CHAPTER 7
CONCLUSIONS

This dissertation has presented four major contributions: an expressive way to encode knowledge as constraints, a general method for creating constrained clustering algorithms, three new constrained clustering algorithms, and a demonstration of the robustness of the approach by applying it to three significant real-world problems. This chapter summarizes each contribution and then discusses several extensions to this work that we envision.

7.1 Major Contributions

Expressive constraint formulation. We have developed an expressive constraint formulation for encoding domain knowledge about a problem. This encoding encompasses soft as well as hard constraints. We have discussed problem domains for which each type of constraint is most appropriate. We demonstrated that this encoding of knowledge effectively transmits information to constrained clustering algorithms and that the constraint formulation is rich enough to be of use on a diverse array of problems. Further, our encoding is novel in that it accommodates *negative* constraints as well as positive ones.

General method for creating constrained clustering algorithms. We have also developed a general method for transforming a regular partitioning clustering algorithm into a constrained clustering algorithm. The resulting algorithm can apply knowledge encoded as instance-level pairwise constraints to guide its selection of the best partition of the data. This transformation is not particular to any single algorithm, and its general applicability is one of its strengths.

Three new constrained clustering algorithms. We have applied this transformation to two common clustering algorithms, k-means and COBWEB, and developed constrained versions of each that are general enough to be applied to a variety of problems. Two of the algorithms, COP-KMEANS and COP-COBWEB, are designed to function with hard constraints, which must be satisfied by the final output partition. The third algorithm, SCOP-KMEANS, accommodates information encoded as soft constraints, which express preferences of varying strengths about the placement of items. All three algorithms represent novel contributions to the body of algorithmic tools available for data analysis.

Results on real-world problems. We have demonstrated the robustness of our approach by applying constrained clustering algorithms to three significant real-world problems and reporting on the improvements obtained when using constraints. First, we used COP-KMEANS to automatically improve road maps, refining them to the resolution of individual lanes. Second, we applied COP-COBWEB to the problem of noun phrase coreference resolution. Finally, we used SCOP-KMEANS to analyze spectral observations of the planet Mars; our results both support existing hypotheses about Mars and uncover interesting new aspects of the data set. The diversity of problems we experimented with is a testament to the general applicability of our techniques.

7.2 Extensions to Constrained Clustering

We have demonstrated the strengths of constrained clustering using several algorithms and applying our techniques to several real-world problems. There are several extensions to this work that we envision as the next logical steps. First, we plan to develop a soft constrained version of COBWEB. We expect that the method we used to develop SCOP-KMEANS can also be applied to develop a SCOP-COBWEB algorithm. We will apply this algorithm to the coreference problem, which will also allow us to encode a broader collection of information, as soft constraints. A comparison between the results of SCOP-COBWEB and the results we have already obtained with COP-COBWEB will further our understanding of the relative benefits of each type of constraint.

In addition, our work has been focused exclusively on partitioning algorithms. There are cases where hierarchical algorithms, which construct an organization of the data at several different levels of generality, are most useful. We anticipate the development of intelligent constrained hierarchical methods modeled after our constrained partitioning work, and we believe that the key lies in the use of soft constraints.

Finally, we have only scratched the surface in terms of interesting information to be discovered in the STIS Mars data set. We plan further experiments, using additional features and constraints, to assist us in further exploration of the planet. For example, we plan to experiment with the contiguity constraints that vary in strength across the planet. We will assign regions known to be fairly homogeneous a stronger contiguity constraint than "borderline" regions (such as the transition to the polar caps). In the process of these experiments, we will demonstrate the power of constrained clustering as a scientific tool. If this tool continues to be effective at confirming existing knowledge about Mars as well as discovering novel information, we hope to encourage its broader use in other scientific endeavors.

BIBLIOGRAPHY

Abu-Mostafa, Y. S. (1995). Hints. *Neural Computation*, 7:639–671.

Abu-Mostafa, Y. S. (2001). Financial model calibration using consistency hints. *IEEE Transactions on Neural Networks*, 12(4):791–808.

Adams, J. B. (1974). Visible and near-infrared diffuse reflectance spectra of pyroxenes as applied to remote sensing of solid objects in the solar system. *Journal of Geophysical Research*, 79:4829–4836.

Adams, J. B. and McCord, T. B. (1969). Mars: Interpretation of spectral reflectivity of light and dark regions. *Journal of Geophysical Research*, 74:4851–4856.

Ambroise, C., Dang, M., and Govaert, G. (1997). Clustering of spatial data by the EM algorithm. In A. Soares et al., editor, *geoENV I-Geostatistics for Environmental Applications*, pages 493–504. Kluwer Academic Publishers.

Anderberg, M. (1973). *Cluster Analysis for Applications.* Academic Press.

Aone, C. and Bennett, W. (1995). Evaluating automated and manual acquisition of anaphora resolution strategies. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pages 122–129. Association for Computational Linguistics.

Appelt, D., Hobbs, J., Bear, J., Israel, D., Kameyama, M., Kehler, A., Martin, D., Meyers, K., and Tyson, M. (1995). SRI International FASTUS system: MUC-6 test results and analysis. In *Proceedings of the Sixth Message Understanding Conference*, pages 237–248, San Francisco, CA. Morgan Kaufmann.

Azzam, S., Humphreys, K., and Gaizauskas, R. (1998). Evaluating a focus-based approach to anaphora resolution. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and COLING-98*, pages 74–78. Association for Computational Linguistics.

Bachar, K. and Lerman, I. (1998). Statistical conditions for a linear complexity for an algorithm of hierarchical classification under constraint of contiguity. In Rizzi, A., Vichi, M., and Bock, H., editors, *Advances in Data Science and Classification*, pages 131–136. Springer-Verlag.

Bagga, A. and Baldwin, B. (1998). Algorithms for scoring coreference chains. In *Proceedings of the LREC 1998 Workshop on Linguistic Coreference*, pages 563–566.

Basu, S., Bannerjee, A., and Mooney, R. (2002). Semi-supervised clustering by seeding. In *Proceedings of the Nineteenth International Conference on Machine Learning.*

Batagelj, V. and Ferligoj, A. (1998). Constrained clustering problems. In Rizzi, A., Vichi, M., and Bock, H., editors, *Advances in Data Science and Classification*, pages 137–144. Springer-Verlag.

Béjar, J. and Cortés, U. (1998). Experiments with domain knowledge in unsupervised learning: using and revising theories. *Computación y Sistemas*, 1(3):136–143.

Bell, J. F. (1992). Charge-coupled device imaging spectroscopy of Mars. *Icarus*, 100:575–597.

Bell, J. F. (1997). Visible and near-infrared spectroscopy. In Shirley, J. H. and Fairbridge, R. W., editors, *Encyclopedia of Planetary Sciences*, pages 911–915. Chapman and Hall.

Bell, J. F., McSween Jr., H. Y., Crisp, J. A., Morris, R. V., Murchie, S. L., Bridges, N. T., Johnson, J. R., Britt, D. T., Golombek, M. P., Moore, H. J., Ghosh, A., Bishop, J. L., Anderson, R. C., Bruckner, J., Economou, T., Greenwood, J. P., Gunnlaugsson, H. P., Hargraves, R. M., Hviid, S., Knudsen, J. M., Madsen, M. B., Reid, R., Rieder, R., and Soderblom, L. (2000). Mineralogic and compositional properties of Martian soil and dust: Results from Mars Pathfinder. *Journal of Geophysical Research*, 105(E1):1721–1755.

Bell, J. F., Wolff, M. J., James, P. B., Clancy, R. T., Lee, S. W., and Martin, L. J. (1997). Mars surface mineralogy from Hubble Space Telescope imaging during 1994-1995: Observations, calibration, and initial results. *Journal of Geophysical Research*, 102(E4):9109–9123.

Bertin, E. (2000). Mining pixels: The extraction and classification of astronomical sources. In Banday, A. J., Zaroubi, S., and Bartelmann, M., editors, *Mining the Sky: Proceedings of the MPA/ESO/MPE Workshop*, pages 353–371. European Southern Observatory, Springer.

Blake, C. L. and Merz, C. J. (1998). UCI repository of machine learning databases. http://www.ics.uci.edu/~mlearn/MLRepository.html.

Blum, A. and Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, pages 92–100.

Bottou, L. and Bengio, Y. (1995). Convergence properties of the k-means algorithm. *Advances in Neural Information Processing Systems*, 7:585–592.

Bradley, P. S., Bennett, K. P., and Demiriz, A. (2000). Constrained k-means clustering. Technical Report MSR-TR-2000-65, Microsoft Research, Redmond, WA.

Bradley, P. S. and Fayyad, U. M. (1998). Refining initial points for k-means clustering. In *Proceedings of the Fifteenth International Conference on Machine Learning*, pages 91–99, San Francisco, CA. Morgan Kaufmann.

Brown, D. E. and Gunderson, L. F. (2001). Using clustering to discover the preferences of computer criminals. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 31(4):311–318.

Cardie, C. (1993). A case-based approach to knowledge acquisition for domain-specific sentence analysis. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pages 798–803, Washington, DC. AAAI Press / MIT Press.

Cardie, C. and Howe, N. (1997). Improving minority class prediction using case-specific feature weights. In *Proceedings of the Fourteenth International Conference on Machine Learning*, pages 57–65, Vanderbilt University, Memphis, TN. Morgan Kaufmann.

Cardie, C. and Pierce, D. (1998). Error-driven pruning of treebank grammars for base noun phrase identification. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and COLING-98*, pages 218–224, University of Montreal, Montreal, Canada. Association for Computational Linguistics.

Cardie, C. and Wagstaff, K. (1999). Noun phrase coreference as clustering. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 82–89, University of Maryland, MD. Association for Computational Linguistics.

Cheeseman, P., Kelly, J., Self, M., Stutz, J., Taylor, W., and Freeman, D. (1988). Autoclass: A Bayesian classification system. In *Proceedings of the Fifth International Workshop on Machine Learning*, pages 54–64, Ann Arbor, MI. Morgan Kaufmann.

Cohn, D., Atlas, L., and Ladner, R. (1994). Improving generalization with active learning. *Machine Learning*, 15(2):201–221.

Cohn, D., Caruana, R., and McCallum, A. (2000). Semi-supervised clustering with user feedback. Unpublished manuscript.

Dagan, I. and Itai, A. (1991). A statistical filter for resolving pronoun references. In Feldman, Y. A. and Bruckstein, A., editors, *Artificial Intelligence and Computer Vision*, pages 125–135. Elsevier Science Publishers, North Holland.

de Carvalho, R. R., Djorgovski, S. G., Weir, N., Fayyad, U., Cherkauer, K., Roden, J., and Gray, A. (1995). Clustering analysis algorithms and their applications to digital POSS-II catalogs. In Shaw, R. A., Payne, H. E., and Hayes, J. J. E., editors, *Astronomical Data Analysis Software and Systems IV*, volume 77 of *ASP Conference Series*, pages 272–275.

Demiriz, A., Bennett, K. P., and Embrechts, M. J. (1999). Semi-supervised clustering using genetic algorithms. In *Proceedings of Artificial Neural Networks in Engineering '99.*

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38.

Djorgovski, S. G., Brunner, R. J., Mahabal, A. A., Odewahn, S. C., de Carvalho, R. R., Gal, R. R., Stolorz, P., Granat, R., Curkendall, D., Jacaob, J., and Castro, S. (2000). Exploration of large digital sky surveys. In Banday, A. J., Zaroubi, S., and Bartelmann, M., editors, *Mining the Sky: Proceedings of the MPA/ESO/MPE Workshop*, pages 305–322. European Southern Observatory, Springer.

Fellbaum, C. (1998). *WordNet: An Electronical Lexical Database.* The MIT Press, Cambridge, MA.

Ferligoj, A. and Batagelj, V. (1982). Clustering with relational constraint. *Psychometrika*, 47(4):413–426.

Ferligoj, A. and Batagelj, V. (1983). Some types of clustering with relational constraints. *Psychometrika*, 48(4):541–552.

Fisher, D. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, 2:139–172.

Fontaine, V., Leich, H., and Hennebert, J. (1994). Influence of vector quantization on isolated word recognition. In Holt, M. J. J., Cowan, C. F. N., Grant, P. M., and Sandham, W. A., editors, *Signal Processing VII, Theories and Applications. Proceedings of EUSIPCO-94. Seventh European Signal Processing Conference*, volume 1, pages 115–118, Lausanne, Switzerland. Eur. Assoc. Signal Process.

Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21:768–769.

Fowlkes, E. B. and Mallows, C. L. (1983). A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association*, 78(383):553–569.

Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company.

Gluck, M. A. and Corter, J. E. (1985). Information, uncertainty, and the utility of categories. In *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*, pages 283–287, Hillsdale, NJ. Lawrence Erlbaum.

Goebel, J., Volk, K., Walker, H., Gerbault, F., Cheeseman, P., Self, M., Stutz, J., and Taylor, W. (1989). A Bayesian classification of the IRAS LRS Atlas. *Astronomy and Astrophysics*, 222:L5–L8.

Gordon, A. D. (1973). Classification in the presence of constraints. *Biometrics*, 29:821–827.

Gordon, A. D. (1981). *Classification*. Chapman and Hall, New York, NY.

Gordon, A. D. (1996). A survey of constrained classification. *Computational Statistics & Data Analysis*, 21:17–29.

Guha, S., Rastogi, R., and Shim, K. (1998). CURE: An efficient algorithm for clustering large databases. In *Proceedings of the ACM-SIGMOD International Conference on Management of Data*, pages 73–84, Seattle, WA.

Gulati, R. K., Gupta, R., Gothoskar, P., and Khobragade, S. (1995). Automated classification of a large database of stellar spectra. In Shaw, R. A., Payne, H. E., and Hayes, J. J. E., editors, *Astronomical Data Analysis Software and Systems IV*, volume 77 of *ASP Conference Series*, pages 253–256.

Hartigan, J. A. (1975). *Clustering Algorithms.* Wiley, New York, NY.

Hunt, G. R. (1980). Electromagnetic radiation: The communication link in remote sensing. In Siegel, B. S. and Gillespie, A. R., editors, *Remote Sensing in Geology*, chapter 2, pages 5–45. Wiley, New York, NY.

Jain, A. and Farrokhnia, F. (1991). Unsupevised texture segmentation using Gabor filters. *Pattern Recognition*, 24(12):1167–1186.

Jain, A. K. and Dubes, R. C. (1988). *Algorithms for Clustering Data.* Prentice Hall.

Jancey, R. C. (1966). Multidimensional group analysis. *Australian Journal of Botany*, 14(1):127–130.

Kinderman, R. and Snell, J. L. (1980). *Markov Random Fields and Their Applications.* American Math Society, Providence, RI.

Klein, D., Kamvar, S. D., and Manning, C. D. (2002). From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering. In *Proceedings of the Nineteenth International Conference on Machine Learning.*

Kleinberg, J. and Tardos, E. (1999). Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. In *Proceedings of the 40th IEEE Symposium on Foundations of Computer Science*, pages 14–23.

Krishna, K. and Murty, M. N. (1999). Genetic k-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 29(3):433–439.

Křivánek, M. (1991). *Algorithmic and Geometric Aspects of Cluster Analysis.* Academia Nakladatelství Československé, Praha.

Lefkovitch, L. P. (1980). Conditional clustering. *Biometrics*, 36:43–58.

Lillesand, T. M. and Kiefer, R. W. (1987). *Remote Sensing and Image Interpretation*, chapter 10 (Digital Image Processing). Wiley, 2nd edition.

Lyman, P. and Varian, H. R. (2001). How much information. Retrieved from http://www.sims.berkeley.edu/how-much-info on November 29.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Symposium on Math, Statistics, and Probability*, volume 1, pages 281–297, Berkeley, CA. University of California Press.

Marroquin, J. and Girosi, F. (1993). Some extensions of the k-means algorithm for image segmentation and pattern recognition. AI Memo 1390, Massachusetts Institute of Technology, Cambridge, MA.

Masson, P. and Pieczynski, W. (1993). SEM algorithm and unsupervised statistical segmentation of satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 31(3):618–633.

McCarthy, J. and Lehnert, W. (1995). Using decision trees for coreference resolution. In Mellish, C., editor, *Proceedings of the Fourteenth International Conference on Artificial Intelligence*, pages 1050–1055.

McCord, T. B., Elias, J. H., and Westphal, J. A. (1971). Mars: The spectral albedo (0.3-2.5μ) of small bright and dark regions. *Icarus*, 14:245–251.

McSween Jr., H. Y., Murchie, S. L., Crisp, J. A., Bridges, N. T., Anderson, R. C., Bell, J. F., Britt, D. T., Bruckner, J., Dreibus, G., Economou, T., Ghosh, A., Golombek, M. P., Greenwood, J. P., Johnson, J. R., Moore, H. J., Morris, R. V., Parker, T. J., , Rieder, R., Singer, R., and Wanke, H. (1999). Chemical, multispectral, and textural constraints on the composition and origin of rocks at the Mars Pathfinder landing site. *Journal of Geophysical Research*, 104(E4):8679–8715.

Merényi, E., Singer, R. B., and Miller, J. S. (1996). Mapping of spectral variations on the surface of Mars from high spectral resolution telescopic images. *Icarus*, 124:280–295.

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill.

Mitkov, R. (1999). Anaphora resolution: the state of the art. Working paper (Based on the COLING'98/ACL'98 tutorial on anaphora resolution).

Mjolsness, E. and DeCoste, D. (2001). Machine learning for science: State of the art and future prospects. *Science*, 293(5537):2051–2055.

Montolio, P., Gasull, A., Monte, E., Torres, L., and Marqués, F. (1992). Analysis and optimization of the k-means algorithm for remote sensing applications. In de la Blanca, N. P., Sanfeliu, A., and Vidal, E., editors, *Pattern Recognition and Image Analysis*, volume 1 of *Series in Machine Perception and Artificial Intelligence*, pages 155–170. World Scientific, Singapore.

Morris, R. V., Lauer Jr., H. V., Lawson, C. A., Gibson Jr., E. K., Nace, G. A., and Stewart, C. (1985). Spectral and other physicochemical properties of submicron powders of hematite (α-Fe_2O_3), maghemite (γ-Fe_2O_3), magnetite (Fe_3O_4), goethite (α-FeOOH), and lepidocrocite (γ-FeOOH). *Journal of Geophysical Research*, 90:3126–3144.

MUC-6 (1995). *Proceedings of the Sixth Message Understanding Conference*. Morgan Kaufmann, San Francisco, CA.

MUC-7 (1998). *Proceedings of the Seventh Message Understanding Conference*. Morgan Kaufmann, San Francisco, CA.

Murtagh, F. (1985). A survey of algorithms for contiguity-constrained clustering and related problems. *The Computer Journal*, 28(1):82–88.

Narendra, P. M. and Goldberg, M. (1977). A non-parametric clustering cheme for Landsat. *Pattern Recognition*, 9:207–215.

Navigation Technologies (1996). *Software Developer's Toolkit.* Navigation Technologies, Inc., Sunnyvale, CA, 5.7.4 Solaris edition.

Ng, V. and Cardie, C. (2002). Combining sample selection and error-driven pruning for machine learning of coreference rules. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing.*

Nigam, K., McCallum, A., Thrun, S., and Mitchell, T. (1998). Learning to classify from labeled and unlabeled documents. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, pages 792–799, Madison, WI. The AAAI Press.

Oliver, M. and Webster, R. (1989). A geostatistical basis for spatial weighting in multivariate classification. *Mathematical Geology*, 21:15–35.

Pelleg, D. and Moore, A. (1999). Accelerating exact k-means algorithms with geometric reasoning. In *Knowledge Discovery and Data Mining*, pages 277–281.

Perruchet, C. (1983). Constrained agglomerative hierarchical classification. *Pattern Recognition*, 16(2):213–217.

Pieters, C. M., Mustard, J. F., and Sunshine, J. M. (1996). Quantitative mineral analyses of planetary surfaces using reflectance spectroscopy. In Dyar, M. D., McCammon, C., and Schaefer, M. W., editors, *Mineral Spectroscopy: A Tribute to Roger G. Burns*, pages 307–325. The Geochemical Society.

Popescu-Belis, A. and Robba, I. (1998). Three new methods for evaluating reference resolution. In *Proceedings of the LREC 1998 Workshop on Linguistic Coreference.*

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning.* Morgan Kaufmann, San Mateo, CA.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(366):846–850.

Roberts, S., Gisler, G., and Theiler, J. (1996). Spatio-spectral image analysis using classical and neural algorithms. In Dagli, C. H., Akay, M., Chen, C. L. P., Fernández, B. R., and Ghosh, J., editors, *Smart Engineering Systems: Neural Networks, Fuzzy Logic, and Evolutionary Programming*, volume 6 of *Intelligent Engineering Systems Through Artificial Neural Networks*, pages 425–430. ASME Press, New York, NY.

Rogers, S., Langley, P., and Wilson, C. (1999). Mining GPS data to augment road models. In *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining*, pages 104–113, San Diego, CA. ACM Press.

Schowengerdt, R. A. (1983). *Techniques for Image Processing and Classification in Remote Sensing*. Academic Press.

Schroedl, S., Rogers, S., , and Wilson, C. (2000). Map refinement from GPS traces. Technical Report RTC Report No. 6, DaimlerChrysler Research and Technology North America.

Selim, S. Z. and Ismail, M. A. (1984). K-means-type algorithms: A generalized convergence theorem and characterization of local optimality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1):81–87.

Selman, B., Mitchell, D. G., and Levesque, H. J. (1996). Generating hard satisfiability problems. *Artificial Intelligence*, 81:17–29.

Shmoys, D. B., Tardos, E., and Aardal, K. (1997). Approximation algorithms for facility location. In *Proceedings of the Twenty-ninth Annual ACM Symposium on the Theory of Computing*, pages 265–274.

Sibson, R. (1973). SLINK: An optimally efficient algorithm for the single-link cluster method. *The Computer Journal*, 16(1):30–34.

Soh, L. and Tsatsoulis, C. (1999). Segmentation of satellite imagery of natural scenes using data mining. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2):1086–1099.

Soon, W. M., Ng, H. T., and Lim, D. C. Y. (2001). A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521–544.

Talavera, L. and Béjar, J. (1999). Integrating declarative knowledge in hierarchical clustering tasks. In *Proceedings of the International Symposium on Intelligent Data Analysis*, pages 211–222, Amsterdam, The Netherlands. Springer-Verlag.

Theiler, J. and Gisler, G. (1997). A contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation. In *Proceedings of SPIE (International Society for Optical Engineering)*, volume 3159, pages 108–118.

Trouilleux, F., Gaussier, E., Bes, G. G., and Zaenen, A. (2000). Coreference resolution evaluation based on descriptive specificity. In *Proceedings of the LREC 2000 Workshop on Linguistic Coreference*.

Tung, A. K. H., Ng, R. T., Lakshmanan, L. V. S., and Han, J. (2001). Constraint-based clustering in large databases. In *Proceedings of the 2001 International Conference on Database Theory*, pages 405–419.

Urquhart, R. (1982). Graph theoretical clustering based on limited neighbourhood sets. *Pattern Recognition*, 15(3):173–187.

Vilain, M., Burger, J., Aberdeen, J., Connolly, D., and Hirschman, L. (1995). A model-theoretic coreference scoring scheme. In *Proceedings of the Sixth Message Understanding Conference*, pages 45–52, San Francisco, CA. Morgan Kaufmann.

Wagstaff, K. and Cardie, C. (2000). Clustering with instance-level constraints. In *Proceedings of the Seventeenth International Conference on Machine Learning*, pages 1103–1110, Palo Alto, CA. Morgan Kaufmann.

Wagstaff, K., Cardie, C., Rogers, S., and Schroedl, S. (2001). Constrained k-means clustering with background knowledge. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 577–584, Williamstown, MA. Morgan Kaufmann.

Wharton, S. W. (1983). A generalized histogram clustering scheme for multidimensional image data. *Pattern Recognition*, 16(2):193–199.

Yoo, J., Gray, A., Roden, J., Fayyad, U. M., de Carvalho, R. R., and Djorgovski, S. G. (1996). Analysis of digital POSS-II catalogs using hierarchical unsupervised learning algorithms. In Jacoby, G. H. and Barnes, J., editors, *Astronomical Data Analysis Software and Systems V*, volume 101 of *ASP Conference Series*, pages 41–44.

Zahn, C. (1971). Graph-theoretical methods for detecting and describing Gestalt clusters. *IEEE Transactions on Computing*, C-20:68–86.

Zhang, T., Ramakrishnan, R., and Livny, M. (1996). BIRCH: An efficient data clustering method for very large databases. In *Proceedings of the ACM-SIGMOD International Conference on Management of Data*.

Zhu, X., Chu, T., Zhu, J., and Caruana, R. (2001). Heuristically inducing a distance metric from user preferences for clustering. Unpublished manuscript.