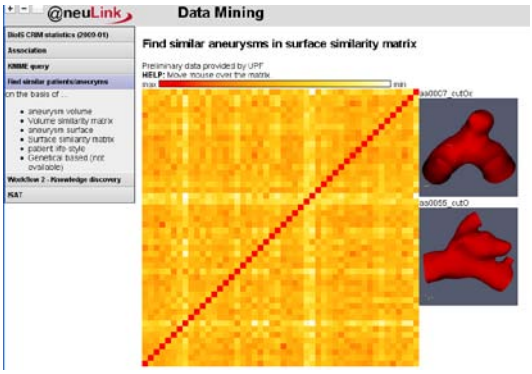
Workflows for Data Mining in Integrated multi-modal Data of Intracranial Aneurysms using KNIME

Roelof Risselada[1], Christoph M. Friedrich[2,*], Christian Ebeling[2], Roman Klinger[2], Anna Bauer-Mehren[3], Manuel Pastor[3], Maria Cruz Villa[4], Jose M. Pozo[4], Alejandro F. Frangi[4], and Martin Hofmann-Apitius[2]

1. Erasmus MC; Medical Informatics; PO Box 2040, 3000 CA Rotterdam; the Netherlands
2. Fraunhofer Institute for Algorithms and Scientific Computing (SCAI); Department of Bioinformatics; Schloss Birlinghoven; D-53754 Sankt Augustin; Germany
3. Research Unit on Biomedical Informatics (GRIB), IMIM/UPF; C/Dr. Aiguader, 88; 08003 Barcelona; Spain
4. Research Group of Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), Universitat Pompeu Fabra, Pg Circumval·lació, 8, 08003 Barcelona; Spain

* Contact author: friedrich@scai.fraunhofer.de

Keywords: Data Mining, Workflow, multi-modal data



Intracranial aneurysms are bulbous expansions of the intracranial vessels that may rupture and lead to subarachnoid haemorrhage, a bleeding in the space lining the brain. This can result in severe disability or death of the affected person. The prediction of the individual rupture risk of a patient based on information from images, haemodynamic simulations, clinical parameters and genetic markers is one of the aims of the European Integrated Project @neurIST. The project developed an architecture [1] which allows the integration of multi-modal data from clinical information systems. Data mining capabilities have been developed in a Knowledge Discovery application suite called @neuLink [2]. Maintenance, re-use of mining strategies and user presentation of data flows is difficult in monolithic mining scripts. To cope with this problem, @neuLink integrates the Konstanz Information Miner (KNIME) [3] as a workflow engine and provides aggregated data mining results based on R scripts and Weka [4]. The advantage of this solution is a better understanding of the workflow, re-usability of data mining strategies and an increased maintainability.

One example sub-workflow is the application of clustering algorithms to find similar aneurysms. This is based on Zernike moments [5] extracted from images of aneurysms. The resulting similarity matrix is depicted in the screenshot.

Several @neurIST partners developed and implemented rupture risk and treatment outcome models as sub-workflows, which are currently being integrated in consensual models.

Acknowledgements

This work has been partially funded in the framework of the European integrated project @neurIST, which is co-financed by the European Commission through the contract no. IST-027703 (see http://www.aneurist.org)

References

1. Rajasekaran, H.; , Iacono, L. L.; Hasselmeyer, P.; Fingberg, J.; Summers, P.; Benkner, S.; Engelbrecht, G.; Arbona, A.; Chiarini, A.; Friedrich, C. M.; Hofmann-Apitius, M.; Kumpf, K.; Moore, B.; Bijlenga, P.; Iavindrasana, J.; Mueller, H.; Hose, R. D.; Dunlop, R., and Frangi, A. *@neurIST – Towards a System Architecture for Advanced Disease Management through Integration of Heterogeneous Data, Computing, and Complex Processing Services;* Proceedings of the 21st IEEE International Symposium on Computer-based Medical Systems, *IEEE*, **2008**, 361-366.
2. Friedrich, C. M.; Dach, H.; Gattermayer, T.; Engelbrecht, G.; Benkner, S., and Hofmann-Apitius, M. *@neuLink: A Service-oriented Application for Biomedical Knowledge Discovery* Proceedings of the HealthGrid 2008, IOS Press, **2008**, 165-172
3. Berthold, M. (2009). *KNIME (Konstanz Information Miner)* webpage; last accessed 2009-02-26 http://www.knime.org/
4. Witten, I. H., and Frank, E. *Data Mining; second edition;* Morgan Kaufmann Publishers, **2005**
5. Millán, R. D.; Dempere-Marco, L.; Pozo, J. M.; Cebral, J. R., and Frangi, A. F. *Morphological characterization of intracranial aneurysms using 3-D moment invariants.* IEEE Trans Med Imaging, **2007**, *26*, 1270-1282