Predicting Apple Bruising using Machine Learning

G.Holmes[1], S.J.Cunningham[1], B.T. Dela Rue[2] and A.F. Bollen[2]

[1] Department of Computer Science, University of Waikato, Hamilton, New Zealand.
[2] Lincoln Technology, Hamilton, New Zealand.

Many models have been used to describe the influence of internal or external factors on apple bruising. Few of these have addressed the application of derived relationships to the evaluation of commercial operations. From an industry perspective, a model must enable fruit to be rejected on the basis of a commercially significant bruise and must also accurately quantify the effects of various combinations of input features (such as cultivar, maturity, size, and so on) on bruise prediction. Input features must in turn have characteristics which are measurable commercially; for example, the measure of force should be impact energy rather than energy absorbed. Further, as the commercial criteria for acceptable damage levels change, the model should be versatile enough to update bruise thresholds from existing data.

Machine learning is a burgeoning technology with a vast range of potential applications particularly in agriculture where large amounts of data can be readily collected [1]. The main advantage of using a machine learning method in an application is that the models built for prediction can be viewed and understood by the owner of the data who is in a position to determine the usefulness of the model, an essential component in a commercial environment.

Machine Learning software recently developed at Waikato University [2] offers potential as a prediction tool for the classification of bruising based on historical data. It gives the user the opportunity to select any number of measured input attributes and determine the influence of that combination on a range of bruise size categories. The user may require a high degree of accuracy in the classification and therefore prune the attributes or bruise classes accordingly, or alternatively seek to discover trends in the dataset (in which case a lower level of accuracy often clarifies implicit structures in the data).

Models such as the theory of elasticity suggest that impact energy and radius of curvature will have a significant effect on the bruise surface area. Cell structure is also thought to contribute to variation in bruise size [3]. The experiment described in this paper uses the machine learning programs C4.5 [4] and M5' [5] to determine the influence of impact energy, radius of curvature and impact site location on bruise area.

1. Introduction

Considerable research effort has gone into the study of damage to fresh produce. The primary objective has been to gain an understanding of failure mechanisms and the relationships between applied forces and the severity of resulting damage in order to i) identify and rectify the source of damaging forces within handling systems or ii) reduce the severity of damage through manipulation of physiological and environmental factors which influence susceptibility.

What the commercial operator requires is an evaluation of the handling system determining the likely percentage of fruit which will be rejected due to impact damage. With this quantification a cost-benefit analysis can be used to justify modifications or at least place a monetary value on damage. Bollen [6] describes a probabilistic approach that enables the prediction of bruise occurrence of a specified size based on impact data.

In order to extend current damage relationships developed for only one site (cheek) on a medium sized apple to other locations on fruit of varying size, it is necessary to consider fruit properties that potentially modulate the bruise susceptibility based on physiological status, apple size and impact location. Two major factors are the local shape of the fruit (curvature) and the physical properties of the fruit tissue at a particular location.

An experiment was conducted to investigate these properties and the results were analysed using two machine learning computer programs called C4.5 and M5'. Both programs attempt to build a classification model of the factors that influence apple bruising. The models are built as a series of rules that describe and generalise from the examples of apple bruising that were recorded in the study. The rules quantify the logical relationships that exist between the factors that influence the bruising. Further, they provide a scientist with some insight into the often complex relationships that exist in their data. The two programs were used because C4.5 performs categorical prediction—prediction of a fixed set of values, and M5' performs continuous prediction—prediction of floating point values.

The greatest challenge in the development of any prediction model is the amalgamation of the many factors influencing bruising to a significant degree, into a form suitably accurate and useable for its intended purpose. Machine learning potentially offers the ability to rapidly scope or detail multi-variable relationships with minimal specialist knowledge.

2. Materials and Methods

2.1 Experimental Details

96 export quality Granny Smith apples of both 100 count size and 150 count size (average weights 195g and 133g respectively) were selected from a commerical packhouse. The fruit were visually inspected and any existing bruises were marked. The 100 count apples were divided into 6 samples of 16 fruit (96 fruit in total) corresponding to the 6 energy levels selected for the experiment. The same procedure was followed for the 150 count apples. The energy levels were predetermined to provide a probability between 0 and 1 for the bruise area exceeding 1 cm^2 and represent drop heights typical of modern packing facilities in New Zealand as determined by handling studies using an Instrumented Sphere [7]. The impact energies are given in Table 1 with equivalent drop heights for 195g apples also shown.

Energy Level	Energy (Joules)	Equivalent drop height (mm)
1	0.019	10
2	0.037	20
3	0.056	30
4	0.075	40
5	0.093	50
6	0.115	60

Table 1. Energy levels used in bruising experiment.

The fruit were impacted against a rigid flat surface using a pendulum impacting test rig. The selected impact site locations on the fruit are shown in Figure 1. Each fruit was impacted 4 times at these locations by rotating the fruit about its polar (stem) axis. For the 100 count fruit this meant that each fruit had 4 impacts around its cheek, 4 impacts around its stem shoulder position and 4 impacts around its calyx shoulder position. The 150 count fruit had 4 impacts around the cheek only. At each energy level there were a total of 64 impact sites for each of the 4 locations (64 impact sites x 6 energies x 4 locations = 1536 sites).

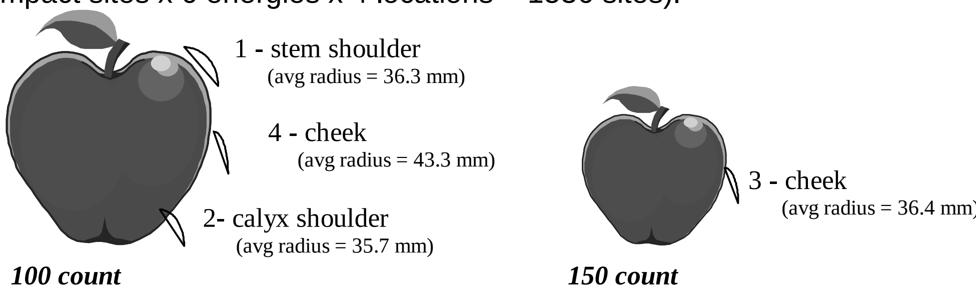

Figure 1 - Location of impact sites on the 100 and 150 count apples.

The surface curvature of the fruit was measured at each location using a 'Tri-probe', a device developed for this experiment. Bruise dimensions were measured within 48 hours of the completion of impacting. Before sectioning, external bruise visibility was assessed. The bruise width in both the polar axis and equatorial axis was then measured.

2.2 Machine Learning

Machine learning techniques were applied to the above data to generate rules to describe the factors that influence apple bruising. The task is one of *classification*: given a set of measured values for the factors that influence apple bruising, predict whether or not an apple will bruise.

2.2.1 Categorical prediction using C4.5

A classification model can be constructed using C4.5 to discover and analyse patterns present in sample data. This data must have the following characteristics:

- It must have an attribute-value format. Here, the factors **Energy Level, Location** and **Curvature** are called the properties or attributes of each case. Each attribute may have a set of discrete values (for example, **Location** values are 1, 2, 3 or 4) or may be continuous—take on continuous values, as for the **Curvature** attribute.

- Each line of data, called a case or instance, must contain a class attribute with discrete values—this is the sense in which the prediction is categorical. For this application that meant imposing a *boundary* on the definition of a bruise. We applied the commercially accepted rule that an apple is bruised if the area of the bruise is greater than or equal to 1 cm^2 (bruised apples are labelled B1), otherwise we declared the sample to be not bruised (labelled B0).

- There must be sufficient data to build and test a classification model. The size of data required will vary between applications and is dependent on the number of attributes and classes and on the complexity of the classification model. A practical rule of thumb is to insist on at least 50 instances.

The following table is part of an example data file (Table 2).

Energy Level	Location	Curvature	Bruise Class
6	2	43	B1
6	2	46	B1
6	2	50	B1
1	3	27	B0
1	3	29	B0

Table 2. Example Machine Learning data file for C4.5

If the sample data is run through C4.5 it will produce a classification model of the data in the form of a set of *production rules*. These take the form IF ‘L’ THEN ‘R’, where L (the left-hand side of the rule) is a conjunction of attribute-based tests and R (the right-hand side of the rule) is a class. These rules, together with a default class (the most frequently encountered class label) are sufficient to predict the class of unseen cases—data with attribute values but no class label. Classification of new data is achieved by finding the first rule (the rules are ordered according to importance) whose left-hand side is satisfied by the case. The predicted class label is then the one nominated by the right-hand side of this rule. If a case *falls through* all of the rules then we designate the class to be the default class.

In practice, the whole data file is used initially to train and test the classification model. If the results look promising then the data file is split into a training set of cases and a test set of cases. In this way the true accuracy of the rules over new data can be estimated. In order to ensure that the rules are truly general and not too specific to one collection of cases, the program is run in *cross-validation* mode. In this mode the data is divided randomly into ten equal parts. Each part is held out in turn and the learning scheme trained on the remaining nine-tenths, then its error rate is calculated on the holdout set. This constitutes one fold of the cross-validation. After a fold new training sets are randomly selected and the process repeated, typically ten times, after which the results are collated and averaged over the number of individual runs (100).

2.2.2 Continuous prediction using M5’

Programs like C4.5 are designed for predicting categories rather than continuous quantities. The same kind of rules can be produced in the continuous case where the right-hand side of a rule contains a continuous value which is the average of all the training set values that the rule applies to. The resulting set of rules represent regression equations, linear models whose weights specify the contribution of an attribute to a classification.

It is possible to combine regression equations with categorical methods similar to those of C4.5 to produce model trees. These model trees have been implemented in a program called M5’. Data for M5’ must conform to the same restrictions as C4.5 except that it requires a real-valued class attribute. Table 3 shows example data for this type of classifier with the actual bruise sizes recorded in the **Bruise Class** column.

Energy Level	Location	Curvature	Bruise Class
6	2	43	2.33
6	2	46	1.65
6	2	50	1.23
1	3	27	0.85
1	3	29	0.22

Table 3. Example Machine Learning data file for M5'

M5' builds similar rules to C4.5 but instead of a categorical classification in the right-hand side of the rule there is either a constant (real) value or linear regression equation. Thus, we can still observe the relationships in the data that are used in making predictions. Once again common practice is to initially train and test on the whole file and confirm results using cross-validation.

One difference between categorical and continuous prediction is in the measure used to establish the error rate. In categorical prediction we can determine the number of new cases correctly classified and then use one minus this figure as the error rate. For continuous prediction several error measures are available: mean squared error or root mean squared error, mean absolute error, relative squared error or root relative squared error, relative absolute error and correlation coefficient. The most commonly used is the mean squared error.

3 Results

First we present rules learned initially from all the data for both C4.5 and M5'. We then present results of cross-validation studies. The research question being addressed in the next section is to find the nature of the relationship between drop height, location and bruise. If a consistent relationship can be uncovered and then suggestions for commercial operations can be made. This study focuses on Granny Smith apples, but if data for other varieties were available then sets of for these varieties could be generated and the commercial operation could be adjusted if necessary to minimise bruising.

3.1 Sample Run using C4.5

The following rules were generated by using all 1536 samples and the **Energy Level** and **Location** attributes to predict the bruise class:

Rule 1: IF **Energy Level** = 1 THEN **Bruise Class** = B0 (96.7%)

Rule 2: IF **Location** = 2 OR **Location** = 4 AND **Energy Level** = 2
THEN **Bruise Class** = B0 (86.6%)

Rule 3: IF **Energy Level** > 2 THEN **Bruise Class** = B1 (91.5%)

Rule 4: IF **Location** = 1 OR **Location** = 3 AND **Energy Level** > 1
THEN **Bruise Class** = B1 (88.6%)

Default Rule: **Bruise Class** = B1

These rules state that 10mm drop heights do not cause major bruising on this apple variety. When that height is increased to 20mm bruising depends on where the impact occurs. Bruising can be avoided in 100 count apples if the apple drops onto either the calyx shoulder or cheek. The 100 apple is predicted to bruise if dropped from 20mm onto the stem shoulder. There are no rules for 150 count apples above a 10mm drop height—all such apples are predicted to bruise. Because rule evaluation is order-dependent, it is possible to read off the relative importance of the attributes. For example, the Energy Level attribute is always evaluated first when determining classification and therefore C4.5 believes that this attribute provides greater discrimination. Note that the default rule will never be reached as rules 1-4 cover all the cases.

An unseen case is classified by matching the conditions on the left hand side of each rule in turn to the values of the attributes of the unclassified case. The first rule encountered that matches these conditions classifies the case. The percentage in brackets after each rule is an estimate of how good the classification will be on unclassified cases. The rules are tested by evaluating unclassified cases to classify the bruise class as shown in Table 4.

Energy Level	Location	Bruise Class
5	1	? (B1 by rule 3)
2	3	? (B1 by rule 4)
1	2	? (B0 by rule 1)
2	4	? (B0 by rule 2)
6	1	? (B1 by rule 3)

Table 4. Table of sample unclassified cases

When the set of rules is evaluated on all the samples the results are as follows:

Rule No.	Error	Times Used	Misclassification	Advantage	Bruise Class
1	3.3%	256	6 (2.3%)	116 (122|6)	B0
2	13.4%	128	30 (23.4%)	68 (98|30)	B0
3	8.5%	1024	80 (7.8%)	0 (0|0)	B1
4	11.4%	128	50 (39.1%)	0 (0|0)	B1

Table 5. Evaluation of individual rules using all samples

Looking at the first line of Table 5 we discover that rule 1 has a predicted error rate of 3.3%, and was used 256 times in classifying the training cases (6 of the 256 cases were misclassified by this rule). The Advantage section estimates the impact of each rule, and is in the form *a (b|c)*. If a given rule were omitted then *b* cases now classified correctly would be classified incorrectly, and *c* cases now classified incorrectly by this rule would be classified correctly by the subsequent rules and the default ($a = b - c$).

Finally, there is a summary and a confusion matrix showing where the misclassifications occur (Table 6). Thus, 348 unbruised apples (B0) were classified correctly, 130 unbruised apples were misclassified as bruised (B1), 36 bruised apples were misclassified as unbruised and 1022 bruised apples were classified correctly. In this application we are looking for rules that have low error rates that minimise the number of bruised apples classified as unbruised.

	Actual B0 cases	Actual B1 cases
Classified as B0	348	36
Classified as B1	130	1022

Table 6. Confusion Matrix

3.2 Results using M5'

Rule sets were generated for the same data with M5' using the **Energy Level** and **Location** attributes to predict the amount of bruising an apple is likely to receive:

Rule 1: IF **Energy Level** = 1 OR **Energy Level** = 2 OR **Energy Level** = 3 THEN
  IF **Energy Level** = 1 THEN LM1
  IF **Energy Level** = 2 OR **Energy Level** = 3 THEN LM2
Rule 2: IF **Energy Level** = 4 OR **Energy Level** = 5 OR **Energy Level** = 6 THEN
  IF **Location** = 2 THEN LM3
  IF **Location** = 1 OR **Location** = 3 OR **Location** = 4 THEN
    IF **Energy Level** = 4 THEN LM4
    IF **Energy Level** = 5 THEN LM5
    IF **Energy Level** = 6 THEN LM5

The right hand sides of these rules are linear models:

LM1: Bruise = 0.251 + 0.193 (Location=1 or 3) + 0.0936 (Location=1)
LM2: Bruise = 0.88 + 0.119 (Location=1 or 3) + 0.277 (Energy Level=3 or 4 or 5 or 6)
LM3: Bruise = 1.62 + 0.2 (Energy Level=6)
LM4: Bruise = 1.75 + 0.101 (Location=1)
LM5: Bruise = 2.18 – 0.181 (Location=1 or 3)
LM6: Bruise = 2.11 + 0.162 (Location=1 or 3) – 0.0853 (Location=1)

Thus, if the drop height is 10mm the apple is predicted to bruise to a maximum of 0.251 if dropped in locations 2 and 4, 150 count apples bruise to a maximum of 0.251 + 0.193, and 100 count apples bruise the most when dropped on their stem shoulder (0.251 + 0.193 + 0.0936).

The interesting case is via LM2 where the constant 0.88 is close to the commercially accepted bruise limit of 1 cm^2. The drop height required to reach LM2 is either 20mm or 30mm. If the drop height is 30mm then the size of the bruise is only just over 1, but only if the location can be restricted to the calyx shoulder or the cheek of the 100 count apple. These results are similar to the categorical case but place a quantifiable measure on the predicted bruise size—an extremely valuable result.

3.3 Results of cross validation

In order to discover the relative importance of each attribute we performed 100 runs for each of six cross-validation studies to classify all bruising and externally visible bruising. In the visible bruising datasets the classification is determined by whether or not a bruise was visible (before sectioning) rather than by the size of the bruise after sectioning. Three sets of attributes were used; the **Energy Level** (E), the **Energy Level** and **Location** (EL), and the **Energy Level**, **Location** and **Curvature** (ELR). Two trials were run, one used the **Visible** attribute (V-E, V-EL, V-ELR) and the other did not (E, EL, ELR). A summary of the results is presented in Table 7.

Trial	Avg Rule Size	Avg Error Rate	Avg no. Misclassifications
E	2.0	12.6%	114
EL	4.7	10.8%	36
ELR	11.3	9.5%	66
V-E	1.2	21.3%	142
V-EL	3.0	20.5%	104
V-ELR	13.9	19.2%	145

Table 7. Results of cross-validation studies using C4.5

These results indicate that for both cases (all bruising and externally visible bruising) while all three attributes (ELR) give the fewest errors, the EL combination is preferred as it significantly minimises the number of bruised apples that are misclassified as unbruised. The average error rate is similar to that of the lowest error rate (ELR) and a small number of rules were used to achieve this. The EL combination is also the most commercially useful combination as these attributes can be measured non destructively and under a degree of commercial control. For the continuous prediction task the cross validation gave a root mean squared error of 0.32 (correlation coefficient of 0.89), which is at a similar performance level to the categorical case.

4. Conclusion

Where bruises were measured after sectioning the classification accuracy was predictably high with the best combination of attributes being the Energy Level and Location. When only externally visible bruises were considered as in commercial practice, the best combination was still Energy Level and Location but the error rate and misclassification level increased considerably. The accuracy of the rules given by the machine learning programs is good given the inherent variability of apple bruising. As a data analysis tool it offers an extremely simple and rapid method for evaluating the significance of measurements from an experiment. Finally, the technique creates the opportunity for a scientist to gain valuable insights into the often complex relationships that exist within their data.

5. References

[1] Garner, S.R., Holmes G., McQueen R.J. & Witten I.H. (1995) Machine learning from agricultural databases: practice and experience *New Zealand J Computing 6* (1a), 69-73, August.

[2] Holmes, G., Donkin A., & Witten I.H. (1994) WEKA: A machine learning workbench, *Proc Second Australia and New Zealand Conference on Intelligent Information Systems.*

[3] Roudot, A.C., Duprat, F., and Wenian, C. (1991) Modelling the Response of Apples to Loads, *J. Agricultural.Engineering Research* 48, 249-259.

[4] Quinlan, J. R., 1993. C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, California.

[5] Wang, Y and Witten, I.H. (1997) Inducing model trees for continuous classes. *Proc of Poster Papers, 9th European Conference on Machine Learning,* Prague, Czech, April.

[6] Bollen, A. F.; Rue, B. T. Dela, (1990). Handling Impacts for Kiwifruit, Asian Pears and Apples. ASAE, 1990 International Summer Meeting. Columbus, Ohio. Paper No. 90-6005, Pages: 1-12

[7] Bollen, A. F. and B. T. Dela Rue, (1995). Application of Sensor Technology to the Assessment of Postharvest Fruit Handling Systems. Transactions of the Institute of Professional Engineers, NZ. Vol 22, No. 1, pg 42-47.