# Discovering Associations in High-Dimensional Data

**Geoffrey I. Webb**

Faculty of Information Technology
Monash University
Clayton, Vic, 3800, Australia
Email: Geoff.Webb@monash.edu

**Abstract**

Association discovery is one of the most studied tasks in the field of data mining (Agrawal et al. 1993). It involves identifying items that occur together in data, and has numerous applications in manufacturing, commerce, administration and science. However, far more attention has been paid to how to discover associations than to what associations should be discovered. In this talk Geoff will provide a highly subjective tour of the field. He will

- highlight shortcomings of the dominant frequent pattern paradigm;
- illustrate benefits of the alternative top-k paradigm; and
- present the self-sufficient itemsets approach to identifying potentially interesting associations as described in Webb & Zhang (2005), Webb (2007, 2008, 2010, 2011).

**References**

Agrawal, R., Imielinski, T. & Swami, A. (1993), Mining associations between sets of items in massive databases, *in* 'Proceedings of the 1993 ACM-SIGMOD International Conference on Management of Data', Washington, DC, pp. 207–216.

Webb, G. I. (2007), 'Discovering significant patterns', *Machine Learning* **68**(1), 1–33.

Webb, G. I. (2008), 'Layered critical values: A powerful direct-adjustment approach to discovering significant patterns', *Machine Learning* **71**(2-3), 307–323. Technical Note.

Webb, G. I. (2010), 'Self-sufficient itemsets: An approach to screening potentially interesting associations between items', *Transactions on Knowledge Discovery from Data* **4**, 3:1–3:20.

Webb, G. I. (2011), 'Filtered-top-k association discovery', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **1**(3), 183–192.
**URL:** *http://dx.doi.org/10.1002/widm.28*

Webb, G. I. & Zhang, S. (2005), 'K-optimal rule discovery', *Data Mining and Knowledge Discovery* **10**(1), 39–79.