UNIVERSITY OF DEBRECEN
Doctoral School of Informatics

Complex exam major subject

Information Technology, Data Science

Syllabus

Supervised data mining models: regression and regularization, kernel method and radial basis function, sparse kernels (SVM and RVM), graphical models and Bayesian networks, high-dimensional problems. With special emphasis on modern stochastic optimization methods, e.g., stochastic gradient descent and Bayesian and nonparametric learning. Unsupervised data mining models: mixtures and EM-algorithm, clustering, Kohonen network, principal components analysis and its variant (kernel-PCA), singular valued decomposition, non-negative matrix faktorization, independent component analysis, multidimensional scaling. Using a data science software, e.g. the Anaconda Python distribution.

Data mining; knowledge discovery in databases (KDD); symbolic data mining methods. Frequent itemsets; frequent association rules. Algorithms for finding frequent itemsets: Apriori, Apriori-Close, Eclat, Charm, Touch. Galois lattices, algorithms for constructing Galois lattices. The Snow algorithm; hypergraphs. Rare itemsets, rare association rules. Levelwise and depth-first algorithms for finding rare itemsets: Apriori-Rare, Walky-G. Case studies; the Coron system.

Bibliography

1. Bishop, C. M., Pattern Recognition and Machine Learning, Springer, 2006.
2. Hastie, T., Tibshirani, R., Friedman, J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, 2009.
3. Koski, T., Noble, J.M., Bayesian Networks. An Introduction. Wiley, 2009.
4. Gorelick, Micha, Ozsvald, Ian, High Performance Python: Practical Performant Programming for Humans (1st ed.). O'Reilly Media, 2014.
5. Tan, P.-N., Steinbach, M., Karpatne, A., Kumar, V.: Introduction to Data Mining, 2nd ed., Pearson, 2018.
6. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2011.
7. Liu, B.: Web Data Mining, Exploring Hyperlinks, Contents, and Usage Data, 2nd ed., Springer 2011.

Compulsory subjects for this major subject

With the approval of the program's leader:

1) Four courses must be selected from the following courses of the program:

- Novel approaches for Internet-based applications (Adamkó Attila)
- Statistical Analysis of the Distributed Systems (Gál Zoltán)
- Advanced data mining methods and applications (Ispány Márton)
- Stochastic data mining (Ispány Márton)
- Symbolic Data Mining (Szathmáry László)
- Statistics with application to Information Technology (Terdik György)
- Statistics and time series with applications (Terdik György)

2) Two courses must be selected from the other programs of the Doctoral School of Informatics.

3) One course must be selected from the programs of the Hungarian Doctoral Schools.

Recommended subjects for this major subject