Compass: A hybrid method for clinical and biobank data mining

We describe a new method for identification of confident associations within large clinical data sets. The method is a hybrid of two existing methods; Self-Organizing Maps and Association Mining. We utilize Self-Organizing Maps as the initial step to reduce the search space, and then apply Association Mining in order to find association rules. We demonstrate that this procedure has a number of advantages compared to traditional Association Mining; it allows for handling numerical variables without a priori binning and is able to generate variable groups which act as “hotspots” for statistically significant associations. We showcase the method on infertility-related data from Danish military conscripts. The clinical data we analyzed contained both categorical type questionnaire data and continuous variables generated from biological measurements, including missing values. From this data set, we successfully generated a number of interesting association rules, which relate an observation with a specific consequence and the p-value for that finding. Additionally, we demonstrate that the method can be used on non-clinical data containing chemical–disease associations in order to find associations between different phenotypes, such as prostate cancer and breast cancer.

General information

State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Copenhagen University Hospital
Authors: Krysiak-Baltyn, K. (Intern), Petersen, T. N. (Intern), Audouze, K. M. L. (Intern), Jørgensen, N. (Ekstern), Ängquist, L. (Ekstern), Brunak, S. (Intern)
Pages: 160-170
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication information

Journal: Journal of Biomedical Informatics
Volume: 47
ISSN (Print): 1532-0464
Ratings:
- BFI (2017): BFI-level 2
- Web of Science (2017): Indexed Yes
- BFI (2016): BFI-level 2
- Scopus rating (2016): SJR 0.857 SNIP 1.521 CiteScore 3.23
- BFI (2015): BFI-level 2
- Scopus rating (2015): SJR 1.281 SNIP 2.243 CiteScore 3.66
- BFI (2014): BFI-level 2
- Scopus rating (2014): SJR 1.069 SNIP 2.372 CiteScore 3.48
- Web of Science (2014): Indexed yes
- BFI (2013): BFI-level 2
- Scopus rating (2013): SJR 1.057 SNIP 1.933 CiteScore 3.07
- ISI indexed (2013): ISI indexed yes
- BFI (2012): BFI-level 2
- Scopus rating (2012): SJR 1.032 SNIP 2.1 CiteScore 3.08
- ISI indexed (2012): ISI indexed yes
- BFI (2011): BFI-level 2
- Scopus rating (2011): SJR 0.851 SNIP 1.91 CiteScore 2.64
- ISI indexed (2011): ISI indexed yes
- BFI (2010): BFI-level 2
- Scopus rating (2010): SJR 0.714 SNIP 1.639
- BFI (2009): BFI-level 2
- Scopus rating (2009): SJR 1.004 SNIP 1.887
- BFI (2008): BFI-level 2
- Scopus rating (2008): SJR 0.856 SNIP 1.542
- Scopus rating (2007): SJR 0.958 SNIP 2.227
- Scopus rating (2006): SJR 0.786 SNIP 2.114
- Scopus rating (2005): SJR 1.162 SNIP 1.999
- Scopus rating (2004): SJR 0.484 SNIP 1.299
- Scopus rating (2003): SJR 0.283 SNIP 0.692
- Scopus rating (2002): SJR 0.3 SNIP 0.776