5th Global Congress on Engineering Education
Kingston, Jamaica, 7-11 February 2005

Artificial Neural Networks and Engineering Education

Uwe Lämmel

Hochschule Wismar
Wismar, Germany

ABSTRACT: In the paper we address the problem of teaching subjects, where almost no engineering approach is available. Especially teaching in the area of artificial neural networks (ANN) or data mining has to deal not only with missing heuristics but with missing data as well. We argue that project work is substantial for a successful education in the area but for serious projects we do need not only data but even analysing results as well. Since we have almost no heuristics in the application of ANN we need a lot of experiments. A competition is a successful way for projects in areas where we lack heuristics. Two years of student project work have shown that a competition not only encourage students to gain best results but improves even the co-operation between teams. The paper reviews the projects and emphasizes the approach where an old competition, the Data Mining Cup 2002, was used again.

INTRODUCTION

The notion "engineering approach" is used in a sense of "engineering method", see e.g. [Arciszewski2002]: The engineering method consists of several steps: analyse the problem, devise a plan to solve it, carry out the plan, and check the solution. In order to develop a certain solution, in a design process heuristics are applied. It results in blueprints and/or models which allow a proof of the desired properties of the final solution, e.g. statics of a building. But even a solution is hard to find if we have almost no heuristics.

Teaching computer science in general has to address the problem that engineering approaches are seldom available. Programming is more a trial and error or an art design process than an engineering method. Nevertheless the problem has been discussed since the end of the sixties and a lot of techniques, methods and tools have been developed in order to implement a culture of software engineering. Substantial development in the field has taken place, but still the development of software is sometimes far away from an engineering approach. If we look at subjects like artificial intelligence, artificial neural networks (ANN), fuzzy logic or data mining the situation is even worse.

In the paper we look at a course in ANN. As a university teacher the author has been looking for methods which promote a systematic development of ANN but still has not found the silver bullet. Solving a given problem artificial neural networks are a kind of last try: If and only if other approaches were not successful then a neural network could be applied. Although today a lot of very successful applications of ANN is known there does not exist an engineering method for the development of a successful application of ANN.

Methods developed for big problems have to be taught using small problems. Often it results in a lack of acceptance of such methods. Bigger Problems can not be solved within a few classroom exercises.

How principles and the development of artificial neural networks can be taught if we can not provide heuristics?
We can roughly determine three problems:

1. An application of an ANN is a kind of a last chance: If and only if other methods, like e.g. design of an algorithm or a rule based approach, have failed then we could try a neural network approach. Therefore small or tiny problems are not convincing. Students often see that such problems can be solved quicker an with less uncertainty by other methods. Only serious problems can show the real advantages of an ANN.
2. In order to apply an ANN we need data. Moreover we need not only some data but a lot of data or even mass of data are required. Especially for teaching purposes we do need not only data as an input but we need a solution as well. Solutions are of great importance because a positive result of a data mining process is not natural but is required in an educational approach. Moreover students should be able to compare their results with a given solution.
3. There are almost no heuristics for the development of a successful ANN beside for some small sophisticated problems.

Real world data mining problems gain more interest but meet the points two and three mentioned above. The paper reviews two years of student projects we have run since 2003 as part of a course in ANN.

NEURAL NETWORK PROJECTS

At the beginning of the course the basics of ANNs are presented. Students learn several architectures of neural networks as well as corresponding learning algorithm. Although the importance of data preparation is mentioned students often do not see these difficulties, due to the tiny examples used within the lectures. Therefore a project is part of the course in order to offer the possibility to handle bigger problems, see www.wi.hs-wismar.de/~laemmel/ for more details. A project work is performed by a team of 4-5 students.. Thus not only neural network knowledge is applied but students learn to work in a team as well.

In 2003 every team had to solve a different problem. We used data provided by the Institute for Statistics of the Ludwig-Maximilian-University Munich and the Special Research Group 386 (www.stat.uni-muenchen.de/service/datenarchiv):

- Determining the solidness of borrowers via credit scoring,
- Brand-name selection of coffee,
- Munich rent standard 1994,
- …

All the teams worked quite hard on the problems but at the end they were not satisfied with their results and neither was the author. The results could not be valuated, nobody had a benchmark we could the results compare to.

In the summer semester 2004 we followed another idea. Since 2000 the Data Mining Cup (see www.data-mining-cup.de) has taken place. It is sponsored by a software company which offers data mining software. Every year a real world problem has to be solved. Real world data are provided as well for a training process as for testing.

Of course we cannot see any results in a running cup but if an annual competition has finished the high score list is published.. Therefore we decided to go back to a former situation: the Data Mining Cup 2002.

In the project of spring 2004 the students had to solve the problem of the Data Mining Cup 2002. Seven teams took part and at the end 5 teams presented results. Since every team had to solve the same problem a competitive atmosphere occurred which substantially influenced the quality of the results.

A power supply company is preparing a mailing action. Customer who most likely would like to cancel their contract should get a special offer. The offer may stop the customer doing a cancellation. A special offer means less income for the company. Therefore only potential cancellers should get an offer. 10,000 data sets containing 1,000 data of cancellers were provided for training. Another 10.000 data for testing were available which of course did not contain information whether a certain data set belongs to a canceller or to a loyal customer.

Figure 1: value matrix

	customer cancel	will: not cancel
customer gets an offer	43,80€	66,30€
customer gets no offer	0,00€	72,00€

The aim was to maximize the income of the company. A matrix of expected income per customer was given dependent whether a customer gets an offer or not and dependent on whether he cancels the contract or not. The task is to train an ANN using the 10.000 training data set in such a way that the ANN is further on able to classify the 10.000 data of the test set. Using the value matrix the quality of the network can be expressed by one figure evaluated by the following function:

$$
\begin{aligned}
f_{eval} = \; & 43{,}80 \cdot \text{number of identified cancellers} \\
& + 72{,}00 \cdot \text{number of identified loyal customers} \\
& + 66{,}30 \cdot \text{number of not identified loyal customers} \\
& - 72{,}00 \cdot \text{number of loyal customers.}
\end{aligned}
$$

The last row is the amount of money the company would earn without a mailing action. So we look only on the lift value.

The Data Mining Cup 2002 had finished long before our course began. Therefore the results were available. Figure 2 shows the results achieved in the Data Mining Cup 2002. Though it was obvious already at the very beginning of our project that substantially results could be achieved by using data mining methods. Using the classification of the test set results in a better performance (income of the company) than without any forecast, e.g. without any mailing action.

Figure 2: top three results of the Data Mining Cup 2002

position	identified canceller	identified customer	sum		lift/result:
1:	523	2673	3196	==>	7671.30€
2:	477	2415	2892	==>	7127.10€
3:	448	2213	2661	==>	7008.30€

Still a question remained: Could an application of an ANN gain similar results? The result list gave good benchmarks and created ambition to gain better results.

At the beginning a project schedule as well as a set of criteria were given:

	schedule
9.4.:	launch
16.4.:	team fixed
29.4.:	work schedule
9.6.:	deadline
15.6.:	presentation (oral and web-site)

	criteria
10%	work schedule and share out of work
30%	Quality of results
30%	Documentation
30%	presentation (oral and web-site)

In the early stage teams were built and teams worked out there work schedule and discussed the share out of work.

A data mining process can be seen as a cycle of data preparation, analysis, and evaluation. So in a next step pre-processing of data had to be done. This caused much more trouble than most of the students expected. Managing 10,000 data is not easy and tools are necessary for data coding. Data pre-processing covers the handling of missing values as well. Different approaches were chosen: One team ignored all training data containing missing values (about 3,000 data sets). The problem is so postponed to the test set only. Missing values were replaced by the average value of a feature or replaced by the value of the feature of the nearest data set.

Most teams normalized the input data into the range [0..1]. This was quite easy because most of the input range between 1 and 8, the data were already coded in some way. Only the feature power consumption had to be classified and then normalized afterwards. A transformation into binary vectors requires much more effort. It came out that the results in this case are not better than in the former case.

Since a training set was provided all teams prepared a feed-forward network. A common feed-forward network consists of an input-laver, a hidden-layer, and an output-layer. It can be trained quite fast, it “learns” by example. Whereas the structure of input and output layers can be easily fixed the size of the hidden layer is still an open question. The size of the input layer depends on the number and the coding of the input values. Every normalized attribute can be implemented by one neuron. All teams used one output neuron only. If the output neuron is activated it identifies a canceller. Different hidden layers were used be the teams. Hidden layers range from 2 to 35 neurons. Figure 3 shows an example network.

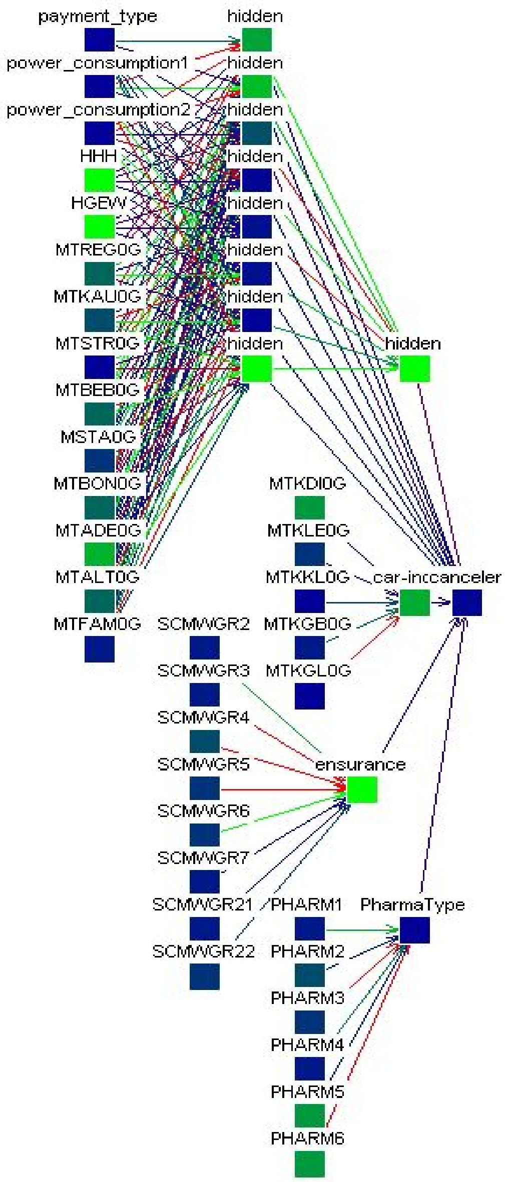

Figure 3 Feed-forward Network (33-19-3-1 layers)

Although several learning algorithms were applied the teams agreed that the standard Backpropagation algorithm showed the best results. We had to realize that a surprisingly small number of training cycles were sufficient. A long training improved the value for the training data a little bit but had a negative influence on the performance for the test data. This well known phenomenon of overfitting was explained within the lectures but personal experiences gained better acceptance.

At the end a critical review of the output of the ANN was the basis for results comparable to the top results of the competition in 2002. One team figured out that the activation of the output neuron differs less than expected. If the bias of the output neuron was decreased a considerably better performance of the network were achieved.

Figure 4 summarizes the results of our teams. It can be seen that comparable or even better results than in the competition in 2002 were achieved. If we value the results we should keep in mind that within the project the results for the test data were known. Although the knowledge was not used for training the network it could be used for the evaluation process of a neural network. Better results are therefore more likely.

Figure 4 Lift value/ result of project teams

1. 8,115€	2. 7,586€
3. 7,378€	4. 4,759€
5. 3,008€	(Using bias shift ca. 5,200 was reached)

CONCLUSION

In order to run a successful student project in the field of neural network applications we need a lot of data, data which have to be freely available. The project work of 2003 showed that not only the input data have to be available but achieved results are desirable as well in order to have good benchmarks for the students. Knowing results, without knowing how they were achieved, can gain more motivation and may raise ambition.

The project work of 2004 has shown that a competition produces not only an atmosphere of competition between the teams but encourages information exchange and even co-operation between teams. Since all the teams address more or less the same problems information exchange is possible without extensive introduction into each others work.

Students especially in 2004 reflect that the project work was fruitful for their understanding. The project convinced the students that ANN can be successful applied. They saw the importance of data pre-processing and output evaluation. A positive side effect was that the students had to apply other knowledge and tools as well. Thus some tools for the pre-processing and the output evaluation were necessary. Programming languages like C++, Java, Perl, or PHP were used to program the utilities. Of course table calculation was applied as well.

We argue that concepts like ANNs should be taught using real world data and moreover that data containing desired results should be available for training. We kindly ask companies and institutions to provide such data for educational purposes.

The project made a deep impact on the students motivation concerning data mining. Some students will go ahead and will take part in the Data Mining Cup 2005. Moreover some students are now working on a problem we just try to solve in co-operation with a German bank. the successful neural network application gave them enough self-confidence to manage now much more real world data without knowing whether results will be possible or not.

Of course we will run the next project in 2005 in a similar way. May be after some years we will have accumulated so many experiences that we will be able to set up some more heuristics for a successful application of neural networks in data mining.

REFERENCES

1. Arciszewski, Tomasz: Engineering Method, IT&E School, George Mason Univ., (IT 894) last access: 2004-10-20: www.civil.gmu.edu/Course%20outlines/IT894Fall04/IT894.04.1.5.ppt
2. Callan, Robert: The essence of neural networks, Pearson Ed, 2002
2. Lämmel, U., Cleve.J.: Lehr- und Übungsbuch Künstliche Intelligenz – Fachbuchverlag – Leipzig, 2004. (in German)
3. Project Documentations 2004, see www.wi.hs-wismar.de/~laemmel/ (in German)