



Thanh Le Van

Diligent and hunger for new knowledge and applications

Interests

Data mining, machine learning, probabilistic modelling
Bioinformatics applications such as cancer subtyping and cancer evolution

Education

- 12/2016 **PhD in Computer Science**, *KU Leuven*, Belgium.
(Expected) Thesis: Rank matrix factorisation and its applications
Advisors: Prof. Luc De Raedt (supervisor), Prof. Siegfried Nijssen (co-supervisor),
Prof. Kathleen Marchal (co-supervisor)
- 2007 **MSc of Information and Communication Technologies**, *Asian Institute of Technology (AIT)*, Thailand, supervised by Prof. Peter Haddawy.
Thesis: Data mining for financial aid optimisation (*excellent grade*)
- 2001 **Bachelor of Computer Science**, *Ho Chi Minh University of Technology*, Vietnam.

Research experience

- Bioinformatics **Cancer subtyping.**
TCGA datasets, i.e., breast cancer.
RNASeq, microarray, mutation, CNV data and biological networks, e.g., STRING network, KEGG networks.
Discover breast cancer subtypes and subtype specific driver pathways, which are recurrently mutated and correlated with aberrant gene expression.
- Data mining **Rank matrix factorisation.**
Introduced a generic *Rank Matrix Factoriation* (RMF) framework based on semiring theory for pattern set mining in *rank data*, i.e., data consists of rankings of items.
Applied the RMF framework in mining different types of patterns in rank matrices, e.g., *Sparse RMF* and *ranked tiling*.
Applied the RMF framework in discovering TCGA breast cancer subtypes.

Department of Computer Science, Celestijnenlaan 200a – Heverlee, Belgium

☎ +32489726430 • ✉ thanh.levan@cs.kuleuven.be; thanhlv@gmail.com

📄 [thanhlevan.github.io](https://github.com/thanhlevan)

- Data mining **Bi-clustering under constraints.**
Developed an algorithm to mine fault-tolerant bi-clusters resembling a staircase.
Developed a Minimum Description Length (MDL) criterion for model selection.
- Optimisation **Constraint Programming (CP) and Integer Linear Programming (ILP).**
Modelled data mining problems, e.g., Sparse RMF, ranked tiling and fault-tolerant bi-clusters, using CP and ILP.
Solved optimisation problems using Gurobi solver, CP solvers such as OscalaR.
- Prob.modelling **Probabilistic graphical models and Bayesian modelling.**
Learned parameters/structures of Bayesian networks.
Predicted prob. of enrollment of applicants using Bayesian networks (MSc thesis).

Computer skills

- O.S Linux, Windows, Unix
Languages Scala, Java, C#, C/C++
Scriptings R, Matlab
Databases Oracle, MySQL
Frameworks JSP, Spring, Hibernate, ASP.NET, PHP

Bioinformatics skills

- Cytoscape for network analysis
R packages for genome science: bioconductor, ggplot, gplots, igraph, surv, snf, icluster+
GISTIC tool

Software engineering experience

- 2012–2016 **Web-based submission system**, *Journal tracks of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases.*
Java Spring, Hibernate, JSP, MySQL
- 2001–2010 **Web-based application development**, *Vietnam Posts and Telecommunications (VNPT) in HoChiMinh City, Vietnam.*
Human resource management system (2002-2010), Customer relation management system (2008-2010)
ASP.NET, C#, Oracle

Teaching experience

- 2015-2016 **Information Structures and Implications**, *Master course*, KU Leuven.
Teaching assistant; responsible for student exercise sessions
- 2014-2016 **Databases**, *Bachelor course*, KU Leuven.
Teaching assistant; responsible for student projects

Awards

- 2016 ECCB 2016 travel grant.
- 2012-2016 KU Leuven scholarship for doctoral study.
- 2008 Asian Institute of Technology (AIT) fellowship for doctoral study (denied).
- 2005 VNPT scholarship for graduate study at AIT, Thailand.
- 2001 HoChiMinh City University of Technology award to the top-20 students (out of ?) in Computer Science.
- 1996-2001 HoChiMinh City University of Technology scholarship for excellent student (a number of years).

Invited visit

- July 2015 Four-day visit to the Theoretical and Applied Computer Science Laboratory led by Prof. Le Thi Hoai An, University of Lorraine, France.
Talk: *Tiling rank matrices and its applications*

References

- 1 **Prof. Luc De Raedt**, (*PhD supervisor*).
Department of Computer Science, KU Leuven
Celestijnenlaan 200A (Room 02.26)
3001 Heverle, Belgium
Phone: +32 16 32 7834
Fax: +32 16 32 79 96
Email: luc.deraedt@cs.kuleuven.be
- 2 **Prof. Siegfried Nijssen**, (*PhD co-supervisor*).
Leiden Institute of Advanced Computer Science, Leiden University
Niels Bohrweg 1, 2333 CA, Leiden, The Netherlands
Phone: +31 715278919
Email: s.nijssen@liacs.leidenuniv.nl
- 3 **Prof. Kathleen Marchal**, (*PhD co-supervisor*).
Dept Information Technology (INTEC, Minds), Ghent University
Zuiderpoort Office Park, Blok C0, Gaston Crommenlaan 8, bus 201, Gent, Belgium
Phone: +32 486909943
Email: kathleen.marchal@intec.ugent.be

Publications

Journal articles:

- 1 **Le Van, T.**, Nijssen, S., van Leeuwen, M., De Raedt, L., Semiring rank matrix factorisation, *IEEE Transactions on Knowledge and Data Engineering*, under revision.
- 2 **Le Van, T.**, van Leeuwen, M., Fierro, A., De Maeyer, D., Van den Eynden, J., Verbeke, L., De Raedt, L., Marchal, K., Nijssen, S. (2016). Simultaneous discovery of cancer subtypes and subtype features by molecular data integration. *Bioinformatics*, 32 (17), i445-i454, Oxford University Press.

Peer-reviewed conference and workshop papers:

- 3 **Le Van, T.**, van Leeuwen, M., Nijssen, S., De Raedt, L. (2015). Rank matrix factorisation. *Proceedings of the 19th Pacific-Asia conference on knowledge discovery and data mining*, pp. 734-746, Springer.
405 submissions, 90 accepted; acceptance rate: 22.2%
- 4 **Le Van, T.**, van Leeuwen, M., Nijssen, S., Fierro, A., Marchal, K., De Raedt, L. (2014). Ranked tiling. *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD) 2014*. pp. 98-113, Springer.
588 submissions, 130 accepted; acceptance rate: 22.1%
- 5 **Le Van, T.**, Fierro, A., Guns, T., van Leeuwen, M., Nijssen, S., De Raedt, L., Marchal, K. (2012). Mining local staircase patterns in noisy data. *International workshop on Co-Clustering and Applications (CoClus'12) in conjunction with IEEE International Conference on Data Mining (ICDM) 2012*, pp. 139-146, IEEE Computer Society.
10 submissions, 6 accepted; acceptance rate: 60%
- 6 **Le Van, T.**, Haddawy, P., Deriving financial aid optimization models from admissions data. *Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007*. FIE '07. 37th Annual, vol., no., pp.F2A-7-F2A-12, 10-13 Oct. 2007, IEEE Computer Society.

Meeting abstracts:

- 7 **Le Van, T.**, Van den Eynden, J., De Maeyer, D., Verbeke, L., Fierro Guti rrez, A., van Leeuwen, M., Nijssen, S., De Raedt, L., Marchal, K. (2015). Ranked tiling based approach to discovering patient subtypes. *10th Benelux Bioinformatics Conference 2015*. Antwerp, 7-8 December 2015. *(accepted for oral presentation)*
- 8 **Le Van, T.**, Fierro, A., Guns, T., van Leeuwen, M., Nijssen, S., De Raedt, L., Marchal, K. (2013). Bi-clustering gene expression data under constraints. *Benelux Bioinformatics Conference*. Brussels - Belgium, 9-10 December 2013. (poster)