

Educational Datamining (EDM)

Seit den 1980er Jahren wird, mangels eines deutschen Begriffes, das Datensammeln als solches, der Prozess der Auswertung und die Konsequenzen als Datamining bezeichnet. Zusammengefasst geht es dabei im Endeffekt um eine große Anzahl von (zum Teil unspezifisch) erfassten Daten, deren (mögliche) Interpretationen und daraus entstehende Konsequenzen. Geschieht dies im Bildungskontext, redet man seit etwa 1995 von Educational Datamining (nach Romero & Ventura, 2007). Die wachsende Verbreitung der Internetzugänge, die Entwicklungen rund um das Web 2.0 und die damit verbundene erhöhten Interaktivität, die zunehmende Nutzung von sozialen Netzwerken und auch die Tendenz, immer mehr Prozesse mit IT zu bearbeiten, führen dazu, dass fast beliebig Daten gesammelt werden, welche Prozesse, wie beispielweise Arbeitsabläufe von der Einlasskontrolle bis zu den täglichen Abläufen, beschreiben (engl. ‚Big Data‘).

“

!

Unter Big Data versteht man eine unüberschaubare Anhäufung von Daten durch die Nutzung verschiedenster webbasierter Dienste zu deren Analyse und Interpretation.

Bei der Datensammlung wird die stark unterschiedliche Struktur der Daten deutlich: Viele Prozesse werden mit festen Strukturen protokolliert. Zum Beispiel das Aufrufen einer Webseite und der Kontakt mit der eigenen IP-Adresse, gegebenenfalls das individuelle Authentifizieren, der Aufruf bestimmter Angebote, das Absolvieren von Tests und bestimmte Auswahlen werden in Datenbanken hinterlegt.

Daneben entstehen zunehmend Texte, die häufig per Chat, E-Mail, in Wikis, Blogs und Foren, oder per Upload in verschiedensten Lernumgebungen produziert und ausgetauscht werden. Immer mehr werden unterrichtliche Angebote nicht mehr nur in Präsenzformaten ‚vor Ort‘ (Seminare, Vorlesungen), sondern auch als Telefon- und Videokonferenzen synchron und als Podcasts asynchron durchgeführt. Hier fallen die Verbindungsdaten und entsprechende Aktivitätszeiten als strukturiertes Datenmaterial an, aus denen dem Aktivitätsanteile und Dominanz- oder Partizipationswerte als Charakteristik der Kommunikation abgeleitet werden können. Aus pädagogischer Sicht ist es darüber hinaus selbstverständlich von höchstem Interesse, was inhaltlich vorgefallen ist. Worüber wurde gesprochen, in welchem Zusammenhang? Die automatische Klassifikation und inhaltliche Zuordnung (semantische Analyse) solcher Daten ist ein

ganz eigenes und wesentlich komplexeres Problem und steckt heute noch in den Kinderschuhen (Spies, 2013), wiewohl es schon erste vielversprechende Ansätze gibt (Softic et al., 2010). Noch viel aufwändiger ist die Untersuchung der Bewegungsmuster und der Mimik.

Das Ziel von EDM ist also, aus einer riesigen Datenmenge heraus überschaubare Typen, Profile, Cluster und darauf bezogen typische inhaltsbezogene Abfolgen und auch kritische Werte zu ermitteln. Pädagogisch geht es darum, Muster in den Daten zu erkennen, um daraus notwendige oder empfehlenswerte Handlungen planen zu können. Dies kann im Einzelfall eine Alarmmeldung sein, die sich auf das globale Verhalten (Engagement, Fleiß, Präsenz) bezieht, oder aber im Detail auf Prozesse, bei denen die weiteren Entfaltungsschritte durch Sackgassen, falsche Arbeitstechniken, Routinen oder Kenntnislücken blockiert sind. Zum Einsatz kommen kann dabei ein automatisches adaptives Online-Hilfsangebot oder konkrete Interventionen von Lehrpersonen, die in kritischen Alarmfällen vorgeschlagen werden. Unter einem eher ökonomischen Aspekt werden durch die Kombination verschiedener Verfahren Zusammenhänge zwischen Eingangs- und Prozessparametern und dem jeweiligen Erfolg ermittelt. Dabei wird versucht, Schulungs- und Bildungsmaßnahmen statistisch zu erklären und Prognosen für den Erfolg und auch Misserfolg konkret einzelner Akteure zu berechnen und natürlich auch Erwartungswerte für deren Gesamtheit zu prognostizieren.

“

!

Das Ziel von Educational Data Mining ist es, aus riesigen Datenmengen überschaubare Typen, Profile, Zusammenhänge, Cluster und darauf bezogen typische Abfolgen, Zusammenhänge und kritische Werte zu ermitteln, um daraus Prognosen zu errechnen und Empfehlungen für sinnvolle pädagogische Handlungen ableiten zu können.

Mit der Einführung des Begriffes „Educational Datamining“ waren und sind natürlich Hoffnungen verbunden, die traditionellen Methoden aus der künstlichen Intelligenzforschung, dem Bereich des Maschinen-Lernen und der Statistik bzw. Mustererkennung zusammenzufassen und im Bildungskontext gezielt einzusetzen (ALMazroui, 2013, 9).

LMazroui (2013) referiert folgende Techniken und Inhaltsbereiche:

- Lokalisierung des Entwicklungsstandes von Individuen und Gruppen, Vorhersagemodelle für Ziele, gewünschte Leistungen: ermittlung des zukünftigen Bedarfs und der Planungsgrundlagen, Ableitungen von Empfehlungen und Feedback für alle Beteiligten
- Modellbildung, Parameterabschätzung für Wahrscheinlichkeitsaussagen, Erzeugung von Gruppen, Clustern mit ähnlichen Eigenschaften und anzunehmenden ähnlichen ,Behandlungs‘verfahren und Konzeption von Kursen

- Untersuchung der Spuren auf sequentielle und hierarchische Abhängigkeiten, Ursache-Wirkungszusammenhänge, einfache Korrelationen, Ermittlung kritischer Werte/Alarme für Eingriffe
- Einsatz von grafischen Verfahren, um Strukturen in großen Datenmengen zu erkennen (Visualisierung)
- Kontrolle von Ausreißerdaten und Einsatz von Interventionen
- Textanalyse mit noch ungeklärtem Potential
- Ermittlung von Parametern für die Interaktion von Personen in Gruppen und Gruppen untereinander durch Soziale Netzwerkanalysen (SNA)

Differenzierung Datamining und Educational Datamining

Es besteht eine gewisse Nähe zwischen kommerziellem Datamining und EDM: Beim einen geht es darum, Kundinnen und Kunden zu beeinflussen, um mehr Profit zu generieren und damit den Verdienst zu erhöhen. Beim EDM dienen die Daten als Grundlage dafür, Lernerfolge zu ermöglichen und Kompetenzen zu vermitteln. Der Erfolg zeigt sich im kommerziellen Umfeld, indem Kundinnen und Kunden ihre Zufriedenheit mit bestimmten Aktionen bekunden. Im EDM wird ermittelt, welche Aktion mit entsprechendem Fortschritt und der Zielerreichung oder dem Gegenteil zusammenhängt.

In der Praxis des Datamining zeigt sich, dass, ausgelöst von akuten Praxisfragen, Datenmengen untersucht und im Forschungsprozess neue Fragen aufgeworfen werden, die bis dahin gar nicht im Raum standen. Wir könnten so zum Beispiel durch die abweichenden Zeitstempel und damit verbundenen Daten feststellen, dass es bei bestimmten

Lehrpersonen immer und ohne systematischen Grund länger dauert, bis die Klasse online präsent ist. Weiterhin lassen sich zum Beispiel Phänomene abbilden, bei denen spezielle Ziele abseits des vollzogenen Lehrplans ohne das Zutun einer Lehrperson plötzlich erreicht werden.

Andererseits muss hier auch betont werden, dass der praktische Einsatz von EDM in Klassenräumen und Lehrsälen überschaubar ist, was daran liegt, dass die oft hochgesteckten Erwartungen nur bedingt erfüllbar sind – also viel technologischer Einsatz einem vergleichsweise bescheidenem Ergebnis gegenübersteht. Auch heute noch verlassen die meisten Entwicklungen nicht die Forschungslabore.

“

?

Denken Sie an Ihre eigene Lernerfahrung. Ließe sich bezogen auf längere Episoden (Schuljahre) Ihr Lernen durch eine systematische Datenerfassung erheblich und nachhaltig optimieren und beschleunigen? Versuchen Sie an einem konkreten Beispiel aus Ihrer Erfahrung darzustellen, welche Daten erhoben werden sollten und wie deren automatisierte Interpretation Ihnen helfen könnte.