Guy Lebanon

www.theanalysisofdata.com
linkedin.com/in/guylebanon
glebanon@gmail.com

May 2018

Biography:

Guy Lebanon is a director at Amazon where he leads the search relevance, query understanding, search labs, and human judgements teams. Previously, he was a director at Netflix where he lead its homepage personalization efforts including its well known movie recommendation system. Before that he had engineering management positions at Amazon and LinkedIn and was a tenured professor at the Georgia Institute of Technology. Guy published several books and over 70 refereed articles in machine learning, and received a PhD from Carnegie Mellon University. He chaired the 2015 AI & Statistics conference and the 2012 ACM CIKM conference and was action editor of Journal of Machine Learning Research during 2013-2018. He won first place in the PASCAL image segmentation competition three times, and received the NSF CAREER Award, the WWW best student paper award, the ICML best paper runner-up award, the Yahoo Faculty Research and Engagement Award, and is a Siebel Scholar.

A. EDUCATION

PhD School of Computer Science Carnegie Mellon University 2005
Dissertation: Riemannian Geometry and Statistical Machine Learning
Advisor: John Lafferty

MS School of Computer Science Carnegie Mellon University 2002
Thesis: Boosting and Maximum Likelihood for Exponential Models
Advisor: John Lafferty

MS Dept of Computer Science Technion – Israel Institute of Technology 2000
Thesis: Optimal Synthesis of Moire Patterns in Computer Vision
Advisor: Alfred Bruckstein

BA Dept of Computer Science Technion – Israel Institute of Technology 1999
Summa Cum Laude (top 3% of graduating students)

B. WORK EXPERIENCE

1. ***Director, Software Engineering and Applied Science at Amazon*** ***2017-***
   I lead several teams of engineers and scientists (search relevance, query understanding, search labs, human judgements) who work on improving Amazon's product search using AI and machine learning.
2. ***Director, Product Management at Netflix*** ***2016-2017***
   I was the product lead for two of Netflix's most important projects: (a) homepage personalization: deciding which movies to show to users and in what order, and (b) asset selection: deciding which assets (images, trailers, etc.) to use when representing our movies.
3. ***Senior Manager, Software Engineering and Applied Science at LinkedIn*** ***2014-2016***
   I lead LinkedIn's news-feed personalization team (second line manager), including 3 sub-teams working on machine learning modeling, offline infrastructure, and online infrastructure. During my tenure, LinkedIn's newsfeed quality experienced significant improvements, including an over +100% increases in feed engagement (as measured by likes) and an over +100% increase in revenue from the feed.
4. ***Senior Manager, Software Engineering and Applied Science at Amazon*** ***2012-2014***
   I lead Amazon's machine learning science team in Seattle and worked on machine learning applications in e-commerce, cloud computing, computer vision, streaming video, and digital devices including Kindle. My team's work resulted in a significantly improved recommendation system for Amazon Instant Video. I also won the ICML best paper runner up award and (with a

student) the WWW best student paper award.

5. ***Associate Professor (with tenure) at Georgia Institute of Technology*** ***2008-2012***

   I conducted research in machine learning including recommendation systems, graphical models, computer vision, and natural language processing. I was granted tenure and promoted in 2011, and in 2012 I was appointed associate director of FODAVA, a national consortium of research institutions in the area of data and visual analytics. I also won first place in three Pascal VOC Image Segmentation Challenges (2010, 2011, 2012).

6. ***Assistant Professor at Purdue University*** ***2005-2008***

   I conducted research in machine learning including natural language processing, graphical models, and statistical models for ranked data. I received the NSF Faculty Early Career Award and a university teaching award.

7. ***Research Assistant and Postdoc at Carnegie Mellon University*** ***2000-2005***

   I conducted research in machine learning including ranking models, classification models, and natural language processing. I was awarded the Siebel scholarship and received the LTI best presentation award.

8. ***Programmer at Technion and Rafael*** ***1997-200***

   I worked on an automatic orientation detection system for the Technion's satellite and automatic target recognition at Rafael.

9. ***Serviceman in the Israel Defense Forces*** ***1993-1996***

   I had a number of roles in the IDF's armored corps including tank commander, operations sergeant, and battalion training officer.

C. HONORS AND AWARDS

1.	ML-Conf 2016 presentation ranked second best by audience (out of 17)	2016
2.	General Chair, AI & Statistics Conference (AISTATS)	2015
3.	Best Student Paper Award The International World Wide Web Conference (WWW)	2014
4.	Action Editor, Journal of Machine Learning Research	2013-2017
5.	Best Paper Runner-Up Award International Conference on Machine Learning (ICML)	2013
6.	First place in the 2012 Pascal VOC Image Segmentation Challenge	2012
7.	Program Chair, ACM CIKM Conference	2012
8.	First place in the 2011 Pascal VOC Image Segmentation Challenge	2011
9.	Yahoo Faculty Research and Engagement Award	2011
10.	First place in the 2010 Pascal VOC Image Segmentation Challenge	2010
11.	Raytheon Faculty Fellowship Award	2010
12.	NSF Faculty Early Career (CAREER) Award	2008
13.	Class of 1969 Teaching Fellow, Georgia Institute of Technology	2008
14.	Teaching for Tomorrow Award, Purdue University	2007
15.	Best Presentation Award, LTI Student Research Symposium	2004
16.	Siebel Scholar, The Siebel Scholars Foundation	2004
17.	Bachelor Degree Awarded Summa Cum Laude, Technion – Israel Institute of Technology (top 3% of student body)	1999
18.	President's Award and Scholarship, Technion – Israel Institute of Technology (top 5% of student body)	1998

D. ADVISORY BOARDS

1.	Headset.io, Seattle WA	2014-
2.	Brand.ai, Seattle WA	2014-
3.	FLAMEL (flamel.gatech.edu, NSF IGERT Program)	2015-2016
4.	Interdisciplinary Master of Science in Analytics Program, Georgia Tech	2014-2016
5.	Peerlyst, San Francisco CA, USA	2012-2013
6.	Reactive Search S.R.L., Trento, Italy	2010-2013
7.	Pave Inc., Atlanta GA, USA	2010-2012
8.	Chatterspike Inc., Indianapolis, IN USA	2008-2009
9.	Vynate Inc., Indianapolis, IN USA	2006-2008

E. GRANTS AND GIFTS

1. *Foundations of Data Analysis and Visual Analytics.* National Science Foundation and US Department of Homeland Security. $3,000,000, 2008-2012.
2. *Fast Algorithms on Imperfect, Heterogeneous, Distributed Data for Interactive Analysis.* US Department of Defense (DARPA). $2,700,000, 2012-2017.
3. *Purdue Regional Visualization and Analytics Center*. US Department of Homeland Security. $750,000, 2006-2008.
4. *CAREER: Multi-Resolution Representations of Documents.* National Science Foundation. $405,548, 2008-2012.
5. *IPS: Decision Theoretic Approaches to Measuring and Minimizing Customized Privacy Risk*. National Science Foundation. $371,625, 2007-2010.
6. *Systematic Control and Management of Data Integrity, Quality and Provenance for Command and Control Applications*. Air Force Office of Scientific Research, $300,000, 2007-2010.
7. *Statistical Inference for Censored Preference Data.* National Science Foundation. $175,881, 2009-2011.
8. *Machine Learning for Tied and Incomplete Preference Data*. US-Israel Binational Science Foundation (BSF). $150,000, 2011-2013.
9. *SBIR Phase I: Software to Aggregate, Correlate, Analyze and Trend data for Knowledge Management in Decision Making.* National Science Foundation. $150,000. 2007-2008.
10. *Assessing the Readability of Documents and Statistical Tools for non-Euclidean Data*. National Science Foundation. $113,000, 2006-2008.
11. *Machine Learning and Visualization for Computational Journalism*. Raytheon Faculty Fellowship Award. Raytheon Corporation through the Georgia Institute of Technology. $53,805, 2010-2011.
12. *Digging into Human Rights Violations: Anaphora Resolution and Emergent Witnesses.* Sub-award from Georgia State University. $50,000, 2012-2012.
13. *Social Satellite Seedling.* US Department of Defense, DARPA. $50,000 2011-2012.
14. *Supplementary CAREER Award Funding.* Office of the Provost, Georgia Institute of Technology. $20,000, 2008-2012.
15. *Isotonic Conditional Random Fields and Local Sentiment Flow.* Purdue Research Foundation. $14,627, *2007-2008.*
16. *The Mood Manifold and Mood Prediction.* Yahoo! $4,000 gift, 2011-2012.
17. *Class of 1969 Teaching Fellowship. Georgia Institute of Technology.* $1,000 gift, 2008-2009.
18. *Purdue Teaching for Tomorrow Award.* Purdue University. $1,000 gift, 2006-2007.

Other Activities Related to Sponsored Research

1. Grant Proposal Reviewer and Panelist for the US Department of Energy
2. Grant Proposal Reviewer and Panelist for the US National Science Foundation
3. Grant Proposal Reviewer for the US National Security Agency

4. Grant Proposal Reviewer for the Maryland Technology Transfer Fund
5. Grant Proposal Reviewer for the Israel Science Foundation
6. Participated in an MSRI Workshop titled *The Mathematics of Visual Analysis* that produced recommendations for a joint call for proposals by the National Science Foundation and the Department of Homeland Security.

F. PROFESSIONAL SERVICE

General Chair and Program Chair

1. AI and Statistics (AISTATS), San Diego CA 2015
2. The Second Amazon Machine Learning Conference, Seattle WA 2014

Program Chair

3. ACM International Conference on Information and Knowledge Management (CIKM), Maui HI USA 2012.
4. Visual Analytics Education Workshop, collocated with the Visual Analytics Community Consortium Meeting. Hyattsville, MD USA 2010.
5. IEEE VisWeek Workshop on Scale and Complexity in Data and Visual Analytics. Salt Lake City, UT USA 2010
6. NIPS Workshop on Statistical Machine Learning for Visual Analytics. Whistler BC Canada 2009.
7. NIPS Workshop on Learning with Orderings. Whistler BC Canada 2009.
8. NIPS Workshop on Algebraic Methods in Machine Learning. Whistler BC Canada 2008.
9. NIPS Mini-Symposium on Algebraic Methods in Machine Learning. Vancouver BC Canada 2008.

Journal Editing

1.	Action Editor, Journal of Machine Learning Research	2013-2018
2.	Associate Editor, STAT	2012-2013
3.	Guest Editor, Data Mining and Knowledge Discovery	2012-2013

Area Chair or Senior Program Committee Member

1. Neural Information Processing Systems (NIPS) 2014
2. International Conference on Machine Learning (ICML) 2014
3. Uncertainty in Artificial Intelligence (UAI) 2013
4. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2013
5. Neural Information Processing systems (NIPS) 2011
6. International Conference on Machine Learning (ICML) 2011
7. Twenty-Fifth Conference on Artificial Intelligence (AAAI) 2011
8. International Joint Conference on Artificial Intelligence (IJCAI) 2011
9. Uncertainty in Artificial Intelligence (UAI) 2009

Program Committee Member

1. AI & Statistics 2010
2. American Statistical Association Conference on Nonparametric Statistics and Statistical Learning 2010
3. Annual ACM SIGIR Conference 2010
4. Best Paper Award Selection Committee, American Statistical Association Section on Statistical Learning and Data Mining 2010.
5. Empirical Methods in Natural Language Processing (EMNLP) 2010
6. International Conference on Computational Linguistics (COLING) 2010
7. International Conference on Machine Learning (ICML) 2010
8. Learning on Cores, Clusters, and Clouds NIPS workshop 2010.
9. AI & Statistics 2009
10. Annual ACM SIGIR Conference 2009
11. International Conference on Machine Learning (ICML) 2009
12. ACM Conference on Information and Knowledge Management (CIKM) 2009
13. SIAM VizMining Workshop 2009
14. The 2nd Midwest Statistics Research Colloquium 2009
15. SIGIR Workshop Learning to Rank for Information Retrieval 2009
16. SIGIR Workshop on Redundancy, Diversity, and Interdependent Document Relevance 2009
17. ACM Conference on Information and Knowledge Management (CIKM) 2008
18. Empirical Methods in Natural Language Processing (EMNLP) 2008
19. International Conference on Machine Learning (ICML) 2008
20. SIGIR Workshop Learning to Rank for Information Retrieval 2008
21. SIGIR Workshop Beyond Relevance Feedback 2008
22. Uncertainty in Artificial Intelligence (UAI) 2008
23. The 23rd AAAI Conference on Artificial Intelligence 2008
24. International Symposium on Artificial Intelligence and Mathematics 2008
25. AI & Statistics 2007
26. Combined European Conference on Machine Learning (ECML) and European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD) 2007
27. International Conference on Data Mining (ICDM) 2007
28. International Conference on Machine Learning (ICML) 2007
29. Joint Conference on Empirical Methods in Natural Language Processing (EMNLP) and Conference on Computational Natural Language Learning (CONLL) 2007
30. SIGIR Workshop Learning to Rank for Information Retrieval 2007
31. Uncertainty in Artificial Intelligence (UAI) 2007
32. International Conference on Data Mining (ICDM) 2006
33. Joint Conference of the Association for Computational Linguistics (ACL) and the International

Conference on Computational Linguistics (COLING) 2006

34. Learning to Compare Examples (NIPS workshop) 2006
35. Uncertainty in Artificial Intelligence (UAI) 2006
36. AI & Statistics 2005
37. Uncertainty in Artificial Intelligence (UAI) 2005

Additional Reviewing for Conference

ACM SIGIR	2007
ACM SIGMOD	2008
AMS-IMS-SIAM Conference on Machine and Statistical Learning	2006
ECML	2002
IEEE Information Visualization Conference	2009
IEEE Symposium on Visual Analytics Science and Technology	2009, 2010
NIPS	2002, 2003, 2004, 2005, 2006, 2007, 2008
NIPS "power reviewer"	2009, 2010
USENIX Workshop on Hot Topics in Parallelism	2011

Journal Reviewing

ACM Journal of Data and Information Quality, Artificial Intelligence Journal, Applied Optics, Computational Statistics and Data Analysis, IEEE Transactions on Evolutionary Computation, IEEE Transactions on Information Theory, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Secure and Dependable Computing, IEEE Transactions on Signal Processing, IEEE Transactions on Visualization and Computer Graphics, Journal of Artificial Intelligence Research, Journal of Machine Learning Research, Journal of the American Statistical Association, Journal of the Optical Society of America, Knowledge and Information Systems, Machine Learning, Mathematical Programming, Neural Computation, Neurocomputing, Optics Letters.

Textbook Reviewing

Cambridge University Press
McGraw Hill
Morgan and Claypool
Springer

G. TEACHING

Evidence of Teaching Effectiveness

1. Received the Class of 1969 Teaching Fellowship and participated in a year long program for enhancing teaching and learning. Georgia Institute of Technology, 2008-2009.
2. Received the 2007 Teaching for Tomorrow Award from the Provost of Purdue University
3. Average instructor rating by students greater than 4.2/5.0

Instructor

1. Data and Visual Analytics, Georgia Institute of Technology Online MS Program (Fall 2016)
2. Data Analysis with R, Minicourse at LinkedIn (Summer 2015)
3. Amazon Machine Learning University (Fall 2013, Winter 2013, Spring 2014, Fall 2014)
4. Computing and Society, Georgia Institute of Technology (Spring 2012)
5. Machine Learning, Georgia Institute of Technology (Fall 2011)
6. Data and Visual Analytics, Georgia Institute of Technology (Spring 2010, Spring 2011)
7. Computational Data Analysis, Georgia Institute of Technology (Fall 2010)
8. Non-Parametric Density Estimation, Georgia Institute of Technology (Spring 2009)
9. Data Structures and Algorithms, Georgia Institute of Technology (Fall 2008)
10. Graphical Models in Machine Learning, Georgia Institute of Technology (Fall 2008)
11. Introduction to Computational Statistics, Purdue University (Fall 2006, Spring 2008)
12. Wavelets and Multiresolution Analysis, Purdue University (Spring 2008)
13. Signals and Systems, Purdue University (Fall 2007)
14. Information Theory, Machine Learning, and Statistics, Purdue University (Spring 2007)
15. Statistical Machine Learning, Purdue University (Spring 2006, Fall 2006)
16. Statistical Theory, Purdue University (Spring 2006)
17. Probabilistic Methods in Electrical and Computer Engineering, Purdue University (Fall 2005)

Teaching Assistant

1. Probability and Statistics for Computer Science, Carnegie Mellon University (Fall 2002)
2. Language and Statistics, Carnegie Mellon University (Spring 2001)
3. Computer Vision, Technion (Spring 2000)
4. Digital Image Processing, Technion (Spring 2000)
5. Introduction to Digital Signal and Image Processing, Technion (Fall 1999)

Other Teaching Activities

1. Participated in a workshop on curriculum design for the new discipline of data and visual analytics.
2. Designed a new core course in computational statistics aimed at statistics graduate students
3. Designed a new PhD qualifying examination in computational statistics
4. Participated in a MSRI workshop on the future of computing in statistics education
5. Attended the following teaching seminars at the Eberly Center for Teaching Excellence at Carnegie Mellon University

    Course and Syllabus Design, Creating a Teaching Portfolio, Overview of Student Cognition, Overview of Student Motivation, Creating Effective Assignments and Exams, Promoting Meaningful and Engaged Knowledge through Service Learning, Encouraging Intellectual Development and Critical Thinking, Conducting Productive and Engaging Discussions.

H. FORMER GRADUATE STUDENTS AND POSTDOC

Former Postdoc

1. Fuxin Li (Assistant Professor at Oregon State University)

Former PhD Students

1. Krishnakumar Balasubramanian (Assistant Professor at UC Davis)
2. Joshua Dillon (Senior Software Engineer at Google)
3. Mohamed Fouad (Software Engineer at Google)
4. Paul Kidwell (VP at Roofstock)
5. Seungyeon Kim (Software Engineer at Google)
6. Joonseok Lee (Software Engineer at Google)
7. Yi Mao (Researcher at Microsoft)
8. Mingxuan Sun (Assistant Professor at Louisiana State University)
9. Yang Zhao (Senior Data Scientist at Google)

Former MS Students

10. Arun Kumar Chithanar (Software Engineer at Indeed)
11. Sanjeet Hajarnis (Software Engineer at Uber)
12. Kaushik Rangadurai (Member of Technical Staff at Passage AI)
13. Yanjun Zhao (High Frequency Trader at Citadel Investment Group)

I. INVITED TALKS AT CONFERENCES, SYMPOSIA, AND WORKSHOPS

Keynote Presentation

1. Being Smart with Art. *The CTO Forum*, San Francisco, February 2017.
2. Being Smart with Art. *MLconf*, San Francisco. November 2016.
3. Visualization and Modeling of Ranked Data. *IEEE Symposium on Large-Data Analysis and Visualization*, Atlanta GA. October 2013

Invited Talks

1. Big Data Skills in Industry. *National Academies Workshop on Training Students to Extract Value from Big Data.* Washington DC. April 2014.
2. Recommendations Systems: Challenges and New Directions. *ACM Conference on Web Search and Data Mining* (WSDM). New York City, NY. February 2014.
3. Machine learning for Information Visualization. Refereed tutorial at *IEEE VisWeek.* Salt Lake City UT. October 2010.
4. Experiences and Conclusions from Teaching Visual Analytics at Georgia Tech. *Visual Analytics Curriculum Workshop, collocated with the annual VAC consortium meeting*. Hyattsville, MD. August 2010 (also workshop organizer).
5. Visualizing Similarities of Search Engines using the Weighted Hoeffding Distance on Permutations. *ASA Conference on Non-Parametric Statistics and Statistical Learning*. Columbus OH. May 2010 (also session organizer).
6. Visualizing Similarities of Search Engines using the Weighted Hoeffding Distance on Permutations. *NIPS Workshop on Learning with Ordering.* Whistler, BC. December 2009 (also workshop organizer)
7. New Directions in Text Visualization. *NIPS Workshop on Statistical Machine Learning for Visual Analytics.* Whistler BC. December 2009 (also workshop organizer).
8. Modeling and Visualization of Missing Preference Data. *The Joint Statistical Meeting*. Washington DC. August 2009.
9. Models on Permutations and Censored Preference Data. *Neural Information Processing Systems Mini-Symposium*. Vancouver BC. December 2008 (also mini-symposium organizer).
10. Riemannian Metrics for Image Spaces. *IEEE Statistical Signal Processing Workshop*. Madison, WI. August 2007.
11. Efficient and Coherent Framework for Aggregating Ranking Data. *39th Symposium on the Interface of Statistics, Computing Science, and Applications*. Philadelphia, PA. May 2007.
12. Expected geometry and statistical translation in text analysis. *SAMSI Workshop on Geometry, Random Matrices, and Statistical Inference.* Research Triangle Park, NC. January 2007.
13. Visualizing Heterogeneous Data. *MSRI Workshop on Mathematics of Visual Analytics*. Berkeley,

CA. October 2006.

14. Conditional Models on the Ranking Poset. *38th Symposium on the Interface of Statistics, Computing Science, and Applications*. Pasadena, CA. May 2006.
15. Information Geometry and Classification of Text Documents. *2nd Symposium on Information Geometry and its Applications*. Tokyo University, Japan. December 2005.
16. A Unifying View of Classification and Ranking. *Workshop on Reductions in Machine Learning*. Toyota Technological Institute. Chicago, IL. September 2003.
17. Conditional Models on the Ranking Poset. *NIPS Beyond Classification and Regression Workshop*. Whistler, BC Canada. December 2002.

J. INVITED TALKS AT COLLOQUIA AND SEMINARS AND PANELS

1. Personalizing the LinkedIn Feed. *Georgia Institute of Technology.* Atlanta GA. September 2015.
2. Local Low-Rank Matrix Factorization. *Twitter*, San Francisco, CA. July 2015.
3. Personalizing the LinkedIn Feed. *Machine Learning Day in Prismatic.* San Francisco, CA. July 2015.
4. Panel Discussion on LinkedIn. *Israeli Executives and Founders Forum*, Mountain View CA May 2015.
5. Local Low-Rank Matrix Factorization. *LinkedIn*, Mountain View, CA. July 2014.
6. Local Low-Rank Matrix Factorization. *Netflix*, Los Gatos, CA. July 2014.
7. Local Low-Rank Matrix Factorization. *Microsoft Research*, Redmond WA. May 2014.
8. Local Low-Rank Matrix Factorization. *University of Washington*, Seattle WA. October 2013.
9. Stochastic m-Estimators and the tradeoff between statistical accuracy and computational complexity. *University of Texas*, Austin TX. September 2013.
10. Unsupervised Supervised Learning: Who Needs Labels Anyway? *University of California,* Santa Cruz CA. October 2012.
11. Unsupervised Supervised Learning: Who Needs Labels Anyway? B*ar-Ilan University,* Tel-Aviv Israel, May 2012.
12. Estimating Probabilities in Recommendation Systems. *Yahoo! Research,* New York City USA, August 2011.
13. Unsupervised Supervised Learning: Who Needs Labels Anyway? *The Hebrew University,* Jerusalem Israel, May 2011.
14. Unsupervised Supervised Learning: Who Needs Labels Anyway? *Technion – Israel Institute of Technology,* Haifa Israel, May 2011.
15. Unsupervised Supervised Learning: Who Needs Labels Anyway? *University of California,* Irvine CA, April 2010.
16. Unsupervised Supervised Learning: Who Needs Labels Anyway? *Google Research,* Mountain View CA, April 2010.
17. Non-Parametric Modeling and Visualization of Partially Ranked Data. School of Computer Science and Engineering, *Hebrew University,* Jerusalem, Israel, June 2008.
18. Non-Parametric Modeling of Partially Ranked Data. Statistics Department, *The Ohio State University*. Columbus, OH, April 2008.
19. The Locally Weighted Bag of Words Representation for Documents. Electrical Engineering Department, *University of Washington,* Seattle, WA, November 2007.
20. The Locally Weighted Bag of Words Representation for Documents. Computer Science Department, *University of California*, Berkeley, CA, November 2007.
21. Beyond k-Anonymity: A Decision Theoretic Approach to Privacy Preservation. *Accenture Technology Labs,* Chicago, IL, November 2007.
22. The Locally Weighted Bag of Words Representation for Documents. Language Technologies Institute, *Carnegie Mellon University,* Pittsburgh, PA, October 2007.
23. Non-Parametric Modeling of Partially Ranked Data. Intelligence Seminar, *Carnegie Mellon*

University, Pittsburgh, PA, October 2007.

24. Sequential Document Visualization. Computer Science Department, *University of Wisconsin,* Madison, WI, August 2007.
25. Visualizing Text using the Lowbow Representation. *Google,* Mountain View, CA, October 2006.
26. Sequential Representation of Documents and Simplicial Curves. *Toyota Technological Institute,* Chicago, IL, May 2006.
27. Information Geometry and Classification of Text Documents. Computer Science and Engineering Department, *Michigan State University,* Lansing MI. November 2005.
28. Riemannian Geometry and Text Classification. Information Retrieval Group, *IBM Research,* Haifa Israel, January 2005.
29. Boosting and Maximum Likelihood for Exponential Models. *Technion – Israel Institute of Technology,* Haifa Israel. July 2003.
30. Boosting and Maximum Likelihood for Exponential Models. *Tel-Aviv University*, Tel-Aviv Israel. July 2003.
31. Boosting and Maximum Likelihood for Exponential Models. School of Computer Science and Engineering, *Hebrew University,* Jerusalem Israel. July 2003.
32. Boosting and Maximum Likelihood for Exponential Models. Machine Learning and Applied Statistics Group, *Microsoft Researc*h, Redmond, WA. August 2002.
33. Conditional Models for Ranked Data. Statistics Department, *The Ohio State University,* Columbus, OH. March 2002.

K. PUBLICATIONS

Theses

[M1] G. Lebanon. *Riemannian Geometry and Statistical Machine Learning*. PhD Dissertation. Carnegie Mellon University, Technical Report CMU-LTI-05-189, 2005.

[M2] G. Lebanon and J. Lafferty. *Boosting and Maximum Likelihood for Exponential Models.* MS Project Report. Carnegie Mellon University, Technical Report CMU-CS-01-144, 2001.

[M3] G. Lebanon. *A Variational Approach to Moiré Pattern Synthesis*. MS Thesis. Technion – Israel Institute of Technology, 2000.

Books and Edited Volumes

[L1] G. Lebanon and S. V. N. Vishwanathan (Editors). *Proceedings of the 18th International AI & Statistics Conference*, JMLR Workshop and Conference Proceedings, Volume 38, 2015.

[L2] G. Lebanon. Riemannian Geometry and Statistical Machine Learning. LAP Lambert Academic Publishing, 2015 (Reprint of 2005 PhD Dissertation).

[L3] G. Lebanon. *Probability.* The Analysis of Data, volume 1, CreateSpace Publishing, 2012.

[L4] X.-W. Chen, G. Lebanon, M. Zaki, and H. Wang (Editors). *Proceedings of the 21st ACM Conference on Information and Knowledge Management*, ACM Press 2012.

Refereed Journal Papers

[J1] K. Balasubramanian, K. Yu, and G. Lebanon. Smooth Sparse Coding via Marginal Regression for Learning Sparse Representations. *Artificial Intelligence* **238**(Sep):83-95, 2016.

[J2] J. Lee, S. Kim, G. Lebanon, Y. Singer, and S. Bengio. LLORMA: Local Low-Rank Matrix Approximation. *Journal of Machine Learning Research* **17**(15):1-24, 2016.

[J3] J. Lee, M. Sun, and G. Lebanon. PREA: Personalized Recommendation Algorithms Toolkit. *Journal of Machine Learning Research* **13**(Sep):2699-2703, 2012.

[J4] M. Sun, G. Lebanon, and P. Kidwell. Estimating Probabilities in Recommendation Systems. *Journal of the Royal Statistical Society C* **61**(3):471-492, 2012.

[J5] K. Balasubramanian, P. Donmez, and G. Lebanon. Unsupervised Supervised Learning II: Margin-Based Classification without Labels. *Journal of Machine Learning Research* **12**(Nov):3119-3145, 2011.

[J6] P. Kidwell, G. Lebanon, and K. Collins-Thompson. Statistical Estimation of Word Acquisition with Application to Readability Prediction. *Journal of the American Statistical Association* **106**(493):21-30, 2011.

[J7] J. V. Dillon and G. Lebanon. Stochastic Composite Likelihood. *Journal of Machine Learning Research* **11**(Oct):2597-2633, 2010.

[J8] P. Donmez, K. Balasubramanian, and G. Lebanon. Unsupervised Supervised Learning I: Estimating Classification and Regression Errors without Labels. *Journal of Machine Learning Research* **11**(April):1323-1351, 2010.

[J9] G. Lebanon, Y. Zhao, and Y. Zhao. Modeling Temporal Text Streams using the Local Multinomial Model. *Electronic Journal of Statistics* **4**:566-584, 2010.

[J10] Y. Mao and G. Lebanon. Generalized Isotonic Conditional Random Fields. *Machine Learning* **77**(2-3):225-248, 2009.

[J11] D. J. Kasik, D. Ebert, G. Lebanon, H. Park, and W. M. Pottenger. Data Transformations and Representations for Information Generation. *Information Visualization* **8**(4):275-285, 2009.

[J12] G. Lebanon, M. Scannapieco, M. R. Fouad, and E. Bertino. Beyond *k*-Anonymity: A Decision Theoretic Framework for Assessing Privacy Risk. *Transactions on Data Privacy* **2**(3):153-183, 2009.

[J13] E. Greenshtein, J. Park, and G. Lebanon. Regularization through Variable Selection and Conditional MLE with Application to Classification in High Dimensions. *Journal of Statistical Planning and Inference* **139**(2):385-395, 2009.

[J14] P. Kidwell, G. Lebanon and W. S. Cleveland. Visualizing Incomplete and Partially Ranked Data. *IEEE Transactions on Visualization and Computer Graphics* (Proc. INFOVIS) **14**(6):1356-1363, 2008.

[J15] G. Lebanon and Y. Mao. Non-Parametric Modeling of Partially Ranked Data. *Journal of Machine Learning Research* **9**(Oct): 2401-2429, 2008.

[J16] G. Lebanon, Y. Mao, and J. Dillon. The Locally Weighted Bag of Words Framework for Document Representation. *Journal of Machine Learning Research* **8**(Oct):2405-2441, 2007.

[J17] Y. Mao, J. Dillon, and G. Lebanon. Sequential Document Visualization. *IEEE Transactions on Visualization and Computer Graphics* (Proc. INFOVIS) **13**(6):1208-1215, 2007.

[J18] G. Lebanon. Metric Learning for Text Documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(4):497-508, 2006.

[J19] G. Lebanon. Axiomatic Geometry of Conditional Models. *IEEE Transactions on Information Theory* **51**(4):1283-1294, 2005.

[J20] J. Lafferty and G. Lebanon. Diffusion Kernels on Statistical Manifolds. *Journal of Machine Learning Research* **6**(Jan):129-163, 2005.

[J21] G. Lebanon and A. Bruckstein. Variational Approach to Moiré Pattern Synthesis. *Journal of the Optical Society of America A* **18**(6):1371-1382, 2001.

Refereed Book Chapters and Papers in Special Volumes

[B1] P. Kidwell and G. Lebanon. Kernel Smoothing for Preference Data using Generating Functions. In M. Vienna and H. Wynn (eds). *Algebraic Methods in Statistics and Probability II.* Contemporary Mathematics Series, The American Mathematical Society 2010.

[B2] G. Lebanon. Axiomatic Geometries for Text Documents. In P. Gibilisco, E. Riccomagno, M.-P. Rogantin, and H. P. Wynn (eds). *Algebraic and Geometric Methods in Statistics.* Cambridge University Press 2009.

[B3] G. Lebanon and A. Bruckstein. On Designing Moiré Patterns. In V. Cantoni, M. Marinaro, and A. Petrosino (eds), *Visual Attention Mechanism*, pages 205-219, Springer 2002.

Refereed Conference Papers

[C1] D. Agarwal, B.-C. Chen, Q. He, Z. Hua, G. Lebanon, Y. Ma, P. Shivaswamy, H.-P. Tseng, J. Yang and L. Zhang. Personalizing LinkedIn Feed. *Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (KDD) 2015.

[C2] S. Kim, J. Lee, G. Lebanon and H. Park. Estimating Temporal Dynamics of Human Emotions. *Proceedings of the 29th AAAI Conference on Artificial Intelligence* (AAAI) 2015.

[C3] S. Kim, J. Lee, G. Lebanon and H. Park. Local Context Sparse Coding. *Proceedings of the 29th AAAI Conference on Artificial Intelligence* (AAAI) 2015.

[C4] J. Lee, S. Bengio, S. Kim, G. Lebanon and Y. Singer. Local Collaborative Ranking. *Proceedings of the 23rd International World Wide Web Conference* (WWW), 2014 *(Best Student Paper Award).*

[C5] K. Balasubramanian, K. Yu, and G. Lebanon. Smooth Sparse Coding via Marginal Regression for Learning Sparse Representations. *Proceedings of the 30th International Conference on Machine Learning* (ICML) 2013. *(Best Paper Runner-Up Award.)*

[C6] J. Lee, S. Kim, G. Lebanon, and Y. Singer. Local Low-Rank Matrix Approximation. *Proceedings of the 30th International Conference on Machine Learning* (ICML) 2013.

[C7] S. Kim, F. Li, G. Lebanon, and I. Essa. Beyond Sentiment: The Manifold of Human Emotions. *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics* (AISTATS) 2013.

[C8] K. Balasubramanian. B. Sriperumbudur, and G. Lebanon. Ultrahigh dimensional feature screening based on RKHS embeddings. *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics* (AISTATS) 2013.

[C9] M. Sun, F. Li, J. Lee, K. Zhou, G. Lebanon, and H. Zha. Learning Multiple Question Decision Trees for Cold-Start Recommendation. *Proceedings of the 6th ACM International Conference on Web Search and Data Mining* (WSDM) 2013.

[C10] F. Li, J. Carreira, G. Lebanon, and C. Sminchisescu. Composite Statistical Inference for Semantic Segmentation. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition* (CVPR) 2013.

[C11] L. Weiss, E. Briscoe, H. Hayes, O. Kemenova, S. Harbert, F. Li, G. Lebanon, C. Stewart, D. Miller Steiger, and D. Foy. A Comparative Study of Social Media and Traditional Polling in the Egyptian Uprising of 2011. *Proceedings of the 6th International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction* (SBP). Lecture Notes in Computer Science, volume 7812, pages 303-310, Springer 2013.

[C12] J. Lee, M. Sun, S. Kim, and G. Lebanon. Automatic Feature Induction for Stagewise Collaborative Filtering. *Advances in Neural Information Processing Systems* (NIPS) 25, 2012.

[C13] K. Balasubramanian and G. Lebanon. The Landmark Selection Method for Multiple Output

Prediction. *Proceedings of the 29th International Conference on Machine Learning* (ICML) 2012.

[C14] L. Li, G. Lebanon, and H. Park. Fast Bregman Divergence NMF using Taylor Expansion and Coordinate Descent. *Proceedings of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (KDD) 2012.

[C15] F. Li, G. Lebanon, and C. Sminchisescu. Chebyshev Approximations to the Histogram Chi-Square Kernel. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition* (CVPR) 2012.

[C16] M. Sun, G. Lebanon, and P. Kidwell. Estimating Probabilities in Recommendation Systems. *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics* (AISTATS) JMLR W&CP 2011.

[C17] K. Balasubramanian, P. Donmez, and G. Lebanon. Unsupervised Supervised Learning II: Margin-Based Classification without Labels. *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics* (AISTATS) JMLR W&CP 2011.

[C18] J. V. Dillon, K. Balasubramanian, and G. Lebanon. Asymptotic Analysis of Generative Semi-Supervised Learning. *Proceedings of the 27th International Conference on Machine Learning* (ICML) 2010.

[C19] M. Sun, G. Lebanon, and K. Collins-Thompson. Visualizing Search Strategies in the World Wide Web. *Proceedings of the 19th International World Wide Web Conference* (WWW) 2010.

[C20] S. Kim and G. Lebanon. Local Space-Time Smoothing for Version Controlled Documents. *Proceedings of the International Conference on Computational Linguistics* (COLING) 2010.

[C21] Y. Mao, K. Balasubramanian, and G. Lebanon. Linguistic Geometries for Unsupervised Dimensionality Reduction. *Proceedings of the International Conference on Computational Linguistics* (COLING) 2010.

[C22] Y. Mao and G. Lebanon. Domain Knowledge Uncertainty and Probabilistic Parameter Constraints. *Proceedings of the 25rd Conference on Uncertainty in Artificial Intelligence* (UAI) AUAI Press 2009.

[C23] P. Kidwell, G. Lebanon, and K. Collins-Thompson. Statistical Estimation of Word Acquisition with Application to Readability Prediction. *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (EMNLP) 2009.

[C24] J. V. Dillon and G. Lebanon. Statistical and Computational Tradeoffs in Stochastic Composite Likelihood. *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics.* (AISTATS) JMLR W&CP 5:129-139, 2009.

[C25] G. Lebanon and Y. Zhao. Local Likelihood Modeling of the Concept Drift Phenomenon. *Proceedings of the 25th International Conference in Machine Learning* (ICML) 2008.

[C26] G. Lebanon and Y. Mao. Non-Parametric Modeling of Partially Ranked Data. *Advances in Neural Information Processing Systems* 20 (NIPS). The MIT Press 2008.

[C27] G. M. Howard, S. Bagchi, and G. Lebanon. Determining placement of intrusion detectors for a distributed application through Bayesian network modeling. *Proceedings of the 11th International Symposium on Recent Advances in Intrusion Detection* (RAID). Lecture Notes in Computer Science,

volume 5230, pages 271-290. Springer, 2008.

[C28] Y. Mao and G. Lebanon. Isotonic Conditional Random Fields and Local Sentiment Flow. *Advances in Neural Information Processing Systems* 19 (NIPS), pages 961-968. The MIT Press 2007.

[C29] J. V. Dillon, Y. Mao, G. Lebanon and J. Zhang. Statistical Translation, Heat Kernels, and Expected Distances. *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence* (UAI) pages 93-100. AUAI Press 2007.

[C30] G. Lebanon. Sequential Document Representations and Simplicial Curves. *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence* (UAI) pages 273-280. AUAI Press 2006.

[C31] G. Lebanon, M. Scannapieco, M. R. Fouad, and E. Bertino. Beyond *k*-Anonymity: A Decision Theoretic Framework for Assessing Privacy Risk. *Privacy in statistical databases.* Lecture Notes in Computer Science, volume 4302, pages 217-232. Springer, 2006.

[C32] G. Lebanon. Information Geometry, the Embedding Principle, and Document Classification. *Proceedings of the 2nd International Symposium on Information Geometry and its Applications*, pages 101-108, 2005.

[C33] G. Lebanon. An Extended Čencov-Campbell Characterization of Conditional Information Geometry. *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* (UAI) AUAI Press 2004.

[C34] G. Lebanon and J. Lafferty. Hyperplane Margin Classifiers on the Multinomial Manifold. *Proceedings of the 21st International Conference on Machine Learning* (ICML). Morgan Kaufmann Press 2004.

[C35] G. Lebanon and J. Lafferty. Conditional Models on the Ranking Poset. *Advances in Neural Information Processing Systems* 15 (NIPS), pages 431-438. The MIT Press 2003.

[C36] J. Lafferty and G. Lebanon. Information Diffusion Kernels. *Advances in Neural Information Processing Systems* 15 (NIPS), pages 391-399. The MIT Press 2003.

[C37] G. Lebanon. Learning Riemannian Metrics. *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence* (UAI). AUAI Press 2003.

[C38] G. Lebanon and J. Lafferty. Cranking: Combining Rankings using Conditional Probability Models on Permutations. *Proceedings of the 19th International Conference on Machine Learning* (ICML). Morgan Kaufmann Press 2002.

[C39] G. Lebanon and J. Lafferty. Boosting and Maximum Likelihood for Exponential Models. *Advances in Neural Information Processing Systems* 14 (NIPS), pages 447-454. The MIT Press 2002.

[C40] G. Lebanon and A. Bruckstein. Designing Moiré Patterns. *Energy Minimization Methods in Computer Vision and Pattern Recognition* (EMMCVPR). Lecture Notes in Computer Science, Volume 2134, pages 185-201. Springer 2001.

Refereed Workshop Papers and Unrefereed Technical Reports

[W1] F. Li, J. Carreira, G. Lebanon, and C. Sminchisescu. CSI: Composite Statistical Inference for Semantic Segmentation. *ICML 2013 Workshop on Inferning: Interactions between Inference and*

Learning, 2013.

[W2] M. Sun, G. Lebanon, and K. Collins-Thompson. Visualizing Spatial Proximity of Search Algorithms. *NIPS Workshop on Learning with Ordering*, 2009.

[W3] M. Fouad, G. Lebanon, and E. Bertino. ARUBA: A Risk-Utility-Based Algorithm for Data Disclosure. *Proceedings of the 5th VLDB Workshop on Secure Data Management* (SDM). Lecture Notes in Computer Science, Springer 2008.

[W4] A. Kamra, E. Bertino, and G. Lebanon. Mechanisms for Database Intrusion Detection and Response. *Proceedings of the 2nd SIGMOD PhD workshop on Innovative database research*, pages 31-36, ACM Press 2008.

[W5] J. V. Dillon, Y. Mao, G. Lebanon, and J. Zhang. Statistical Translation, Heat Kernels, and Expected Distances. *NIPS workshop on Learning to Compare Examples*, 2006.

[W6] Y. Mao and G. Lebanon. Sequential Models for Sentiment Prediction. *ICML workshop on Learning in Structured Output Spaces*, 2006.

[W7] Z. Zhang, M. Gupta, S. Yang, G. Lebanon, Y. C. Hu, and S. Midkiff. *Extracting Source Level Program Similarities from Dynamic Behavior*. Technical Report TR-ECE-08-08, Purdue University, 2008.

[W8] G. Lebanon. *Computing the Volume Element of a Family of Transformations on the Multinomial Simplex.* Technical Report CMU-CS-03-145, Carnegie Mellon University 2003.