

APPLICATION OF ARTIFICIAL INTELLIGENCE FOR PREDICTING BEER FLAVOURS FROM CHEMICAL ANALYSIS

C.I. Wilson & L.Threapleton

Coors Brewers, Technical Centre, P.O Box 12, Cross Street, Burton-on-Trent,
DE14 1XH, UK
email christopher.wilson@coorsbrewers.com, lee.threapleton2@coorsbrewers.com

Keywords: Flavour, Sensory, Analytical, Models, Artificial Intelligence, Neural Networks, Genetic Algorithms.

INTRODUCTION

We all work in an industry where the consumer is king. We are constantly trying to evolve our products to satisfy the consumer's changing requirements whilst at the same time always looking for the opportunity to develop niche products for new markets. However the relationship between beer flavour and its chemical analysis is poorly understood.

Should it prove possible to predict final beer flavours according to their chemical composition, then it would open up the possibility of 'tuning' such products to meet the expectations of the consumer. The challenge is "Can Beer Flavour Be Predicted From Analytical Results ?"

Substantial empirical data exists, in disparate data sources, concerning product chemical and sensory analysis. However, currently there is no mechanism for linking them to each other. Any such relationships are undoubtedly complex and highly non-linear. In order to identify such relationships we have turned our attention to the modern techniques of artificial intelligence, and specifically neural networks and genetic algorithms.

The former is associated with machine learning whilst the latter is associated with biological evolution. The development of both these fields can be traced back to the 1960s. However it is only recently, with the rapid expansion in computing power combined with the availability of packaged software solutions that these techniques have been moved from the computer science laboratory into industry.

Neural Networks

Neural networks can be visualised as a mechanism for learning complex non-linear patterns in data. A key differentiator from other computer algorithms is that to a very limited extent, they model the human brain. This allows them to learn from experience; i.e. training, rather than being programmed. However training does require significant quantities of data.

When we were at school we were taught to visualise data by plotting it on a graph and joining up the data points. We then progressed to using a technique called linear regression which allowed us to calculate the best gradient and intercept parameters for a straight line such that the sum of the errors was minimised. Finally, in an attempt to achieve a better fit we may have used a polynomial curve fitting programme. The

previously described techniques are particularly suited to simple relationships involving a very limited number of input variables.

In contrast to this a neural network model can handle multiple inputs. These can be associated with multiple outputs which are mapped via non-linear relationships. The process by which a neural network model is developed to provide a best fit function between an input and output data set is known as training. During this training process the network modifies its own internal parameters, known as weights, so as to minimise the difference between the value of the output data set and the values predicted by the network. A key requirement during training is that over training should be avoided thus ensuring that only generalised models are developed which perform equally well on both in sample, and out of sample data. This was achieved using a technique called 'Cross Validation'.

Further information concerning neural networks and their application can be found in references [1], [2] and [3].

Genetic Algorithms

These provide a means of solving complex mathematical models where we know what a good solution looks like but which can not be solved using conventional algebra. The basis of this technique is very simple, Darwin's theory of evolution, and specifically survival of the fittest. Much of the terminology is borrowed from biology. A population is made of a series of chromosomes with each chromosome representing a possible solution. A chromosome is made up of a collection of genes which are simply the variables to be optimized.

A genetic algorithm creates an initial population (a collection of chromosomes), evaluates this population, and then evolves the population through multiple generations. At the end of each generation the fittest chromosomes, i.e. those that represent the best solution, from the population are retained and are allowed to crossover with other fit members. The idea behind crossover is that the newly created chromosomes may be fitter than both of the parents if it takes the best characteristics from each of the parents. Thus over a number of generations, the fitness of the chromosome population will increase with the genes within the fittest chromosome representing the optimal solution. The whole process is similar to the way in which a living species will evolve to match its changing environment.

Introductory information concerning genetic algorithms may be found in reference [4] whilst more advanced material concerning their application may be found in reference [5].

THE FLAVOUR MODEL

Coors Brewers Limited is fortunate enough to have a significant amount of final product analytical data which has been accumulated over a period of years. This has been complimented by sensory data which has been provided by the trained in-house testing panel. The range of analytical and sensory measures available is shown in table 1.

Analytical Data - Inputs

OG
 PG
 FG
 FR (Max)
 Alcohol
 Colour
 CO2 Keg
 pH
 HPLC Isoacids
 HPLC Tetra
 Calculated Bitterness
 Diacetyl
 Chloride
 Sulphate
 Acetaldehyde (Max)
 DMS
 2-Me Butanol
 3-Me Butanol
 Total IAA
 Ethyl Acetate
 Iso Butyl Acetate
 Ethyl Butyrate
 Iso Amyl Acetate
 Ethyl Hexanoate

Sensory Data - Outputs

Alcohol
 Estery
 Malty
 Grainy
 Burnt
 Hoppy
 Toffee
 Sweet
 DMS
 Warming
 Bitter
 Thick

Table 1: Available Analytical Inputs and Sensory Outputs

Initial attempts at modelling the relationship between the analytical and sensory data were restricted to a single quality and flavour and focussed on mapping all available inputs through a single neural network as shown in figure 2.

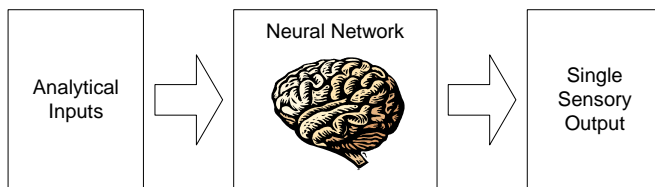


Figure 2: Simple Network

The available data consisted of 350 records which were divided into training (80%) and cross validation (20%) data sets. The neural network was based on Multilayer Perceptron (MLP) architecture with two hidden layers. All data was normalised within the network thereby enabling the results for the various sensory outputs to be compared. Training was terminated automatically when no improvement in the network error was observed during the last one hundred epochs. In all cases training was carried out fifty times to ensure that a significant mean network error could be calculated for comparison purposes. Prior to each training run the source data records

were randomised to ensure a different training and cross validation data set was presented, thereby removing any bias.

The neural network was based on a package solution supplied by NeuroDimension (www.nd.com).

Results using this technique were poor. This was thought to be due to two major factors. Firstly by concentrating on a single product quality the amount of variation in the data was low. This therefore presented the neural network with a very limited opportunity to exact useful relationships from the data. Secondly it was likely that only a subset of the available inputs would impact on the selected beer flavour. Those inputs which had no impact on favour were effectively contributing noise, thus hindering the performance of the neural network.

The first factor was readily addressed by extending the training data to cover a more diverse product range.

Identification of Relevant Analytical Inputs

The problem with identifying the most significant analytical inputs was more challenging. This was addressed by means of a software switch, see figure 3, which enabled the neural network to be trained on all possible combinations of inputs. The premise behind using a switch is that if a significant input is disabled then we would expect the network error to increase, while conversely if the disabled input was insignificant then the network error would either remain unchanged or reduce, due to the removal of noise. Such an approach is known as an exhaustive search since all possible combinations would be evaluated. Although the technique was conceptionally simple it was quickly realised that with the present twenty-four inputs the number of possible combinations, at 16.7 million per flavour was computationally impractical.

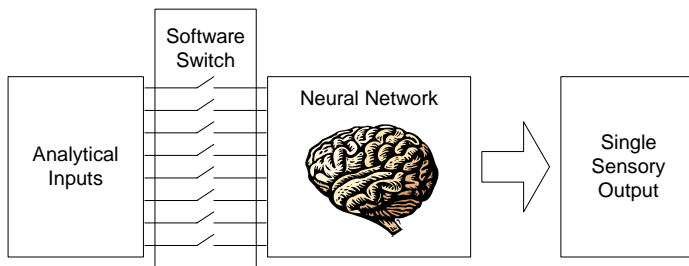


Figure 3: Network with Switched Inputs - Exhaustive Search

What was required was a more efficient method of searching for the relevant inputs. The solution to the problem was to use a genetic algorithm, see figure 4, which would manipulate the various input switches in response to the error term from the neural network. The goal of the genetic algorithm was to minimise the network error term. The switch settings made when this minimum was reached would identify those analytical inputs which could best be used to predict the flavour.

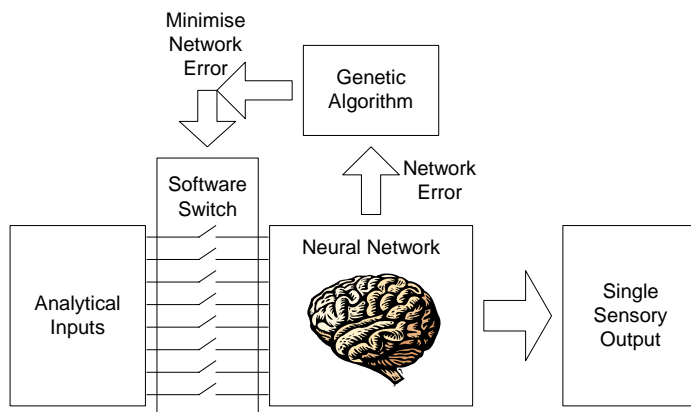


Figure 4: Network with Switched Inputs Controlled by a Genetic Algorithm

The results of this work are summarised in table 5.

Analytical Input	Sensory Output											
	Alcohol	Estery	Malty	Grainy	Burnt	Hoppy	Toffee	Sweet	DMS	Warming	Bitter	Thick
Iso Butyl Acetate	No	No	No	No	No	No	No	No	No	No	No	No
Alcohol	No	No	No	No	No	No	No	No	Yes	No	No	No
Diacetyl	No	No	No	No	No	No	Yes	No	No	No	Yes	No
Ethyl Acetate	No	No	No	Yes	No	No	No	No	No	Yes	No	No
FG	No	No	Yes	No	No	Yes	No	No	Yes	No	No	No
FR (Max)	No	No	No	No	No	No	No	Yes	No	Yes	Yes	Yes
HPLC Isoacids	No	No	Yes	Yes	No	No	No	No	No	Yes	Yes	No
2-Me Butanol	No	No	No	Yes	No	Yes	Yes	Yes	No	No	No	No
Iso Amyl Acetate	No	Yes	Yes	No	No	Yes	No	No	Yes	No	No	No
Ethyl Hexanoate	No	No	Yes	No	No	Yes	Yes	No	Yes	No	No	No
pH	No	Yes	No	No	Yes	Yes	Yes	No	Yes	No	No	Yes
Chloride	No	No	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	No
3-Me Butanol	Yes	No	No	Yes	No	No	Yes	No	No	Yes	Yes	Yes
Total IAA	No	No	No	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes
OG	Yes	No	No	No	Yes	Yes	Yes	No	Yes	No	Yes	Yes
PG	Yes	Yes	No	Yes	No	Yes	No	No	Yes	Yes	Yes	No
Sulphate	Yes	No	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Acetaldehyde (Max)	Yes	Yes	No	No	No	Yes	No	Yes	Yes	No	Yes	Yes
Ethyl Butyrate	No	No	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Colour	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No
CO2 Keg	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No
HPLC Tetra	Yes	No	Yes	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes
Calculated Bitterness	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	No	Yes
DMS	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Figure 5: Relevant Analytical Inputs as a Function of Sensory Output

The above results suggest that in some instances, i.e. Iso Butyl Acetate there was no discernable relationship between the analytical input and any flavour whilst in other cases, i.e. DMS, the input may impact on a large number of flavours. It was also evident that typically any one flavour may be influenced by a large number of inputs. For example the DMS flavour was found to be influenced by fourteen of the total of twenty four available inputs. Although this work identified which inputs were relevant it did not allow the relative significance of each input to be calculated.

Prediction of Beer Flavour

Having determined which inputs were relevant it was now possible to identify which flavours could be more ably predicted. This was done by training the network, using the relevant inputs previously identified multiple times. Prior to each training run the network data was randomised to ensure that a different training and cross validation data set was used. After each training run the network error was recorded. A good flavour predictor should have both a small network error and associated standard deviation. The results, see figure 6, indicated that it should be possible to predict the ‘Burnt’ and ‘DMS’ flavours and yet would only poorly predict those flavours with low scores such as the alcohol flavour.

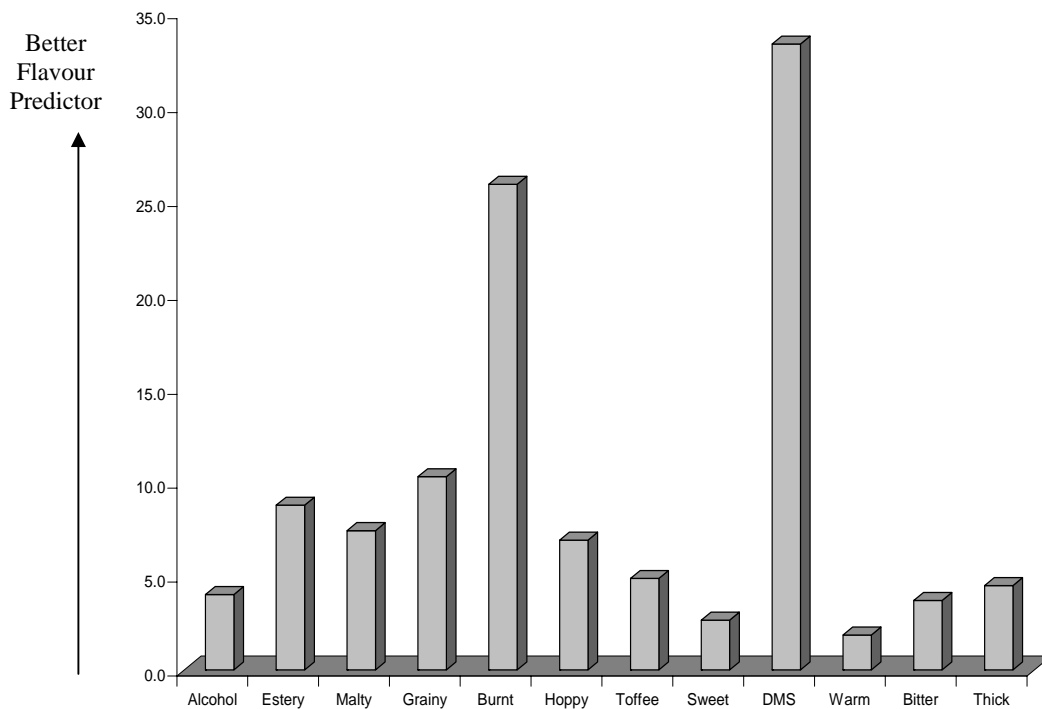


Figure 6: Estimate of Quality of Prediction

However the acid test is “Can the flavour be predicted based on out of sample data”? To answer this question the available analytical and sensory data was divided into three unequal sets. These were used respectively for training, cross validation and testing. The network was trained using the training and cross validation sets. The testing set, which was comprised of approximately eighty records of out of sample data, was used for assessing the performance of the trained network.

Firstly we turn our attention to the ‘Burnt’ flavour. A correlation coefficient of 0.87 was achieved showing good correlation between the predicted burnt flavour from the neural network and the flavour as determined by the sensory results, see figure 7. However there are still shortfalls in predicting peak sensory values. Nevertheless this model does show a degree of robustness.

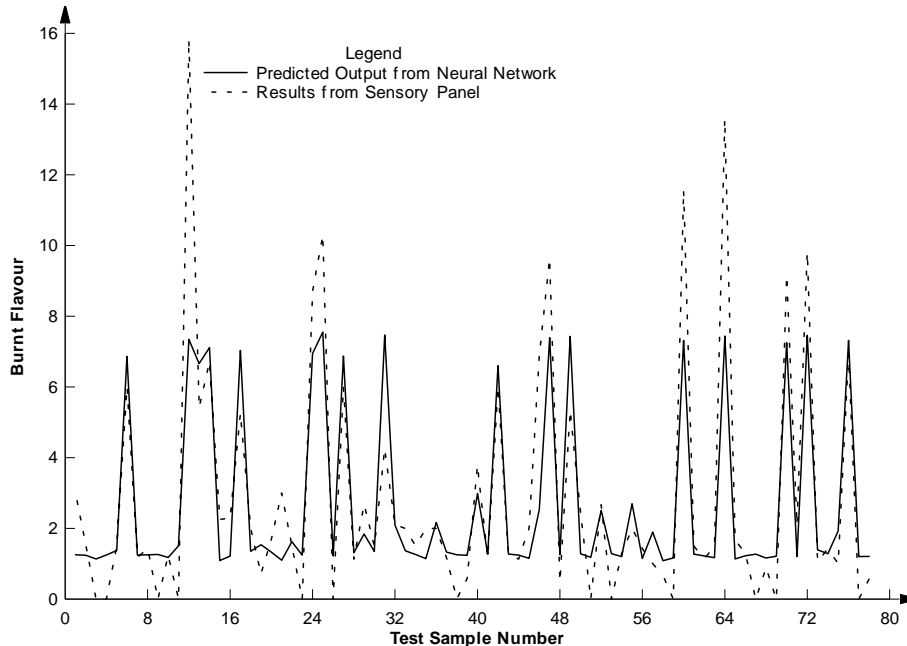


Figure 7: Neural Network Burnt Flavour Prediction Vs Sensory Results

Unfortunately one of the shortcomings of neural networks is that they do not explain their results nor do they provide a readily available mathematical equation. This disadvantage can be addressed to a limited degree by probing the model to understand which analytical inputs are important. This process is generally known as sensitivity analysis. For the ‘Burnt’ flavour each analytical input was individually ‘disturbed’ by ten percent and the change in output, the predicted ‘Burnt’ flavour was measured and expressed as a percentage. As can be seen from figure 8 an increase in Carbon Dioxide was found to decrease the ‘Burnt’ flavour whilst increasing IAA would tend to promote the flavour. On a cautionary note it should be appreciated that neural networks simply recognise patterns in data and therefore such sensitivity results do not necessarily imply a cause and affect relationship.

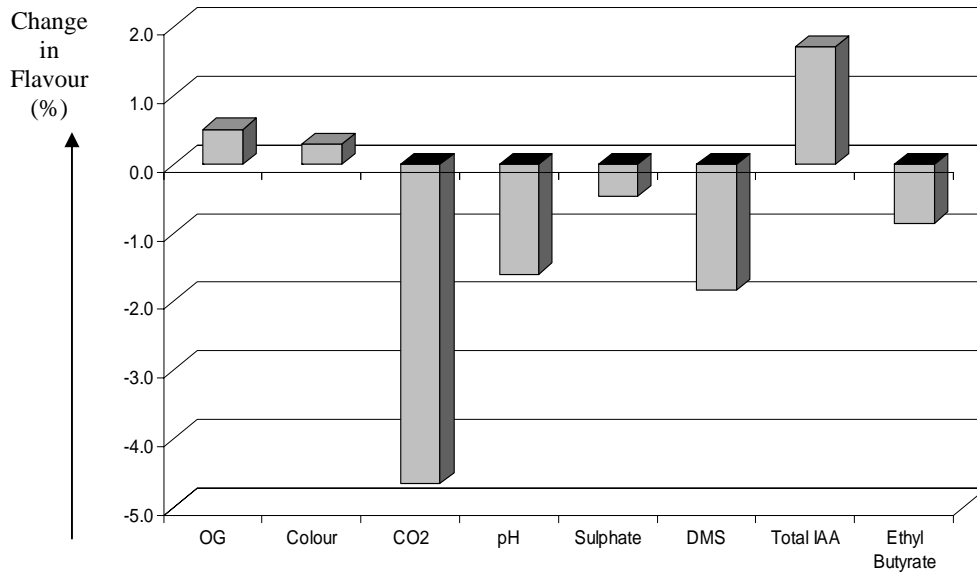


Figure 8: Sensitivity Analysis for the Burnt Flavour

Earlier work suggested that we would be most able to predict the ‘DMS’ attribute. This is borne out in practice, see figure 9, with out of sample testing showing a correlation coefficient of 0.92. This time the network is accurately able to predict the low and mid range values but still lacks the ability to predict the very high extremes.

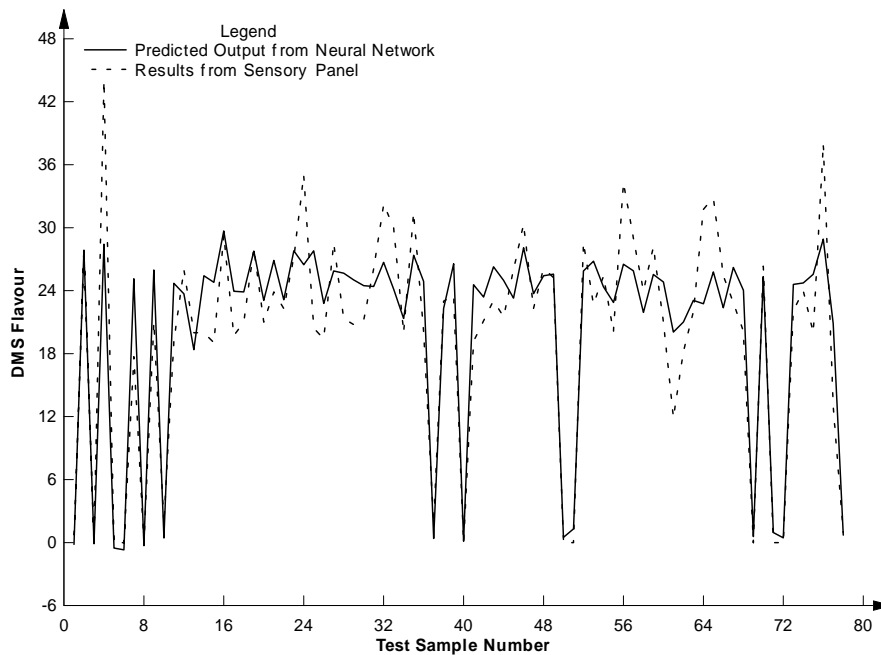


Figure 9: Neural Network DMS Flavour Prediction Vs Sensory Results

Beer Flavour Optimisation

Currently two neural network models have been built, which are to a reasonable degree able to predict the 'Burnt' and 'DMS' characteristics. In total these models have sixteen inputs of which six are shared by both characteristics. The limitation of these models is that they only predict in one direction. That is, they will only predict sensory flavours from the analytical inputs. It would perhaps be more useful if they could be reversed so that given a target sensory characteristic they would calculate the required analytical inputs. This problem can not be solved by conventional algebra. However it is known what a good solution would look like, i.e. when the predicted and target sensory values are identical and therefore it is possible to solve this problem using a genetic algorithm, see figure 10.

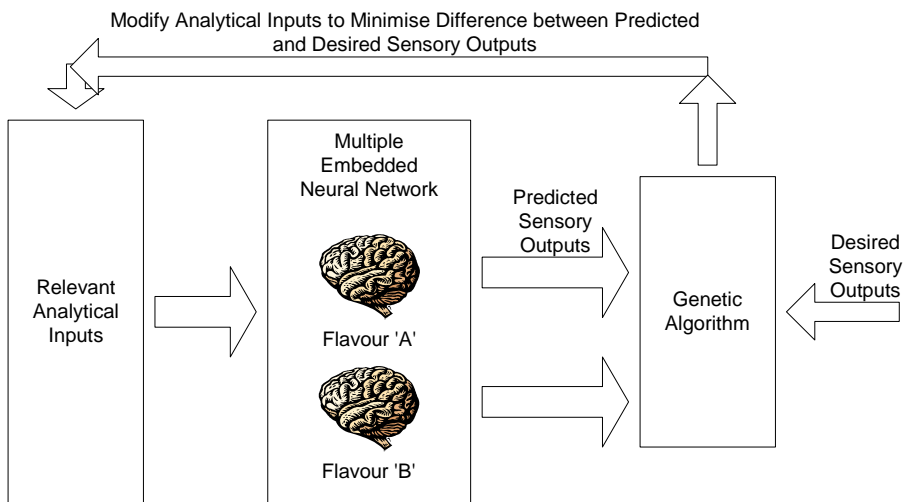


Figure 10: Flavour Optimiser

CONCLUSIONS

Can Beer Flavour Be Predicted From Analytical Results ? Today the answer is a conditional yes, but only for a very limited number of flavours. Sensory response is extremely complex, with many potential interactions and hugely variable sensitivity thresholds, from % to parts per trillion. Standard instrumental analysis tends to be of gross parameters and many flavour active compounds are simply not measured for practical or economical reasons. The relationship of flavour and analysis can only be effectively modelled if a significant number of flavour contributory analytes are measured. What is more, it is not just the obvious flavour active materials but also mouthfeel and physical contributors to the overall sensory profile that should be considered.

With further development of the input parameters the accuracy of the neural network models will improve.

FURTHER WORK

However what is most exciting, is that these techniques show much potential. They have demonstrated an ability to mine data across disparate data sources and develop credible models. Such models, which can represent complex relationships, can be used as the basis for process optimisation.

This paper has concentrated on sensory and analytical data. However our business is much wider than this. Even limiting ourselves to the supply chain, many breweries have substantial quantities of information relating to:

1. Raw Materials
2. Process Conditions
3. Analytical Results
4. Sensory and Consumer Preference Data

There are some broad understandings of relationships, but a poor understanding across the whole process. The use of neural networks and genetic algorithms offers the possibility of modelling across the whole process, from raw materials and process parameters to the preferences of the consumer.

REFERENCES

1. Swingler K., Applying Neural Networks - A Practical Guide, ISBN 0126791708 - Morgan Kaufman Publishers.
2. Callan R., The Essence of Neural Networks, ISBN 013908732X - Prentice Hall Europe.
3. Principe J. Euliano N. Lefebvre C., Neural and Adaptive Systems - Fundamentals Through Simulation, ISBN 0471351679 - John Wiley & Sons.
4. Mitchell M., An Introduction to Genetic Algorithms, ISBN 0262631857 - MIT Press.
5. Gen M. Cheng R., Genetic Algorithms & Engineering Design, ISBN 0471127418 - Wiley Interscience Publications.

This article has been reprinted with the permission of Coors Brewers Limited and the European Brewery Convention, from the Proceedings of the 29th EBC Congress - Dublin 2003.