

Movie Enjoyment Predictor

Jonathan Williams (1790472)

Probleemomschrijving:

Mijn vader is geobsedeerd met het kijken van films, alleen hij raakt heel snel teleurgesteld in sommige films die hij gezien heeft. Nu heeft hij behoefte aan een systeem die op basis van zijn preferenties kan bepalen of hij een film leuk gaat vinden of niet. Dit systeem moet rekening kunnen houden met welke acteurs mijn vader leuk vindt en welke niet. Het systeem moet ook zichzelf aanpassen op basis van hoe leuk mijn vader de films vond die hij aangeboden kreeg van het systeem. Hiermee hoopt mijn vader minder tijd te verspillen met films die het toch niet waard zijn, door eerst te kijken wat het programma er van vindt.

Eisen:

- Het programma moet kunnen voorspellen wat voor score de gebruiker zou geven aan een film
- De gebruiker moet verschillende acteurs kunnen bekijken en beoordelen met een cijfer
- De gebruiker moet zelf een ensemble van acteurs kunnen samenstellen. Het programma moet de kans op succes kunnen berekenen van deze ensemble acteurs.

Algoritme:

Om te berekenen of de gebruiker van een film gaat houden of niet wordt er gebruik gemaakt van een weighted pattern structuur gedefiniëerd in sectie 2 van "[Algorithms for Weighted Matching](#)", door Leena Salmela en Jorma Tarhio van de Helsinki University of Technology. De acteurs van een film worden geplaatst in een weighted pattern, en ieder acteur krijgt een reeks invloed scores gebaseerd op de mogelijke posities in de aftiteling (Indien de aftiteling op volgorde is van prominentie, wat bij IMDb het geval lijkt te zijn). De invloed scores van een acteur met een beginnende positie hebben meer effect op het geheel aangezien deze acteurs prominenter voorkomen in de film dan invloed scores van een acteur met een positie richting het einde. De invloed scores worden ook nog beïnvloed door de preferenties van de gebruiker en de resultaten van vorige gokken. De uitkomst kan worden berekend met de score berekening (Definitie 2 van de paper). De formule in definitie 4 van de paper (p-value) kan gebruikt worden voor een extra modus waar de gebruiker een lijst van acteurs kan samenstellen, en dan krijgt te zien wat de kans is op succes voor hun zelf samengestelde film.

Evaluatie algoritme:

Het gebruik van de Weighted Pattern voelde een beetje omslachtig soms, maar het was wel fijner om een data structuur te gebruiken met goed gedefiniëerde regels. Voor tijdens het testen met verschillende waardes was het gebruik van een Weighted Pattern wel fijn, aangezien het vullen van een rij meteen een hele groep waardes geeft om te kijken of alles logisch in elkaar zit. Wel een bonus van het gebruik van de Weighted Pattern is het berekenen van de p-value. Het maakte voor een leuke extra modus in de programma en ook voor een leuke uitdaging aangezien de formule nogal complex was. Het is een recursieve formule. Waar ik later achter kwam echter is dat het aantal recursieve stappen afhangt op het aantal acteurs in de lijst, in een exponentiële relatie. Het aantal recursieve stappen is namelijk n^n , waarbij n het aantal acteurs is. Ik het hiervoor de gebruiker moeten limiteren op maximaal 7 acteurs (met een reken tijd van ~10 seconden), aangezien het bij 8 acteurs meer dan 3 minuten kost. Het was het nog wel waard om deze modus te laten in de applicatie.