Finding Rules for Extracting Chinese Medicines with FormRules

Background

Momordica cochinchinensis is a perennial squash which grows in tropical regions. Its seeds are used in Chinese medicine as a topical treatment for ulcers, carbuncles and fistulae, and in Vietnam to treat inflammation. The roots are used in traditional Chinese medicine as an expectorant and an anti-inflammatory drug.

Saponins known as momordins have been isolated from both the roots and the seeds of this plant.

The extraction of biologically active constituents from medicinal plants depends on factors like the plant materials, the solvents, the equipment used, and experimental conditions. In order to understand the process, though, it is necessary to carry out some systematic experimentation, and then to look for cause-and-effect relationships within the data, and express these as intelligible rules.

Extraction Process Data

The data used in this study have been provided by Professor Dang Van Giap of the Faculty of Pharmacy in Medicine and Pharmacy University of Ho Chi Minh City, Vietnam.

Dang and his colleagues carried out 16 different experiments, using seeds collected from ripe *M. cochinchinensis* fruits. They varied the experimental conditions for the extraction process, shown in the box below.

Input variables

- Heating method – whether using a heating device, or over live charcoals
- Ethanol concentration (as percentage)
- Extraction temperature (either 30^{o} or 80^{o})
- Seed/ethanol ratio

Properties that were measured were:

- Total dried extracts
- Fat-soluble residue
- Saponins

These data were used in **FormRules**, a data mining program based on neurofuzzy logic. **FormRules** develops 'sparse' models that show which input variables affect each of the properties, and expresses these in terms of rules.

Models and Rules for the Extraction

The model for the amount of dried extracts showed that all variables except the heating method played a role in determining this property, as illustrated in Figure 1.

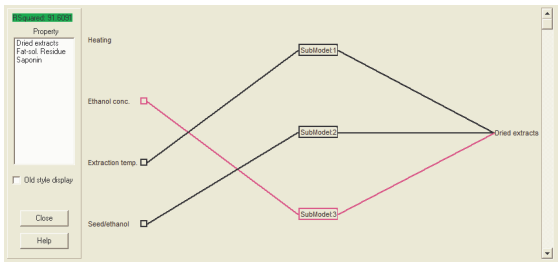

Figure 1. Model (showing sub-models) for total dried extracts.

The most important variable (shown by the purple line) is the ethanol concentration. The rules obtained from each set of sub-models are as follows:

IF Ethanol conc. is LOW THEN Dried extracts is HIGH (1.00)
IF Ethanol conc. is HIGH THEN Dried extracts is LOW (1.00)

IF Seed/ethanol is LOW THEN Dried extracts is HIGH (1.00)
IF Seed/ethanol is MID THEN Dried extracts is HIGH (1.00)
IF Seed/ethanol is HIGH THEN Dried extracts is LOW (1.00)

and
IF Extraction temp. is LOW THEN Dried extracts is HIGH (0.55)
IF Extraction temp. is HIGH THEN Dried extracts is HIGH (1.00)

For these, the values in parentheses are 'confidence levels' indicating whether the value is high, low, or in the middle of the range

Intelligensys Ltd, Springboard Business Centre, Stokesley Business Park, Stokesley, N. Yorks. TS9 5JZ, UK
Tel: (+44) 1642 715380 Fax: (+44) 1642 715301 e-mail postmaster@intelligensys.co.uk

(as indicated by a confidence level around 0.5). The ANOVA statistics for this model show an R^2 of 91.6%, showing a very good fit to the data.

The fat-soluble residue is determined largely by the ethanol concentration, as shown in Figure 2 below.

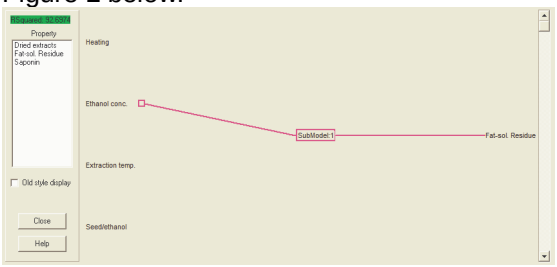

Figure 2. Model for fat-soluble residue

Here, the ANOVA statistics give R^2 of 92.7%, an excellent fit to the data. Therefore, it is clear that the heating method, the extraction temperature, and the seed/ethanol ratio play virtually no role in determining this property.

The rules underlying this model are:

IF Ethanol conc. is LOW THEN Fat-sol. Residue is LOW (0.97)
IF Ethanol conc. is MID THEN Fat-sol. Residue is LOW (1.00)
IF Ethanol conc. is HIGH THEN Fat-sol. Residue is HIGH (0.81)

The rule shown in red makes the largest negative contribution to the model, while the one in blue makes the largest positive contribution.

Looking at the graphical plot for this model elucidates further, as shown in Figure 3.

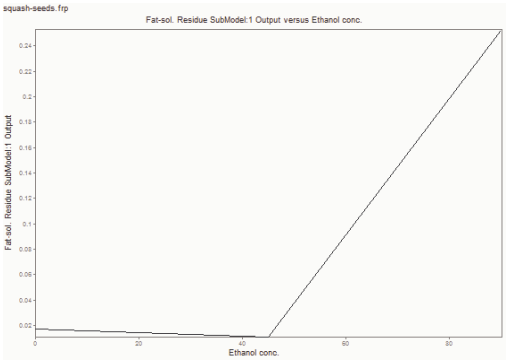

Figure 3. Output plot for Model for Fat-soluble Residue

When the ethanol concentration is low to medium, then it has virtually no effect on the fat-soluble residue, which is low under those circumstances. As the ethanol concentration increases above the mid-point, the fat-soluble residue increases.

The saponins, like the total dried residue, depend on all variables except the heating method. This is shown in Figure 4.

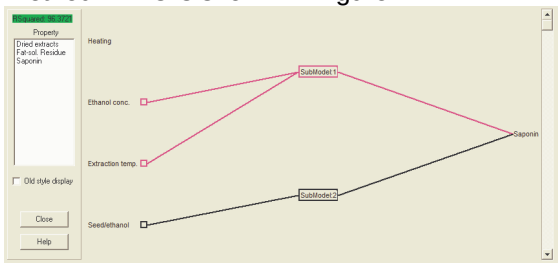

Figure 4. Model for saponins, showing sub-models

R^2 here is 96.3%, indicating that an excellent model has been obtained. As the purple line shows, ethanol concentration and extraction temperature are the most important variables, and the corresponding rules are:

IF Extraction temp. is LOW AND Ethanol conc. is LOW THEN Saponin is HIGH (1.00)
IF Extraction temp. is LOW AND Ethanol conc. is HIGH THEN Saponin is LOW (1.00)
IF Extraction temp. is HIGH AND Ethanol conc. is LOW THEN Saponin is HIGH (1.00)
IF Extraction temp. is HIGH AND Ethanol conc. is HIGH THEN Saponin is LOW (1.00)

The second sub-model, which plays a less important role, involves the seed/ethanol ratio, and its rules are

IF Seed/ethanol is LOW THEN Saponin is HIGH (1.00)
IF Seed/ethanol is MID THEN Saponin is HIGH (1.00)
IF Seed/ethanol is HIGH THEN Saponin is LOW (1.00)

Conclusions

From a limited number of experiments, very good models could be developed. These showed that the heating method (whether using a heating device or live charcoals) had little if any effect on the observed properties.

The other 3 variables – ethanol concentration, extraction temperature, and seed/ethanol ratio – were important for the total dried residue and for the amount of saponins.

The fat-soluble residue, however, depended only on the ethanol concentration used during the extraction.

In this way, key variables for the process can be identified.

For further information on **FormRules** *and on applying neurofuzzy logic to your problems, contact us at the address below.*

Intelligensys Ltd, Springboard Business Centre, Stokesley Business Park, Stokesley, N. Yorks. TS9 5JZ, UK
Tel: (+44) 1642 715380 Fax: (+44) 1642 715301 e-mail postmaster@intelligensys.co.uk