

Deelverslag 3 Groepsproject

Collectieve Intelligentie

Groepsleden:

- Rabie Afqir (**11606843**)
- Björn Out (**12567930**)
- Frank Tamer (**12248738**)
- Gino Pennasilico (**12393819**)

3.0 Update Taakverdeling

Er zijn geen veranderingen ten opzichte van vorige week.

3.1 Methode

SVM

Omdat een SVM een *classifier* is, ondersteunt deze geen berekeningen als MSE, RMSE en MAPE. Daarom wordt de evaluatie van dit algoritme grotendeels op basis van *precision-recall curves* uitgevoerd. Voor de drie verschillende parameters 'gamma-waarde', 'nu-waarde', en 'kernel algoritme' zijn ieder de resultaten voor vier verschillende waarden in *precision-recall plots* geplot, waarbij de andere parameters op vaste waarden zijn gezet (gamma = 0.000001, nu = 0.2, kernel = 'rbf'). Aan de hand hiervan is per parameter de optimale waarde bepaald. Vervolgens is aan de hand van één laatste *precision-recall curve*, met alle optimale waarden voor de parameters, het algoritme vergeleken met de *baselines* 'random predictions' en 'item means predictions', is het algoritme vergeleken met de andere omschreven aanpakken. Ten slotte is een functie toegepast om op basis van onderlinge *Jaccard similarity* aanbevelingen te selecteren uit een lijst met potentieel aan te bevelen bedrijven. De hieruit volgende gemiddelde *Jaccard score* is vergeleken met een gemiddelde *Jaccard similarity* die resulteert uit willekeurige selectie.

Cold Start Approach

De *Cold Start approach* kijkt naar de *lower confidence bound* van de beoordelingen van een bedrijf. De reden hiervoor is omdat wij een bedrijf met een consistent hoge beoordeling meer vertrouwen dan een inconsistente beoordeling, bijvoorbeeld omdat de aanbevelingen hierdoor resistenter zijn tegen botaccounts en hun activiteiten (McGlohon *et al.*, 2010). De aanbevelingen zijn op individuele wijze vergeleken. De *lower bound* wordt elke keer met de daadwerkelijke beoordelingen vergelijkt. De *baseline* hierbij is het maken van *random predictions* voor de bedrijven die aanbevolen zijn. De baseline wordt met de methode

vergeleken aan de hand van de *root mean square error* (RMSE), *mean squared error* (MSE) en de *mean absolute percentage error* (MAPE). Om de diversiteit van aanbevelingen te beoordelen wordt de gemiddelde Jaccard similarity van de gesuggereerde bedrijven berekend.

Item-based Collaborative Filtering

Voor de evaluatie van onze *item-based filtering* hebben wij gebruik gemaakt van de evaluatie maat MSE. We hebben bij het testen bijvoorbeeld de *mean squared error* genomen van de '*mean per item*', en deze vergeleken met wanneer er *random* aanbevelingen gedaan zouden worden.

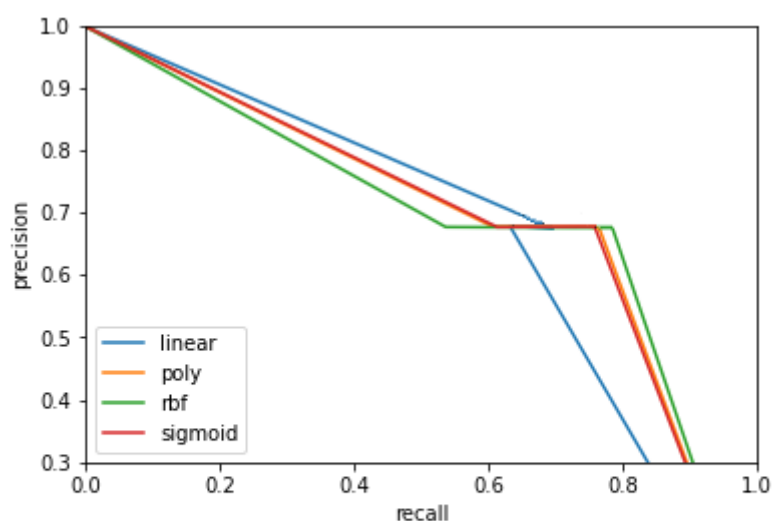
De *baselines* zijn de *random prediction* en de *item-means prediction* methoden. We hebben de prestaties van onze algoritmen in beeld gebracht met een *precision-recall curve*. Bij het *item-based* algoritme hebben we de resultaten met onze *baseline* vergeleken aan de hand van de MSE. Onze resultaten zijn verkregen uit het testen op een kleinere subset van de data van één bepaalde stad, Brooklyn. Onze train- en testset van de data zijn gesplitst in een 25/75 verhouding, zonder een 'random state'. De evaluaties zijn gedaan op deze testset, die bestond uit 435 reviews. De training set bestond uit 1249 reviews.

3.2 Resultaten

SVM

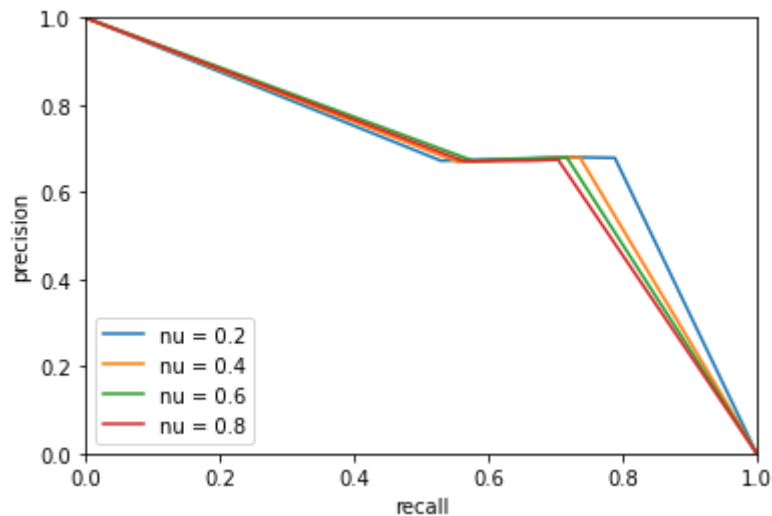
Kernel type

De *precision* en *recall* waarden van diverse thresholds zijn te zien in onderstaand plot. Aangezien alle kernels altijd hetzelfde presteren op het gebied van *precision*, is de kernel die potentieel het beste presteert op het gebied van *recall* het beste. In dit geval is dit de kernel 'rbf'.



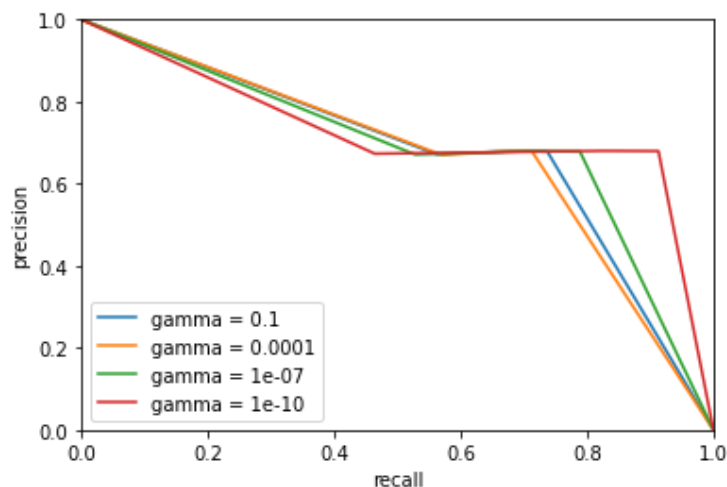
Nu-waarde

Onderstaande afbeelding toont enkele *curves* voor verschillende nu-waarden. Hier is wederom sprake van geen verschil in *precision* voor alle waarden, waardoor de nu-waarde die het beste kan presteren op het gebied van *recall* het beste moet zijn. Dit is $\nu = 0.2$.



Gamma-waarde

De beste gamma-waarde is te vinden bij $\text{recall} = 0.9$. De optimale waarde voor gamma is 0.0000000001. Dit is wederom de optimale waarde, omdat alleen recall wordt beïnvloed door dit parameter en de optimale potentiële *recall* wordt bereikt bij $\gamma = 0.0000000001$.



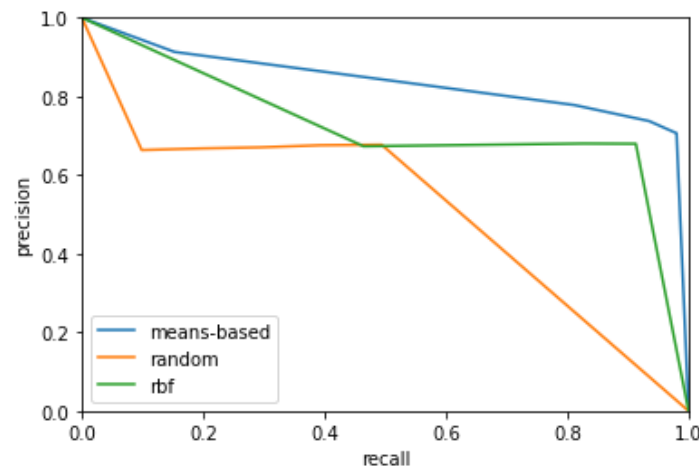
Jaccard

Het Jaccard selectie onderdeel van het algoritme heeft geen parameters, maar blijkt na experimentele tests altijd ongeveer gelijkwaardige prestaties te leveren. Bij voldoende opties zal de gemiddelde Jaccard van de suggesties altijd rond de 0 uitkomen (d.w.z. geen gelijke features). Bij *random* selectie ligt deze rond de 0.05. In beide gevallen is er sprake van redelijk diverse aanbevelingen.

SVM vs. Baseline

In de onderstaande *precision-recall curve* is de optimale SVM-oplossing uitgezet tegen twee baselines: *random predictions* (ieder bedrijf-gebruiker paar krijgt een willekeurige voorspelde *rating*) en *item-based mean predictions* (voorspel altijd de gemiddelde *rating* van een bedrijf

als voorspelde *rating* voor dat bedrijf). Te zien is dat de SVM-oplossing weliswaar beter presteert dan de *random baseline*, maar dat de *item-based mean baseline* altijd beter presteert dan de SVM-oplossing.



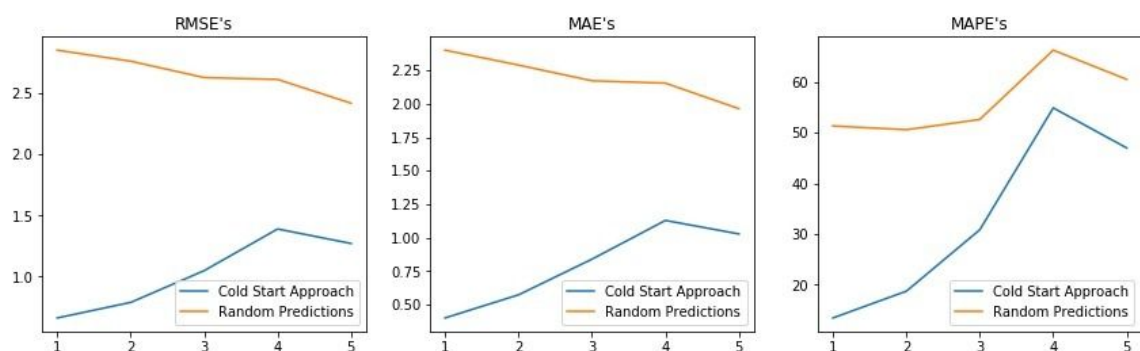
Cold Start Approach

Zoals te zien is in de grafieken, is zowel de RMSE, de MAE en de MAPE beter dan het willekeurig toewijzen van de beoordelingen. De X-as is genummerd in overeenstemming met het Dataframe. Daarnaast heeft het aanbod dat de *Cold Start approach* een variatie van 0,00365, gemeten in Jaccard Similarity.

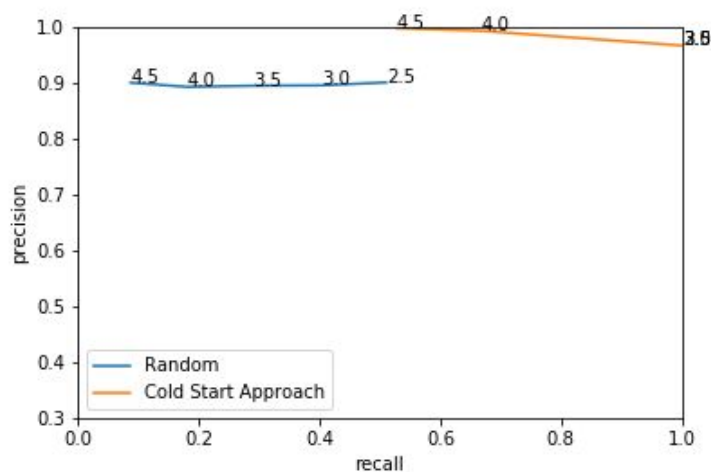
	Business ID	Predicted Rating	RMSE	MAE	MAPE
1	Xg5qEQiB-7L6kGJ5F4K3bQ	4.750960	0.660154	0.401922	13.430810
2	mDR12Hafvr84ctpsV6YLa	4.605711	0.787607	0.574419	18.715428
3	mz9ltimeAly2c2qf5ctijw	4.328782	1.047926	0.840471	30.831568
4	t-o_Sraneime4DDhWrQRBA	3.932575	1.389513	1.129438	54.882167
5	WYw3Uf56DT5lwpalNnCH5Q	3.764400	1.270705	1.028700	46.979958

Text(1, 1.1, "RMSE, MAE and MAPE's of the predictions (Jaccard Similarity = 0.00365):")

RMSE, MAE and MAPE's of the predictions (Jaccard Similarity = 0.00365):

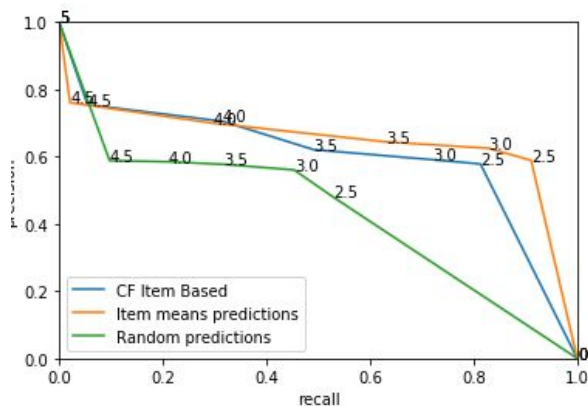


Daarnaast heeft de *Cold Start Approach* ook een betere *precision* en *recall* dan de willekeurige ratings zoals te zien is in de onderstaande grafiek.



Item-based Filtering

Hieronder is de *precision-recall curve* te zien waarin het *item-based* algoritme wordt afgezet tegen een algoritme dat simpelweg willekeurige voorspellingen doet voor een bedrijf, en tegen een algoritme dat voorspellingen doet op basis van de gemiddelde ratings van bedrijven. Zoals te zien in de grafiek vallen zowel *precision* als *recall* van deze implementatie consistent lager uit dan het algoritme dat voorspelt op basis van gemiddelde ratings van bedrijven.



3.3 Bespreking

Qua uniformiteit valt er wellicht nog wat te verbeteren. We hebben namelijk geen rekenkracht in ons bezit gehad om alle steden mee te nemen in onze algoritmen, en dus in onze evaluaties. Omdat de Cold Start approach wel de hele data mee neemt, en de item-based en SVM methode enkel één stad kiezen, is het vergelijken tussen de drie moeilijker dan zou hoeven.

Mean item predictions

Wel is duidelijk dat een aanpak met mean item predictions betere voorspellingen geeft dan de aanpakken gebaseerd op SVM's en item-based collaborative filtering. Voor collaborative filtering is dit terug te zien in zowel de MSE vergelijkingen als in de precision-recall curves. Aangezien het bij SVM's om classificaties gaat is de vergelijking tussen SVM en de baseline alleen te zien in de precision-recall curves. Voor beide algoritmes valt er te verklaren waarom er bij de de item mean predictions betere resultaten zijn. Bij item-based

collaborative filtering komt dit doordat de data ten eerste erg sparse is. Tenslotte zorgt dit voor onnauwkeurige voorspellingen voor een bepaalde gebruiker.

Bij de aanpak met support vector machines is sprake van een klassiek ' $P \gg N$ '-probleem. Met andere woorden: het aantal features is groter dan het aantal bedrijven waarop een SVM wordt getraind. Wanneer hier sprake van is, is er altijd een hyperplane te vinden. Deze generaliseert echter niet goed naar nieuwe, niet eerder geziene data in verband met *overfitting* (Rasmussen & Williams, 2006). Een mogelijke oplossing hiervoor is het toepassen van L1 *regularization*, waarbij sprake is van *feature selection* en dus features verdwijnen (Ng, 2004). Een ander alternatief is het gebruiken van een machine-learning algoritme dat minder of niet gevoelig is voor dergelijke problemen, zoals *neural networks* (Kukačka et al., 2017).

De Item-Based Collaborative Filtering aanpak werkt minder goed omdat veel gebruikers slechts enkele reviews hebben gegeven. Dit effect wordt versterkt door het splitsen van de data in training en test sets. Hierdoor wordt het moeilijker om nauwkeurige ratings te bepalen, omdat het berekenen van de cosine distances vaak niet mogelijk is. De similarity matrix bevat hierdoor veel lege waarden. In de praktijk zou dit naarmate een gebruiker meer reviews plaatst voor deze gebruiker vanzelf worden opgelost. Er kan ook worden onderzocht of het mogelijk is een combinatie van *user-based* en *item-based collaborative filtering* toe te passen, zodat het één terug kan vallen op het ander wanneer nodig.

Conclusie

Hoewel de aanpakken middels SVM's en item based collaborative filtering goede, diverse aanbevelingen doen, is het aanzienlijk eenvoudiger simpelweg de *item mean rating* te gebruiken om aanbevelingen te doen. Er moet nader worden onderzocht of het gebruik van bijvoorbeeld *regularization* of *neural networks* resulteert in betere prestaties op deze dataset. Wél is de inzet van Cold Start Approach mogelijk om algehele aanbevelingen te doen aan niet-ingelogde of recent aangemaakte gebruikers. Deze approach biedt een aantal goede aanbevelingen die voor vele gebruikers aantrekkelijk zullen zijn. Ook kan de techniek van selectie op basis van Jaccard similarity gebruikt worden om diverse selecties van bedrijven te maken uit een reeds bestaande lijst met mogelijkheden.

Bronnen

- Kukačka, J., Golkov, V., & Cremers, D. (2017). Regularization for deep learning: A taxonomy. *arXiv preprint arXiv:1710.10686*.
- McGlohon, M., Glance, N., & Reiter, Z. (2010). Star quality: Aggregating reviews to rank products and merchants. *Fourth international AAAI conference on weblogs and social media*, 114-121.
- Ng, A. Y. (2004). Feature selection, L1 vs. L2 regularization, and rotational invariance. *Twenty-first international conference on Machine learning - ICML '04*, 78. doi: 10.1145/1015330.1015435
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Geraadpleegd van <http://www.gaussianprocess.org/gpml/>