

De kracht van recommender systems

Rory Sie, AI masterstudent aan de VU
(rllsie@few.vu.nl)

De goede reputatie van recommender systems heeft ertoe geleid dat populaire websites als Amazon (<http://www.amazon.com>) en Internet Movie Database (<http://www.imdb.com>) deze systemen hebben geïntegreerd in hun websites. De potentie van recommender systemen wordt nog eens benadrukt door het feit dat NetFlix een competitie (<http://www.netflixprize.com>) is begonnen met een prijs van \$1,000,000, waarbij het doel een verbetering is van tenminste tien procent ten opzichte van het huidige algoritme, dat klanten DVD's te huur aanbiedt die dicht bij hun interesse liggen.

De term recommender systems werd in 1997 voorgesteld door Resnick en Varian, die stelden dat deze een manier waren om aanbevelingen van mensen te verzamelen en door te spelen naar andere mensen. In andere woorden, de consument werd een platform geboden waarop ze een waardering konden geven aan bepaalde items, zoals boeken, zodat deze later konden worden verzameld en

gebruikt om andere mensen een aanbeveling te doen. Een goed voorbeeld daarbij is de website van Amazon, waarop boeken, maar ook andere producten worden verkocht. Wanneer de consument een product koopt, wordt dit opgeslagen in de database van Amazon. Op deze manier krijgt Amazon dus een goed idee wat de klant koopt en interessant vindt, en kan die informatie ook gebruiken om andere klanten mee te adviseren. Dit gebeurt nu in de vorm van een blokje met interessante producten onderaan de pagina, aangeduid met "customers who bought this item also bought". Ik vertel verderop wat de redenen voor het gebruik van een recommender system kunnen zijn, maar eerst laat ik even in het kort zien hoe de techniek achter een dergelijk systeem werkt.

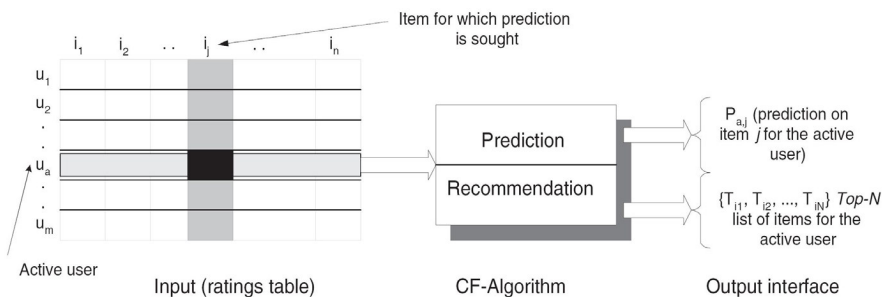
Recommender systems - in het Nederlands vertaald: aanbevelingssystemen, maar we gebruiken toch de Engelse term – hebben de laatste jaren getoond dat ze van grote waarde kunnen zijn voor bedrijven die hun producten aan consumenten willen aanbieden. Zo'n product kan een boek of film zijn, maar eigenlijk kan het elk soort verkoopbaar product zijn. De kracht van recommender systems wordt onderkend door grote bedrijven, maar wat zijn het eigenlijk? En mogelijk veel belangrijker: voor wie en waarom kunnen ze van waarde zijn?

Techniek

Er zijn verschillende manieren om items aan te bevelen aan klanten. De meest noemenswaardige manieren zijn *content-based recommendation*, *collaborative filtering* en *demographic filtering*, die zowel afzonderlijk van elkaar als gecombineerd zijn te gebruiken. Content-based filtering gebeurt op basis van information retrieval-technieken. Information retrieval houdt in dat je zoekt naar één of meerdere documenten die jouw zoekterm bevatten. Een voorbeeld hiervan is een zoekmachine,

waarbij je zoekt naar een document (lees: website) dat jouw zoekterm bevat.

Collaborative Filtering, wellicht de meest interessante van de drie, houdt in dat je waarderingen die mensen aan producten geven (zowel expliciet als impliciet!) in een 'ratings matrix' stopt, waarbij elke regel een klant voorstelt, en elke kolom een item voorstelt. Een voorbeeld hiervan is te zien in figuur 1.



Figuur 1. Een zogeheten 'ratings matrix' (uit [Sarwar2001])

Je kunt dan regels gaan vergelijken door middel van een k-nearest neighbor-techniek, wat inhoudt dat je kijkt welke andere klanten het meest overeenkomen met de klant die op dat moment op jouw website zit. De items die de huidige klant niet heeft bezocht, maar

vergelijkbare klanten wel, zou je dan kunnen aanbevelen. Demographic Filtering is een techniek die minder vaak wordt gebruikt, omdat je daarbij afhankelijk bent van informatie die de gebruiker over zichzelf prijsgeeft, zoals leeftijd, geslacht en beroep. Je gebruikt deze informatie om gebruikers te vergelijken. Je zou dan kunnen denken aan clustering. Clustering houdt in dat je de klanten gaat classificeren. Je gaat ze dus in groepen samenbrengen, waarbij de klanten in die groepen op de een of andere manier met elkaar verbonden zijn, hetzij door leeftijd, of door beroep, maar dat hangt sterk af van het product

1. Expliciete waardering van items kan zijn in de vorm van cijfers geven aan een product, terwijl impliciete waardering inhoudt dat je bijhoudt wat een klant heeft gekocht, of welke pagina's hij binnen jouw website heeft bezocht.



Rory Sie is een Masterstudent aan de Vrije Universiteit Amsterdam. Hij studeert Kunstmatige Intelligentie met als specialisatie Knowledge Technology and Intelligent Internet Applications, waarbij zijn interesse uitgaat naar intelligente systemen op het web. Hij loopt op dit moment dan ook zijn stage bij een bedrijf dat naast financieel nieuws en artikelen, ook trainingen en cursussen aanbiedt op haar website. Hij zal een recommender system bouwen om trainingen en cursussen aan te bieden op basis van het surfgedrag van ruim 31.000 bezoekers op de website. Over een paar maanden, wanneer hij zijn studie heeft afgerond, wil hij zijn eigen bedrijf starten op het gebied van deze, maar ook andere intelligente systemen op het web.

dat je wilt aanbieden.

De kracht

Nu we de techniek globaal hebben beschreven, kunnen we denken aan toepassingen van recommender systems, want waarom zouden we ze eigenlijk willen gebruiken? Mocht je achter de computer zitten tijdens het lezen van dit artikel, wat mij niet onwaarschijnlijk lijkt, dan heb je misschien al hebben gekeken op de websites van Amazon en Netflix. Wat je daarbij opvalt is dat beide websites een immens grote keuze aan producten hebben. Amazon verkoopt bijna alle soorten producten die je mogelijk zou willen hebben, Netflix biedt je de keuze uit meer dan 70.000 te huren DVD titels in de Verenigde Staten. Het probleem is dat je van de klant niet kunt verwachten dat ze even tienduizenden items gaan doorzoeken. Wat doe je dan? Je biedt je items aan in categorieën. In tegenstelling tot wat veel andere, kleinere bedrijven doen, hebben Netflix en Amazon naast die categorisatie, gekozen voor een recommender system. Kort gezegd bieden ze de klant een gepersonaliseerde website aan, waarbij de klant door het systeem een kleine set met interessante items aangeboden krijgt die van tevoren uit een veel grotere keuze gefilterd zijn. Deze items zijn gebaseerd op de waarderingen of het surfgedrag van de gebruiker zelf, en zijn dus vrij betrouwbaar. Om het idee van gepersonaliseerde websites te ondersteunen, een quote van Jeff Bezos, CEO van Amazon.com: *“If I have 3 million customers on the Web, I should have 3 million stores on the Web.”*

Waar het op neerkomt, is dat je niet meer één website hebt voor meerdere gebruikers, maar dat je nu opeens websites hebt voor elke gebruiker apart, want iedere gebruiker krijgt in eerste instantie alleen maar voor zijn of haar interesse relevante items te zien. Dit heeft een aantal voordelen (Schafer, 1999):

- Gebruikers zijn eerder geneigd om iets te kopen, omdat ze beter geadviseerd worden en geassisteerd worden in het zoeken naar producten op je website. Dus waar ze eerder alleen op je website kwamen voor informatie en weer weggingen, zullen ze met een recommender system sneller een product kopen, wat natuurlijk het uiteindelijke doel is.

- Wanneer je een recommender system gebruikt, kun je verbanden leggen naar andere producten, zodat mensen die een bepaald product vinden, niet alleen het product zelf zien, maar ook producten die daar op de een of andere manier mee te maken hebben. Je probeert er dus voor te zorgen dat ze meerdere producten per keer kopen.

- Gebruikers krijgen meer vertrouwen in je website, omdat je in eerste instantie alleen maar content aanbiedt die voor hen relevant zijn. Ze komen dus vaker terug naar je website, omdat jouw website biedt waar zij naar op zoek zijn. Daarnaast kun je gebruikers met elkaar in contact brengen. Hierdoor leren mensen gelijkgestemden kennen en zullen ze ook weer vaker naar je website terugkeren, omdat ze nu een “band” met de website hebben, althans, met andere gebruikers op dezelfde website.

Mogelijke toepassingen van recommender systems zijn dan ook websites waarbij relatief veel informatie wordt getoond aan de gebruiker, terwijl misschien maar een klein deel daarvan geschikt is voor de gebruiker. Hierbij kun je denken aan websites waar iets verkocht wordt (bijvoorbeeld je eigen webwinkel) of websites waar het belangrijk is dat mensen maar een klein deel van lezen, zoals nieuwssites, maar ook sites waar een overzicht op staat van concurrerende goederen of diensten, zoals restaurants, autogarages, televisieprogramma's, noem het maar op. Er is dus nog een grote markt om te veroveren. ∅

Referenties:

- [Sarwar2001] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. *Item-based collaborative filtering recommendation algorithms. Proceedings of the tenth international conference on World Wide Web, pages 285–295, 2001*
- [Schafer1999] J.B. Schafer, J. Konstan, and J. Riedl. *Recommender systems in e-commerce. Proceedings of the 1st ACM conference on Electronic commerce, pages 158–166, 1999*
- Link: Houd de website www.nerdhouse.nl in de gaten voor meer informatie.**