

Schwarzfleckigkeit von Kartoffeln

Bestimmung mit Hilfe von Neuronalen Netzen auf einem Rechner-Cluster

Die Vorhersage der Schwarzfleckigkeit dient zur Qualitätssteigerung des Agrarprodukts Kartoffel. Die Bestimmung der relevanten Einflussfaktoren ist entscheidend für die Erstellung eines leistungsfähigen Neuronalen Netzes (NN). Zur Ermittlung der besten Einflussfaktoren ist die Erzeugung einer großen Anzahl an NN notwendig, die eine sehr hohe Rechenleistung erfordert. Die Kopplung mehrerer Rechner zu einem Cluster stellt die Lösung dieses Problems dar.

Die Schwarzfleckigkeit bei Kartoffeln ist ein Qualitätsmangel, der nur durch destruktive Methoden messbar ist. Der Einzelhandel, das verarbeitende Gewerbe und der Konsument sind nicht in der Lage, den Qualitätsmangel Schwarzfleckigkeit durch optische Methoden zu erfassen. Eine leistungsfähige Vorhersage reduziert die benötigten Proben auf ein Minimum und muss trotzdem in der Lage sein, eine befriedigende Genauigkeit zu liefern. Vor diesem Hintergrund wurde im Institut für Agrartechnik Bornim e.V. (ATB) eine Methode zur Vorhersage von Schwarzfleckigkeit bei Kartoffeln entwickelt, die auf der Lernfähigkeit von NN beruht. Im ATB werden die Inhaltsstoffe und der Anteil der Schwarzfleckigkeit bei jeweils 240 Kartoffeln der drei Sorten Likaria, Adretta und Koretta seit 1995 gemessen. Die in der Online-Datenbank gesammelten Informationen über die Inhaltsstoffe der Kartoffeln werden als Eingabe für die NN benutzt, während die tatsächlich gemessene Schwarzfleckigkeit den Zielparameter darstellt. Die Abweichung zwischen Vorhersage und gemessener Schwarzfleckigkeit dient als Fehler und wird in das NN zur Korrektur der internen Gewichte der einzelnen Neuronen verwendet. Da wir als NN ein Backpropagation Netz verwenden, wird der Fehler von den Ausgabe-Neuronen über die Neuronen in der sogenannten Hidden-Layer bis hin zu den Eingabe-Neuronen verteilt. Die Berechnung eines NN, das auf der Eingabe aller Einflussfaktoren basiert, führt zu einer sehr ungenauen Vorhersage, da die Einflussfaktoren, die keinen Einfluss auf die Güte der Vorhersage der Schwarzfleckigkeit haben, das NN verfälschen. Auch ist es mit nur einem NN nicht wahrscheinlich, die richtige Vorverarbeitung der Eingabedaten auf Anhieb herauszufinden. Daher ist es sinnvoller, sehr viele NN zu berechnen, die sich hinsichtlich der Eingabeparameter und der Vorverarbeitung unterscheiden.

Varianten trainiert, die sich hinsichtlich der Vorverarbeitung der Einflussfaktoren unterscheiden. Der von uns verwendete Neuronale Netzwerk-Typ – Backpropagation – kann nur einen linearen Einfluss der Eingabe-Neuronen auf die Ausgabe-Neuronen abbilden. Daher ist es notwendig, die Eingabedaten vorzuverarbeiten und Varianten der NN zu erzeugen, die sich nur hinsichtlich der Vorverarbeitung der Daten unterscheiden. Wir haben uns auf die vier Vorverarbeitungsoperatoren x^{-2} , x^{-1} , x^1 und x^2 beschränkt. Bei einem NN aus zwei Einflussfaktoren müssen somit 16 Varianten durchtrainiert werden. Die Kombinationen aller Einflussfaktoren ergeben eine große Summe an zu berechnenden NN. Der Rechenaufwand ist dabei abhängig von der Iterationstiefe des Lernprozesses, der Anzahl der Hidden-Layer Schichten und der Anzahl der Neuronen in der Hidden-Layer. Hinzu kommt, dass Kartoffeln einem degenerativen Prozess unterliegen, und daher die jetzigen Kartoffelsorten durch andere ersetzt und infolgedessen die NN fortlaufend angepasst werden müssen. In der Praxis bedeutet dies, dass die NN einem ständigem Lernprozess ausgesetzt sein müssen, um eine gleichbleibende Qualität zu erreichen. Jedoch ist der Rechenaufwand so erheblich, dass das mit der damals zur Verfügung stehenden PC-Technik nicht praktikabel erschien. Da die Berechnungen der NN in sich abgeschlossen sind, war der Gedanke naheliegend, die Berechnung verschiedener NN auf mehreren Rechnern parallel auszuführen und die Ergebnisse abschließend zusammenzuführen. Die Verteilung der Messdaten und die Synchronisierung der Ergebnisse sollen automatisch erfolgen. Da die Software in einem dynamischen Entwicklungsumfeld steht, soll eine automatische Abgleichung der Daten zwischen den beteiligten Computern nicht nur für den Austausch der Daten gelten, sondern darüber hinaus die verwendete Software konsistent halten.

Sascha Richter ist Doktorand und Dr.-Ing. Klaus Gottschalk wissenschaftlicher Mitarbeiter des ATB, Abt. 3, Max-Eyth-Allee 100, 14469 Potsdam.

Schlüsselwörter

Schwarzfleckigkeit bei Kartoffeln, Neuronale Netze, Cluster

Keywords

Blackleg of potatoes, neuronal networks, cluster

Varianten von Neuronalen Netzen

Damit die Güte eines Einflussfaktors hinsichtlich des Einflusses auf die Schwarzfleckigkeit bewertet werden kann, wird eine Klassifikation der NN vorgenommen. Jedes NN wird darüber hinaus in verschiedenen

Konzeption des Clusters

Der Cluster wird durch einen zentralen Computer geleitet. Dieser Computer, der Leitwolf, hat vielfältige Aufgaben.

- Benutzerinterface

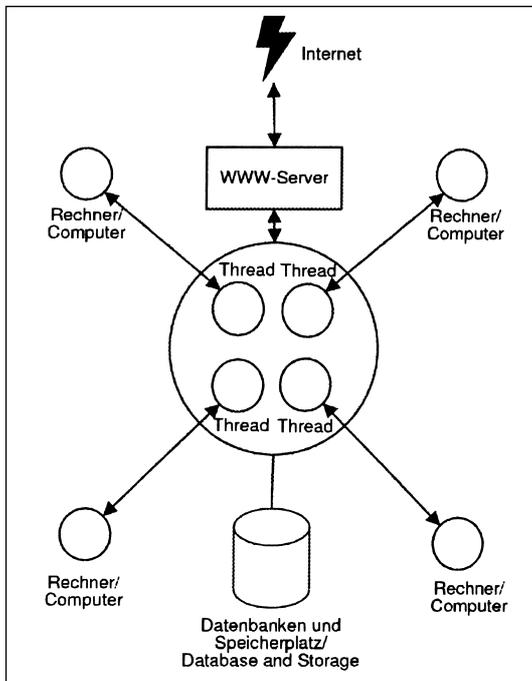


Bild 1: Struktur des Clusters

Fig. 1: Structure of the cluster

- Verwaltung der Cluster-Software
- Datenbankverwaltung
- Verwaltung der Benutzer-Software
- Aufgabenverteilung auf die einzelnen Rechner im Cluster

Benutzerinterface

Ein zeitgemäßes Benutzerinterface setzt auf die Verwendung der Internet Browser-Technologie, den die Benutzer eines Clusters sehr leicht bedienen können, da viele schon Erfahrung in der Bedienung von Web-Browsern wie dem Internet Explorer oder dem Netscape Communicator haben und der Cluster nicht anders zu bedienen ist als eine herkömmliche interaktive Web-Site. Auch bietet das HTTP, das Protokoll, das die Web-Inhalte für die Web-Browser transportiert, schon die meisten für einen erfolgreichen Client-Server-Betrieb notwendigen Technologien. Die notwendigen Erweiterungen wurden in JAVA/RMI/Corba entwickelt.

Angewandte Methoden

Die Kommunikation zwischen den Rechnern und den Modulen, die in den Programmiersprachen C++ und Java geschrieben wurden, wird über das Objektmodell Corba durchgeführt. Corba ermöglicht transparente Funktionsaufrufe über Rechengrenzen hinweg.

Die eigentlichen Programme, die auf dem Cluster ablaufen, werden in einer, der Programmiersprache Prolog angelehnten, Ent-

wicklungsumgebung programmiert. Wobei Prolog um clusterspezifische Befehle erweitert wurde.

Die Programmiersprache Java findet im Bereich der Benutzerschnittstelle Verwendung. Auf dem Web-Server sind die Projekt- und Benutzerverwaltung als sogenannte Servlets und JSP-Seiten eingebunden.

Applets werden als Programme auf den Web-Browser des Benutzers übertragen. Sie können mit dem Web-Server kommunizieren, um zum Beispiel Datenbestände grafisch aufzubereiten.

Cluster

Ein Cluster ist ein aus mehreren Rechnern zusammenschaltetes System, das eine wesentlich höhere Rechenleistung als ein herkömmliches System bietet. Die Leistung eines

Cluster ist abhängig von der Netzwerk-Geschwindigkeit. Ein homogener Cluster besteht aus Rechnern mit gleichem Netzwerk-durchsatz und gleicher Rechenleistung. Von der Aufgabe des Clusters hängt es ab, ob der Cluster von diesem Vorgaben abweichen darf, ohne zu viel von seiner Gesamtleistung zu verlieren. Aufgaben, die sich nur in sehr kleine Teile zerlegen lassen, können viel ihrer Leistung durch eine langsame Netzwerk-Verbindung verlieren, da über das Netzwerk die Daten ausgetauscht werden und das Netz somit zum Flaschenhals wird. Sind in einem heterogenen Netz die Unterschiede in der Leistung groß, dann kann es passieren, dass das schwächste Mitglied ein wichtiges Zwischenergebnis liefern muss, ohne das die restlichen Mitglieder nicht weiterarbeiten können. Die schnelleren Computer verbringen ihre Zeit dann mit unnötigem Warten. Ein Cluster eignet sich für Aufgaben:

- die sich nebenläufig berechnen lassen
- die sich, am Rechenaufwand gemessen, gut in große Teile zergliedern lassen

Die Kommunikation zwischen dem Zentralrechner und den Rechnern im Cluster erfolgt über leichte Prozesse, auch Threads genannt. Diese Threads laufen im Programmkontext des Zentralrechners. Jeder dieser Prozesse kommuniziert mit einem Rechner (Bild 1).

Gestaltung des Leitwolf-Clusters

Der Leitwolf-Cluster besteht aus sechs Rechnern, die über einen 100 MBit-Hub mittels Twisted-Pair Verkabelung vernetzt sind. Die Rechner im Cluster verfügen über 400 MHz CPU und 64 MB RAM. Als Betriebssystem wird die jeweils neueste SuSE-Linux-Distribution eingesetzt. Ein Web-Ser-

ver dient als Schnittstelle zwischen den Clients und dem Cluster. Es wird von uns der Open-Source Web-Server Apache mit der Tomcat Java-Server-Page-Erweiterung verwendet. Generell wurde bei der Zusammenstellung des Clusters darauf geachtet, preiswerte Standard-Komponenten zu verwenden und auf kommerzielle Software zu verzichten. Es wurde nur Open-Source Software verwendet, etwa MICO als Corba-Implementierung.

Verteilung der Daten im Cluster

Der Leitwolf als zentraler Anlaufpunkt in der Kommunikation zwischen dem Benutzer und dem Rechen-Cluster hat folgende Aufgaben wahrzunehmen:

- Interaktion mit dem Benutzer
- Synchronisierung der Cluster-Software
- Übersetzen der Benutzer-Programme
- Verteilung der übersetzten Benutzerprogramme im Cluster
- Ablaufsteuerung der Benutzerprogramme im Cluster
- Zusammenführung der Ergebnisse
- Benutzer- und Projektverwaltung

Literatur

- [1] Informationen zu der Programmiersprache JAVA, JSP, Servlets und RMI: <http://java.sun.com>
- [2] Informationen zu Corba: <http://www.omg.org>
- [3] Informationen zu Cluster-Konzepten : <http://www.beowulf.org>

Vorschau

In der August-Ausgabe Ihrer LAND-TECHNIK finden Sie:

- DGPS-gestütztes Sicherheitssystem für Landmaschinen
- Rotierende Reinigung im Mährescher
- Wassergehaltsbestimmung in Holzbrennstoffen
- Aufbereitung von Knick- und Schwachholz
- Einsatzerfahrungen mit automatischen Melksystemen
- Staubbinderung in eingestreuten Rinderställen