

Joachim Breitner

Seminarkurs am Schickhardt-Gymnasium Herrenberg

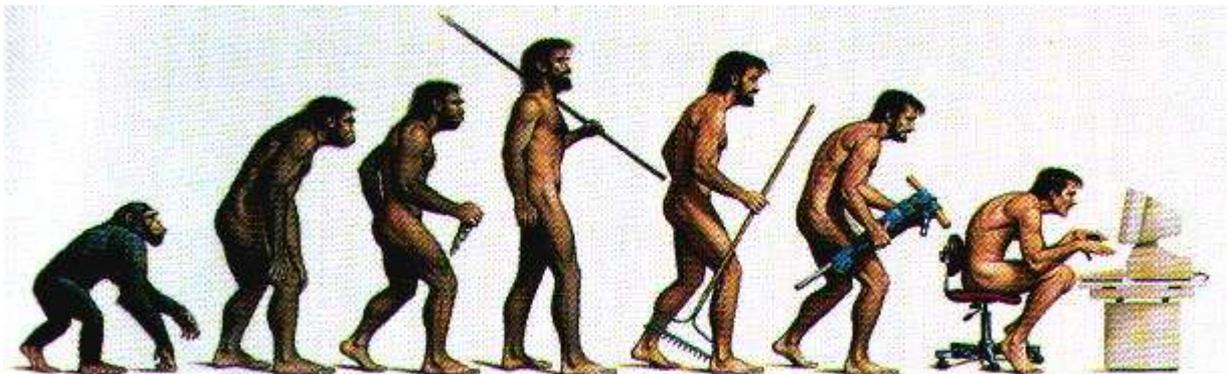
2002/2003 Halbjahr 12.2

Modellbildung und Simulation

am Beispiel der

Evolution

und dem Programm „EvoLab“



Thema 2:

Der genetische Algorithmus

Inhaltsverzeichnis

Einführung.....	2
Definitionen.....	2
Genom.....	2
Universalität.....	2
Zielfunktion.....	3
Der genetische Algorithmus.....	3
Flussdiagramm.....	4
Die Startpopulation.....	4
Der Fitnessstest.....	4
Die Auslese.....	4
Reproduktion.....	5
Ergebnis.....	5
Der Zufall.....	5

Einführung

Dieses Dokument ist die schriftliche Form der Präsentation von Joachim Breitner am Montag, dem 23.6.2003 zu diesem Thema. Sie reiht sich ein in die Folge der insgesamt vier Beiträge zu diesem Themenbereich, und setzt den ersten Beitrag, „Modellierung der 'Ameisen' und ihrer Welt in EvoLab“ voraus. Das Thema dieses Dokumentes ist die Theorie hinter dem Programm EvoLab, genauer gesagt der genetische Algorithmus.

Definitionen

Genom

Eigentlich bezeichnet das Wort „Genom“ die Summe der Gene eines Lebewesens. Im unserem Kontext jedoch wird diese Bedeutung auf das, was man per Evolution zu verbessern versucht, verallgemeinert. Das kann von Aufgabe zu Aufgabe sehr unterschiedlich sein. In EvoLab ist es die Übergangstabelle eines Zustandsautomaten, in technischen Anwendungen kann es ein Konstruktionsplan, eine Materialzusammensetzung oder Ähnliches darstellen. Inbegriffen ist weiterhin die natürliche Bedeutung als Summe der Gene eines natürlichen Lebewesens. Im Folgenden wird stets der Begriff „Genom“ in seiner allgemeinen Form verwendet.

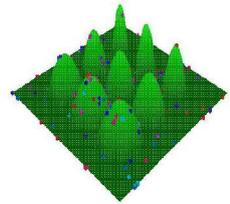
Universalität

Die Theorie der Universalität besagt, dass der genetische Algorithmus auf eine Vielzahl von Anwendungen neben ihrer ursprünglichen, der natürlichen Evolution, übertragbar ist. Eine kleine Auswahl sei hier zusammengestellt:

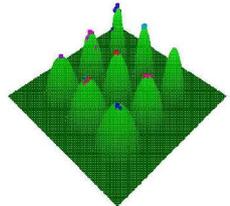
- Die Formoptimierung von Fahr- oder Flugzeugen auf möglichst geringen Luftwiderstand
- Die Steuerung von autonomen oder zentralcomputergesteuerten Robotern
- Die „Intelligenz“ der Gegner in Computerspielen
- Das Problem des Handlungsreisenden
- Die optimale Art, Kisten oder Ähnliches auf begrenztem Raum unterzubringen
- Die Entwicklung von Pfandfindermodellameisen in einem zweidimensionalen torusförmigen Universum mit digitaler, schachbrettartiger Oberfläche.

Zielfunktion

Die Zielfunktion sieht je nach gestellter Aufgabe anders aus, und ihre Parameter hängen von der jeweiligen „Umwelt“ ab. Allgemein gesagt ist die Zielfunktion die Funktion, die man zu Maximieren oder zu Minimieren versucht. Im Rahmen der Evolution nennt man die Zielfunktion auch Fitnessfunktion, da der Wert, den sie jedem Individuum und damit auch jedem Genom, zuweist, beschreibt, wie „fit“ dieses Individuum ist. Beispielsweise ist die Zahl der erreichten Etappen eines Genomes in EvoLab seine Fitness, und diese hängt stark vom gewählten Parcours ab.



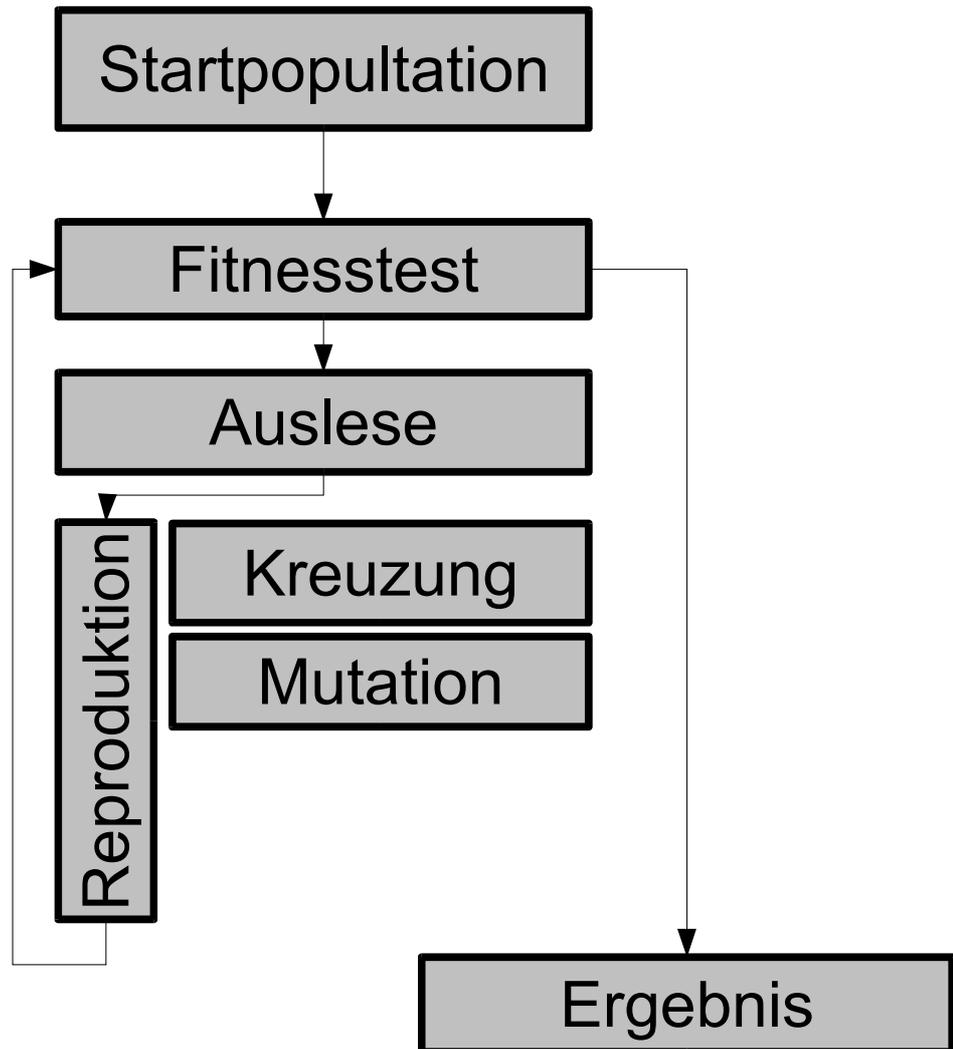
Das Schaubild der Zielfunktion bezeichnet man als „Fitnesslandschaft“. Diese ist meist nicht darstellbar, da die Zielfunktion Genome als Eingabe erwartet, von der es Abermilliarden Mögliche gibt. Die beiden Graphiken hier sind stark vereinfachte Modelle: Auf der Ebene sind die möglichen Genome, deren Fitness nach oben hin dargestellt ist. Die Punkte sind die Genome einer Generation – einmal am Anfang des Prozesses, und einmal gegen Ende. Gut erkennbar sind die lokalen Maxima, in denen sich auch Genome sammeln. Unter Umständen kann es sein, dass die Genome sich ausschließlich in lokalen Maxima befinden und nicht mehr herauskommen. In dem Fall ist die Evolution gescheitert.



Der genetische Algorithmus

Die Definition des Algorithmus lautet „Ein Algorithmus ist ein Plan zur Lösung gleichartiger Probleme mit Hilfe einer endlichen Folge eindeutiger, ausführbarer Schritte.“ In unserem Fall lautet das Problem: „Finde das optimale Genom.“ Dabei kann „Genom“, wie bereits angesprochen, vieles bedeuten. Was es im konkreten Fall bedeutet, hängt vom Problem ab. Der genetische Algorithmus wurde von der natürlichen Evolution hergeleitet, die bereits seit Millionen von Jahren sehr erfolgreich gewesen ist – schuf sie doch ein Wesen, das Texte über die natürliche Evolution verfassen kann.

Flussdiagramm



Die Startpopulation

Am Anfang der Evolution steht die Startpopulation. Diese ist eine größere Anzahl an Individuen, die die erste Generation stellen. Ihre Herkunft ist je nach Aufgabe unterschiedlich:

In Evolab werden die Genome der Startpopulation zufällig ermittelt. In technischen Anwendungen entweder ebenso, oder man geht von einem Prototypen oder existierendem Modell aus, das man verbessern möchte. In der Natur ist der Ursprung umstritten. Es könnte sich die DNS zufällig in der sogenannten Ursuppe gebildet haben; sie könnte aus dem Weltall per Meteorit auf die Erde gekommen sein oder es gab einen Schöpfungsakt durch ein höheres Wesen.

Der Fitnesstest

Jede Generation wird einem Fitnesstest unterzogen. Das heißt, man wendet die Zielfunktion auf jedes Individuum bzw. Genom an, um seine Fitness festzustellen. Was die Fitness im jeweiligen Fall bedeutet, hängt von der Aufgabe ab:

In EvoLab ist sie die Anzahl der erreichten Etappen. In technischen Anwendungen kann sie beispielsweise den Luftwiderstand oder die Tragfähigkeit darstellen. In der Natur ist ein Individuum dann „fit“, wenn es überleben und sein Erbgut weitergeben konnte – je öfter desto besser.

Die Auslese

Nur eine gewisse Anzahl der Genome einer Generation, nämlich die fittesten, überleben und bilden die Eltern der nächsten Generation. Die Zahl der überlebenden Genome ist bei technischen Anwendungen wie auch bei EvoLab festgelegt. In der Natur jedoch ist sie dynamisch und hängt von den Ressourcen sowie von den Feinden und Gefahren einer Spezies ab.

Reproduktion

Die ausgewählten Elterngenome werden zu einer neuen Generation vervielfältigt. Das geschieht in EvoLab – in Anlehnung an die natürliche Evolution – in zwei Schritten:

Zuerst werden zwei zufällig ausgewählte Elterngenome gekreuzt, das heißt jede Spalte der Übergangstabelle kommt entweder vom „Vatergenom“ oder vom „Muttergenom“. Daraufhin werden noch ein bestimmter, zufällig ausgewählter, Anteil der Gene zufällig mutiert, das heißt abgeändert.

Ergebnis

War beim Fitnessstest ein Genom dabei, dass das Ziel erreicht hat, war der Algorithmus erfolgreich und dieses Genom ist das Ergebnis. In EvoLab lautet das Ziel, alle Etappe zu erreichen. In den technischen Anwendungen ist das Ziel auch meist vorgegeben. Da die Evolution nicht immer den gewünschten Erfolg bringt, wird in EvoLab wie auch in technischen Anwendungen die Anzahl der Generationen beschränkt. Wird diese Schranke erreicht, bevor ein Ergebnis vorliegt, so ist die Evolution gescheitert und man muss die Parameter oder die Methode überdenken. In der Natur endet die Evolution nie, es sei denn eine Spezies stirbt aus, da sie sich ständig an neue Umgebungsbedingungen anpassen muss.

Der Zufall

Der Zufall spielt eine wichtige Rolle bei der Evolution. Ausgehend vom Modell in EvoLab, wirkt der Zufall an drei Punkten:

- Die Genome der Startpopulation sind zufällig gewählt.
- Die Elterngenome eines neuen Genomes werden zufällig ausgewählt.
- Bei der Mutation werden zufällig ausgewählte Genome zufällig verändert.

Trotz dem großen Einfluss des Zufalls bei der Evolution ist diese wesentlich leistungsfähiger als der reine Zufall. Das kann man an zwei großen Zahlen verdeutlichen:

- Bei 400 Schritten, die eine Ameise in EvoLab zurücklegt, hat sie jeweils 4 mögliche Richtungen. Das ergibt $4^{200} = 2,5 \times 10^{123}$ mögliche Wege. Von diesen erreicht nur ein verschwindend geringer Teil sämtliche Etappen.
- Jede Zelle in der Entscheidungsmatrix in EvoLab in er Standardeinstellung hat 4 Zustände mal 4 Richtungen, also 16 mögliche Werte, und das 4 Zeilen mal 8 Spalten oft ergibt $(4 \cdot 4)^{4 \cdot 8} = 3,4 \times 10^{38}$ mögliche Genome. Auch hier ist die Anzahl der erfolgreichen Genome extrem klein.

