



SCC5895 – Análise de Agrupamento de Dados

Apresentação do Curso

Prof. Ricardo J. G. B. Campello

PPG-CCMC / ICMC / USP



Tópicos do Curso

- **Conceituação de Análise de Agrupamento**
 - Conceitos e definições básicas, motivação, aplicações, ...
- **Representação de Dados**
 - Tipos de dados, normalizações, medidas de similaridade e dissimilaridade, ...
- **Métodos Hierárquicos**
 - Métodos aglomerativos, métodos divisivos, formulações parametrizadas, relação com teoria dos grafos, ...



Tópicos do Curso

- **Métodos Particionais**

- Métodos de partições rígidas, métodos de partições com sobreposição, ...

- **Avaliação de Agrupamentos**

- Índices de validação externos, internos e relativos, técnicas de validação de estruturas hierárquicas e particionais, qualidade de grupos individuais, tendência de agrupamento, estimação do número de grupos, ...



Tópicos Complementares*

- **Eventuais Tópicos Avançados / Seminários**

- Técnicas de agrupamento paralelo e distribuído de dados
- Estruturas de dados para escalabilidade computacional de algoritmos
- Agrupamento semi-supervisionado
- Combinações de agrupamentos (clustering ensembles)
- Análise de estabilidade de agrupamento
- Agrupamento de dados em sub-espacos (biclustering, redução de atributos, ...)
- Agrupamento de fluxo de dados (data streams)
- Métodos estatísticos para comparação de algoritmos
- Agrupamento de sequências
- ...



Calendário (Sujeito a Alterações)

- 07/08 – Aula
- 14/08 – Aula
- 21/08 – Aula
- 28/08 – Aula*
- 11/09 – Aula
- 18/09 – **PROVA I**
- 25/09 – Aula**
- 02/10 – Aula

- 09/10 – Aula
- 16/10 – Aula
- 23/10 – Aula
- 30/10 – Aula
- 06/11 – **PROVA II**
- 13/11 – Seminários
- 20/11 – Seminários
- 27/11 – Seminários

(*) Congresso SCC

(**) ECML-PKDD



Avaliação

- **Nota Final:** $NF = (7.5 * N_{pV} + 2.5 * N_S) / 10$
 - onde N_{pV} e N_S são as notas de prova e seminário
 - a nota de prova (N_{pV}) é uma média aritmética das notas da PI e PII



Bibliografía Principal

- Jain, A. K., Dubes, R. C., **Algorithms for Clustering Data**, Prentice Hall, 1988.
- Xu, R., Wunsch, D., Clustering, IEEE Press, 2009.
- Gan, G., Ma, C., Wu, J., Data Clustering: Theory, Algorithms, and Applications, SIAM Series on Statistics and Applied Probability, 2007.
- Kogan, J., Introduction to Clustering Large and High-Dimensional Data, Cambridge University Press, 2006.
- Everitt, B. S., Landau, S., Leese, M., Cluster Analysis, Hodder Arnold Publication, 2001.
- Arabie, P., Hubert, L. J., Soete, G., Clustering and Classification, World Scientific Publ., 1996.
- Höppner, F., Klawonn, F., Kruse, R., Runkler, T., Fuzzy Cluster Analysis, 1999.
- Kaufman, L., Rousseeuw, P. J., Finding Groups in Data – An Introduction to Cluster Analysis, Wiley, 1990.
- Hartigan, J. A., Clustering Algorithms, John Wiley & Sons, 1975.
- Anderberg, M. R., Cluster Analysis for Applications, Academic Press, 1973.
- ...
- **Artigos de periódicos especializados...**