DISCOVERING INTERACTION PATTERNS

Martin McGrane
School of Information Technologies

FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES

INTRODUCTION

Herbal Medicine in Traditional Chinese Medicine relies on interactions between the ingredients of a prescription. The combination is chosen to promote desirable interactions. Analysing these interactions is an important step in quantitatively analysing TCM outcomes. However, existing methods to find interactions focus on association and do not have an adequate quantitative basis for evaluation. Here we introduce a method to analyse data such as prescription records for synergistic interactions between herbs.

CONTRIBUTION

In this work we demonstrate a method to find the complex interactions in a dataset and identify those that can be shown to be significant. With this method it is possible to find both the strength of an interaction and the reliability of the measurement.

INTERACTIONS

An interaction exists when the measured outcome of combining components is different than expected from the parts. A well known example is that smoking and asbestos exposure together **doubles** the risk of lung cancer (Berry and Liddell, 2004).

		Smoking	
		Non-smoker	Smoker
Asbestos exposure	Not exposed	0.1%	1%
	Exposed	0.3%	**2%**

The expected probability of lung cancer is

1% + 0.3% - 0.1% = **1.2%**

The actual probability found is **2%.**

SYNERGY

Synergy, or mutual reinforcement, is an interaction that produces a greater effect than the sum of the parts. This can be expressed as the difference between the expected value calculated from the parts and the actual measured result.

$F(1,1) = \mathbb{E}\{F(1,1)\} = F(0,1) + F(1,0) - F(0,0)$

If the actual value is greater than expected with no interaction we call this synergy.

$Fn - \mathbb{E}\{Fn\} > 0$

In terms of increase in probability, this becomes

$$P_{11} = \frac{(1-R_{11})(1-R_{00})}{(1-R_{01})(1-R_{10})}$$

METHODS

Pre-processing

Data is first converted to a dichotomised form. Both attributes and outcomes (e.g. herbs and health improvement) are dichotomised.

Finding Interactions

We use the dichotomised figures to find the probabilities of a 'good' outcome for each attribute combination. From this we find the expected combined effect. E.g.

$R_{i\{1 \ldots n\}} = P(\text{outcome} \mid \text{col } i_1 = x_1, \text{col } i_2 = x_2)$

$R_{00} = 0.1, R_{01} = 0.2$

$R_{10} = 0.2, R_{11} = 0.5$

$$P_{11} = \frac{(1-0.5)(1-0.1)}{(1-0.2)(1-0.2)}$$

$$P_{11} = 0.30¶$$

Thus there is a synergistic interaction of 30%.

Analysing Significance

Once we have found a potential interaction it is tested for reliability. To do this we test if this result is statistically significant. The variance of the data is determined, then used to calculate the expected deviation from no interaction.

Z = syn{columns}/sd(syn{columns})

N(Z, sd) < 0.05 implies significance

Continuing the example, we now calculate the significance of this combination.

$$V = V_{00}\left(\frac{1-P_{11}}{1-R_{00}}\right)^2 + V_{01}\left(\frac{1-P_{11}}{1-R_{01}}\right)^2 + V_{10}\left(\frac{1-P_{11}}{1-R_{10}}\right)^2 + V_{11}\left(\frac{1-P_{11}}{1-R_{11}}\right)^2$$

Thus variance = 0. 32

N(Z, sd) = 0.26 > 0.05.

The interaction appeared strong but the statistical test indicated it was not reliable.

CASE STUDY: TREATMENT OUTCOMES IN CLINICAL TREATMENT OF DIABETES

Source

A clinical data warehouse was developed by Zhou, et al. 2010. Here we use a subset of this dataset on the treatment of diabetes. The dataset contains the results of 1915 diabetes outpatients. The prescription was selected from 312 possible components.

Results

In total, 21 two-way, 27 three-way and 1 four-way significant interactions were found. The most reliable three and four-way results are shown.

3-way interactions

Herbs	Synergy	p
baical skullcap root (黄芩), amur cork-tree (黄柏), anemarrhena rhizome (知母)	0.482	0.00002
achyranthus root (牛膝), rhubarb (大黄), anemarrhena rhizome (知母)	0.458	0.00021
achyranthes root (牛膝), baical skullcap root (黄芩), dried ginger (干姜)	0.413	0.00068
cassia bark (肉桂), dried ginger (干姜), asiatic cornelian cherry (山萸肉)	0.458	0.00543
baical skullcap root (黄芩), earthworm (地龙), dried ginger (干姜)	0.344	0.00582

4-way interactions

Factors	Synergy	p
baical skullcap root (黄芩), trichosanthis (天花粉), kudzuvine root (葛根), anemarrhena rhizome (知母)	0.295	0.0130

Analysis of the results has already been of interest to researchers studying the biochemical pathways involved in Chinese Medicine, and the application to clinical practice.

REFERENCES

Zhou, X., Chen, S., Liu, B., Zhang, R., Wang, Y., Li, P., Guo, Y., Zhang, H., Gao, Z. and Yan, X. *Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support.* Artif. Intell. Med., 48, 2-3 (2010), 139-152.

Berry, G. and Liddell, F. D. K. *The Interaction of Asbestos and Smoking in Lung Cancer: A Modified Measure of Effect.* Ann Occup Hyg, 48, 5 (2004), 459-462.

CONCLUSION

The system we have developed can analyse data for complex interactions in multiple dimensions. The strength of the interaction is determined and validated statistically. Only interactions that we can show are significant are returned.

The measure we have developed identifies information that is not accessible to standard analysis tools.

http://sydney.edu.au/engineering/it/~itcm/