DATA MINING BREAD QUALITY AND PROCESS DATA IN A PLANT BAKERY

A.J. Wilson[1], M.P.Morgenstern[1], B Pfahringer[2], C Leschi[3]

[1] Crop & Food Research Ltd, Lincoln, New Zealand
[2] University Of Waikato, Hamilton, New Zealand
[3] Institut National des Sciences Appliquées de Lyon, France

INTRODUCTION

In modern automated plant bakeries a large amount of data is collected on the operation of the plant. When this data is combined with product quality data such as loaf colour, appearance, consumer complaints, sales data etc, it has the potential to be used to improve processing efficiency, final product quality, and product marketability. However the huge volume of this data means it is often ignored as being too hard to analyse in any meaningful way. Data mining, which is a combination of techniques that produces information from large data sets has the potential to be applied to this data to extract useful information.

Data mining is a relatively recent approach to data analysis, which uses techniques from computer science, statistics and data visualisation to search for patterns and relationships in large data sets. While finding its main application in the financial and marketing worlds the techniques have been applied in processing industries, and more recently in the baking industry (Rousu et al 2003).

In this paper we describe a pilot scale data mining project we carried out at a plant bakery in New Zealand. A bakery was selected that had as much automation as possible to allow automatic collection of data, and long runs of one variety to allow the process to stabilize.

METHODS

Bakery description

The bakery process consisted of two MDD mixers alternatively feeding a single divider, check weigher, rounder and intermediate proover. The dough pieces were processed with two separate moulders, each feeding opposite sides of 6x2 tin straps creating two parallel streams of dough/bread. The dough/bread was prooved, baked and cooled in two streams before being sorted into three streams for slicing. These changes from one to two to three streams plus tin strap order reversal in entering and leaving the oven meant it was impossible to track individual dough pieces/bread through the process.

Data collection

Data were collected at the bakery using four portable PC's with data acquisition hardware and software and one data logger. The PC's recorded divider check weight, proof height, loaf height, digital photos, and data from the PLC control system (Table 1). As not all the bakery's equipment was on the PLC system some values had to be recorded by other means (data logger, bakery paperwork and some manually). Due to computer constraints none of the mixer data (Flour weight, water weight, water temperature, flour temperature, total work input, mixer efficiency, mix time, recipe number, divider dough temperature, divider level) could be recorded.

Table 1. Bakery properties with measuring method and time between measurements

Property	Measuring method	Time between measurements
Flour report	Manual	~ daily
Dough temperature	Manual	30 mins
Dough piece weight	Check weigher	0.6 secs
Proover temperature, humidity	Data logger	5 mins
Proof height	Laser triangulation	1.2 secs
Oven temperature (3 zones)	PLC*	1 min
Cooler temperature, humidity (in/out)	PLC	1 min
Bread room humidity, exhaust temperature	PLC	1 min
Bread despatch temperature	PLC	1 min
Loaf height, top collapse	Rotary encoder	1.2 secs
Loaf quality shape, colour, four piecing	Manual from photo	~2 mins**
Loaf quality score texture, holes, cores, temperature	Manual from loaf	30 mins
Loaf weight	Check weigher	0.6 secs
Outside temperature, humidity, pressure, wind	Internet	Daily

* PLC data were recorded every minute but the individual items were read sequentially with 1 second between items to avoid overloading the data highway. The 48 properties read included set points, and counter values and not all were used in the analysis.

** Photo taken randomly every 125 loaves with at least 25 loaves between photos (to allow the flash to recharge). Properties were manually scored from the photo.

The project ran for approximately two and a half months with data collection taking place for 4 days each week. At the end of each week the files were uploaded from the bakery to a computer network where scoring of the photos, data cleaning and processing took place.

Data cleaning and preparation

Raw data from the various sources (PLC, data loggers, sensors) went through an extensive process to delete outliers, fill in missing data and link them together with a unique identifier. Data was split into tables representing various stages along the production line (e.g. proover, oven or slicers). Averaging over 1 minute intervals provided tables with properties of clusters of dough pieces/loaves versus date and time with values typically averaged over 100 values (6000 loaves/hr). Travelling times between the different stages in the bread making process was determined (and assumed constant throughout a day). Using these times (Table 2) the average values of properties and parameters that the particular cluster of dough experienced during processing was determined. This produced a table with properties and parameters versus cluster identifier (Table 3). This final data set was used for subsequent analysis.

Table 2. Clusters tracked through the process

ClusterID	Time	ProoverIn	ProoverOut	OvenIn	OvenOut	CoolerIn	CoolerOut	SlicersIn
3710	3:55:21	4:03:28	5:02:13	5:04:30	5:26:43	5:29:58	7:00:01	7:14:20
3711	3:56:21	4:04:28	5:03:14	5:05:31	5:27:43	5:30:58	7:01:01	7:15:20
3712	3:57:22	4:05:29	5:04:15	5:06:32	5:28:44	5:31:59	7:02:02	7:16:21
3713	3:58:23	4:06:30	5:05:16	5:07:33	5:29:45	5:33:00	7:03:03	7:17:22

Table 3. Properties and parameters versus cluster identifier

ClusterID	RecipeID	DivWeight	TProover	RHProover	ProofHeight	TOven	TCooler	LoafHeight	TopCollapse	SlicerWeight
9148	2	816.5	40.3	87.5	157.0	229.2	37.9	108.9	1.9	715.2
9150	2	813.0	40.5	87.4	159.4	229.4	37.9	121.3	2.7	715.0

9290	2	808.8	41.9	76.2	179.8	225.6	39.1	131.2	5.1	702.6
9291	2	810.4	41.6	75.4	178.5	225.6	39.1	131.5	4.8	702.4

The use of standard process times to determine what stage the bread had reached in the process meant that the data was only valid when the plant was running without stoppages. This was determined by looking at the values of the dough piece/loaf counters with data being marked as invalid if the counter rate dropped significantly below the standard production rate.

Selected data were extracted from the database using SQL queries. Extracted data were then converted to the correct format for loading into the WEKA data mining workbench (Witten and Frank 2000).

ANALYSIS

Overview

The data were analysed in two ways. Firstly each process step where individual dough pieces or loaves were measured was looked at as a time series. For these data patterns could be measured that reflected the performance of divider pockets, tin strap position, oven position etc. Secondly the data were averaged over 1-minute intervals to give average conditions for each 1-minute block of dough/bread as it passed through the plant.

Time series analysis

Data on every dough/loaf was measured by the divider check weigher and half the loaves by the proof height and loaf height sensors. Because of the way the conveyors operate in the bakery the gap between loaves was set by the equipment. The divider drops five dough pieces on the conveyor at a time, and there is a small delay before the next five are dropped. By looking at the time between dough pieces it was possible to determine what pocket a dough piece came from. A similar situation existed for tin strap position at the proof height sensor, and oven position and tin strap position at the loaf height sensor.

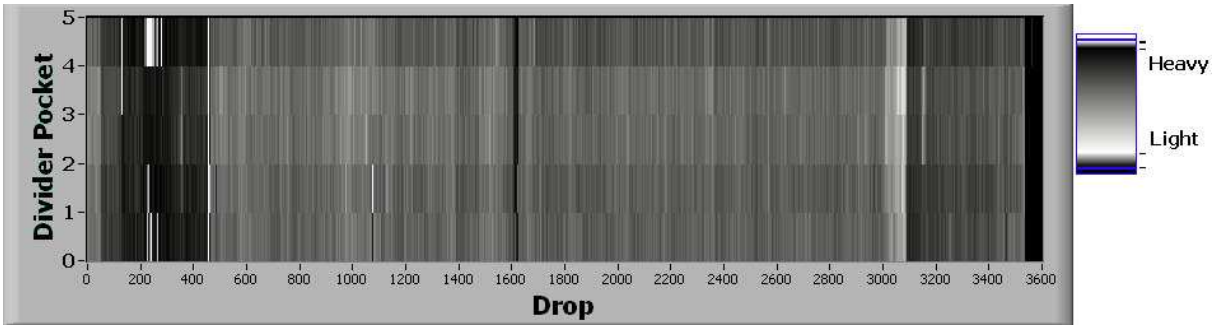

Figure 1 Dough piece weight by divider pocket and drop where the greyscale indicates the relative weight.

Figure 1 demonstrates how this data is used through visualisation. There is a clear difference between the weights produced by the different divider pockets with pockets 1, 2 and 5 weighing heavier than 3 and 4. Plotting the data in this way provides a tool for the engineer to adjust individual pocket weight. A similar approach could be used for loaf height and top collapse. Both showed patterns indicating effects of position in the tin and position in the oven.

Data mining

The analysis approach was to use classification and visualisation to look for patterns in the data. Algorithms applied included neural networks, regression, nearest neighbour, and decision trees.

Top collapse (where the top surface of a lidded loaf bows in) is an issue in many bakeries and the cause is not always clear. By dividing the top collapse data into categories (high, low) we could apply a tree classifier (J48, Witten and Frank 2000). A relationship was found (Figure 2.) with proof height, proover humidity and cooler temperature.

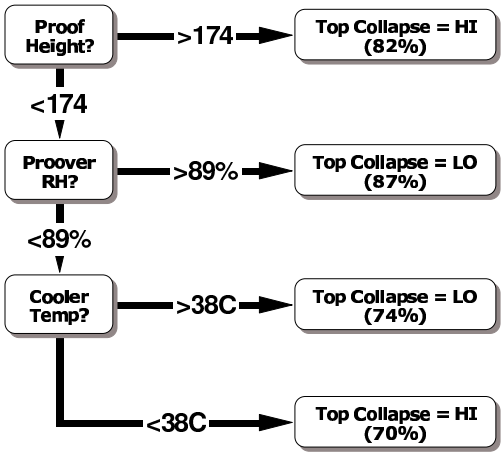

Figure 2. Top collapse predicted from processing parameters using a tree classifier

Two issues arise from analysing top collapse data. First predicting is not always accurate. In this case about 80% accurate. This may be sufficient for process control. It may also be an important factor was not taken into account. Experiments with larger numbers of sample would show this. The second issue is that processing parameters that did not change much during the trial cannot contribute to any model of top collapse. Even though a parameter might be very important, this effect will not be visible in the data. An example of this is oven temperature, which did not fluctuate, but is likely to have an effect on top collapse.

CONCLUSION

Data mining can find relationships between bakery parameters. These relationships can be used for better process control, trouble shooting, as a training tool, and if linked to other quality data (e.g. customer complaints) can give another tool to help satisfy customer wishes. However a large amount of data is required to find robust relationships. In particular you can't predict conditions that have not occurred. To acquire the large amount of data, automation of the data collection process is required.

REFERENCES

Juho Rousu, Laura Flander, Marjaana Suutarinen, Karin Autio, Petri Kontkanen and Ari Rantanen. (2003) Novel computational tools in bakery process data analysis: a comparative study, Journal of Food Engineering, Volume 57, Issue 1, Pages 45-56

Data Mining: Practical machine learning tools with Java implementations," by Ian H. Witten and Eibe Frank, Morgan Kaufmann, San Francisco, 2000.