# OPEN EDUCATION, BIG DATA EN LEARNING ANALYTICS

door **Sander Latour** en **Robert Schuwer**

De opkomst van MOOC´s in 2012 heeft gezorgd voor een grote interesse in de mogelijkheden van open en online onderwijs bij instellingen voor hoger onderwijs in Nederland. Omdat daarin veelal duizenden tot tienduizenden lerenden participeren, komen grote hoeveelheden gebruiks- en gebruikersdata beschikbaar. Tegelijkertijd zien we het aanbod van deze en andere open leermaterialen stijgen. Het inzetten van learning analytics op deze data biedt kansen om het open en online onderwijs te verbeteren, maar biedt ook enkele uitdagingen. In dit artikel worden twee kansen en een uitdaging beschreven.

## Wat is learning analytics?

SoLAR, de internationale gemeenschap voor onderzoek op dit gebied, definieert learning analytics als het 'verzamelen, analyseren en rapporteren van data over studenten en hun omgeving ten behoeve van het begrijpen en verbeteren van het onderwijs en de omgeving waarin dit onderwijs plaatsvindt'. Een gangbare opvatting binnen de gemeenschap is dat dit een cyclisch proces behelst, zoals beschreven in (Clow, 2012). Dit cyclische proces is te vergelijken met een docent die een vraag stelt met als doel om op basis van het geanalyseerde antwoord een didactische beslissing te kunnen nemen. Door de toenemende schaal van het onderwijs is het in de praktijk echter vaak lastig voor een docent om middels deze cyclus zijn of haar onderwijs optimaal te verbeteren. Tegelijkertijd manifesteert een steeds groter deel van het leerproces zich in digitale data. Learning analytics maakt gebruik van deze digitale data om dezelfde cyclus op grote schaal toe te kunnen passen om het leren te ondersteunen.

Learning analytics richt zich op het ondersteunen van studenten en docenten. Aanverwante vakgebieden maken voor een deel gebruik van dezelfde data en technieken als learning analytics, maar richten zich op andere vragen (Chatti, 2012). Waar learning analytics zich bezig houdt met het ondersteunen van het leerproces, richt *educational data mining* zich meer op de technische mogelijkheden om waarde uit de grote hoeveelheid data te extraheren (Ferguson, 2012). *Academic analytics* tot slot past technieken uit *business intelligence* toe op data op instellingsniveau om het studierendement te verhogen (Long, 2011).

## Uitdaging: vervuilde data

Een van de beloften van MOOC's is dat door er de grote aantallen studenten die eraan deelnemen, een schat aan data ontstaat waaruit met learning analytics technieken informatie kan worden gehaald over verbeterpunten van de cursus. Tevens kunnen dergelijke analyses ook meer inzicht geven in hoe mensen leren (zie bijvoorbeeld http://blog.socrato.com/learning-analytics-and-moocs/).

MOOC's kennen echter een hoog uitvalpercentage. Deze uitval wordt deels veroorzaakt doordat deelnemers niet beginnen met de intentie de MOOC tot het einde te

**Sander Latour** (sanderlatour@gmail.com) is sinds 2012 actief in learning analytics en betrokken bij vele initiatieven. Tot februari 2014 was hij in dienst bij de Universiteit van Amsterdam en voorzitter van het kernteam van de special interest group Learning Analytics van SURF. Hij is nog steeds lid van het kernteam en blijft actief in het veld.

**Robert Schuwer** (robert.schuwer@ou.nl) werkt bij de Open Universiteit. Sinds 2006 is hij betrokken geweest bij vele OER-projecten, zowel aan de Open Universiteit als daarbuiten. Hij is voorzitter van het kernteam van de special interest group Open Education van SURF.

gaan volgen, maar slechts geïnteresseerd zijn in onderdelen van de cursus of gewoon beginnen uit nieuwsgierigheid (Clow, 2013). Door de lage drempel van inschrijven en de geringe consequenties die kleven aan stoppen met een MOOC wordt de prioriteit ook eerder elders gelegd (bijvoorbeeld een deadline voor het werk). Dit zorgt potentieel voor vervuiling van de data: de reden van stoppen of overslaan van delen van cursussen is immers voor het grootste deel niet bekend en hoeft niet gerelateerd te zijn aan de inhoud van de cursus.

Traditioneel richten veel analytics-technieken in het hoger onderwijs zich op het kunnen voorspellen van studiesucces op basis van gedrag van studenten uit het verleden (Siemens, 2013). De aanname hierbij is dat studenten op ongeveer hetzelfde niveau binnen komen en naar hetzelfde punt toe moeten. Door de grote verschillen in de intenties en omstandigheden van MOOC-deelnemers wordt het echter een stuk moeilijker om betekenisvolle correlaties te vinden in de data en die te gebruiken. Het monitoren van de voortgang in de groep is op zijn minst ingewikkelder dan wanneer er een grotere mate van uniformiteit zou zijn. Immers, een student zou perfect op schema kunnen liggen voor een persoonlijk leerdoel, maar in relatie tot het halen van het eindtentamen achterop raken.

Deze realiteit stelt hogere eisen aan het gebruik van dergelijke technieken dan in een traditionele cursus het geval is, iets wat de toepassing ervan in MOOC's mogelijk problematisch maakt (Clow, 2013). Initieel kan (bijvoorbeeld via een korte enquête) gevraagd worden naar de intentie van de lerenden, maar deze intentie kan gedurende het proces wijzigen door oorzaken die buiten de cursus liggen. Een andere aanpak is om automatisch te proberen de deelnemers in subgroepen te classificeren (Kilzilcec, 2013). Vervolgens zou je per subgroep learning analytics technieken toe kunnen passen in de hoop dat er binnen de subgroep meer uniformiteit bestaat.

Er zijn echter ook learning analytics toepassingen die van de datavervuiling weinig tot geen last hebben. Voorbeelden hiervan worden genoemd in (Fournier, 2011; Ferguson, 2012). De data van een individuele student kunnen bijvoorbeeld gevisualiseerd worden om reflectie te ondersteunen. De toepassingen van learning analytics die last hebben van een heterogene populatie lijken dan ook relevanter te zijn voor een xMOOC dan voor een cMOOC (Schuwer, 2013b).

### Kans 1: bepalen van de kwaliteit van open leermaterialen

Kwaliteit van open leermaterialen wordt bepaald door veel criteria (Schuwer, 2013a). Een aantal van die criteria heeft te maken met de context van (her)gebruik van open

leermaterialen. Denk bijvoorbeeld aan kenmerken van de doelgroep, de te gebruiken didactiek en de reeds aanwezige voorkennis over het onderwerp. Wanneer een docent open leermaterialen zoekt voor hergebruik zijn gegevens over juist die criteria veelal niet of slechts in beperkte mate aanwezig (via *reviews* van eerdere gebruikers). Dat maakt het lastig voor een docent om te bepalen of de gevonden open leermaterialen te gebruiken zijn binnen zijn of haar context en welke aanpassingen er eventueel nog moeten worden aangebracht.

Learning analytics zou kunnen helpen door een deel van deze context automatisch aan te vullen. Een koppeling met de bestaande data uit een *learning management system* levert bijvoorbeeld al mogelijkheden voor de automatische generatie van metadata voor open leermaterialen (Duval, 2004). Een relatief simpel voorbeeld hiervan is het vinden van patronen in de metadata van andere leermiddelen die in combinatie met de open leermaterialen zijn gebruikt. Naarmate meer gedetailleerde data uit leeromgevingen gekoppeld kunnen worden met verzamelingen van open leermaterialen kunnen zelfs leerkarakteristieken van de studenten die de open leermaterialen gebruikt hebben worden toegevoegd. Op die manier zou een docent al meer informatie hebben over de geschikte doelgroep. Learning analytics zou zelfs de opgeslagen karakteristieken met die van de beoogde studenten kunnen vergelijken om tot een rangorde te komen.

Ook aan het gebruik zelf valt relevante informatie te ontlenen. Zo zou je de aandacht in tijd kunnen meten die lerenden geven aan de open leermaterialen (Ochoa, 2006). Daarnaast ligt het in de lijn der verwachting dat learning analytics op den duur in toenemende mate in staat zal zijn om uitspraken te doen over de effectiviteit van open leermaterialen. Open leermaterialen worden immers vaak toegepast in een context waarin ook toetsen worden afgenomen. Als de data uit deze toetsen samengebracht kunnen worden met de data van het gebruik van de open leermaterialen, dan kunnen statistische modellen schattingen gaan maken van het effect dat de open leermaterialen hebben gehad op de prestaties. Deze schattingen kunnen in de lijst met open leermaterialen aangeboden worden aan de zoekende docent of als input dienen voor een algoritme dat automatisch leermiddelen aan een student aanbiedt.

Tot slot zou het voor een auteur van open leermaterialen nuttig kunnen zijn om feedback te krijgen in de vorm van kwalitatieve informatie over de interacties tussen studenten en het leermateriaal. Voor een instructievideo kan het bijvoorbeeld bijzonder informatief zijn om te weten waar studenten terug spoelen of pauzeren.

De genoemde toepassingen van learning analytics kunnen beter functioneren naarmate er een infrastructuur is die de terugkoppeling ondersteunt vanuit de leeromgeving waar de open leermaterialen in gebruikt worden naar een databank met open leermaterialen. De standaarden die nodig zijn zodat systemen over deze gegevens met elkaar kunnen communiceren, beginnen nu vorm te krijgen. Recente ontwikkelingen binnen standaarden platform IMS (Learning Measurement for Analytics Whitepaper, 2013) duiden er op dat de verschillende bestaande standaarden ook steeds beter geïntegreerd zijn. Een centrale databank die dergelijke informatie over open leermaterialen verwerkt, analyseert en presenteert is er echter nog niet, al zijn de drie onderdelen er in toenemende mate al wel.

### Kans 2: op weg naar massamaatwerk

Massamaatwerk is een aanpak waarbij efficiënt een aangeboden leersituatie zo goed als mogelijk wordt aangepast aan de eigenschappen en wensen van de individuele lerende. Door de steeds grotere beschikbaarheid van open leermaterialen en MOOC's nemen de potentiële mogelijkheden van het kunnen aanbieden van massamaatwerk

toe (Yuan et al., 2008). Te denken valt dan aan het aanbieden van diverse alternatieven op het niveau van cursus- of leereenheid of het kunnen laten configureren van een individueel leerpad door een curriculum. Voorbeelden van alternatieven zijn keuzevakken, binnen een cursus het variëren van cases of het aanbieden van opgaven op verschillende moeilijkheidsniveaus. De schaal van het onderwijs is echter te groot voor een docent om handmatig op individueel niveau deze alternatieven aan te bieden.

Er lijkt hier een rol voor de techniek weggelegd en die biedt zich in grofweg twee varianten aan. Binnen de learning analytics gemeenschap wordt vooral gezocht naar manieren om een beslissing voor een student of docent makkelijker te maken, terwijl de *educational data mining* gemeenschap meestal meer in de hoek zit van systemen die een beslissing al automatisch doorvoeren door middel van adaptiviteit (Siemens, 2012). Uit een verkennende studie  bleek dat de beschikbare datasets uit onder andere *learning object repositories* in ieder geval al bruikbaar zijn voor het geven van automatische aanbevelingen voor relevant lesmateriaal (Verbert, 2011). Dergelijke data worden overigens ook toegepast om docenten tijdens het auteursproces te ondersteunen (Verbert, 2012).

Binnen de wereld van de *adaptive (educational) hypermedia* valt de taak van het personaliseren met open leermaterialen onder het opencorpus-probleem (Brusilovsky, 2007; Kravčík, 2013). Een onderdeel van dit toegepaste opencorpus-probleem is het vraagstuk hoe je leermiddelen moet representeren om onderlinge links te kunnen genereren. Een geprobeerde aanvliegroute hiervoor is vanuit de hoek van het *semantic web* (Henze, 2004). Ook voor het personaliseren van een curriculum is er met *semantic web* representaties gewerkt (Baldoni, 2011). Voor deze voorbeelden geldt echter wel dat een behoorlijk deel van de benodigde informatie expliciet toegevoegd moet worden aan het systeem. De grote omvang van het aantal open leermaterialen maakt dat tot een grote uitdaging.

Learning analytics zou juist deze grote omvang als voordeel kunnen inzetten, door de grote hoeveelheid gebruiksdata van open leermaterialen te gebruiken voor aanbevelingen van lesmateriaal (Duval, 2011), zonder daarnaast nog veel informatie handmatig te moeten toevoegen aan de leermiddelen. Ook in de adaptieve hoek zijn er alternatieven voor de kennisintensieve aanpak. De grote toename van data maakt het bijvoorbeeld mogelijk om onderdelen van *intelligent tutoring systems* automatisch uit de data te extraheren (Koedinger, 2013).

In het verhaal van massamaatwerk waar de lerende centraal staat, is het van belang dat de data aan de lerende gekoppeld zijn en niet aan een systeem. De lerende moet zijn gebruiksdata mee kunnen nemen naar andere systemen om ook daar een gepersonaliseerd aanbod te kunnen krijgen. Een risico bij massamaatwerk is dat de lerende gevaar loopt vast te komen zitten in dezelfde soort aanbevelingen waardoor nooit nieuwe paden worden bewandeld.

## Conclusie

In dit artikel zijn binnen open onderwijs twee kansen en een uitdaging voor learning analytics besproken. De samenwerking tussen beide gebieden lijkt vooralsnog veelbelovend, maar tegelijkertijd ligt een aantal genoemde toepassingen ook nog echt in de toekomst en kunnen ze wellicht zelfs speculatief worden genoemd. Onderzoek op deze terreinen is nu nog gefragmenteerd en daardoor minder efficiënt en effectief. Voor deze en andere onderzoeksthema's binnen de wereld van open educational resources (OER) pleiten we daarom voor een nationale onderzoeksagenda OER.

Door de toevoegingen van learning analytics neemt de waarde van data toe. Er is een behoefte aan een flexibele infrastructuur waarin deze data getransporteerd kunnen worden. Het verwezenlijken van deze infrastructuur is echter niet alleen een technische uitdaging, maar ook een economische. Desalniettemin beschouwen we open education en learning analytics als een gouden combinatie.

## Bronnen


- Baldoni, M., Baroglio, C., Brunkhorst, I., Henze, N., Marengo, E., & Patti, V. (2011). Constraint modelling for curriculum planning and validation. Interactive Learning Environments, 19(1), 81-123.
- Brusilovsky, P., & Henze, N. (2007). Open corpus adaptive educational hypermedia. In The Adaptive Web (pp. 671-696). Springer Berlin Heidelberg.
- Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. International Journal of Technology Enhanced Learning, 4(5), 318-331.
- Clow, Doug (2012). The learning analytics cycle: closing the loop effectively. In: LAK12: 2nd International Conference on Learning Analytics & Knowledge, 29 April - 2 May 2012, Vancouver, BC.
- Clow, Doug (2013). MOOCs and the funnel of participation. In: Third Conference on Learning Analytics and Knowledge (LAK 2013), 8-12 April 2013, Leuven, Belgium.
- Duval, E. (2011). Attention please!: learning analytics for visualization and recommendation. In Proceedings of the 1st International Conference on Learning Analytics and Knowledge (pp. 9-17). ACM.
- Ferguson, Rebecca (2012). Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5/6) pp. 304-317.
- Fournier, H., Kop, R., & Sitlia, H. (2011). The value of learning analytics to networked learning on a Personal Learning Environment.
- Henze, N., Dolog, P., & Nejdl, W. (2004). Reasoning and Ontologies for Personalized E-Learning in the Semantic Web. Educational Technology & Society, 7(4), 82-97.
- Learning Measurement for Analytics Whitepaper. (2013). Retrieved December 18, 2013 from www.imsglobal.org/IMSLearningAnalyticsWP.pdf.
- Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. Educause Review, 46(5), 30-32.
- Kizilcec, R. F., Piech, C., & Schneider, E. (2013, April). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 170-179). ACM.
- Koedinger, K.R., Brunskill, E., Baker, R.S.J.d., McLaughlin, E.A., Stamper, J. (in press) New potentials for data-driven intelligent tutoring system development and optimization. To appear in AI Magazine (Vol 43, no 3).
- Kravčík, M., & Wan, J. (2013). Towards Open Corpus Adaptive E-learning Systems on the Web. In Advances in Web-Based Learning–ICWL 2013 (pp. 111-120). Springer Berlin Heidelberg.
- Ochoa, X., & Duval, E. (2006). Use of contextualized attention metadata for ranking and recommending learning objects. In Proceedings of the 1st international workshop on Contextualized attention metadata: collecting, managing and exploiting of rich usage information (pp. 9-16). ACM.
- Schuwer, R. (2013a). Kwaliteit van open leermaterialen. SURF, Utrecht. Beschikbaar op: www.surf.nl/binaries/content/assets/surf/nl/kennisbank/2013/artikel-kwaliteit-van-open-leermaterialen-07-11-2013.pdf.
- Schuwer, R. (2013b). Massive Open Online Courses, www.surfspace.nl/artikel/1024-massive-open-online-courses-moocs/.
- Siemens, G. (2013). Learning Analytics The Emergence of a Discipline. American Behavioral Scientist, 57(10), 1380-1400.

- Siemens, G., & Baker, R. S. D. (2012, April). Learning analytics and educational data mining: Towards communication and collaboration. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 252-254). ACM.
- SoLAR: www.solaresearch.org/mission/about/
- Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., & Duval, E. (2011, February). Dataset-driven research for improving recommender systems for learning. In Proceedings of the 1st International Conference on Learning Analytics and Knowledge (pp. 44-53). ACM.
- Verbert, K., Ochoa, X., Derntl, M., Wolpers, M., Pardo, A., & Duval, E. (2012). Semi-automatic assembly of learning resources. Computers & Education, 59(4), 1257-1272.
- Yuan, L., MacNeill, S., & Kraan, W. (2008). Open educational resources—Opportunities and challenges for higher education. Gevonden op 18 december 2013 op: http://learn.creativecommons.org/wp-content/uploads/2008/09/oer_briefing_paper.pdf.