiengang hat einen klaren Ablauf und selbst wenn ein solcher vorhanden ist, ist er oft auf einen Studierenden ausgelegt, der kein Modul wiederholen muss. Außerdem können meist nicht alle Lebensentwürfe und –umstände berücksichtigt werden. In diesem Beitrag wird ein Konzept für ein Empfehlungssystem vorgestellt, welches aufgrund von Studienverläufen Abhängigkeiten zwischen den Modulen identifiziert. Hierbei werden sowohl die Voraussetzungen für die Module als auch mögliche Präferenzen der Studierenden bei einer Empfehlung berücksichtigt.

Keywords: Empfehlungssysteme, Studienplan-Optimierung, recommender system, Modulauswahl

1 Einleitung

An den Universitäten ist IT-Unterstützung heute nicht mehr wegzudenken. Den Studierenden werden unterschiedliche Dienste angeboten, wie beispielsweise Online-Bewerbungen oder E-Mail-Accounts.

Viele Studierende bringen bereits zum Studienanfang ihre eigene IT-Umgebung mit und verknüpfen die universitären Dienste mit ihrer eigenen Infrastruktur. Diese Entwicklung wird von den Universitäten gefördert, um den Studierenden ein größtmögliches Maß an Komfort zu bieten.

Trotzdem ist der Verwaltungsaufwand während eines Studiums groß. Unterschiedliche Studienordnungen schaffen Regeln, an die sich die Studierenden der unterschiedlichen Fachrichtungen zu halten haben. Module müssen gesichtet und einzelne Vorlesungen ausgewählt werden, um die nötigen Anforderungen letztlich erfüllen zu können. Dabei ist es nicht immer einfach, die genauen Anforderungen der Prüfungsordnung oder auch nur eines bestimmten Moduls zu erfassen. Selbst Studiengänge, die einen relativ klaren Studienverlaufsplan vorschlagen, können nicht weiterhelfen, wenn z.B. eine Prüfung nicht bestanden wurde oder nur ein Teil der Prüfungen angetreten werden konnte. Bisher existieren wenige Systeme, die es den Studierenden leichter machen, herauszufinden, welche Module von ihnen noch zu belegen sind, und welche Voraussetzungen dafür mitzubringen sind. So müssen sich die meisten Studierenden selbst informieren, was ihnen überhaupt noch zum Abschluss ihres Studiums fehlt.

¹ Technische Universität Berlin, tubIT – IT-Service-Center der TU-Berlin, Einsteinufer 17, 10587 Berlin, patrick.bittner@tu-berlin.de

Für die Entwicklung eines Systems, welches die Studierenden an dieser Stelle unterstützt, müssen die Rahmenbedingungen in Form von Studienordnungen und Modulhandbüchern bekannt sein. Um allerdings tatsächlich auf einzelne Studierende abgestimmte Empfehlungen geben zu können, muss dem System bekannt sein, welche Vorlieben der Studierende hat, und welche der möglichen Module diesen Vorlieben am ehesten entgegenkommen. Dabei können beispielsweise die Studienverläufe anderer Studierender herangezogen werden. Je ähnlicher sich die Studierenden sind, desto wahrscheinlicher erscheinen ähnliche Modulvorlieben.

In dieser Arbeit werden wir zunächst bisherige Unterstützungsansätze für Studierende vorstellen, sowie gängige Empfehlungsansätze und Probleme bei solchen benennen. Anschließend stellen wir das Konzept für ein Empfehlungssystem für Module vor, welches auf den bisherigen Belegungshistorien der Studierenden basiert, und über Belegungszusammenhänge Empfehlungen generiert. Abschließend geben wir einen Ausblick über mögliche, weitere Ausbaumöglichkeiten dieses Systems.

2 Verwandte Arbeiten

Es existieren bereits einige Systeme, die den Studierenden bei der Organisation ihres Studiums helfen sollen. Im Folgenden werden beispielhaft zwei derartige Systeme genannt, die sich bereits im Einsatz befinden.

[FH09] beschreibt die Einführung eines Campus Management Systems in der Universität Hamburg. Das System beinhaltet auch die Optimierung der Modulplanung, um zeitliche Überschneidungen zu minimieren. Auf diese Weise können Studierende Module, welche für das gleiche Semester geplant sind, auch im gleichen Semester belegen. Dabei liegt der Schwerpunkt aber auf den Zeit- und Raumbelegungen. Eine Information an die Studierenden, welche Module für sie empfehlenswert sein könnten, erfolgt auf diese Weise jedoch nicht.

In [Ri09] wird ein im System integriertes „Self Monitoring“ angesprochen. Dieses ermöglicht es den Studierenden, sich selbst einzuschätzen, indem sie ihre Noten im Vergleich zu anderen Studierenden einsehen können. Ein zusätzliches „Self Assessment“ dient zur besseren Einschätzung der eigenen Stärken und Schwächen und berücksichtigt dabei den bisherigen Studienverlauf. Die Daten werden nicht dazu verwendet, konkrete Modulempfehlungen zu generieren. Unser Ansatz hingegen hat genau dieses Ziel, verzichtet aber bisher auf die Bereitstellung eines Tools zur Selbsteinschätzung.

2.1 Verbreitete „Recommender“-Ansätze

Um Empfehlungen zu generieren gibt es diverse unterschiedliche Ansätze, die seit langer Zeit Thema der Forschung sind. In [GT05], [He04], [Bu02] und [SK09] sind einige davon beschrieben.

Die meisten Ansätze für Empfehlungssysteme lassen sich in eine der folgenden Kategorien einordnen:

- *Demographische Ansätze* versuchen, über die statischen Eigenschaften der User Ähnlichkeiten zu finden. Dabei wird angenommen, dass User mit ähnlichen Eigenschaften ähnliche Präferenzen haben, woraus Empfehlungen generiert werden können.
- Bei *Inhalts- (Content-) basierten Ansätzen* werden Ähnlichkeiten zwischen den Eigenschaften der Items gesucht. Die Annahme besteht hierbei, dass User Items der gleichen „Art“ bevorzugen.
- *Wissens- (Knowledge-) basierte Ansätze* verwenden ebenfalls die Eigenschaften der Items, besitzen aber zusätzlich das Wissen darum, was der User konkret braucht. Dieses Wissen kann z.B. über konkrete Abfragen gewonnen werden.
- *Collaborative Ansätze* gehen davon aus, dass User, die bisher ähnliche Präferenzen hatten, also ähnliche Bewertungen abgegeben haben, auch weiterhin die gleichen Items bevorzugen.
- *Modellbasierte Ansätze* versuchen, über Trainingsdaten Muster im Wahlverhalten zu erkennen und auf diese Weise Empfehlungen zu generieren. Dabei werden verschiedene Techniken verwendet, wie beispielsweise Matrix-Faktorisierung oder Bayessche Netze.

Je nach Ansatz ist ein Empfehlungssystem mit unterschiedlichen Herausforderungen konfrontiert. Es gibt einige gängige Probleme im Zusammenhang mit Empfehlungssystemen:

- Das *New-User Problem* bezeichnet die Problematik beim Anlegen eines neuen Users. Da ein neuer Benutzer anfangs noch keine oder nur sehr wenige Bewertungen abgegeben hat, können noch keine Präferenzen identifiziert werden. Erst wenn die Vorlieben des Users besser bekannt sind, können für ihn passende Items empfohlen werden.
- Als *New-Item Problem* bezeichnet man die analoge Problematik beim Anlegen eines neuen Items. Hier können zwar Item-Eigenschaften beim Erstellen bekannt sein, aber auch hier gibt es anfangs keine oder nur wenige Bewertungen. Dadurch können diese neuen Items in den Ansätzen, die Präferenzen berücksichtigen, nicht als Empfehlungen identifiziert werden, bis genug Daten vorhanden sind.
- Das *Cold-Start Problem* ist hierbei noch allgemeiner. Es bezeichnet den Umstand, dass ein System erstellt wird, aber keinerlei Bewertungen, also Verknüpfungen zwischen Usern und Items, vorhanden sind. Neu erstellte

Systeme ohne vorhandene Bewertungen können keinerlei Präferenzen identifizieren. Da jegliche Präferenzen in diesem Fall unbekannt sind, ist eine Generierung von Empfehlungen durch viele Ansätze so nicht möglich, bis eine Datenbasis geschaffen wurde.

Eine gängige Praxis, um den genannten Problemen entgegenzuwirken, ist die Hybridisierung der Ansätze. Hier wird versucht, die Ansätze zu kombinieren. Dadurch kann beispielsweise der Effekt fehlender Daten minimiert werden, um auf diese Weise trotzdem eine möglichst adäquate Empfehlung abgeben zu können. In [Bu02] werden einige Hybridisierungsansätze untersucht und verglichen.

3 Empfehlungsansatz

An den Universitäten wird eine Vielzahl an unterschiedlichen Modulen angeboten, welche sich wieder in verschiedene Veranstaltungen aufteilen können. Allerdings ist nicht jedes Modul für jeden Studiengang sinnvoll oder auch nur verwendbar. Eine erste Vorauswahl der Module ist also bereits durch eine Berücksichtigung der Studienordnung möglich. Das heißt, dass für einen Studierenden nur Module vorgeschlagen werden können, wenn diese ihn entsprechend seiner Studienordnung überhaupt weiterbringen können.

Diese Einschränkung ist jedoch sehr allgemein und bringt dem Studierenden kaum einen Mehrwert, da er sich dessen im Allgemeinen bereits bewusst ist.

Um eine auf einen einzelnen Studierenden abgestimmte Empfehlung geben zu können, muss zuerst klar sein, was über den Studierenden und über die Module überhaupt bekannt ist.

Im Folgenden stellen wir einen eigenen Ansatz vor und betrachten die Stärken und Schwächen auch im Vergleich zu anderen Ansätzen.

3.1 Modulhistorien der Studierenden als Input für das Empfehlungssystem

Um sein Studium abzuschließen, muss ein Studierender Module wählen und belegen. Daher ist ein Empfehlungssystem, welches im universitären Umfeld Module vorschlägt, nicht mit einem System zu vergleichen, welches zum Beispiel Kaufempfehlungen ausspricht. Der Zwang, tatsächlich etwas wählen zu müssen, sowie der Umstand, dass sämtliche bereits gewählten Module bekannt sind, unterscheiden diese spezielle Domäne von vielen anderen.

Die bisherigen Modulbelegungen der Studierenden sind als Datenbestand vorhanden. Über diese Belegungsdaten lässt sich eine Art Historie für jeden einzelnen Studierenden erzeugen, also eine Aufstellung, in welchem Semester ein Studierender welches Modul belegt hat. Diese Historien können als Input verwendet werden, um Empfehlungen zu

generieren. Ein Studierender im ersten Semester würde hier mit einer leeren Historie beginnen.

3.2 Vorgängermodule

Unser Ansatz geht davon aus, dass die Reihenfolge, in der Module gewählt werden, von den Präferenzen der Studierenden und deren Wissensstand abhängig ist. Ein bestimmtes Modul wird gewählt, wenn es durch die Studienordnung vorgegeben wird oder der Studierende ein Interesse für das Modul aufweist. Der Zeitpunkt der Belegung ist davon abhängig, ob der Studierende bereits alle nötigen Vorkenntnisse hat. In Modulbeschreibungen sind oftmals bereits solche Anforderungen an die Studierenden hinterlegt, aber nicht immer lässt sich damit das gesamte Spektrum abdecken.

In diesem Ansatz analysieren wir die Studierendendaten, um herauszufinden, welche Module als Vorgänger für welche anderen Module angesehen werden können. Dabei wird für jedes Modul festgehalten, wie viele Studierende dieses Modul gewählt haben, und welche Module von wie vielen Studierenden gewählt wurden, bevor das untersuchte Modul gewählt wurde. Über den Anteil der Studierenden, die ein bestimmtes Modul vorher gewählt haben, lässt sich das Ausmaß der Verbindung zwischen den Modulen abschätzen. Ein Modul, welches von einem großen Teil der Studierenden vor einem anderen gewählt wurde, kann als potentiell Vorgängermodul angesehen werden. Dazu wird ein Minimalanteil festgesetzt. Übersteigt der Anteil der Studierenden, welche ein Modul vor dem untersuchten Modul besucht haben, diesen Minimalanteil, so gilt das Modul als Vorgängermodul für das untersuchte Modul.

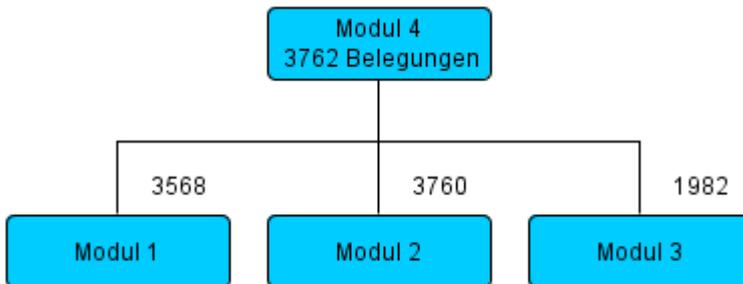


Abbildung 1: beispielhafte Darstellung des Moduls 4 mit seinen Vorgängermodulen

Wir gehen davon aus, dass man, wenn man sämtliche Vorgängermodule belegt hat, potentiell auch an dem eigentlichen Modul interessiert ist und auch die nötigen Kenntnisse erworben hat, um dieses Modul zu belegen.

Im Beispiel aus Abbildung 1 wird das Modul 4 von 3.762 Studierenden belegt. Bei der Betrachtung des Moduls wird festgestellt, dass laut den Historien der Studierenden, die

das Modul belegt haben, drei Module vor dem Modul 4 gewählt wurden. Diese drei Module sind potentielle Vorgängermodule. Während fast 100% (3760 von 3762) der Studierenden, die das Modul gewählt haben, vorher das Modul 2 gewählt haben, war der Anteil derer, die vorher das Modul 3 wählten, nur bei rund 55%. Das Modul 1 wurde von ca. 95% der Studierenden vor Belegung des Moduls 4 besucht. Je nach Parametrisierung, also Wahl des Minimalanteils, können die Module 1-3 nun als Vorgängermodul für das Modul 4 identifiziert werden.

3.3 Kombinationsvorgänger

Die vorgestellte Vorgängerbeziehung stellt einen direkten Zusammenhang zwischen zwei Modulen her. Diese zeigt an, dass ein Großteil der Studierenden vor dem untersuchten Modul das Vorgängermodul belegt hat. Es erscheint daher sinnvoll, Module nur zu empfehlen, wenn vorab sämtliche Vorgängermodule belegt wurden.

Allerdings können auch Module, die den in 3.2 angesprochenen Minimalanteil an Studierenden nicht erreichen (und somit keine Vorgängermodule sind) entscheidend dafür sein, ob ein Modul empfohlen werden sollte. Dies wird z.B. deutlich, wenn in mehreren Modulen ähnliche Grundlagen vermittelt werden. Wenn nun ein Modul auf diesen Grundlagen aufbaut, wird erwartet, dass mindestens eines der Grundlagenmodule gewählt wurde. Diese tauchen aber ggf. nicht als Vorgängermodul auf, da der Anteil an Studierenden, die das Grundlagenmodul vor dem untersuchten Modul belegt haben, ggf. recht klein ist.

An dieser Stelle kann eine weitere Analyse helfen, die feststellt, ob ein kombinierter Zusammenhang aus mehreren Modulen besteht. Darüber lassen sich Aussagen der Art „es ist sinnvoll, mindestens ‚X‘ dieser Module vor Antritt des Moduls belegt zu haben“ treffen, die auch zur Einschränkung der Empfehlungen verwendet werden können.

3.4 Nachfolgemodule

Analog zur Analyse der Module auf Vorgänger ist auch eine Überprüfung auf Nachfolger denkbar. Diese wäre analog zur Vorgänger-Analyse aufzubauen.

Die Aussagen von Nachfolger- und Vorgängerbeziehungen sind hierbei nicht als redundant zu betrachten. Während eine Vorgängerbeziehung Auskunft darüber gibt, dass ein Großteil der Studierenden das Vorgängermodul bereits belegt hat, sagt sie nichts darüber aus, wie viele der Studierenden, welche das Vorgängermodul besucht haben, sich anschließend für das betrachtete Modul entschieden haben. Diese Lücke wird durch die Analyse auf Nachfolge geschlossen. Dabei ist zu erwarten, dass die Nachfolge-Analyse tendenziell die Präferenzen der Studierenden aufdeckt, während die Vorgänger-Analyse sich auch für die Aufdeckung von Voraussetzungen eignet.

Eine Kombination aus Vorgänger- und Nachfolgeanalyse erhält somit ein Potential,

fundierte Empfehlungen generieren zu können.

3.5 Vor- und Nachteile

Der von uns vorgestellte Empfehlungsansatz arbeitet mit dem Umstand, dass die Modulwahl als Sequenz aufgebaut ist, also eine Reihenfolge besitzt. Kein Modul kann mehrmals (erfolgreich) belegt werden und im Regelfall wird keine explizite Bewertung vorgenommen. Die angesprochene Reihenfolge wird von unserem Ansatz aufgegriffen, um daraus direkt Empfehlungen zu generieren. Auf diese Weise können fachliche und interessenmäßige Zusammenhänge erkannt und berücksichtigt werden. Außerdem bietet sich die Möglichkeit, die Module zu analysieren, wodurch die Lehrkräfte in die Lage versetzt werden, die von ihnen angebotenen Module noch besser im Gesamtkontext einschätzen zu können.

Der Ansatz baut auf den bisherigen Modulbelegungen auf. Sind diese nicht vorhanden, können keine Vorgänger oder Nachfolger-Beziehungen zwischen den Modulen identifiziert werden. Der Ansatz ist also anfällig für das in 2.1 beschriebene Cold-Start Problem.

Das New-User Problem betrifft diesen Ansatz nicht. Ein neuer Studierender ist ein User mit leerer Historie. Diesem User können direkt Module vorgeschlagen werden, welche keine Vorgänger besitzen oder der Nachfolger von keinem Modul sind. Dies sind im Wesentlichen die Module, welche im ersten Semester gewählt werden, und somit auch tendenziell für neue Studierende geeignet sind.

Bei Verwendung unseres Ansatzes in Reinform wird ein neu erstelltes Modul vom System nicht empfohlen werden. Dies begründet sich dadurch, dass das Modul in diesem Fall noch von niemandem gewählt wurde, und somit kein Vorgänger oder Nachfolger ist.

Wir gehen davon aus, dass in einer Hochschule die Studienverlaufspläne der einzelnen Studierenden vorliegen und als Daten verwendet werden können. Somit ist das Cold-Start Problem für den vorgestellten Ansatz nur von theoretischer Bedeutung: Unter Berücksichtigung des Datenschutzes sind die bisherigen Studienverläufe aller Studierenden die zugrundeliegende Datenbasis.

Das New-Item Problem ist hier problematischer. An dieser Stelle könnte zum Beispiel über händische Einträge oder eine Hybridisierung dem Effekt entgegengewirkt werden.

3.6 Hybridisierungsmöglichkeiten

In [Bu02] wurde dargelegt, dass eine kombinierte Lösung aus mehreren Ansätzen potentiell die Schwächen der einzelnen Ansätze beseitigen kann.

Der von uns dargestellte Ansatz ist in der Reinform zwar verwendbar, hilft aber vor

allem bei Modulen, welche nur von einzelnen Fachrichtungen belegt werden. Und auch dann nur, wenn bereits einige entsprechende Vorgängermodule bekannt sind, sodass das System auf diesen Modulen aufbauen kann.

Um eine präzisere Empfehlung generieren zu können, bietet sich eine Kombination mit anderen Ansätzen an. Im Folgenden werden einige beispielhafte Hybridisierungen kurz skizziert.

Bereits am Anfang dieses Kapitels wird erwähnt, dass es sinnvoll ist, nur die Module vorzuschlagen, welche laut der Studienordnung des Studierenden möglich sind. Diese Einschränkung ist ein Beispiel für die Kombination mit einem Knowledge-basierten Ansatz, da hier das Wissen um die Bedürfnisse des Studierenden herangezogen wird. Auch eine Berücksichtigung der Leistungspunkte, die ein Modul bringt bzw. ein Studierender in einem bestimmten Bereich noch benötigt, fällt in diesen Bereich.

Ein anderes Beispiel wäre die Hybridisierung mit dem demographischen Ansatz. So könnte beispielsweise bereits vor der Analyse auf Vorgänger oder Nachfolger die Menge der betrachteten Studierenden eingeschränkt werden. Verwendet man nur die Daten der Studierenden, die beispielsweise den gleichen Studiengang wie der User belegen und sich innerhalb der letzten 6 Jahre eingeschrieben haben, erhält man einen deutlich spezifischeren Überblick der Vorgänger und Nachfolger. Nimmt man andere Kriterien als „Studiengang“ und „Einschreibungszeitpunkt“, lassen sich so diverse Personengruppen eingrenzen, sodass der Benutzer ggf. eine auf ihn abgestimmte Teilgruppe aller Studierenden als Referenzgruppe erhalten würde. Dabei ist als Kriterium alles denkbar, was dem System über die User bekannt ist.

Um dem New-Item Problem entgegenzuwirken wäre eine Kombination mit einem gegen dieses Problem resistenten Ansatz, wie dem Content-basierten Ansatz denkbar. Hierbei wird die Ähnlichkeit der Items unter Berücksichtigung vorher definierter Eigenschaften der Items verwendet, um eine Empfehlung auszusprechen. Werden beide Ansätze kombiniert, erhält das Item dadurch eine Möglichkeit, empfohlen zu werden, und so selbst Vorgänger- und Nachfolgerbeziehungen aufzubauen.

4 Ausblick

Das in diesem Beitrag vorgestellte Konzept ist in dieser Form noch nicht implementiert. Eine Umsetzung wäre vor allem in Hinblick auf eine Evaluierung der Präzision des Ansatzes sinnvoll.

Auch eine genauere Betrachtung der möglichen Hybridisierungsmöglichkeiten ist wünschenswert. Dabei müsste evaluiert werden, welche Hybrid-Ansätze potentiell effektiv sind, also möglichst zutreffende Empfehlungen generieren. Anschließend müsste der genaue Effekt der Hybrid-Ansätze auf die Qualität der Empfehlungen verglichen und analysiert werden. Auch eine Kombination mehrerer Hybridansätze ist

dabei denkbar.

Außerdem ist die Entwicklung eines Sandbox-Systems denkbar. Hier könnte ein Studierender sehen, welche Auswirkungen eine Modulauswahl auf die weiteren Empfehlungen durch das System hat.

Literaturverzeichnis

- [FH09] Fischer, H.; Hartau, C.: STiNE an der Universität Hamburg – Zur Einführung eines integrierten Campus-Management-Systems. 9. Internationale Tagung Wirtschaftsinformatik
- [Ri09] Rieger, B.; Haarmann, T.; Höckmann, E.; Lüttecke, S.: Data Warehouse basierte Entscheidungsunterstützung für das Campus-Management von Hochschulen. 9. Internationale Tagung Wirtschaftsinformatik
- [Bu02] Burke, R. D.: Hybrid Recommender Systems: Survey and Experiment. *User modeling and user-adapted interaction* 12.4 (2002): S. 331-370.
- [SK09] Xiaoyuan Su; Taghi M. Khoshgoftaar: A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence*, vol. 2009, Article ID 421425, 19 pages, 2009. doi:10.1155/2009/421425
- [He04] Herlocker, J.L. et.al.: Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)* 22.1 (2004): 5-53.
- [GT05] Adomavicius, G.; Tuzhilin, A.: Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *Knowledge and Data Engineering, IEEE Transactions on* 17.6 (2005): 734-749.