
Contents

Preface	v
Contributors	ix

Part I Clustering

1 Cluster-Preserving Dimension Reduction Methods for Document Classification	
<i>Peg Howland, Haesun Park</i>	3
2 Automatic Discovery of Similar Words	
<i>Pierre Senellart, Vincent D. Blondel</i>	25
3 Principal Direction Divisive Partitioning with Kernels and k-Means Steering	
<i>Dimitrios Zeimpekis, Efstratios Gallopoulos</i>	45
4 Hybrid Clustering with Divergences	
<i>Jacob Kogan, Charles Nicholas, Mike Wiacek</i>	65
5 Text Clustering with Local Semantic Kernels	
<i>Loulwah AlSumait, Carlotta Domeniconi</i>	87

Part II Document Retrieval and Representation

6 Vector Space Models for Search and Cluster Mining	
<i>Mei Kobayashi, Masaki Aono</i>	109
7 Applications of Semidefinite Programming in XML Document Classification	
<i>Zhonghang Xia, Guangming Xing, Houduo Qi, Qi Li</i>	129

Part III Email Surveillance and Filtering

8 Discussion Tracking in Enron Email Using PARAFAC

Brett W. Bader, Michael W. Berry, Murray Browne 147

9 Spam Filtering Based on Latent Semantic Indexing

Wilfried N. Gansterer, Andreas G.K. Janecek, Robert Neumayer 165

Part IV Anomaly Detection

**10 A Probabilistic Model for Fast and Confident Categorization
of Textual Documents**

Cyril Goutte 187

11 Anomaly Detection Using Nonnegative Matrix Factorization

*Edward G. Allan, Michael R. Horvath, Christopher V. Kopek, Brian T. Lamb,
Thomas S. Whaples, Michael W. Berry* 203

12 Document Representation and Quality of Text: An Analysis

*Mostafa Keikha, Narjes Sharif Razavian, Farhad Oroumchian,
Hassan Seyed Razi* 219

Appendix: SIAM Text Mining Competition 2007 233

Index 237

Survey of Text Mining II
Clustering, Classification, and Retrieval
Berry, M.W.; CASTELLANOS, M. (Eds.)
2008, XVI, 240 p., Hardcover
ISBN: 978-1-84800-045-2