

1 Genetische Algorithmen

1.1 Zusammenfassung

Genetische Algorithmen oder auch Evolutionäre Algorithmen beschreiben eine Gruppe von Verfahren die zur Lösung von Optimierungsproblemen genutzt werden kann.

Hierzu wird wie in vielen Machine Learning Verfahren ein Optimum durch die Methode des Statistic Gradient Decent gesucht. Die Suche basiert dabei auf der Evolutionstheorie und der Genetik.

Hierbei wird die vereinfachte Anschauung von Darwin verwendet:

Der Stärkere überlebt und trägt sein Erbgut weiter, der schwächere stirbt aus. Auf diese Weise soll dafür gesorgt werden, dass es nach einer gewissen Zeit oder einer gewissen Anzahl an Generationen nur noch starke Individuen gibt.

1.2 Grundbegriffe

1.2.1 Gen

Ein Gen stellt eine Stelle in der DNA dar und enthält eine einzelne Erbinformation und somit eine Teileigenschaft eines Individuums. Diese Erbinformation kann auf verschiedene Arten codiert werden. Es sollte eine günstige Codierung für das entsprechende Problem gewählt werden.

1.2.2 Genotyp die DNA

Die DNA oder in der Genetik der Genotyp stellt die Eigenschaften oder Verhaltensweisen eines Individuums in einer vom Computer leicht zu verarbeitenden und für das Problem geeigneten Form da. Dabei besteht die DNA aus einer Reihe von Genen.

1.2.3 Phänotyp das Individuum

Das Individuum oder in der Genetik der Phänotyp ist die Interpretation der DNA als möglicher Lösungskandidat des zu lösenden Problems. Das kann im speziellen bedeuten, dass die DNA als Eigenschaften, Verhaltensweisen oder Handlungsabläufe des Individuums interpretiert werden können.

1.2.4 Allel

Allel ist die Bezeichnung für eine spezielle Ausprägung eines Gens, also eine festes Wort aus dem gewählten Alphabet der Codierung.

1.2.5 Population

Die Population ist die Menge aller Individuen. Sie hat in der regel eine feste Populationsgröße.

1.2.6 Generation

Eine Generation bezeichnet die Population zu einem bestimmten Iterationsschritt innerhalb des Algorithmus. Bei jedem Iterationsschritt innerhalb des Algorithmus spricht man von einer neuen Generation.

1.2.7 Fitness

Die Fitness gibt an wie gut ein Individuum das zu lösende Problem löst. Sie wird mithilfe der Fitnessfunktion für jedes Individuum berechnet. Sie ist dann Basis für die Entscheidung welche Individuen überleben.

1.2.8 Fitnessfunktion

Die Fitnessfunktion dient zur Berechnung der Fitness eines Individuums. Sie ist die Quantitative Formulierung des zu optimierenden Problems. Hierbei wird der Mehrdimensionale Problemraum in einen skalaren Raum transformiert. Wird diese Transformation falsch formuliert wird nicht das eigentlich gewollte Problem gelöst sondern das Problem das durch die Fitnessfunktion dargestellt wurde. Die Fitnessfunktion stellt damit den wichtigsten Grundbaustein des Genetischen Algorithmus dar. Gerade bei komplexen Problemen ist sie aber auch oft schwer zu formulieren.

1.3 Ablauf

1.3.1 Übersicht

Beim Genetischen Algorithmus werden folgende Schritte der Evolution nachgestellt:

- 1) Bevölkerung / Population: Es wird eine zufällige Population erstellt
- 2) Bewertung / Ranking / Fitness: Die Population wird anhand einer Bewertungsfunktion bewertet.
- 3.1) Überleben / Survival: Anhand der Bewertung wird eine prozentuale Überlebenswahrscheinlichkeit der einzelnen Individuen ermittelt.
- 3.2) Auswahl / Selektion: es wird zufällig aber gewichtet nach der Überlebenswahrscheinlichkeit aus den Individuen der Grundstamm für die nächste Population ausgewählt.
- 4.1) Fortpflanzung / Crossover: Es werden Nachkommen aus der zufälligen Kreuzung des Grundstamms erzeugt.
- 4.2) Mutation / Mutation: Die Nachkommen werden mit einer geringen Wahrscheinlichkeit zufällig leicht verändert.

--> Es wird wieder 2) die Bewertung der neuen Individuen durchgeführt.

Hierbei gilt zu beachten, dass man zwischen der DNA des Individuums und dem Individuum selbst häufig eine Trennung vollzieht.

1.3.2 Bevölkerung

Zu Beginn des Genetischen Algorithmus muss eine Anfangspopulation erzeugt werden, welche als Generation 0 bezeichnet wird. Hierzu werden meist zufällige DNA Stränge erzeugt. Es besteht allerdings auch die Möglichkeit, die Generation 0 mit einem gewissen Vorwissen zu initialisieren, hier besteht allerdings die Gefahr, dass das Vorwissen nur ein lokales Maximum der Problemfunktion beschreibt. In diesem Fall würde der Algorithmus nur das lokale Maximum finden, nicht aber das globale Maximum.

1.3.3 Bewertung

Es muss in jedem Iterationsschritt eine Bewertung der Individuen erfolgen. Hierzu wird mithilfe der Fitnessfunktion die Fitness jedes Individuums berechnet. Die Fitnessfunktion muss hierbei das zu lösende Problem so abbilden, dass das globale Maximum der Fitnessfunktion mit dem globalen Maximum der Problemfunktion übereinstimmt. Des Weiteren sollte auch der Gradient der Fitnessfunktion möglichst immer in Richtung des globalen Maximum zeigen, da sonst die Gefahr besteht, nur ein lokales Maximum zu finden. Es gehört also zu einer der Hauptaufgaben des Entwicklers eines Genetischen Algorithmus, eine passende Fitnessfunktion zu finden.

1.3.4 Survival

Anhand der Fitness der einzelnen Individuen wird eine Überlebenswahrscheinlichkeit berechnet. Hier können verschiedene Methoden zur Berechnung der

Überlebenswahrscheinlichkeit eingesetzt werden. Die Überlebenswahrscheinlichkeit bestimmt letztendlich wie Wahrscheinlich es ist, dass ein Individuum für die nächste Generation als Elternteil gewählt wird. Die Überlebenswahrscheinlichkeiten aller Individuen zusammen addiert ergibt 1.

1.3.5 Selektion

Bei der Selektion werden nun zufällig aber gewichtet nach der Überlebenswahrscheinlichkeit Individuen ausgewählt. Diese dienen als Grundlage für die neue Generation, sind also sozusagen die Eltern der neuen Generation. Je nachdem nach welchem Verfahren hier die Individuen gewählt werden, kann sich die Zusammensetzung der nächsten Generation ändern.

1.3.6 Fortpflanzung

Bei der Fortpflanzung werden die Gene zweier Eltern miteinander gemischt. Hierzu werden je zwei oder auch mehr Individuen ausgewählt die gekreuzt werden sollen und ein genetischer Operator auf sie angewendet. Durch diese Paarung sollen neue Bereiche der Problemfunktion erkundet werden, und er dient somit als einer der wichtigsten Suchoperatoren. Der genetische Operator bestimmt hierbei wie eine Kreuzung durchgeführt wird, dass bedeutet welche Teile der beiden Eltern für das Kind genutzt werden.

1.3.7 Mutation

Die Mutation erzeugt aus einem Elternteil ein zufälliges neues Individuum das sich anders nicht aus den Eltern bilden hätte können. Sie bringt also ganz neue Erbinformationen in die neue Generation ein. Auch hier können verschiedene Mutationsoperatoren verwendet werden. Generell hat die Mutation in vielen Genetischen Algorithmen nur eine Hintergrundrolle, so wird sie nur mit einer sehr geringen Mutationswahrscheinlichkeit durchgeführt.

1.4 Umsetzungen

1.4.1 Fitnessfunktion

Die Fitnessfunktion ist sehr Problemspezifisch für jedes neue Problem muss zunächst eine geeignete neue Fitnessfunktion gefunden werden. Dies ist Aufgabe des Entwicklers und kann viel Zeit in Anspruch nehmen da nicht für jedes Problem eine Fitness einfach berechenbar ist. So müssen für viele Anwendungen erst aufwendige Simulationen durchgeführt werden bevor die Fitness berechnet werden kann.

1.4.2 Survival

So wird die Überlebenswahrscheinlichkeit im einfachsten Fall linear zur Fitness gewählt. Es sind aber auch andere Funktionen möglich um die Überlebenswahrscheinlichkeit zu berechnen. So könnte sie auch zum Quadrat der Fitness oder Exponentiell zur Fitness gewählt werden. In diesen Fällen haben Individuen mit höherer Fitness eine wesentlich höhere Wahrscheinlichkeit zu überleben als Individuen mit geringer Fitness. Somit konvergiert der Genetische Algorithmus schneller zu einem Optimum, womöglich aber auch zu einem lokalen Optimum. Die

1.4.3 Selektion

Die Selektion wird hierbei häufig durch ein Glücksrad umgesetzt. Jedes Feld auf dem Rad ist hierbei einem Individuum zugeordnet, die Größe des Feldes ist entsprechend der Überlebenswahrscheinlichkeit gewählt. Das Rad wird nun so oft gedreht bis die gewünschte Anzahl an Eltern ausgewählt wurde. Das stochastische universelle Samplingrate ist eine Abwandlung des Glücksrad, hier wird nur einmal gedreht, dafür gibt es so viele Zeiger wie Individuen gezogen werden sollen. Die Zeiger haben alle den gleichen Abstand voneinander, hierdurch wird vermieden, dass sehr viele Individuen von nur einer Sorte gezogen werden können. Die Varianz zwischen den Individuen ist also höher und es besteht somit eher die Möglichkeit neue Bereiche der Problemfunktion zu erkunden. Ebenfalls denkbar ist nur die Wahl der N Besten Individuen, hier kann die Konvergenzgeschwindigkeit bestimmt werden, es besteht aber auch wieder die Gefahr einer zu schnellen Konvergenz.

1.4.4 Fortpflanzung

1.4.4.a Auswahl

Die Auswahl der Elternpaare kann zufällig geschehen und für alle Eltern durchgeführt werden, hier gibt es dann zudem eine Wahrscheinlichkeit die bestimmt ob die Eltern überhaupt gekreuzt werden sollen, oder unverändert in die neue Generation übernommen werden. Hier hat sich eine Kreuzungswahrscheinlichkeit von 0.6 etabliert.

Es können alternativ auch nur 60 Prozent der Eltern zufällig zur Kreuzung ausgewählt werden und alle Verbleibenden bleiben unverändert.

1.4.4.b Genetische Operatoren

Als Genetische Operatoren kann eine Vielzahl an Operationen vollzogen werden, so kann ein DNA Strang einfach in der Mitte geteilt werden. Und die Teile der beiden Eltern werden für das Kind genutzt. Es ist auch eine Teilung an einer Zufällig bestimmten Stelle denkbar (1 Punkt Crossover), oder die DNA wird an mehreren Stellen geteilt (n Punkt Crossover). Bei einer DNA die aus R^2 Genen besteht ist auch eine Teilung wie beim Menschen

möglich. Das bedeutet die Komponenten der Vektoren werden unter den Eltern getauscht. Beim Uniform Crossover wird für jedes Gen einzeln bestimmt ob es getauscht werden soll oder nicht. Hier kann mit einer Austauschwahrscheinlichkeit gearbeitet werden, die häufig bei 50% liegt. Es können auch mehrere dieser Operatoren Kombiniert werden.

1.4.5 Mutation

Bei der Mutation können verschiedene Operatoren angewandt werden. Diese werden mit einer meist geringen Wahrscheinlichkeit von beispielsweise 10% durchgeführt. Beim Random Resetting werden hier den einzelnen Genen zufällige neue Werte gegeben. Bei der Greep Mutation wird die quantitative Reihenfolge der DNA weitgehend beibehalten, hierzu wird ein kleiner zufälliger Wert zu den entsprechenden Genen dazu addiert bzw. subtrahiert.