

Charu C. Aggarwal

Work Address

Charu C. Aggarwal
19 Skyline Drive,
Hawthorne, NY 10532
Email: charu@us.ibm.com

Personal Address

Charu C. Aggarwal
1744 Strawberry Road,
Mohegan Lake, NY 10547
Email: charu@mit.edu

Research Interests

- Privacy preserving techniques for data management.
- Algorithms for fast streams of data.
- Algorithmic Methods in Electronic Commerce.
- Algorithms for searching, querying, indexing and crawling the world wide web.
- Application of information retrieval algorithms to world wide web resource discovery and personalization.
- Development of indexing, searching, querying, and mining techniques for massive commercial databases.
- Data and text mining algorithms.

Education

Massachusetts Institute of Technology

Cambridge, MA

Ph. D. in Operations Research, April 1996.

Thesis Title: Faster Algorithms for Some Network Flow Problems.

Indian Institute of Technology, Kanpur

Kanpur, India

Bachelor of Technology in Computer Science and Engineering, May 1993.

Graduated with grade point average of 10.0 out of 10.0

Received award for Best Finishing Undergraduate in the department of Computer Science.

Experience

IBM T. J. Watson Research Center

Yorktown Heights, New York

(June 10, 1996-present)

- Research staff member at IBM Research. Primary research interests in data mining, databases, information retrieval, privacy, and data streams.

Massachusetts Institute of Technology

Cambridge, MA

(August 1993- June 1996)

- Research and teaching assistant
- Research in Network Flows, Dynamic Data Structures, and Genetic Algorithms for Combinatorial Optimization
- Twice served as Teaching Assistant for the Course 15.082 (Network Flow Algorithms)

Honors

- IBM Research Division Award for contributions to stream mining in System S
- IBM Outstanding Innovation Award for scientific contributions to privacy technology, 2008
- IBM Corporate Award for contributions to Epispire Project on Bio-terrorist threat detection from high dimensional data streams of medical and pharmaceutical records, 2003.
- Designated **Master Inventor** at IBM Research (2000-2003, 2003-2006, 2006-2010).
- Invited Keynote Speaker, Workshop on Privacy-Preserving Data Mining, (Held in conjunction with ICDE, Istanbul, Turkey), 2007
- Invited Keynote Speaker, European Conference on Machine Learning, 2006.
- Invited Keynote Speaker, WEBKDD, 2005.
- Invited guest at the IBM Academy of Science, Stuttgart Germany, October 2000.
- IBM Seventeenth Plateau Invention Achievement Award (2008)
- IBM Sixteenth Plateau Invention Achievement Award (2007)
- IBM Fifteenth Plateau Invention Achievement Award, (2006)
- IBM Fourteenth Plateau Invention Achievement Award (2005)
- IBM Thirteenth Plateau Invention Achievement Award (2004)
- IBM Twelfth Plateau Invention Achievement Award (2004)
- IBM Eleventh Plateau Invention Achievement Award (2003)
- IBM Tenth Plateau Invention Achievement Award (2002)
- IBM Ninth Plateau Invention Achievement Award (2001)
- IBM Eighth Plateau Invention Achievement Award. (2000)
- IBM Seventh Plateau Invention Achievement Award. (2000)
- IBM Sixth Plateau Invention Achievement Award. (1999)
- IBM Fifth Plateau Invention Achievement Award. (1998)
- IBM Fourth Plateau Invention Achievement Award. (1998)
- IBM Third Plateau Invention Achievement Award. (1997)
- IBM Second Plateau Invention Achievement Award. (1997)
- IBM First Plateau Invention Achievement Award. (1996)
- First Patent Application award, IBM Research. (1996)
- Sloan School of Management Fellowship, Massachusetts Institute of Technology (1995).
- Indian Institute of Technology award for best finishing undergraduate in the Department of Computer Science (1993)
- 1991 first prize for academic excellence (IIT).

- 1990-1991 academic proficiency award (Indian Institute of Technology).
- 1989-1990 academic proficiency award (Indian Institute of Technology).

—Publications (Journals/Highly Refereed Conferences) —

- Charu C. Aggarwal, On Classification and Segmentation of Massive Audio Data Streams, *Knowledge and Information Systems*, accepted to appear.
- Charu C. Aggarwal, A Framework for Clustering Massive-Domain Data Streams, *ICDE Conference*, 2009.
- Charu C. Aggarwal, On High-Dimensional Projected Clustering of Uncertain Data Streams, *ICDE Conference*, 2009.
- Dina Thomas, Rajesh Bordawekar, Charu Aggarwal, Philip S. Yu, On Efficient Query Processing of Stream Counts on the Cell Processor, *ICDE Conference*, 2009.
- Charu C. Aggarwal, On Segment-Based Stream Modeling and its Applications, *ICDE Conference*, 2009.
- Charu C. Aggarwal, Philip S. Yu. A Survey of Uncertain Data Algorithms and Applications, *IEEE Transactions on Knowledge and Data Engineering*, to appear, 2009.
- Jianyong Wang, Youzhu Zhang, Lizhu Zhou, George Karypis, Charu C. Aggarwal. CONTOUR: An Efficient Algorithm for Discovering Discriminating Subsequences, *Data Mining and Knowledge Discovery Journal*, Feb 2009.
- Charu C. Aggarwal, Philip S. Yu. A Framework for Condensation-based Anonymization of String Data, *Data Mining and Knowledge Discovery Journal*, July 2008.
- Charu C. Aggarwal, Philip S. Yu. On Indexing High Dimensional Data with Uncertainty, *SIAM Conference on Data Mining*, 2008.
- Charu C. Aggarwal, Philip S. Yu. Outlier Detection with Uncertain Data. *SIAM Conference on Data Mining*, 2008.
- Charu C. Aggarwal, Philip S. Yu. On High Dimensional Indexing of Uncertain Data. *ICDE Conference*, 2008.
- Charu C. Aggarwal. On Unifying Privacy and Uncertain Data Models. *ICDE Conference*, 2008.
- Charu C. Aggarwal, Philip S. Yu. LOCUST: An Online Analytical Processing Frame-

work for High Dimensional Classification of Data Streams. *ICDE Conference*, 2008.

- Charu C. Aggarwal, Philip S. Yu. On Static and Dynamic Methods for Condensation-based Privacy-Preserving Data Mining. *ACM TODS Journal*, March 2008.
- Charu C. Aggarwal, Philip S. Yu. A Framework for Clustering Uncertain Data Streams. *ICDE Conference*, 2008.
- Kun-Lung Wu, Philip S. Yu, Bugra Gedik, Kirsten Hildrum, Charu C. Aggarwal, Eric Bouillet, Wei Fan, David George, Xiaohui Gu, Gang Luo, Haixun Wang: Challenges and Experience in Prototyping a Multi-Modal Stream Analytic and Monitoring Application on System S. *Very Large Databases (VLDB) Conference*, pp. 1185-1196, 2007.
- Charu C. Aggarwal. Towards Exploratory Test Instance Centered Diagnosis in High Dimensional Classification. *IEEE Transactions on Knowledge and Data Engineering*, August, 2007.
- Charu C. Aggarwal, Philip S. Yu. On String Classification in Data Streams. *ACM KDD Conference*, 2007.
- Charu C. Aggarwal, Na Ta, Jianyong Wang, Jianhua Feng, Mohammed Zaki. XProj: A Framework for Projected Structural Clustering of XML Documents. *ACM KDD Conference*, 2007.
- Charu C. Aggarwal. A Framework for Classification and Segmentation of Massive Audio Data Streams. *ACM KDD Conference*, 2007.
- Charu C. Aggarwal. On Point Sampling versus Space Sampling for Dimensionality Reduction. *SIAM Conference on Data Mining*, 2007.
- Charu C. Aggarwal, Philip S. Yu. On Privacy-Preservation of Text and Sparse Binary Data with Sketches. *SIAM Conference on Data Mining*, 2007.
- Charu C. Aggarwal, Philip S. Yu. On Anonymization of String Data. *SIAM Conference On Data Mining*, 2007.
- Jianyong Wang, Yuzhou Zhang, Lizhu Zhou, George Karypis, Charu C. Aggarwal. Discriminating subsequence discovery for sequence clustering. *SIAM Conference on Data Mining*, 2007.
- Charu C. Aggarwal. On Randomization, Public Information, and the Curse of Di-

dimensionality. *IEEE International Conference on Data Engineering*, 2007.

- Charu C. Aggarwal. On Density Based Transforms for Uncertain Data Mining. *IEEE International Conference on Data Engineering*, 2007.
- Charu C. Aggarwal. On Biased Reservoir Sampling in the Presence of Stream Evolution. *Very Large Databases (VLDB) Conference*, pp. 607–618, 2006.
- Charu C. Aggarwal, Jian Pei, Bo Zhang. On Privacy Preservation against Adversarial Data Mining. *ACM Knowledge Discovery and Data Mining (KDD) Conference*, pp. 510–516, 2006.
- Charu C. Aggarwal. On Futuristic Query Processing in Data Streams. *European Conference on Database Technology (EDBT)*, pp. 41–58, 2006.
- Charu C. Aggarwal, Chen Chen, Jiawei Han. On the Inverse Classification Problem and its Applications. *IEEE International Conference on Data Engineering*, 2006.
- Charu C. Aggarwal, Jiawei Han, Jianyong Wang, Philip S. Yu. A Framework for On Demand Classification of Evolving Data Streams. *IEEE Transaction on Knowledge and Data Engineering*, 18(5), pp. 577–589, 2006.
- Charu C. Aggarwal. On the use of human-computer interaction for meaningful high-dimensional nearest neighbor search. *Data Mining and Knowledge Discovery Journal*, 13(1), pp. 89–117, 2006.
- M. Zaki, C. Aggarwal: XRules: An effective algorithm for structural classification of XML data. *Machine Learning Journal* 62(1-2): 137-170, 2006.
- Charu C. Aggarwal. A Framework for Local Supervised Dimensionality Reduction of High Dimensional Data, *SIAM Conference on Data Mining*, pp. 360–371, 2006.
- Charu C. Aggarwal. Representation is Everything: Towards Efficient and Adaptable Similarity Measures for Biological Data , *SIAM Conference on Data Mining*, pp. 210–221, 2006.
- Charu C. Aggarwal, Philip S. Yu. A Framework for Clustering Massive Text and Categorical Data Streams, *SIAM Conference on Data Mining*, pp. 479–483, 2006.
- Charu C. Aggarwal. On k -anonymity and the curse of dimensionality. *Very Large Databases (VLDB) Conference*, pp. 901–909, 2005.

- Charu C. Aggarwal. Towards Exploratory Test Instance Centered Algorithms for High Dimensional Classification. *Knowledge Discovery and Data Mining (KDD) Conference*, pp. 526–531, 2005.
- Charu C. Aggarwal, Jiawei Han, Jianyong Wang, Philip S. Yu. On High Dimensional Projected Clustering of Data Streams. *Data Mining and Knowledge Discovery Journal*, 10(3), pp. 251–273, 2005.
- Charu C. Aggarwal, Philip S. Yu. An efficient and effective algorithm for high dimensional outlier detection. *VLDB Journal*, 14(2), pp. 211–221, 2005.
- Charu C. Aggarwal. On the use of Wavelet Decomposition for String Classification, *Data Mining and Knowledge Discovery Journal*, 10(2), pp. 117–139, 2005.
- Charu C. Aggarwal, Philip S. Yu. Online Analysis of Community Evolution in Data Streams. *SIAM Conference on Data Mining*, pp. 56–67, 2005.
- Charu C. Aggarwal, Philip S. Yu. On Variable Constraints in Privacy Preserving Data Mining Algorithms. *SIAM Conference on Data Mining*, pp. 115–125, 2005.
- Charu C. Aggarwal. On Abnormality Detection in Spuriously Populated Data Streams. *SIAM Conference on Data Mining*, pp. 80–91, 2005.
- Charu C. Aggarwal. On Change Diagnosis in Evolving Data Streams. *IEEE Transactions on Knowledge and Data Engineering*, 17(5), pp. 587–600, 2005.
- Charu C. Aggarwal. On Leveraging User Access Patterns for Topical Specific Crawling. *Data Mining and Knowledge Discovery Journal*, 9(2), pp. 123–146, 2004.
- Charu C. Aggarwal, Philip S. Yu. A Condensation Approach to Privacy Preserving Data Mining. *European Conference on Database Technology (EDBT)*, pp. 183–199, 2004.
- Charu C. Aggarwal, Jiawei Han, Jianyong Wang, Philip S. Yu. On Demand Classification of Data Streams. *ACM Knowledge Discovery and Data Mining (KDD) Conference*, pp. 503–508, 2004.
- Charu C. Aggarwal, Jiawei Han, Jianyong Wang, Philip S. Yu. A Framework For Projected Clustering of High Dimensional Data Streams. *Very Large Databases (VLDB) Conference*, pp. 852–863, 2004.
- Charu C. Aggarwal. An Efficient Subspace Sampling Framework for High Dimen-

sional Data Reduction, Selectivity Estimation, and Nearest Neighbor Search. *IEEE Transactions on Knowledge and Data Engineering*, 16(10), pp. 1247–1262, 2004.

- Charu C. Aggarwal. A Human Computer Interactive Method for Projected Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 16(4), pp. 448–460, 2004.

- Charu C. Aggarwal, Stephen C. Gates, Philip S. Yu. On Using Partial Supervision for Text Categorization. *IEEE Transactions on Knowledge and Data Engineering* 16(2) pp. 245–255, 2004.

- Srinivasan Parthasarathy, Charu C. Aggarwal. On the Use of Conceptual Reconstruction for Mining Massively Incomplete Data Sets. *IEEE Transactions on Knowledge and Data Engineering*, 15(6), pp. 1512–1521, 2003.

- Charu C. Aggarwal. A Framework for Diagnosing Changes in Evolving Data Streams. *ACM SIGMOD Conference on Management of Data*, pp. 575–586, 2003.

- Charu C. Aggarwal, Jiawei Han, Jianyong Wang, Philip Yu. CluStream: A Framework for Clustering Evolving Data Streams. *Very Large Databases (VLDB) Conference*, pp. 81–92, 2003.

- Charu C. Aggarwal. Towards Systematic Design of Distance Functions for Data Mining Applications. *ACM Knowledge Discovery and Data Mining (KDD) Conference*, pp. 9–18, 2003.

- M. J. Zaki, Charu C. Aggarwal. XRules: An Effective Structural Classifier for XML Data. *ACM Knowledge Discovery and Data Mining (KDD) Conference*, pp. 316–325, 2003.

- Charu C. Aggarwal, Dakshi Agrawal. On Nearest Neighbor Indexing of Nonlinear Trajectories. *ACM Principles on Database Systems (PODS Conference)*, pp. 252–259, 2003.

- Charu C. Aggarwal, Philip S. Yu. An Automated System for Web Portal Personalization. *Very Large Databases (VLDB) Conference*, pp. 1031–1040, 2002.

- Charu C. Aggarwal. On Effective Classification of Strings with Wavelets. *ACM Knowledge Discovery and Data Mining (KDD) Conference*, pp. 163–172, 2002.

- Charu C. Aggarwal. Collaborative Crawling: Mining User Experiences for Topical Resource Discovery. *ACM Knowledge Discovery and Data Mining (KDD Conference)*, pp. 423–428, 2002.

- Charu C. Aggarwal. Towards Meaningful High Dimensional Nearest Neighbor Search by Human-Computer Interaction. *IEEE International Conference On Data Engineering*, pp. 593–604, 2002.
- Charu C. Aggarwal. An Intuitive Framework for Understanding Changes in Evolving Data Streams. *IEEE International Conference on Data Engineering*, 2002.
- Charu C. Aggarwal, Philip S. Yu. A System for Personalization of Web News Feeds. *IEEE International Conference on Data Engineering*, 2002.
- Charu C. Aggarwal. Towards Effective and Interpretable Data Mining by Visual Interaction. *ACM SIGKDD Explorations*, 3(2), pp. 11–22, 2002.
- Charu C. Aggarwal. Hierarchical Subspace Sampling: A Unified Framework for High Dimensional Data Reduction, Selectivity Estimation and Nearest Neighbor Search. *ACM SIGMOD Conference on Management of Data*, pp. 452–463, 2002.
- Charu C. Aggarwal, Philip S. Yu. Redefining clustering for high dimensional applications. *IEEE Transactions on Knowledge and Data Engineering*, 14(2), pp. 210–225, 2002.
- Charu C. Aggarwal, Cecilia Procopiuc, Philip S. Yu. Finding Localized Associations in Market Basket Data. *IEEE Transactions on Knowledge and Data Engineering*, 14(1), pp. 51–62, 2002.
- Charu C. Aggarwal, James B. Orlin. On Multiroute Maximum Flows in Networks. *Networks*, 39(1), pp. 43–52, 2002.
- Charu C. Aggarwal, Zheng Sun, Philip S. Yu. Fast Algorithms for Online Generation of Profile Association Rules. *IEEE Transactions on Knowledge and Data Engineering*, 14(5), pp. 1017–1028, 2002.
- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu. Adaptive Piggybacking Schemes for Video-on-demand Systems. *Multimedia Tools and Applications*, 16(3), pp. 231–250, 2002.
- Charu C. Aggarwal. Re-designing distance functions and distance based applications for high dimensional data. *ACM SIGMOD Record*, 30(1), pp. 13–18, March 2001.
- Charu C. Aggarwal. On the Effects of Dimensionality Reduction on High Dimensional Similarity Search. *ACM Principles on Databases Systems (PODS Conference)*,

pp. 256–266, 2001.

- Dakshi Agrawal, Charu C. Aggarwal. On the Design and Quantification of Privacy Preserving Data Mining Algorithms. *ACM Principles on Databases Systems (PODS Conference)*, pp. 247–255, 2001.
- Charu C. Aggarwal. A Human-Computer Cooperative System for Effective High Dimensional Clustering. *ACM Knowledge Discovery and Data Mining (KDD) Conference*, pp. 221–226, 2001.
- Charu C. Aggarwal, Srinivasan Parthasarathy. Mining Massively Incomplete Data Sets by Conceptual Reconstruction. *ACM Knowledge Discovery and Data Mining (KDD) Conference*, pp. 227–232, 2001.
- Charu C. Aggarwal. Towards Long Pattern Generation in Dense Databases. *ACM SIGKDD Explorations*, 3(1), pp. 20–26, 2001.
- Charu C. Aggarwal, Philip S. Yu. On Effective Conceptual Indexing and Similarity Search in Text Data. *IEEE International Conference on Data Mining (ICDM Conference)*, pp. 3–10, 2001.
- Sang-Wook Kim, Charu C. Aggarwal, Philip S. Yu. Effective Nearest Neighbor Indexing with the Euclidean Metric. *ACM CIKM Conference*, pp. 9–16, 2001.
- Charu C. Aggarwal, Fatima Al-Garawi, Philip Yu. On the Design of a Learning Crawler for Topical Resource Discovery. *ACM Transactions on Information Systems*, 19(3), pp. 286–309, 2001.
- Charu C. Aggarwal, Philip S. Yu. Outlier Detection for High Dimensional Data. *ACM SIGMOD Conference on Management of Data*, pp. 37–46, 2001.
- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu. The Maximum Factor Queue Length Batching Scheme for Video On Demand Systems. *IEEE Transactions on Computers*, 50(2), pp. 97–110, 2001.
- Charu C. Aggarwal, Philip S. Yu. A New Approach to Online algorithms for generation of association rules. *IEEE Transactions on Knowledge and Data Engineering*, 13(4), pp. 527–540, 2001.
- Charu C. Aggarwal, Philip S. Yu. Mining Associations with the Collective Strength Approach. *IEEE Transactions on Knowledge and Data Engineering*, 13(6), pp. 863–873, 2001.

- Ramesh C. Agarwal, Charu C. Aggarwal, V. V. V. Prasad. A Tree Projection Algorithm for Generation of Frequent Itemsets. *Journal of Parallel and Distributed computing*. 61(3), pp. 360–371, 2001.
- Charu C. Aggarwal, Fatima Al-Garawi, Philip S. Yu. Intelligent Crawling on the World Wide Web with arbitrary predicates. *World Wide Web (WWW) Conference*, pp. 96–105, 2001.
- Charu C. Aggarwal, Alexander Hinneburg, Daniel A. Keim. On the surprising behavior of distance metrics in high dimensional space. *International Conference on Database Theory (ICDT Conference)*, pp. 420–434, 2001.
- Alexander Hinneburg, Charu C. Aggarwal, Daniel A. Keim. What is the nearest neighbor in high dimensional space? *Very Large Databases (VLDB) Conference*, pp. 506–515, 2000.
- Charu C. Aggarwal, Philip S. Yu. The IGrid Index: Reversing the dimensionality curse for similarity indexing in high dimensional space. *ACM Knowledge Discovery and Data Mining (KDD) Conference*, pp. 119–129, 2000.
- Ramesh C. Agarwal, Charu C. Aggarwal, V.V.V. Prasad. Depth First Generation of Long Patterns. *ACM Knowledge Discovery and Data Mining (KDD) Conference*, pp. 108–118, 2000.
- Charu C. Aggarwal, Philip S. Yu. Finding Generalized Projected Clusters in High dimensional spaces. *ACM SIGMOD Conference on Management of Data*, pp. 70–81, 2000.
- Charu C. Aggarwal, Philip S. Yu. Data Mining Techniques for Personalization. *IEEE Data Engineering Bulletin*, 23(1), pp. 4–9, 2000.
- Charu C. Aggarwal, Joel L. Wolf, Kun-Lung Wu, Philip S. Yu. Horting Hatches an Egg: A New Graph Theoretic Approach to Collaborative Filtering. *ACM Knowledge Discovery and Data Mining (KDD) Conference*, pp. 201–212, 1999.
- Charu C. Aggarwal, Stephen C. Gates, Philip S. Yu. On the merits of building categorization systems by supervised clustering. *ACM Knowledge Discovery and Data Mining (KDD) Conference*, pp. 352–356, 1999.
- Charu C. Aggarwal, Cecilia Procopiuc, Joel L. Wolf, Philip S. Yu, Jong Soo Park. Fast Algorithms for Projected Clustering. *ACM SIGMOD Conference on Management*

of Data, pp. 61–72, 1999.

- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu. A new method for similarity indexing of market basket data. *ACM SIGMOD Conference on Management of Data*, pp. 407–418, 1999.
- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu. Using unbalanced Trees for indexing multidimensional objects. *Knowledge and Information Systems Journal*, 1(3), pp. 157–192, 1999.
- Charu C. Aggarwal, Joel L. Wolf, and Philip S. Yu. Design and Analysis of Permutation-Based Pyramid Broadcasting. *Multimedia Systems Journal*, 7(6), pp. 439–448, 1999.
- Charu C. Aggarwal, Joel L. Wolf, and Philip S. Yu. Caching on the world wide web. *IEEE Transactions of Knowledge and Data Engineering*, 11(1) pp. 95–107, 1999.
- Charu C. Aggarwal, Joel L. Wolf, and Philip S. Yu. Optimization Issues in Multimedia Systems. *International Journal of Intelligent Systems on Multimedia Computing Systems*, December 1998. John Wiley and Sons.
- Charu C. Aggarwal, Philip S. Yu. Online Algorithms for Finding Profile Association Rules. *ACM CIKM Conference*, pp. 86–95, 1998.
- Charu C. Aggarwal, Zheng Sun, Philip S. Yu. Online Generation of Profile Association Rules. *Knowledge Discovery and Data Mining (KDD) Conference*, pp. 129–133, 1998.
- Charu C. Aggarwal, Joel L. Wolf and Philip S. Yu, A Framework for the Optimizing of WWW Advertising, *Trends in Distributed Systems for Electronic Commerce*, edited by W. Lamersdorf and M. Merz, Springer-Verlag, Lecture Notes in Computer Science, Vol. 1402, pp. 1-10, 1998.
- Charu C. Aggarwal and Philip S. Yu. Online Generation of Association Rules. *IEEE International Conference on Data Engineering*, pp. 402–411, 1998.
- Charu C. Aggarwal, and Philip S. Yu, Mining large itemsets for association rules. *IEEE Data Engineering Bulletin*, 21(1), pp. 23–31, 1998.
- Charu C. Aggarwal, and Philip S. Yu, A new framework for itemset generation. *ACM Symposium on Principles of Database Systems (PODS Conference)*, pp. 18–24, 1998.

- Charu C. Aggarwal, Ravindra K. Ahuja, Jianxiu Hao, and James B. Orlin, Diagnosing Infeasibilities in Network Flow Problems, *Mathematical Programming*, 81, pp. 263–280, 1998, *Elsevier Science*, Amsterdam, Netherlands.
- Charu C. Aggarwal and Philip S. Yu. On disk caching of web objects in proxy servers. *International Conference on Information and Knowledge Management*, pp. 238–245, 1997.
- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu, and Marina Epelman. The S-Tree: An efficient index for multi-dimensional objects. *Symposium on Spatial Databases*, Berlin, Germany, 1997. Paper also appears in *Lecture Notes in Computer Science*, pp. 350-376, *Springer-Verlag*, Berlin, Germany.
- Charu C. Aggarwal, James B. Orlin, and Ray P. Tai, Optimized Crossover for the Independent Set Problem. *Operations Research*, 45(2), pp. 226-234, 1997.
- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu, On optimal piggyback merging policies in video-on-demand systems. *ACM SIGMETRICS Conference*, pp. 200–209, 1996.
- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu, On optimal batching policies in video-on-demand storage systems. *IEEE International Conference on Multimedia Computing and Systems*, pp. 253–258, 1996.
- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu, A permutation based pyramid broadcasting scheme for video-on-demand storage systems. *IEEE International Conference on Multimedia Computing and Systems*, pp. 118–126, 1996.
- Charu C. Aggarwal, Nitesh Jain, and Phalguni Gupta, An efficient selection algorithm on the pyramid, *Information Processing Letters*, 53(1) pp. 37-47, 1995, *Elsevier Science*, Amsterdam, Netherlands.

Book Chapters

- Charu C. Aggarwal, Stephen C. Gates, Philip S. Yu. On Supervised Clustering for Creating Categorization Segmentations. In *Constrained Clustering: Advances in Theory, Algorithms and Applications*. Ed. Sugato Basu, Ian Davidson, Kiri Wagstaff, Chapman and Hall, 2008.
- Charu C. Aggarwal, Philip S. Yu. An Introduction to Privacy-Preserving Data Mining. *Privacy-Preserving Data Mining: Models and Algorithms*. Ed. Charu Aggarwal, Philip S. Yu, Springer Inc., 2008.

- Charu C. Aggarwal, Philip S. Yu. A General Survey of Privacy-Preserving Data Mining Models and Algorithms. *Privacy-Preserving Data Mining: Models and Algorithms*. Ed. Charu C. Aggarwal, Philip S. Yu, Springer, 2008.
- Charu C. Aggarwal, Philip S. Yu. A survey of randomization algorithms for privacy-preserving data mining. *Privacy-Preserving Data Mining: Models and Algorithms*. Ed. Charu C. Aggarwal, Philip S. Yu, Springer, 2008.
- Charu C. Aggarwal. On the Curse of Dimensionality for Privacy-Preserving Data Mining. *Privacy-Preserving Data Mining: Models and Algorithms*. Ed. Charu C. Aggarwal, Philip S. Yu, Springer, 2008.
- Charu C. Aggarwal, Philip S. Yu. Privacy-Preserving Data Mining: A Survey. *Handbook on Database Security*, ed. Michael Gertz, Sushil Jajodia, Springer, 2008.
- Charu C. Aggarwal. An Introduction to Data Streams. *Data Streams: Models and Algorithms*. Ed. Charu Aggarwal, Springer Inc., 2007.
- Charu C. Aggarwal, Philip S. Yu. A Survey of Synopsis Construction in Data Streams. *Data Streams: Models and Algorithms*. Ed. Charu Aggarwal, Springer Inc., 2007.
- Charu C. Aggarwal. A Survey of Change Diagnosis Methods in Data Streams. *Data Streams: Models and Algorithms*, Ed. Charu Aggarwal, Springer Inc., 2007.
- Charu C. Aggarwal, Jiawei Han, Jianyong Wang, Philip Yu. On Clustering Data Streams: A Summarization Paradigm. *Data Streams: Models and Algorithms*. Ed. Charu Aggarwal, Springer Inc., 2007.
- Charu C. Aggarwal, Philip S. Yu. On Clustering Techniques for Change Diagnosis in Data Streams. *Lecture Notes in Computer Science*, vol. 4198, 2006.
- Charu C. Aggarwal. On Learning Strategies for Topic Specific Web Crawling. *Next Generation Data Mining Applications*, Ed. J. Zurada, M. Kantardzic, John Wiley Inc., 2005.
- Charu C. Aggarwal, Philip S. Yu. Data Mining Techniques for Associations, Clustering, and Classification, *Lecture Notes in Computer Science*, Vol. 1574, 1999.

Books

- Charu C. Aggarwal, Philip S. Yu. Privacy-Preserving Data Mining: Models and Algorithms, Springer, 2008. The work contains surveys on privacy-preserving data mining from well known researchers.

- Charu C. Aggarwal (Editor). *Data Streams: Models and Algorithms*, Springer, 2007. The work contains surveys on data streams by well known researchers in the field.
- R. K. Ahuja, T. L. Magnanti, J. B. Orlin and C. C. Aggarwal. *Solutions manual to Network Flows: Theory, Algorithms and Applications*. Published by *Prentice Hall*, Englewood Cliffs, NJ. Approximately 200 pages. The work is a guide for teaching instructors.

Patents

60 patents have been filed. Of these, 39 have already been granted by the US patent office. The remaining are pending processing at the patent office. I have also been granted several invention plateaus by IBM for my patents. I have twice been designated **Master Inventor** by IBM Research for the commercial value of my patents. A list of the 37 granted patents is below:

- Charu C. Aggarwal. Method and apparatus for flexible data reduction of High Dimensional data, US Patent 7,277,893, 2007.
- Charu C. Aggarwal. Method and Apparatus for Reduction of High Dimensional Data, US Patent 7,236,638, 2007.
- Charu C. Aggarwal. Method and Apparatus of Generating Test data sets in accordance with user feedback. US Patent 7,085,981, August 2006.
- Charu C. Aggarwal. Methods and Apparatus for User-Centered Similarity Learning. US Patent 6,970,884, November 2005.
- Charu C. Aggarwal, Philip S. Yu. System and Method of Similarity Indexing and Searching in High Dimensional Space. US Patent 6,922,700, July 2005.
- Charu C. Aggarwal. Method and Apparatus for Classifying Time Series Data using wavelet based approach. US Patent 6,871,165, March 2005.
- Charu C. Aggarwal. System and Method for Mining Unstructured Data Sets. US Patent 6,847,955, January 2005.
- Charu C. Aggarwal. Method and Apparatus for User-Centered Class Supervision. 6,804,669, October 2004.
- Charu C. Aggarwal. System and Method of Determining and Searching for Patterns in A Large Database. US Patent 6,799,175, September 2004.
- Charu C. Aggarwal, Philip S. Yu. System and Method of Flexible Indexing of Text

for Use in Similarity Searches. US Patent 6,785,669, August 2004.

- Charu C. Aggarwal, Philip S. Yu. System and Method for Classification using Time Sequences. US Patent 6,721,719, April, 2004.
- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu. Method for Targeted Advertising on the Web based on accumulated self learning data, clustering users, and semantic node graph techniques. US Patent 6,714,975, March 2004.
- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu. Method for optimizing profits in electronic delivery of digital objects. US Patent 6,613,413, October 2003.
- Charu C. Aggarwal, Philip S. Yu. Methods and Apparatus for Performing an Affinity Based Similarity Search. US Patent 6,587,848, July 2003.
- Charu C. Aggarwal, Philip S. Yu. Method and Apparatus for Similarity Text Search Based on Conceptual Indexing. US Patent 6,542,889, April 2003.
- Charu C. Aggarwal, Philip S. Yu. Method and Apparatus for Performing Dimensionality Reduction in a Supervised Application Domain. US Patent 6,505,207, January 2003.
- Charu C. Aggarwal, Joel Wolf, Philip S. Yu. System and Method of Collaborative Filtering with Applications to Ecommerce. US Patent 6,487,541, November 2002.
- Charu C. Aggarwal, Philip S. Yu. Semantic based collaborative Filtering. US Patent 6,487,539, November 2002.
- Charu C. Aggarwal, J. Sethuraman, M. Squillante, J. Wolf, P. Yu. Optimizing method for digital content delivery in a multicast network. US Patent 6,477,180, November 2002.
- Charu C. Aggarwal, Philip S. Yu. System and Method of Using Clustering to Find Personalized Associations. US Patent 6,408,295, June 2002.
- Ramesh C. Agarwal, Charu C. Aggarwal, V. V. V. Prasad. Depth First Method for Generating Itemsets. US Patent 6,389,416, May 2002.
- Charu C. Aggarwal, Stephen C. Gates, Philip S. Yu. System and Method for Generating Taxonomies with Applications to Content Based Recommendations. US Patent 6,360,227, March 2002.
- Charu C. Aggarwal, Philip S. Yu. Content Based Method for Product Peer Filtering. US Patent 6,356,879, March 2002.

- Charu C. Aggarwal, Philip S. Yu. System and Method for Detecting Clusters of Information with Applications to Ecommerce. US Patent 6,349,309, February 2002.
- Ramesh C. Agarwal, Charu C. Aggarwal, V. V. V. Prasad. System and Method of Generating Associations. US Patent 6,311,179, October 2001.
- Charu C. Aggarwal, Joel L. Wolf, Philip Yu. System and Method for Detecting Clusters of Information. US Patent 6,307,965, October 2001.
- Charu C. Aggarwal, B. Hailpern, Joel L. Wolf, Philip Yu. System and Method for Similarity Searching in High Dimensional Space. US Patent 6,289,354, September 2001.
- Charu C. Aggarwal, Philip S. Yu. Finding Collective Baskets and Inference Rules for Internet Mining. US Patent 6,263,327, July 2001.
- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu. System and Method for Searching Databases with Applications such as Peer Groups, Collaborative Filtering and Ecommerce. US Patent 6,236,985, May 2001.
- Charu C. Aggarwal, Philip S. Yu. Methods for Performing Large Scale Auctions and Online Negotiations. US Patent 6,151,589, November 2000.
- Charu C. Aggarwal, Philip S. Yu. Online Mining of quantitative association rules. US Patent 6,092,064, July 2000.
- Charu C. Aggarwal, Philip S. Yu. Finding Collective Baskets and inference rules for internet or intranet mining for large databases. US Patent 6,094,645, July 2000.
- Charu C. Aggarwal, M. Epelman, Joel L. Wolf, Philip S. Yu. System and Method for caching objects of non-uniform size using multiple LRU stack partitions into a range of sizes. US Patent 6,012,126, January 2000.
- Charu C. Aggarwal, Philip S. Yu. Eliminating Redundancy in Generation of Association Rules for Online Mining. US Patent 5,943,667, August 1999.
- Charu C. Aggarwal, P. Malkin, B. Schloss, Philip S. Yu. Collaborative caching of a requested object by a lower level node as a function of the caching status of the object at a higher level node. US Patent 5,924,116, July 1999.
- Charu C. Aggarwal, Philip S. Yu. Online Mining of Association Rules. US Patent 5,920,855, July 1999.

- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu. Permutation Based Pyramid Block Transmission Scheme for Broadcasting in Video-On-Demand Storage Systems. US Patent 5,751,336, May 1998.
- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu. System and Method for Construction of a Data Structure for Indexing Multi-Dimensional Objects. US Patent 5,781,906, July 1998.
- Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu. Maximum Factor Selection Policy for Batching VOD Requests. US Patent 5,631,694, May 1997.

— Research Prototypes and Product Transfers —

- Developed micro-clustering prototype for fast data streams. Applied technique to intrusion detection. Designed a prototype on a distillery cluster for fast processing of incoming data. The approach can be flexibly used for any data mining or streaming application which requires summarization.
- Developed a number of real time algorithms for abnormality detection in data streams. Developed a prototype implementation which can detect possible alarms for terrorist attacks from medical data streams. Subsequently, the approach was scaled up to increase its speed to very fast data streams. The contributions to this project was recognized by an **IBM Corporate Award for Environmental Excellence** in 2003.
- Interactive Miner: The prototype was adapted from a research paper in the ICDE Conference, 1998. This is a technique for online and interactive generation of association rules with user specified constraints. Uses an OLAP-like preprocess once-query many methodology. Implemented and conceived main ideas behind the prototype. The prototype was released by IBM as an alphaworks release in order to allow the technology to receive public exposure.
- Profile Miner: My research on profile association rule mining was used to create this prototype. The prototype computes profiles of customers which show different kinds of buying behaviors. This approach is useful in e-commerce applications with fast click streams. An example would be an e-commerce merchant with large volumes of buying behavior. This approach provides interactive mining ability with the use of a preprocess-one query-many paradigm. The prototype was released by IBM as part of its alphaworks suite of releases.
- Portal Assistant: My research on text categorization was used to develop a system to personalize web portals. The particular portal which was implemented using

this scheme was the GMAC financial web site of the General Motors Corporation. I implemented the text categorizer and recommendation system in this IGS offered service. The system provides effective personalization of a web portal, and improves the browsing experience of the customers.

Professional activities

Conference Executive Committees/Editorships/Keynotes:

- Editorial Board, *Journal of Database Management Research*, 2009-present.
- Editorial Board, *Knowledge and Information Systems*, 2009-present.
- Associate Editor, *ACM SIGKDD Explorations*, 2008-present
- Associate editor, *IEEE Transactions on Knowledge and Data Engineering*, 2003-2007.
- Action Editor, *Data Mining and Knowledge Discovery Journal*, 2005-present
- Program vice-chair, *ICDM Conference*, 2007.
- Program vice-chair, *SIAM Conference on Data Mining*, 2007.
- Program vice-chair (co-chair of Data Mining Track), *WWW Conference*, 2009
- Keynote Speaker at the *European Conference on Machine Learning*, 2006.
- Keynote Speaker of the *WEBKDD Workshop*, 2005.
- Keynote Speaker, *Workshop on Privacy-Preserving Data Mining*, (held in conjunction with *ICDE*), Istanbul, Turkey, 2007.
- Sponsorship chair for *ACM KDD 2002*
- Workshop Chair for *ACM KDD 2003*
- Program Chair for *DMKD 2003*.

Conference Program Committees:

- Program Committee: *IEEE ICDE Conference 2002*
- Program Committee: *IEEE ICDM Conference 2002*
- Program Committee: *IEEE ICDE Conference 2004*
- Program Committee: *WWW Conference 2004*
- Program Committee: *ACM CIKM Conference 2004*
- Program Committee: *ICDM Conference 2004*
- Program Committee: *WWW Conference 2005*
- Program Committee: *ACM CIKM Conference, 2005*
- Program Committee: *ACM KDD Conference, 2005*
- Program Committee: *ICDM Conference, 2005*
- Program Committee: *ACM KDD Conference 2006*
- Program Committee: *SIAM Conference on Data Mining 2006*
- Program Committee: *IEEE ICDE Conference 2006*
- Program Committee: *ECML Conference 2006*
- Program Committee: *ACM CIKM Conference 2006*
- Program Committee: *IEEE ICDM Conference 2006*
- Program Committee: *SIGMOD Conference 2007*

- Program Committee: ICDE Conference 2007
- Program Committee: KDD Conference 2007
- Program Committee: WWW Conference 2007
- Program Committee: SIGMOD Conference, 2007
- Program Committee: ECML Conference, 2007
- Program Committee: AAI Conference, 2007
- Program Committee: CIKM Conference, 2007
- Program Committee: SIAM Data Mining Conference, 2008
- Program Committee: SIGMOD Conference, 2008
- Program Committee: WWW Conference, 2008
- Program Committee: ICML Conference, 2008
- Program Committee: ICDE Conference, 2008
- Program Committee: KDD Conference, 2008
- Program Committee: ECML Conference, 2008
- Program Committee: ICDE Conference, 2009