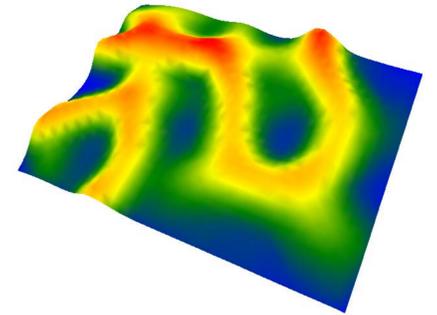


Entwurf und Implementierung einer Applikation zur Visualisierung von Lernvorgängen bei Selbstorganisierenden Karten

Diplomarbeit, vorgelegt von Benjamin Hoepner



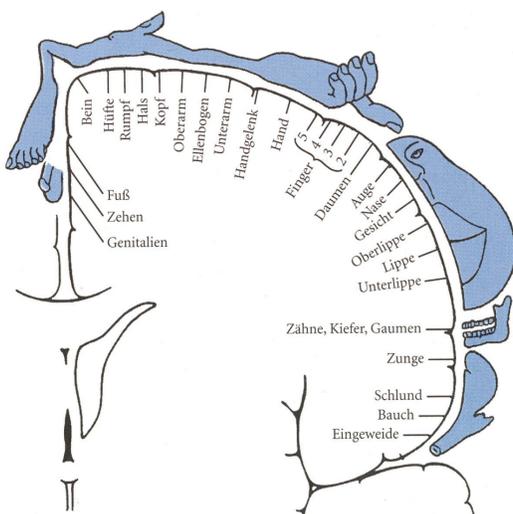
Aufgabenstellung

Selbstorganisierende Karten stellen eine besondere Form von künstlichen neuronalen Netzen dar, die sich unüberwacht trainieren lassen. Ziel der Arbeit ist die Konzeption und Implementierung einer Anwendung zum Training von selbstorganisierenden Karten. Schwerpunkt ist hierbei die Darstellung des Lernverlaufs und die Visualisierung der Karte.

Wesentliche Eigenschaften selbstorganisierender Karten sollen abstrahiert und austauschbar gestaltet werden, um es zu ermöglichen, auf einfache Weise neue Funktionen zu integrieren. Besonderer Wert wird dabei auf die Wiederverwendbarkeit von Modulen und die Erweiterbarkeit durch neue Module gelegt.

Selbstorganisierende Karten

Mit selbstorganisierenden Karten (self-organizing maps, SOM) wird versucht, die Funktion eines somatosensorischen Kortex nachzubilden. In ihm werden die taktilen Sinneszellen der Hautoberfläche auf Regionen auf der Gehirnoberfläche abgebildet. Es existiert also eine Art Landkarte der Hautoberfläche, die benachbarte taktiler Bereiche auch möglichst benachbart abbildet.



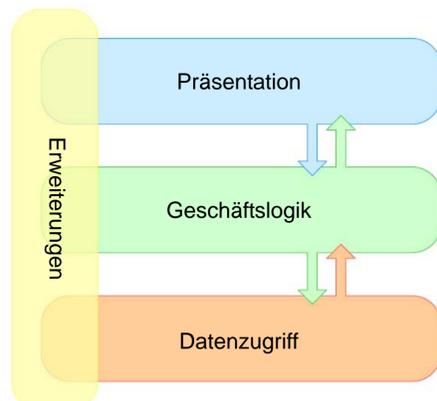
Der somatosensorische Kortex des Menschen (aus Schandry, Rainer: Biologische Psychologie, 2003)

Die selbstorganisierenden Karten imitieren diesen biologischen Vorgang. Trainierte (oft hochdimensionale) Daten werden topologieerhaltend auf einer zweidimensionalen Karte abgebildet. Eine solche Map lässt sich auf verschiedene Arten visualisieren und anschließend inspizieren. So können Bereiche entdeckt werden, in denen sich ähnliche Daten ansammeln. Diese Häufungen werden Cluster genannt und spielen beim Data Mining eine wichtige Rolle. Es wird ermöglicht, Zusammenhänge zwischen den Trainingsdaten zu erkennen, die zuvor unbekannt waren.

Das Training der SOM erfolgt unüberwacht, der Vorgang wird Unsupervised Learning genannt. Es existiert keine Zielvorgabe beim Lernvorgang, die Strukturen auf der Karte bilden sich von selbst nach einem bestimmten Algorithmus. Die Güte der errechneten SOM lässt sich trotzdem über den

sogenannten Quantisierungsfehler bestimmen. Häufig werden Parameter für den Trainingsalgorithmus eingesetzt, die eine Grobordnung der Daten auf der Map zu Beginn des Lernvorgangs erlauben, während die Ausprägung der Details zum Ende hin stattfindet.

Die selbstorganisierenden Karten kommen für viele Anwendungsbereiche in Frage. Einige Beispiele sind die Analyse von DNS und Aktienkursen, sowie die Unterstützung medizinischer Diagnosen. SOM können auch für die Lösung des Travelling Salesman Problems eingesetzt werden.



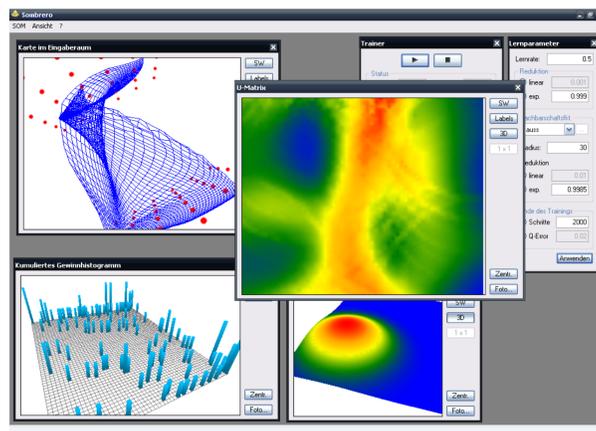
Die Systemarchitektur basiert auf der 3-Tier-Architecture und kapselt einige Bereiche für Erweiterungen

Systemarchitektur

In dieser Arbeit wurde ein Programm entwickelt, das selbstorganisierende Karten, deren Training und einige ihrer Darstellungsformen implementiert.

Die Architektur der Applikation soll Modularität und Erweiterbarkeit verschiedener Aspekte der SOM gewährleisten. Deshalb wird auf die klassische 3-Tier-Architecture zurückgegriffen, die die drei Hauptaufgaben vieler Programme in Schichten kapselt: den Zugriff auf Daten, die Geschäftslogik und die Präsentation für den Nutzer.

Diese Architektur wurde so umgestaltet, dass bestimmte Funktionalitäten der Applikation in der sogenannten Erweiterungsschicht gekapselt werden. Dort ist es möglich, Plugins in das Programm einzubinden und so z.B. neue Visualisierungen, Netztopologien oder Datenquellen anzukoppeln.



Die entwickelte Applikation "Sombbrero" ermöglicht das Training von selbstorganisierenden Karten und bietet vielfältige Visualisierungstechniken

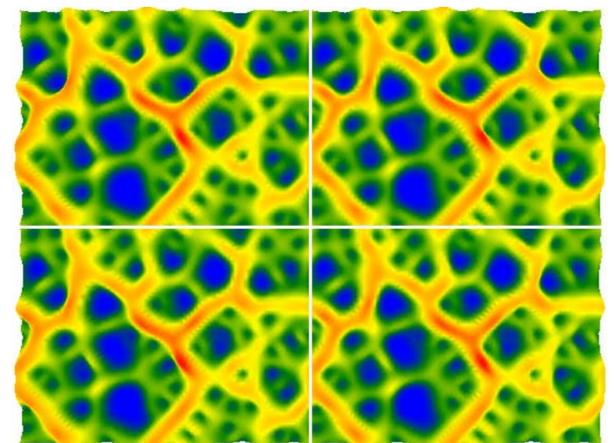
Die Zielapplikation "Sombbrero"

Das im Rahmen dieser Arbeit entwickelte Programm trägt den Namen "Sombbrero". Es basiert auf der beschriebenen Architektur und erfüllt einen großen Aufgabenbereich.

Der Trainingsprozess von selbstorganisierenden Karten ist hochgradig parametrisierbar implementiert. Der Lernvorgang kann direkt beobachtet und sogar zur Anpassung der Parameter angehalten und wieder fortgesetzt werden.

Es sind insgesamt 23 Plugins entstanden, durch die erst der volle Funktionsumfang des Programms zur Verfügung steht. Darunter sind auch sieben Kartenvisualisierungen. Eine davon ist die U-Matrix, die Datenhäufungen (Cluster) mit Tälern und größere Distanzen mit Bergzügen darstellt.

Die Applikation ist in der Programmiersprache C# geschrieben und nutzt das .NET Framework 2.0 von Microsoft als Laufzeitumgebung. Für die dreidimensionalen Visualisierungen wird die DirectX-Erweiterung eingesetzt. Einige Rechenprozesse werden parallel ausgeführt und deshalb auf mehrere Threads ausgelagert.



Mit der U-Matrix-Visualisierung lassen sich Strukturen wie Datenhäufungen und deren Grenzen gut erkennen

Zusammenfassung

Die Arbeit bietet einen Einblick in Aufbau und Funktionsweise selbstorganisierender Karten. Dabei wird auf biologische Vorbilder eingegangen und die abgeleiteten Modelle werden beleuchtet. Der Trainingsalgorithmus und seine Parameter werden ebenfalls ausführlich behandelt. Ein wichtiger Aspekt sind auch die SOM-Visualisierungsformen.

In diesem Zusammenhang ist eine sehr umfangreiche Applikation entstanden, die trotz ihrer Optionsvielfalt benutzerfreundlich aufgebaut ist. Erweiterbarkeit ist ein wichtiger Aspekt des Programms und wurde daher sorgfältig umgesetzt.

Am Ende der Arbeit werden einige Experimente beschrieben, die in den Umgang mit dem Programm einführen und so auch seine Funktionstüchtigkeit nachweisen.

Selbstorganisierende Karten stellen sich als wichtige Methode zur Analyse großer Datenmengen heraus.