

2010/02/23
長谷川 秀彦

データマイニング課題

データマイニングの最終レポートが二転三転して申し訳ありません。すでにクラスタリングの宿題をやってもらっているのと、出題が遅くなったことで、評価はかなり甘くします。「以下の課題から2題を選択して、3月8日(月)正午までに」7D211 研究室に提出してください。レポートをスキャンしてメールで hasegawa@slis.tsukuba.ac.jp まで送ってもかまいません。紙のサイズは A4・片面で統一してください。

1. データマイニングとはどのようなものか? データ解析や、機械学習と対応付けて、1ページ以内で説明しなさい。

2. 自分で用意した適当な件数のデータを用いて、LSI (Latent Semantic Indexing) を実行しなさい。実行結果に対して考察を加えることも忘れずに。

参考: **Understanding Search Engines: Mathematical Modeling and Text Retrieval** by M.W. Berry and M. Browne, [SIAM](#) Book Series: Software, Environments, and Tools, Second Edition (May 2005), ISBN: 0-89871-581-4.

3. 自分で用意した適当な件数のデータを用いて、Google 流の PageRank を求めなさい。実行結果に対して考察を加えることも忘れずに。

参考: **Google's PageRank and Beyond: The Science of Search Engine Rankings** by Amy N. Langville & Carl D. Meyer, Princeton University Press, 2006, ISBN: 978-1-4008-3032-9. (Honorable Mention, 2006 AAP Award for Best Professional/Scholarly Book in Computer & Information Science)

4. Web から SVM (Support Vector Machine) のソフトウェアを探し出し、そのソフトウェアに適当なデータを与えて実行し、簡単な実行ガイドを作成しなさい。

- 課題は友人と協力・分担・相談して実行してもよい(むしろ、協力してやることを推奨)。ただし、「協力者の氏名をレポートに明記すること」、「レポートは自分で作成すること」の2点を守ってください
- Web 上の情報資源を活用してください。そして、レポートにはきちんと参考文献として URL などを明記してください。