

## Predicting the Speed of Beer Fermentation in Laboratory and Industrial Scale

Juho Rousu<sup>1</sup>, Tapio Elomaa<sup>2</sup>, and Robert Aarts<sup>1\*</sup>

<sup>1</sup> VTT Biotechnology and Food Research, P.O. Box 1500  
FIN-02044 VTT, Finland, juho.rousu@vtt.fi

<sup>2</sup> Department of Computer Science, P.O. Box 26  
FIN-00014 University of Helsinki, Finland, elomaa@cs.helsinki.fi

**Abstract.** Characteristic of the beer production process is the uncertainty caused by the complex biological raw materials and the yeast, a living organism. This uncertainty is exemplified by the fact that predicting the speed of the beer fermentation process is a non-trivial task.

We employ neural network and decision tree learning to predict the speed of the beer fermentation process. We use two data sets: one that comes from laboratory-scale experiments and another that has been collected from an industrial scale brewing process. In the laboratory-scale experiments a neural network that employs characteristics of the ingredients and the condition of the yeast, could predict the fermentation speed within 2% of the true value. Decision trees for classifying whether the speed of fermentation will be slow or fast were constructed from the same data. Astonishing simple decision trees were able to predict the classes with 95%–98% accuracy. In contrast to the neural net experiment, even the highest accuracy could be reached by utilizing only standard brewery analyses.

We then set out to check the utility of these methods in a real brewery environment. The setting in the brewery is more complex and unpredictable than the laboratory in several ways. Regardless, reasonably good results were obtained: the neural network could, on average, predict the duration of the fermentation process within a day of the true value; an accuracy that is sufficient for today's brewery logistics. The accuracy of the decision tree in detecting slow fermentation was around 70%, which is also a useful result.

### 1 Introduction

The art of producing beers has developed over 5000–8000 years. Nevertheless, the complexity of the process still provides challenges to the brewers. Both the complexity of the ingredients and the unpredictable nature of the yeast, a living organism, contribute to the uncertainty that the breweries are forced to live with.

---

\* Current address: Nokia Telecommunications, P.O. Box 370, FIN-00045 Nokia Group, Finland, robert.aarts@nokia.ntc.com.

From the production management point of view, the ability of predict the duration of the fermentations would be a useful one [3]. In practise, the fermentation times in seemingly equivalent settings can vary considerably, which hinders efficient scheduling of the plants. Moreover, the breweries are forced to make daily measurements to observe the course of the fermentations, in order to make the decision when to stop the process. With a good predictor for the fermentation speed, one could manage with fewer measurements.

In this paper we study how two predictor families, neural nets and decision trees suit this problem. The task of the neural net is to predict the fermentation time and the task of the decision tree is to classify the batches as slow or fast. The neural net prediction gives continuous classification while the decision tree is understandable even to the brewers. We perform two sets of tests. The first set is performed with data from laboratory tests. The second data set is collected from a real brewery.

The rest of this paper is organized as follows. First, Section 2 briefly explains the beer fermentation process. Section 3 reviews the results that were obtained on the laboratory-scale data. Section 4 goes through the results that were obtained on the brewery data. Finally, Section 5 presents the conclusions of the current work.

## 2 The beer fermentation process

The main ingredients of beer are malt, water and hops. The main phases of the brewing process are *wort production* and *fermentation*.

The wort production starts with crushing the malt into coarse flour, which is then mixed with water. The resulting porridge-like mash is heated according to a carefully selected temperature program which encourages the malt enzymes to partially solubilize the ground malt. The resulting sugar-rich aqueous extract, *wort*, is then separated from the solids and boiled with hops. The wort is then clarified and cooled.

The fermentation process starts with aerating the cooled wort and adding yeast to it. The yeast starts to consume the nutrients contained in wort, in order to stay alive and grow. At the same time, the yeast produces alcohols and esters. Fermentation is controlled by regulating the temperature, oxygen content, and the pitch rate; i.e., the amount of yeast put into the fermentation tank. Temperature has a great effect on both the speed of fermentation and the flavour of beer. The growth of yeast can be controlled by the oxygen content. The pitch rate affects the fermentation speed but not as much as the temperature. However, the effects of pitch rate on flavor are small which permits larger changes without altering the flavor profile.

In addition, the course of fermentation is affected by other factors, such as the wort composition and the yeast condition. Ideally, these factors should be constant, so that the predictability of fermentation is maintained. In practice, neither the wort composition nor the yeast condition is static. The natural vari-

ation of malt induces some variation to the wort composition, although such variations can be diminished by re-planning the mashing recipes [1, 2].

The condition of the yeast is a more complicated issue. Traditionally, the breweries have observed the *viability*, i.e. the percentage of live cells in the batch by laboratory analyses. However, these methods do not tell anything about the *vitality* of the yeast, i.e. the fermentation rate of the cells. The yeast used in brewing is grown by the brewery and recycled many times before disposal. The ability of the yeast to ferment is greatly dependent on the history of the yeast. For example, new yeast typically behaves differently from yeast that has been recycled many times. Also, yeast that has been stored long periods between fermentation is often less vital.

Ideally, the brewery should be able to modify the fermentation recipes so that the variability of the yeast and wort would be canceled out. So, if the vitality of the yeast is low, the brewery could increase the pitch rate or elevate the temperature or oxygen content slightly. A fermentation recipe planner, such as the Sophist system [8] is well suited to this task. A reliable estimate of the yeast vitality is needed for such an approach, though. However, as one can expect from the above introduction, no single analysis exists that would permit predicting the time of fermentations to any reasonable degree.

### 3 Results using laboratory-scale data

A set of 100 fermentations [4] was used for both the artificial neural net (ANN) and the decision tree experiments. This data set contains fermentations with recycled yeast (up to 4 cycles) and fermentations with freshly propagated yeast. The worts used in these experiments were all made according one recipe using a single lot of malt extract. Hence the worts were all very similar indeed. Yeast viability was assessed by *methylene blue* (MB) and *methylene violet* (MV) staining, both at the end and at the start of a fermentation. In addition the *trehalose content* of the yeast, which is a stress indicator, was measured before pitching. The pitching rate was constant. As a fourth yeast condition measurement the *acidifying power* (AP) was recorded. Cropped yeast was aerated for 0, 3 or 5 hours before pitching. The percentage of apparent fermentation—the percentage of sugars consumed—was calculated from daily measurements of the *specific gravity* (SG) of the wort. A review of these measurements is given, e.g., by Londesborough [5].

#### 3.1 Neural net results

The first approach was to train ANN on this data. In the work presented here an ANN was trained to predict the relative degree of fermentation at 72 and 130 hours. Several sets of inputs were used, in order to see what analysis contribute to the quality of prediction.

A number of neural nets to estimate the apparent degree of fermentation at 72 and 130 hours were trained. For each net approximately 75% of the available

**Table 1.** The error of prediction of degree of fermentation of neural nets using different measurements. The errors are given in absolute percentages, e.g. the difference between the predicted value and the actual measured value was never more than the given error. "Prev. adf" means the measured degree of fermentation of the batch that the yeast was cropped from. This value is not available when freshly propagated yeast is used.

MB	MV	Trehalose	Aerat. time	SG	Wort O <sub>2</sub> /pH/temp	Prev. adf	Error ADF <sub>72</sub>	Error ADF <sub>130</sub>
X	X		X	X	X	X	±0.4%	±0.7%
X	X		X	X	X		±10.1%	±12.6%
X	X	X	X	X	X		±8.5%	±6.4%
X			X			X	±8.1%	±13.8%
X	X		X	X		X	±3.4%	±4.9%

data was used for training and 25% for validation. The nets differed in the input measurements used. Table 1 lists the inputs to the nets that were constructed and the prediction errors of these nets.

It can be seen that information about the behavior of the yeast in previous batches is rather useful, inclusion of this data reduces the error of prediction significantly. For freshly propagated yeast such data is not available and it is therefore more difficult to predict the behavior of such yeast. Adding information about the physiological condition of the yeast in the form of the trehalose measurement helped prediction in this case.

### 3.2 Decision tree results

Another approach to prediction of starting speed of fermentation is to classify a batch as slow, normal or fast, based on the descent of specific gravity in the first 72 hours. We assigned these classes according to "natural" clusters that were seen in the data. These kind of classification tasks are particularly suitable to symbolic learning methods such as decision tree or rule inducers. A benefit of these kind of methods is that the predictor, i.e. the rule, is understandable to humans. We used the latest version of the well-known C4.5 decision tree learning package [6, 7] to build the classification rules.

In the decision tree experiments, the predictive accuracies of the trees were estimated by 10x10-fold cross-validation method that works by dividing the data into 10 subsets and then uses each subset at a time as test set while the other 9 subsets are used for training. The whole process was repeated 10 times and the accuracies and standard deviations reported below are averages over the 10 iterations. Cross-validation accuracies of this kind are considered reliable estimates of the performance of the prediction method on unseen cases. In our first test, all measurements of yeast condition and wort quality were available to the learning algorithm to choose from. In addition, the performance of the yeast in the previous batch, that is, whether the start of the fermentation in the previous batch was slow, normal or fast.

Of this set only two measurements appeared in the tree (Table 2) induced from the whole data, namely methylene blue measurement, and somewhat surprisingly, the specific gravity at the start of the fermentation. The training accuracy of the depicted tree, as well as that of the trees on Tables 3 and 4, is 98%. The cross-validation accuracy (i.e. the estimated performance on unseen cases) of this scheme is  $97.8\% \pm 0.4\%$ , meaning that circa 2% of new batches would be misclassified using this rule.

**Table 2.** Rule induced from the whole data.

```
if MB > 8.8 then slow
else if SG <= 1.04353 then normal
else if MB <= 3 then fast
else normal
```

Next experiment was to exclude measurements that were chosen at the first round, one at a time, in order to see how dependent the predictability was of the two preferred measurements. Excluding methylene blue measurement had the effect of bringing the acidifying power (AP) measurement into the tree (Table 3). The original gravity was present this tree also. The cross-validation accuracy of this setting dropped to 95%, suggesting that methylene blue is more robust measurement for this task than the acidifying power. A minor surprise was that methylene violet measurement did not appear in the rule, even though it was deemed useful in the neural net experiments.

**Table 3.** Rule induced from the data where methylene blue measurement was excluded.

```
if AP <= 2.1425 then slow
else if SG <= 1.04353 then normal
else if AP <= 2.515 then normal
else fast
```

Our third experiment was to exclude the original specific gravity from the set of available measurements. The effects were parallel to the second experiment: the specific gravity was replaced with pH of the original wort. The cross-validation accuracy was  $94.8\% \pm 0.7\%$ , again pointing out that the original gravity is a more informative measurement in this setting.

The immediate conclusion to be drawn of these decision tree experiments is that predicting whether a batch will be slow or not can be done with surprisingly little information about the wort quality and the yeast condition. Only two measurements of both appear in the three decision trees that, moreover, all have very high accuracy.

**Table 4.** Rule induced from the data where original specific gravity measurement was excluded.

```
if MB > 8.8 then slow
else if pH > 5.53 then normal
else if AP <= 2.515 then normal
else fast
```

One important question arose from these experiments: why seems the original specific gravity of the wort to be necessary in predicting the speed of fermentation. This finding seems peculiar since the wort was of very even quality in the different batches. Another question was why replacing SG measurement with pH obtains almost as good results. A technical answer to this question is that the two measurements are quite strongly inversely correlated ( $r=0.7818$ ) in this data. Still, answer to the fundamental question why either of these measurements is relevant, remains unclear.

## 4 Results on brewery data

We set out to validate the laboratory results in industrial scale. For that end, we collected the data of 118 fermentations from a brewery.

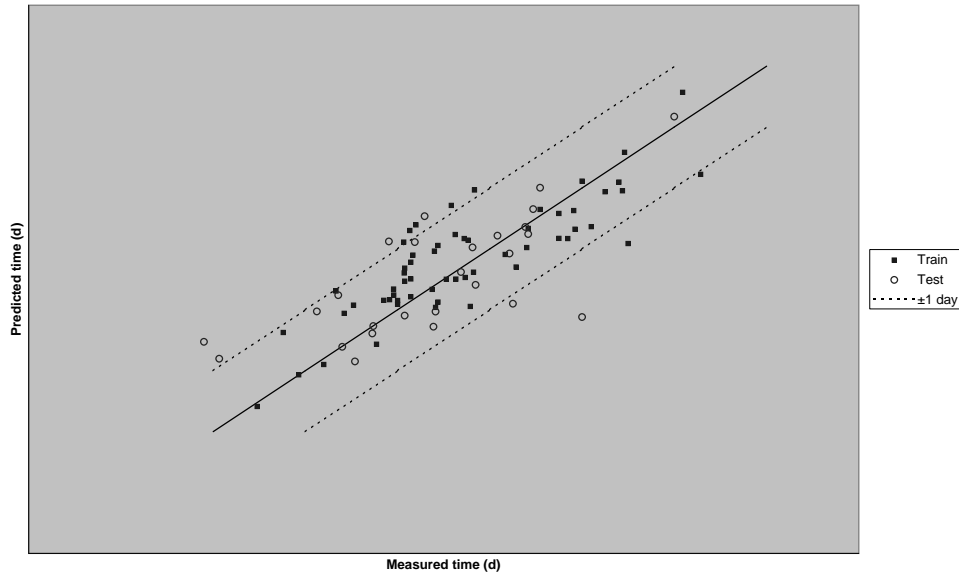
The set of attributes in the data was different from the one used in the laboratory tests: The brewery uses a online capacitance measuring device for assessing the viability of the yeast mass rather than the staining methods. The benefit of the former approach is that the whole yeast mass becomes measured instead of a small sample. In addition, the volume of the pitched yeast was used as an additional measurement.

Since the history of the yeast was found important in the laboratory data set, the fermentation time of the previous round of the yeast was included. Also, the length of the history of the yeast as the number of fermentations was included. The propagations where the yeasts originated from were more numerous which contributed to variation that could not be coded into the data.

The fermentation tank was filled with two brews that entered the tank in intervals of varying length (several hours in each case). We found it necessary to include the first SG measurement performed from the full tank into the data set in order to manage this complication, in addition to the average of the original specific gravities of the two brews. The interim time between the two brews was also included into the data set.

### 4.1 Neural net results

We trained a neural network using the backpropagation learning algorithm using the momentum term to avoid local minimas. The data was split into training (70%) and test sets (30%). The size of network was decided by manually checking



**Fig. 1.** Neural network prediction results. Each dot represents one prediction: the squares correspond to data items that were included into the training set and the circles represent predictions in fresh cases. The solid line corresponds to the correct prediction and the dashed lines are the one-day error margins.

the predictive accuracy on the test set. A network of 2x4 hidden units was found to give a good result when all yeast strains were present in the data. In contrast, as small as 1x3 unit network was found to generalize well when the data was restricted to include just one strain.

The predictions given by the network of 1x3 hidden units are depicted in the Figure 1, which plots the fermentation speed predictions given by the network on the training and test data against the measured duration. A correct prediction falls on the solid line. The dashed lines represent error margins of one day. It can be seen that most predictions are within the one-day error limit. The average deviation of the predictions is 0.6 days (14.4 hours), which is clearly worse than the best results (1 hour and 6.5 hours) obtained on the laboratory-scale data. However, taking the more complicated real-world setting into account the result is satisfactory.

#### 4.2 Decision tree results

Again the ten times repeated cross-validation testing was applied with the C4.5 decision tree learning algorithm [6, 7]. The intention was not to predict the actual

speed of fermentation, but simply to classify whether the fermentation is slow or fast. The axis-parallel single-attribute value categorization carried out by C4.5 does not have the chance of producing as good results in the very noisy real-world data as neural networks do. Nevertheless, the intelligible classifiers produced are of interest.

Depending very much on the attribute set used quite different looking decision trees were produced. Quite heavy pruning was needed to delete the many apparent dependencies produced by noise. Despite all, the prediction accuracy level using decision trees is quite stable: in repeated cross-validation 67–70% prediction accuracy was obtained with many different settings. In contrast, the overall prediction accuracy of the neural network on the brewery data is approximately 73%.

## 5 Conclusions

In light of the laboratory-scale experiments, it is well possible to predict the behavior of beer fermentations. If accurate predictions are desired it is necessary to have detailed information about the wort and the history of the yeast. Neural nets can be trained on such data, probably one net for each strain. If an early warning for exceptionally slow (or fast) fermentations suffices, it is possible to use a simple decision tree that employs only a very small set of routine measurements.

In the brewery data, however, there exists variation that cannot be directly encoded into the input variables. For example, the propagations where the yeast originated from were distinctly different from each other. Therefore, the results cannot reach the same exactness as in the laboratory-scale experiments. Nevertheless, the obtained results would seem to be applicable in the breweries.

## References

1. Aarts, R., Sjöholm, K., Home S., Pietilä K.: Computer-Planned Mashing. In *Proceeding of the Twenty-Fourth Congress, European Brewery Convention*. IRL Press, Oxford (1993) 655–662
2. Aarts, R., Rousu, J.: Towards CBR for Bioprocess Planning. In Smith, I., Faltings, B. (eds.): *Advances in Case-Based Reasoning, Proceedings of the Third European Workshop, EWCBR-96*. Lecture Notes in Artificial Intelligence, Vol. 1168. Springer-Verlag, Berlin Heidelberg New York (1996) 16–27
3. Cummins, S., Plant, N., Kelleher, P., O'Connor, J.B.: Optimisation of Brewery Operations Using Fuzzy Logic and Simulation Tools. *Proceedings of the International Symposium on Automatic Control of Food and Biological Processes*. SIK. Göteborg, Sweden (1998) 459–467
4. Kataja, K.: *Yeast Recycling in Main Fermentation of Beer* (in Finnish). Master's thesis, Department of Chemical Technology, Helsinki University of Technology, Finland (1997)
5. Londesborough, J.: The Measurement of Yeast Viability in Breweries (in Finnish). *Mallas ja Olut* 5 (Oct. 1998) 139–148

6. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, Calif. (1993)
7. Quinlan, J.R.: Improved Use of Continuous Attributes in C4.5. *J. Artif. Intell. Res.* **4** (1996) 77–90
8. Rousu, J., Aarts, R.: Case-Based Planning Methods in Biotechnical and Food Processes. Proceedings of the International Symposium on Automatic Control of Food and Biological Processes. SIK. Göteborg, Sweden (1998) 215–224