



HoGent

Faculteit Bedrijf en Organisatie

Vergelijken voorspellingsalgoritmen met games

Matthias Seghers

Scriptie voorgedragen tot het bekomen van de graad van
professionele bachelor in de toegepaste informatica

Promotor:
Nathalie Declercq

Instelling: ToThePoint

Academiejaar: 2016-2017

Tweede examenperiode

Faculteit Bedrijf en Organisatie

Vergelijken voorspellingsalgoritmen met games

Matthias Seghers

Scriptie voorgedragen tot het bekomen van de graad van
professionele bachelor in de toegepaste informatica

Promotor:
Nathalie Declercq

Instelling: ToThePoint

Academiejaar: 2016-2017

Tweede examenperiode

Samenvatting

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus.

Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Voorwoord

Inhoudsopgave

1	Inleiding	9
1.1	Stand van zaken	9
1.2	Probleemstelling en Onderzoeksvragen	10
1.3	Opzet van deze bachelorproef	11
2	Methodologie	13
2.1	Keuze algoritmen	13
2.1.1	Superviseerd vs Ongesuperviseerd leren	13
2.2	Gesuperviseerde classificatie algoritmen	14
2.2.1	Logistische regressie	14
3	Uitwerking	17
3.1	Fase 1	17
3.1.1	Logistische regressie	17

4	Conclusie	19
	Bibliografie	21

1. Inleiding

De inleiding moet de lezer alle nodige informatie verschaffen om het onderwerp te begrijpen zonder nog externe werken te moeten raadplegen (Pollefliet, 2011). Dit is een doorlopende tekst die gebaseerd is op al wat je over het onderwerp gelezen hebt (literatuuronderzoek).

Je verwijst bij elke bewering die je doet, vakterm die je introduceert, enz. naar je bronnen. In \LaTeX kan dat met het commando `\textcite{}` of `\autocite{}`. Als argument van het commando geef je de “sleutel” van een “record” in een bibliografische databank in het Bib \TeX -formaat (een tekstbestand). Als je expliciet naar de auteur verwijst in de zin, gebruik je `\textcite{}`. Soms wil je de auteur niet expliciet vernoemen, dan gebruik je `\autocite{}`. Hieronder een voorbeeld van elk.

Knuth (1998) schreef een van de standaardwerken over sorteer- en zoekalgoritmen. Experts zijn het erover eens dat cloud computing een interessante opportuniteit vormen, zowel voor gebruikers als voor dienstverleners op vlak van informatietechnologie (Creager, 2009).

1.1 Stand van zaken

De onderzoeken en technieken om de noden van gebruikers te voorspellen kent de laatste jaren een grote opmars. Google en Facebook zijn dan ook volop bezig met eigen onderzoekscentra en technologieën te ontwikkelen. FAIR is de afkorting voor Facebook Artificial Intelligence Research er zijn 3 labo's in de wereld die constant opzoek zijn naar nieuwe mogelijkheden binnen AI. RankBrain is een algoritme van Google die a.h.v. artificiële intelligentie de ranking op de zoekpagina bepaald. De persoonlijke advertenties die

Google toont zijn veelal gegenereerd met algoritmes. Maar ook gerelateerde producten of "wat jouw ook kan interesseren-lijsten worden dikwijls door machine learning gegenereerd. Tensorflow is een open source API ontwikkelt door Google die je uiteraard gratis kan gebruiken om zelf artificiële toepassingen te maken. De API is ontwikkelt in Python, deze taal is dan ook een van de veel gebruikte in de data science.

In deze bachelorproef zullen er voorspellingen gemaakt worden op een arcademachine. ToThePoint zal het eerste bedrijf zijn die AI in een arcademachine zal verwerken. Aangezien er nog geen andere gelijkaardige voorbeelden te vinden zijn zal er van bij het begin onderzoek moeten worden hoe we het best zouden aanpakken en verklaren. De spelletjes worden bestuurd door zes knoppen en een joystick. Aan de hand van de snelheid, aantal keer een knop of joystickbeweging is gedaan zullen we kunnen onderscheiden welk spel de gebruiker aan het spelen is. Er bestaan al veel verschillende soorten algoritmen maar deze zijn niet allemaal geschikt voor deze casus en daar zal dus aan gewerkt moeten worden.

Artificiële Intelligentie is hot en zo komen er allerlei beginnende frameworks naar boven die reeds geïmplementeerde algoritmen bevatten of het eenvoudiger maken om ermee te starten. Tensorflow is een populair framework ontwikkelt door Google gemaakt in Python. Dit wordt gebruikt in verschillende producten van Google zoals, Google's stem herkenning, Google vertaler, het zoeken op afbeeldingen, etc. Eenvoudig uitgelegd werkt TensorFlow met deep learning of neurale netwerken. De voorbeelden van Tensorflow gaan bijna uitsluitend over image recognition, het is mogelijk om andere applicaties ermee te maken. Dit is dan ook de reden dat Google dit framework open source heeft gemaakt zodat ze kunnen zien waar er nog verbeteringen kunnen aangebracht worden en wat er nog allemaal mogelijk is. Een ander interessant framework is het accord-framework die volledig ontwikkelt is in C#. Hierin zitten veel geïmplementeerde algoritmen in die dan kan gebruikt worden door derden om een eigen applicatie te ontwikkelen.

1.2 Probleemstelling en Onderzoeksvragen

Machine learning is momenteel aan het boomen. Dit wordt nu zeer veel toegepast voor online advertising met Google en Facebook als de leiders. Maar ook meer en meer bedrijven beginnen zich te verdiepen in artificieel intelligentie. ToThePoint is zich hierop ook aan het voorbereiden. Doormiddel van een funproject willen ze zoveel als mogelijk bijleren op allerlei gebieden binnen de informatica.

Men heeft een arcade machine gekocht die ze volledig gaan customizen met verschillende technologieën. Op deze manier kunnen ze al hun kennis loslaten. Inclusief de kennis die ze zullen opdoen via deze bachelorproef. Hierdoor zullen ze dus iets bijleren over artificiële intelligentie en dit dan ook kunnen toepassen. Verder zal de machine geplaatst worden op allerlei jobbeurzen om gegevens van studenten op te slaan bijvoorbeeld. Het nut van de machine is niet alleen om bij te leren maar hij zal ook dienen als referentie naar klanten toe. Op deze manier kan ToThePoint aantonen tot wat ze instaat zijn.

In deze bachelorproef wordt er een vergelijkende studie gemaakt over twee verschillende

algoritmen. Het beste algoritme zal dan uiteindelijk geïmplementeerd worden in de arcade machine. Je kan hier heel ver in gaan. De voorspellingen kunnen bijvoorbeeld gemaakt worden door wat de meest gebruikte knop is of joystick beweging. Maar door de snelheid dat knoppen bedient worden en de frequentie van dezelfde knop kan bepaald worden welk spel er gespeeld wordt. Als het nog verder uitgewerkt wordt kan er naar toetsencombinaties gekeken worden maar dit is te vergaand voor deze bachelorproef. Er zal vooral gefocust worden op hoe een goed algoritme gekozen wordt. Het stappenplan die uitgewerkt zal worden kan dan ook toegepast worden op toekomstige projecten.

1.3 Opzet van deze bachelorproef

De rest van deze bachelorproef is als volgt opgebouwd:

In Hoofdstuk 2 wordt de methodologie toegelicht en worden de gebruikte onderzoekstechnieken besproken om een antwoord te kunnen formuleren op de onderzoeksvragen.

In Hoofdstuk 4, tenslotte, wordt de conclusie gegeven en een antwoord geformuleerd op de onderzoeksvragen. Daarbij wordt ook een aanzet gegeven voor toekomstig onderzoek binnen dit domein.

2. Methodologie

2.1 Keuze algoritmen

Om goede resultaten te behalen is het uiterst belangrijk dat de juiste algoritmen gebruikt worden. Zo zijn er bepaalde algoritmen die helemaal niet bruikbaar zouden kunnen zijn voor deze casus. Er zal hier verder uitgelegd worden hoe we een mogelijk algoritme kunnen vinden.

2.1.1 Superviseerd vs Ongesuperviseerd leren

Machine learning kan onderverdeelt worden in verschillende types van algoritmen. Deze zijn gesuperviseerd leren, ongesuperviseerd leren en reinforcement leren. Dit laatste type is niet van toepassing voor deze casus. Met dit type wordt er geleerd op basis van positieve signalen (beloningen). Er wordt ook niet gebaseerd op een dataset en aangezien wij in bezit zijn van datasets is dit type overbodig om verder te onderzoeken.

Superviseerd leren

Dit type heeft als doel om een hypothese te bekomen die dan zal kunnen gebruikt worden om nieuwe ongekende input toe te wijzen aan een label die voorkwam in de eerdere trainingsdataset. De trainingsdataset bestaat uit verschillende parameters en een label. Onder dit type kan je algoritmen vinden die voor zowat alle cases gebruikt kunnen worden. Deze zijn echter wel nog opgedeeld in drie verschillende categorieën. Zo kan je een nieuwe waarde voorspellen op basis van vroegere resultaten, dit wordt regressie genoemd. Verder heb je classificatiealgoritmen hiermee kan een input toegewezen worden aan een

bepaalde klasse met een label. En als laatste bestaan er clusteringsalgoritmen waarmee je ook onderverdelingen kunt maken in klassen maar deze hebben geen label. Clustering en classificatie lijken op elkaar maar met classificatie weet een onderzoeker ook precies wat de data voorstelt.

Ongesuperviseerd leren

Verschillend met gesuperviseerd leren beschikt een ongesuperviseerd algoritme over een ongelabelde dataset. Er zijn verschillende inputs maar die behoren niet tot een specifieke klasse. De meest gebruikte techniek is dus clustering. Als we dit bekijken voor deze casus is dit geen optimale manier. We kunnen wel bepaalde besturingsevents clusteren waardoor je tot een x aantal clusters kunt komen maar we hebben geen idee of de ene cluster Pac-Man of Mortal Combat is bijvoorbeeld.

misschien een voorbeeldje uitwerken?

2.2 Gesuperviseerde classificatie algoritmen

In deze casus beschikken we over een gelabelde dataset. Het doel is om met een gegeven input een concreet spel te krijgen als output. Gesuperviseerde classificatiealgoritmen zijn dus de meest geschikte voor deze in dit onderzoek. Doordat we een gelabelde dataset hebben en er moeten verdelingen gemaakt worden op basis van labels. Op deze manier kunnen we ongekende input plaatsen bij één bepaald spel.

2.2.1 Logistische regressie

Logistische regressie is het eerste algoritme die we zullen bespreken in deze bachelorproef. Zoals u uit de naam kan afleiden maakt dit algoritme gebruik van een logistische functie. Aan de hand van een vergelijking kunnen er voorspellingen worden gemaakt. Gedurende het trainingsproces wordt deze vergelijking geoptimaliseerd. Er zullen meerdere spelletjes voorzien zijn op de arcademachine dus ook meerdere klassen. Omdat we van nul af aan starten zal ook de binaire logistische regressie gebruikt worden en vervolgens twee mogelijke methoden om met meerdere klassen te werken.

Binaire logistische regressie

Zoals de naam al doet vermoeden wordt met dit algoritme een classificatie gemaakt tussen twee klassen. Er wordt gestart met een lineaire regressie als hypothese. De waarde y zal altijd binnen de grenzen 0 en 1 liggen. Door middel van de sigmoïdfunctie krijgen we altijd een waarde tussen 0 en 1 exclusief de grenzen. Het functievoorschrift is als volgt.

$$y(x) = \frac{1}{1 + e^{-x}}$$

Theorie moet nog verder uitgewerkt worden

Multiklasse logistische regressie

3. Uitwerking

3.1 Fase 1

Om goede resultaten te behalen is het uiterst belangrijk dat de juiste algoritmen gebruikt worden. Zo zijn er bepaalde algoritmen die helemaal niet bruikbaar zouden kunnen zijn voor deze casus. Er zal hier verder uitgelegd worden hoe we een mogelijk algoritme kunnen vinden.

3.1.1 Logistische regressie

4. Conclusie

Curabitur nunc magna, posuere eget, venenatis eu, vehicula ac, velit. Aenean ornare, massa a accumsan pulvinar, quam lorem laoreet purus, eu sodales magna risus molestie lorem. Nunc erat velit, hendrerit quis, malesuada ut, aliquam vitae, wisi. Sed posuere. Suspendisse ipsum arcu, scelerisque nec, aliquam eu, molestie tincidunt, justo. Phasellus iaculis. Sed posuere lorem non ipsum. Pellentesque dapibus. Suspendisse quam libero, laoreet a, tincidunt eget, consequat at, est. Nullam ut lectus non enim consequat facilisis. Mauris leo. Quisque pede ligula, auctor vel, pellentesque vel, posuere id, turpis. Cras ipsum sem, cursus et, facilisis ut, tempus euismod, quam. Suspendisse tristique dolor eu orci. Mauris mattis. Aenean semper. Vivamus tortor magna, facilisis id, varius mattis, hendrerit in, justo. Integer purus.

Vivamus adipiscing. Curabitur imperdiet tempus turpis. Vivamus sapien dolor, congue venenatis, euismod eget, porta rhoncus, magna. Proin condimentum pretium enim. Fusce fringilla, libero et venenatis facilisis, eros enim cursus arcu, vitae facilisis odio augue vitae orci. Aliquam varius nibh ut odio. Sed condimentum condimentum nunc. Pellentesque eget massa. Pellentesque quis mauris. Donec ut ligula ac pede pulvinar lobortis. Pellentesque euismod. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent elit. Ut laoreet ornare est. Phasellus gravida vulputate nulla. Donec sit amet arcu ut sem tempor malesuada. Praesent hendrerit augue in urna. Proin enim ante, ornare vel, consequat ut, blandit in, justo. Donec felis elit, dignissim sed, sagittis ut, ullamcorper a, nulla. Aenean pharetra vulputate odio.

Quisque enim. Proin velit neque, tristique eu, eleifend eget, vestibulum nec, lacus. Vivamus odio. Duis odio urna, vehicula in, elementum aliquam, aliquet laoreet, tellus. Sed velit. Sed vel mi ac elit aliquet interdum. Etiam sapien neque, convallis et, aliquet vel, auctor non, arcu. Aliquam suscipit aliquam lectus. Proin tincidunt magna sed wisi. Integer blandit

lacus ut lorem. Sed luctus justo sed enim.

Morbi malesuada hendrerit dui. Nunc mauris leo, dapibus sit amet, vestibulum et, commodo id, est. Pellentesque purus. Pellentesque tristique, nunc ac pulvinar adipiscing, justo eros consequat lectus, sit amet posuere lectus neque vel augue. Cras consectetur libero ac eros. Ut eget massa. Fusce sit amet enim eleifend sem dictum auctor. In eget risus luctus wisi convallis pulvinar. Vivamus sapien risus, tempor in, viverra in, aliquet pellentesque, eros. Aliquam euismod libero a sem.

Nunc velit augue, scelerisque dignissim, lobortis et, aliquam in, risus. In eu eros. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Curabitur vulputate elit viverra augue. Mauris fringilla, tortor sit amet malesuada mollis, sapien mi dapibus odio, ac imperdiet ligula enim eget nisl. Quisque vitae pede a pede aliquet suscipit. Phasellus tellus pede, viverra vestibulum, gravida id, laoreet in, justo. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Integer commodo luctus lectus. Mauris justo. Duis varius eros. Sed quam. Cras lacus eros, rutrum eget, varius quis, convallis iaculis, velit. Mauris imperdiet, metus at tristique venenatis, purus neque pellentesque mauris, a ultrices elit lacus nec tortor. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent malesuada. Nam lacus lectus, auctor sit amet, malesuada vel, elementum eget, metus. Duis neque pede, facilisis eget, egestas elementum, nonummy id, neque.

Bibliografie

- Creeger, M. (2009). CTO Roundtable: Cloud Computing. *Communications of the ACM*, 52(8), 50–56.
- Knuth, D. E. (1998). *The art of computer programming, volume 3: (2nd ed.) sorting and searching*. Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc.
- Polleffiet, L. (2011). *Schrijven van verslag tot eindwerk: do's en don'ts*. Gent: Academia Press.

Lijst van figuren

Lijst van tabellen