
Neuronale Netze im Data Mining: Ein Studenten-Projekt

Uwe Lämmel

*Hochschule Wismar – University of Technology, Business and Design, Fachbereich Wirtschaft
Philipp-Müller-Straße 21, D-23952 Wismar, Deutschland*

Projektarbeiten sind Bestandteil vieler Lehrveranstaltungen auf dem Gebiet der Informatik. Insbesondere für Techniken, die eine große Zahl von Daten erfordern, sind praktisch relevante Aufgabenstellungen schwer auszumachen. Die erneute Bearbeitung der Aufgabenstellung des Data Mining Cup 2002 in studentischen Projekten im Rahmen einer Lehrveranstaltung hat sich als sehr vorteilhaft erwiesen. Die Arbeit beschreibt die Aufgabenstellung und die erzielten Resultate. Auf die positive Wirkung der Aufgabenstellung auf die Arbeiten und Ergebnisse der Studierenden wird ebenso eingegangen wie auf die Vorteile und Risiken, die die Herangehensweise für den Lehrenden in sich birgt.

EINEINFÜHRUNG

Im Diplomstudiengang Wirtschaftsinformatik wird die Spezialisierung Wissensmanagement/Wissensbasierte Systeme als Wahlpflichtfach angeboten. Bestandteil dieser Spezialisierung ist die Lehrveranstaltung *Wissensextraktion mittels neuronaler Netze*. Diese findet im vierten Semester statt. Den Studierenden werden Aufbau, Funktionsweise und Einsatzmöglichkeiten künstlicher neuronaler Netze vermittelt. Ein Schwerpunkt aus Sicht der betriebswirtschaftlichen Anwendungen neuronaler Netze stellt deren Einsatz im Data Mining dar. Techniken werden hierbei notwendigerweise an kleinen Beispielen eingeführt und erläutert. Das führt immer wieder zu Akzeptanzproblemen, da der Sinn des Einsatzes künstlicher neuronaler Netze für die Analyse von 10-20 Datensätzen natürlich nicht offensichtlich ist [1].

Um die Motivation zu erhöhen und die Kenntnisse anwendungsbereit zu vertiefen, wird eine Projektaufgabe bearbeitet, bei der unter möglichst realistischen Bedingungen gearbeitet werden soll. Das größte Problem bei der Formulierung von Aufgabenstellungen ist die Bereitstellung geeigneter Daten. Für den sinnvollen Einsatz künstlicher neuronaler Netze werden große Datenmengen, mehrere tausend Datensätze, benötigt. So entstand die Idee, im Projekt des Jahres 2004 auf die Aufgabenstellung des Data Mining Cup [2] aus dem Jahre 2002 zurückzugreifen, siehe www.data-mining-cup.de/2002/
Eine derartige Aufgabenstellung als Projektaufgabe

zu vergeben, hat sich aus mehreren Gründen als vorteilhaft erwiesen:

- Es stehen große Datenmengen zur Verfügung: Für die Aufgabe des Data Mining Cup 2002 sind 10.000 Datensätze zum Trainieren und 10.000 Datensätze zum Testen gegeben.
- Da der Wettbewerb bereits abgeschlossen ist, stehen Resultate zur Verfügung – die Machbarkeit ist gesichert.
- An den Resultaten lassen sich die eigenen Ergebnisse messen. Der Vergleich erzeugt einen deutlichen Motivationsschub.

Im Folgenden wird das Projekt vorgestellt und insbesondere die Ergebnisse verglichen. Ausführliche Informationen zum Projekt, die Daten, die neuronalen Netze und die Ergebnisse sind auf den Web-Seiten des Autors zur Lehrveranstaltung Neuronale Netze zu finden: www.wi.hs-wismar.de/~laemmel/

Das Studenten-Projekt

Die Studierenden haben sich für die Bearbeitung der Aufgabe selbstständig zu Projektteams zusammengefunden. Aufgestellt wurden sieben Teams mit 3-4 Mitarbeitern, wobei ein Team dann die Arbeit gar nicht begonnen hat, ein Team reduzierte sich auf drei, ein weiteres auf zwei Bearbeiter. Sinnvolle Ergebnisse liegen von fünf Teams vor.

Für die Lösung der Aufgabe werden künstliche

neuronale Netze eingesetzt. Eine spezielle Software wird nicht vorgegeben. In der Lehrveranstaltung wird der SNNS - Stuttgarter Neuronale Netze Simulator [3] bzw. der JavaNNS benutzt [4]. SNNS ist unter Linux/Unix verfügbar, der JavaNNS arbeitet zudem unter Windows-Betriebssystemen. Diese Software wurde von allen Projektgruppen zur Analyse benutzt.

Den Studierenden wurde folgende Richtlinie für die Bewertung der Projektarbeit gegeben:

- Arbeitsplan: 10%
- Ergebnisse: 30%
- Dokumentation: 30%
- Präsentation: 30%

Diese Verteilung stellt sicher, dass nicht einseitig auf das Ergebnis hin gearbeitet wird. Für Projekte ist ein Arbeitsplan (Zeitplan, Arbeitsteilung) wichtig. Die Zusammenstellung der Ergebnisse, so dass diese nachvollziehbar sind (Dokumentation), sowie die öffentliche Darstellung im Vortrag und als WWW-Seite (Präsentation), sind mindestens ebenso wichtig, wie das erzielte Resultat.

Diese Aufteilung erregt bei den Studierenden zwar Verwunderung, wird aber akzeptiert, da die Regelung im Vorfeld bekannt gegeben und begründet wurde [5].

Die Aufgabe

Ein Energieversorger möchte Kunden, die möglicherweise kündigen wollen, einen günstigeren Tarif, einen Rabatt, anbieten. Damit sollen potenzielle Kündiger gehalten werden, denn trotz des Rabattes erhöht der Kunde den Deckungsbeitrag. Die treuen Kunden haben nicht die Absicht zu kündigen und sollen deshalb von diesem Tarif nichts erfahren, da der Rabatt-Tarif letztendlich einen geringeren Gewinn für das Unternehmen zur Folge hat.

Die Daten

Der Energieversorger hat eine Trainingsmenge von 9.000 treuen Kunden und 1.000 Kündigern des letzten Jahres bereitgestellt, da ca. 10% potenzielle Kündiger erwartet werden. Ein Datensatz der Trainingsmenge besteht aus 34 Merkmalen. Neben einem Datensatz-Schlüssel sind der Jahresverbrauch in kWh, die Zahlungsart sowie Angaben zum Haus und der Umgebung vorhanden. Darüber hinaus sind auch Kundeneinschätzungen vorhanden. Für die Datensätze der Trainingsmenge ist zudem angegeben, ob der Kunde gekündigt hat (Tabelle 1).

Tabelle 1: Merkmale der gegebenen Datensätze ohne Schlüssel [6].

1	payment_type	Text	Art der Bezahlung
2	power_consumption	Integer	letzter jährlicher Stromverbrauch in kWh
3	HHH	Integer	Anzahl Haushalte im Haus
4	HGEW	Integer	Anzahl Gewerbe im Haus
5	MTREGOG	Integer	Regionaltyp
6	MTKAUOG	Integer	Kaufkraft
7	MTSTRÖG	Integer	Straßentyp
8	MTBEBÖG	Integer	Bebauungstyp
9	MTSTAOG	Integer	Status
10	MTBONOOG	Integer	Prüfungsgrad Bonität
11	MTADEOG	Integer	Anteil Deutscher
12	MTALTOG	Integer	Altersstruktur
13	MTFAMOG	Integer	Familienstand
14	... 18		PKW – Indizes
19	... 26		Versicherungstyp des Kunden
27	... 32		Angaben zum Gesundheitstyp
33	canceler	yes/no	Ziel: yes- Kündiger ; no-Kunde

Bewertungsfunktion

Als Grundlage für die Optimierung sind die Einnahmen pro Kunde und Jahr je nach Kunde beziehungsweise Kündiger in einer Kundenwertmatrix gegeben (Tabelle 2).

Tabelle 2: Kundenwertmatrix.

	Kunde kündigen	will: nicht kündigen
Kunde erhält Angebot	43,80€	66,30€
Kunde erhält kein Angebot	0,00€	72,00€

Die Einnahmen, die ohne eine Rabatt-Aktion erzielt werden betragen:

$$f_{Grund} = 72,00 \cdot \text{Anzahl Nichtkündiger}$$

Ziel des Data Mining ist ein Mechanismus, der als Vorbereitung für die Rabatt-Aktion, die Kunden möglichst gut in potenzielle Kündiger und treue Kunden, Nichtkündiger, klassifiziert, um so ein verbessertes Ergebnis zu erzielen.

Die Zielfunktion unter Beachtung von f_{Grund} ist somit:

$$f_{Ziel} = 43,80 \cdot \text{Anzahl erkannter Kündiger} + 72,00 \cdot \text{Anzahl erkannter treuer Kunden} + 66,30 \cdot \text{Anzahl nicht erkannter Nichtkündiger} - f_{Grund}$$

Für die praktische Auswertung wurde f_{Grund} dargestellt als:

$$f_{Grund} = 72,00 \cdot (\text{Anzahl erkannter Nichtkündiger} + \text{Anzahl nicht erkannter Nichtkündiger})$$

Bei einer Menge von 10.000 Datensätze, die 10% Kündiger enthält, ist der Ausgangswert

$$f_{Grund} = 9.000 \cdot 72,00 = 648.000$$

Die Aufgabe lautet nun, die Daten der Testdatenmenge so zu klassifizieren, dass ein möglichst hohes zusätzliches Ergebnis für f_{Ziel} erzielt wird. Das Ergebnis f_{Ziel} gibt an, wie viel Einnahmen über den Grundwert durch eine Rabatt-Aktion gemäß der erarbeiteten Klassifizierung zu erwarten sind.

Zeitplan

Die Aufgabe wurde Anfang April gestellt und die Ergebnisse Anfang Juni eingereicht. Die Resultate wurden von jeder Gruppe in einem Vortrag Mitte Juni präsentiert. Die Bearbeitungszeit betrug somit insgesamt 10 Wochen.

In der ersten Phase haben sich die Studenten selbstständig zu Projektgruppen von drei bis vier Personen zusammengeschlossen und einen Projektleiter bestimmt. Als Abschluss dieser Phase war ein Arbeitsplan vorzulegen.

Die Arbeitsergebnisse wurden in einer Dokumentation dargestellt. Dokumentation und Daten waren eine Woche vor der Projekt-Verteidigung einzureichen.

Datenvorverarbeitung

Obwohl auf die Bedeutung der Datenvorverarbeitung in der Lehrveranstaltung besonders hingewiesen wurde, ist den Studierenden erst bei der Beschäftigung mit den Daten richtig deutlich geworden, wie viel Aufwand für diese Teilaufgabe notwendig ist. Ziel der Datenvorverarbeitung ist eine Datenmenge, die von einem künstlichen neuronalen Netz verarbeitet werden kann. Typische Eingaben für ein neuronales Netz sind numerische Werte aus dem Intervall (0,1) bzw. (-1,1). Zahlwerte, die die Zugehörigkeit zu einer Klasse repräsentieren, können z.B. auch als 0-1-Vektor codiert werden.

Neben der Verwendung von *Excel* haben sich alle Teams Hilfsprogramme für die Datenvorverarbeitung erarbeitet. Zum Einsatz kam PHP, Python, Perl oder auch Java.

Auswahl

Es ist durchaus üblich, nicht alle vorgegebenen Merkmale für die Datenanalyse zu verwenden. Naturgemäß wird das Schlüsselmerkmal weggelassen. Darüber hinaus wurde allerdings von den Studierenden keine weitere Auswahl getroffen, sondern alle Merkmale beachtet.

Ein Team hat im Vorfeld eine Tabelle der Korrelationswerte erarbeitet, um somit die lineare Abhängigkeit aller Merkmalspaare einschätzen zu können. Dieses führte zu der Erkenntnis, dass einfache lineare Abhängigkeiten in den Daten nicht vorhanden sind. Eine lineare Abhängigkeit war zwar nicht zu erwarten, die Untersuchung sollte aber Bestandteil einer jeden Vorbetrachtung zu analysierender Daten sein, um unnötige, aufwändige Untersuchungen gegebenenfalls zu vermeiden.

Behandlung fehlender Angaben

In der Trainingsmenge sind etwa 3.000 der 10.000 Datensätze nicht vollständig mit Werten belegt. Darunter gibt es Datensätze, die nur den Energieverbrauch, die Zahlungsart und die Angabe zur Kündigung enthalten, somit nur 3 von 33 Angaben. Erschwerend kommt hinzu, dass durchaus auch gleichartige Merkmalskombinationen zum einen zu einem Kündiger gehören andererseits auch ein Datensatz mit denselben Werten einen treuen Kunden charakterisieren. Für die Behandlung fehlender Daten wurden verschiedene Verfahren gewählt:

- Wert außerhalb der Wertebereiche, meist -1;
- Eintrag des Mittelwertes aller Werte des Merkmals über alle Datensätze;
- Finden des *nächsten* Datensatzes durch Suche nach dem Datensatz mit dem geringsten Abstand (Manhattan, Euklid) und Übernahme der Werte;

Die Mittelwert-Strategie wurde häufig verwendet. Das Team mit dem Einsatz der Euklidischen Distanz-Strategie erzielte jedoch das beste Ergebnis.

Codierung

Es wurde überwiegend eine Abbildung der Werte in den Bereich [0,11] vorgenommen. Der Energieverbrauch mit einem Wertebereich von [0,27.500] innerhalb der Trainingsdaten stellt dabei eine besondere Herausforderung dar, wobei nur ca. 5% der Werte größer als 6.000 sind. Alle anderen Merkmale liegen in einer ganzzahligen Codierung vor, meist aus dem Bereich [0,10].

Ein Team hat eine binäre Codierung verwendet, wobei jedem Merkmal so viele Neuronen zugeordnet werden, wie das Merkmal Ausprägungen hat. Der Wert 3 eines Merkmals mit dem Wertebereich 1..5 wird dann als 0-1-Vektor codiert: (0,0,1,0,0). Durch diese Codierung wird die Anzahl der Eingabewerte für das neuronale Netz extrem erhöht.

Neuronale Netze

Die Verwendung eines neuronalen Netzes für die Aufgabenlösung ergibt sich aus dem Thema der Lehrveranstaltung und war somit zwingend vorgegeben.

Architektur

Alle Teams haben vorwärts gerichtete Netze (feed forward network) eingesetzt. Dies ist aufgrund der vorliegenden Daten, es sind Trainingsdaten gegeben, die Ergebnisse enthalten, nahe liegend.

Die konkrete Architektur variiert jedoch von Team zu Team. Die Größe der Eingabeschicht ergibt sich aus der verwendeten Codierung: 32 Eingabe-Neuronen, falls die Normierung auf (0,1) als Codierung der 32 Merkmale gewählt wurde. Etwa 200 Eingabe-Neuronen sind notwendig, falls eine binäre Codierung wie oben beschrieben zum Einsatz kommt (Abbildung 1).

Die Größe der Zwischenschicht kann zwar vorher experimentell abgeschätzt werden, meist wurde jedoch mit einer Zwischenschicht etwa halb so groß wie die Eingabe-Schicht begonnen und danach mit kleineren

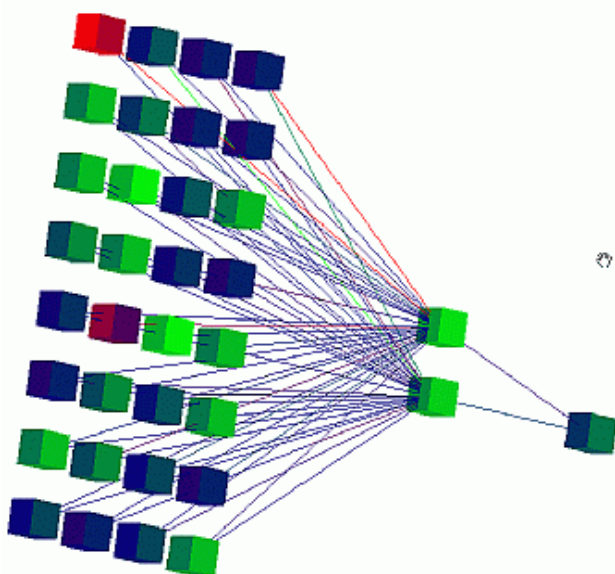


Abbildung 1: Ein vorwärts gerichtetes Netz mit Eingabe-Schicht (32 Neuronen), verdeckter Schicht (2 Neuronen) und einem Ausgabe-Neuron (aus [7]).

Schichten experimentiert. Die Ausgabe-Schicht besteht bei allen Teams aus einem Neuron: Die Aktivierung des Neurons gibt an, ob ein Kündigungserkannt wurde.

Lernverfahren

Alle Teams haben mit mehreren Lernverfahren Erfahrungen gesammelt: Standard-Backpropagation, Backpropagation mit Momentum, QuickProp, Resilient Propagation (RPROP). Die Erfahrungen können wie folgt zusammengefasst werden:

- Standard Backpropagation liefert die besten und stabilsten Ergebnisse;
- QuickProp führt während des Trainings zu einer stark schwankenden Fehlerkurve und zu keinem signifikanten Absinken des Fehlers;
- Eine eher geringe Anzahl von Lernzyklen (10-300) führt bereits zu einem guten Ergebnis;
- Ein zu langes Trainieren führt zu einem geringeren Fehler für die Trainingsdatensätze, lässt aber dann den Fehler für die Testdaten wieder ansteigen.

Auf das Problem des Auswendiglernens (overfitting) und der damit verbundenen geringeren Generalisierungsfähigkeit eines Netzes wurde in der Lehrveranstaltung mehrfach hingewiesen, aber erst die eigene Beobachtung hat das Problem deutlich werden lassen.

Auswertung der Netze

Das Ergebnis der Analyse eines neuronalen Netzes lässt sich an der Ausgabe-Schicht, hier an dem einen Ausgabe-Neuron, ablesen. Während es für die kleinen Beispiele im Rahmen der Lehrveranstaltung völlig ausreichte, sich visuell ein Bild zu machen, ist das für 10.000 Datensätze nicht mehr möglich. Darüber hinaus besteht hier die Besonderheit, dass es nicht in erster Linie darauf ankommt, die Erkennungsrate zu erhöhen, sondern der Wert der Zielfunktion ist zu maximieren. Das Problem lässt sich leicht aus der Kundenwertematrix (Tabelle 2) ableiten. Während ein nicht erkannter Kündigungserkannt einen Verlust von 43,80€ zur Folge hat, bedeutet ein falsch angeschriebener treuer Kunde einen Verlust von 5,70€.

Bestimmung des Gewinns für Trainings- und Testmenge

Um die Qualität eines Netzes einschätzen zu können, musste die Testdatenmenge verarbeitet werden, die Ergebnisse mit den gewünschten Ergebnissen der

Testdatenmenge verglichen und die Zahl der richtig respektive falsch erkannten Kündigung respektive treue Kunden ermittelt werden. Daraus lässt sich der Wert der Zielfunktion errechnen.

Von den Teams wurden für diese Aufgabe Analyseprogramme in C, Perl, PHP, Python oder auch Java entwickelt.

Ergebnisse

Waren die ersten Ergebnisse recht bescheiden, so entwickelte sich doch ein Wettbewerb, der dann zu Ergebnissen führte, die in das Spitzenfeld der Ergebnisse des Data Mining Cup 2002 eingeordnet werden können. Die erzielten Ergebnisse der Projektgruppen in absteigender Reihenfolge:

1. 8.115•
2. 7.586•
3. 7.378•
4. 4.759•
5. 3.008•

Entscheidend für die Qualität des Ergebnisses ist die Erkenntnis, dass 0,5 als Standard-Wert des Schwellwertes des Ausgabe-Neurons ungeeignet ist. Ein Team erkannte dies durch den Einsatz der Analyse-Komponente *Network Analyzer* des SNNS. Die Erkenntnis wurde von anderen Projektgruppen übernommen. Ein Zeichen dafür, dass trotz Wettbewerb auch kooperiert wurde. Das Festlegen eines geringeren Schwellwertes führt zu einem Anstieg des Ergebnisses. Die Gruppe mit dem kleinsten Ergebnis (3.008) hat dieses ohne Anpassung des Schwellwertes erzielt. Mit einer Anpassung wurde ein Wert von etwa 5.000 erreicht.

Vergleich mit 2002

Die Ergebnisse des Projektes entsprechen den im Data-Mining-Cup 2002 erzielten Ergebnissen. Auf der einen Seite war die Situation im Projekt 2004 einfacher, da die Resultate für die Testdatenmenge zur Verfügung standen und somit eine Bewertung der erzielten Klassifikation des Netzes vorgenommen werden konnte. Aus didaktischer Sicht war dies sehr vorteilhaft, da eine direkte Einordnung der erzielten Ergebnisse möglich wurde. Sowohl der Vergleich mit den im Jahre 2002 erzielten Ergebnissen, als auch der Vergleich der Projektteams untereinander hat Ehrgeiz geweckt und war mit ausschlaggebend für das insgesamt sehr gute Ergebnis (Tabelle 3).

Auf der anderen Seite können die Ergebnisse nur bedingt verglichen werden, da in diesem Projekt

Tabelle 3: Siegerwerte im Data Mining Cup 2002 [2].

Platz	Erkannte Kündiger	Erkannte Kunden	Gesamt		Zielwert
1:	523	2673	3196	==>	7671.30
2:	477	2415	2892	==>	7127.10
3:	448	2213	2661	==>	7008.30
4:	440	2156	2596	==>	6982.80
5:	454	2269	2723	==>	6951.90
6:	456	2287	2743	==>	6936.90
7:	379	1731	2110	==>	6733.50
8:	411	1991	2402	==>	6653.10
9:	446	2274	2720	==>	6573.00
10:	418	2059	2477	==>	6572.10

einschränkend nur neuronale Netze zum Einsatz kommen durften.

Für die Studierenden ist das Ergebnis sehr motivierend: Die Praxistauglichkeit neuronaler Netze wurde unter Beweis gestellt, damit ist indirekt auch eine Bestätigung für die Wahl der Lehrveranstaltung verbunden. Darüber hinaus wurden Ergebnisse erzielt, die in die Spitzengruppe der Resultate einzuordnen sind, das gibt den Studierenden die Sicherheit, dass die eigenen Leistungen mit denen anderer sehr gut vergleichbar sind.

Optimierungen

Ein Team hat sich insbesondere mit der Optimierung des neuronalen Netzes auseinander gesetzt und dabei äußerst beachtenswerte, überraschende Ergebnisse erzielt.

So wurde gezeigt, dass auf eine innere Schicht verzichtet werden kann, ohne dass sich das Endergebnis entscheidend verändert. Dieses Netz, nun ein Perzeptron, bestehend aus 24 Eingabe-Neuronen und einem Ausgabe-Neuron, erzielte einen Zielwert von 8.367, siehe Abbildung 2.

Dieses Netz wurde vom Team schrittweise automatisch aufgebaut, indem Schritt für Schritt das wesentlichste Merkmal bestimmt und dann in das Netz aufgenommen wurde. Noch beeindruckender ist, dass bereits ein Netz aus 4 Eingabe-Neuronen einen Zielwert von über 7.000 erreicht (Abbildung 3).

Diese Arbeiten liefern die Erkenntnis, dass die folgenden vier der 32 Merkmale für eine Klassifikation entscheidend sind: Energieverbrauch, Zahlungsweise, Anzahl der Gewerbe sowie die Affinität zur Direktwerbung. Auf diese Erkenntnis ließe sich jetzt aufbauen und eine regelbasierte Klassifikation entwickeln. Ein klassischer Fall einer Wissensextraktion.

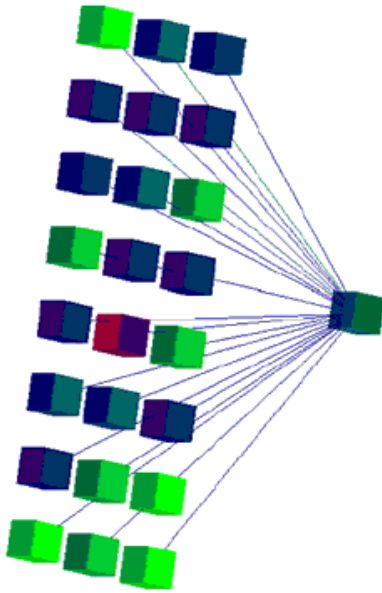


Abbildung 2: Optimiertes Netz [7] mit 24 Eingabe-Neuronen und einem Zielwert von 8.367.

SCHLUSSFOLGERUNGEN UND AUSBLICK

Im Jahre 2003 wurden von den damaligen Teams jeweils unterschiedliche Aufgaben unter Nutzung der im Datensatz-Archiv des Instituts für Statistik der Universität München vorhandenen Daten vergeben [8]. Das hatte den Nachteil, dass die erzielten Ergebnisse nicht vergleichbar waren, weder untereinander noch mit Bearbeitern von anderen Hochschulen. Genau diese Probleme wurden mit dem diesjährigen Ansatz vermieden.

Über die Vertiefung praktischer Fähigkeiten im Umgang mit großen Datenmengen hinaus, sorgte die Wettbewerbssituation für Konkurrenz der Projektgruppen. Es wurden hervorragende Ergebnisse erzielt, wie ein Vergleich mit den Resultaten des Data Mining Cup 2002 aufzeigt.

Eine praktische Aufgabenstellung, deren prinzipielle Lösbarkeit mit gegebenen Resultaten gegeben ist, erzeugt den Anreiz für intensive Bearbeitung des Problems. Wenn die Ergebnisse sich dann in die Resultatslisten eines deutschlandweiten Wettbewerbs einordnen lassen, erzeugt das Stolz auf das Erreichte,

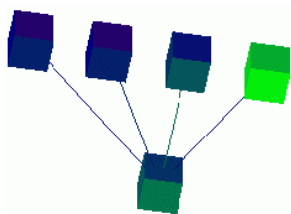


Abbildung 3: Perzeptron mit vier ausgewählten Eingaben erreicht einen Zielwert von 7058,70.

dieses wirkt wiederum positiv auf die Studienmotivation für das Fach im Einzelnen aber auch darüber hinaus. So werden sich einige Studierende im nächsten Jahr am dann aktuellen Data Mining Cup beteiligen.

Das Projekt kann als voller Erfolg betrachtet werden. Es darf jedoch nicht verschwiegen werden, dass für den Lehrenden am Anfang eine gehörige Portion Unsicherheit mit dem Projekt verbunden war:

Da die im Wettbewerb erzielten Ergebnisse nicht mit neuronalen Netze erreicht wurden, war unklar, ob sich mit neuronalen Netzen vergleichbare Ergebnisse erzielen lassen.

Im nächsten Jahr, werden wieder Aufgaben aus dem Data-Mining-Cup vergeben werden. Aus welchem Jahr? Das bleibt vorerst noch ein Geheimnis, um eine vorzeitige und damit längere Bearbeitung des Themas auszuschließen. Ganz sicher aber nicht erneut die Aufgabe des Jahres 2002. Bleiben wird die Regelung zur Bewertung der studentischen Leistungen.

Die Studierenden der Lehrveranstaltung *Wissensextraktion mittels neuronaler Netze* sind nach der erfolgreichen Projektbearbeitung in der Lage, auch Aufgaben zu bearbeiten, deren Ergebnis nicht derartig messbar ist. Die erarbeiteten Vorgehensweisen bei der Datenvorverarbeitung, der Analyse mittels neuronaler Netze, der Dokumentation der Experimente sowie die Auswertung der Netzausgaben sind Kompetenzen, die bei der Bearbeitung weiterer Praxis-Aufgaben eingesetzt werden können [9-11]. Im Projekt DaMEN – Data Mining Engineering – analysieren derzeit einige Studierende Daten des Praxispartners, eines deutschen Bankinstitutes, zum Kundenverhalten im Online-Banking.

In umfangreicheren Arbeiten wie zum Beispiel Diplomarbeiten oder zukünftig Master-Arbeiten können diese oder auch andere Data-Mining-Aufgabenstellungen aus der Praxis bearbeitet werden.

REFERENZEN

1. Lämmel, U. und Cleve, J., *Lehr- und Übungsbuch Künstliche Intelligenz* (2. Auflage). Leipzig: Fachbuchverlag Leipzig (2004).
2. Data Mining Cup (2004), www.dmc.de
3. www-ra.informatik.uni-tuebingen.de/software/
4. www-ra.informatik.uni-tuebingen.de/software/
5. Brätz, M., Buse, M., Gerwien, N. und Thiel, S., *Neuronale Netze-Stromversorgerprojekt*. Hochschule Wismar, Fachbereich Wirtschaft (2004), www.wi.hs-wismar.de/~laemmel/
6. www.data-mining-cup.de/2002/

7. Heuer, A., Andersch, C., Staack, C. und Klumbies, F., Wissensextraktion mittels Neuronaler Netze. Projektdokumentation, Hochschule Wismar, Fachbereich Wirtschaft (2004), www.wi.hs-wismar.de/~laemmel/
8. www.stat.uni-muenchen.de/service/datenarchiv/
9. Jahnke, J., Seidel, J., Straus, S. und Scholz, F., Wissensextraktion mittels Neuronaler Netze, Projektdokumentation, Hochschule Wismar, Fachbereich Wirtschaft (2004), www.wi.hs-wismar.de/~laemmel/
10. Meier, K., Pälecke, M. und Pfaff, C., Wissensextraktion mittels Neuronaler Netze. Projektdokumentation, Hochschule Wismar, Fachbereich Wirtschaft (2004), www.wi.hs-wismar.de/~laemmel/
11. Thinius, C., Fentroß, M. und Kalfack, A., Wissensextraktion mittels Neuronaler Netze. Projektdokumentation, Hochschule Wismar, Fachbereich Wirtschaft (2004), www.wi.hs-wismar.de/~laemmel/

BIOGRAPHIE



Prof. Dr.-Ing. Uwe Lämmel wurde 1956 geboren. Von 1976-1981 studierte er Mathematik an der Universität Rostock und promovierte 1985 auf dem Gebiet der Informatik. Als Mitarbeiter am Lehrstuhl Programmiersprachen des Fachbereich Informatik der Universität Rostock befasste er sich mit attribuierten Grammatiken und logischer Programmierung. Seit 1996 ist er am Fachbereich Wirtschaft der Hochschule Wismar berufen für Grundlagen der Informatik/Künstliche Intelligenz. Schwerpunkt seiner Arbeiten sind künstliche neuronale Netze und deren Anwendungen. Hierzu war und ist er an mehreren Projekten beteiligt. Prof. Lämmel ist Mitarbeiter am Gottlob-Frege-Zentrum der Hochschule Wismar.

**Conference Proceedings of the
7th UICEE Annual Conference on Engineering Education
under the theme: *Educating for the Global Community***

edited by Zenon J. Pudlowski

The 7th UICEE Annual Conference on Engineering Education, held under the theme of *Educating for the Global Community*, was organised by the UNESCO International Centre for Engineering Education (UICEE) and was staged in Mumbai, Maharashtra State, India, between 9 and 13 February 2004.

This volume of Proceedings includes papers submitted to this Conference and offers a diverse compendium of articles that detail various international approaches to engineering education research and development related to the Conference theme, as well as other specific activities.

The 47 published papers, representing 21 countries, offer an excellent collection of works that tackle fundamental issues, concepts and achievements of individual researchers, as well as the concerns and challenges regarding engineering and technology education in different cultures.

The papers have been organised into the following groups:

- Opening and Keynote addresses
- Multimedia and the Internet in engineering education
- Quality issues and improvements in engineering education
- Innovation and alternatives in engineering education
- International examples of engineering education and training
- New trends and approaches to engineering education
- Important issues and challenges in engineering education
- Specific engineering education programmes

The variation of subjects, concepts, ideas and international backgrounds in this volume of Proceedings demonstrate the global nature of UICEE-run Conferences, as well as its relevance within the worldwide affairs related to engineering and technology education.

In order to ensure the high quality and value of the Proceedings into the future, all of the papers have undergone assessment by independent international peer referees and have been professionally edited. As such, it is envisaged that this volume will become a useful source of information on research and development activities in engineering and technology education, seen within the context of educating future engineers for the global community.

In order to purchase a copy of the Proceedings, a cheque for \$A100 (+ \$A10 for postage within Australia, and \$A20 for overseas postage) should be made payable to Monash University - UICEE, and sent to: Administrative Officer, UICEE, Faculty of Engineering, Monash University, Clayton, Victoria 3800, Australia. Tel: +61 3 990-54977 Fax: +61 3 990-51547