Determinantal Point Processes for Machine Learning

Determinantal Point Processes for Machine Learning

Alex Kulesza

University of Michigan
USA
kulesza@umich.edu

Ben Taskar

University of Pennsylvania
USA
taskar@cis.upenn.edu

the essence of knowledge

Boston – Delft

Foundations and Trends® in Machine Learning

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is A. Kulesza and B. Taskar, Determinantal Point Processes for Machine Learning, Foundation and Trends® in Machine Learning, vol 5, nos 2–3, pp 123–286, 2012.

ISBN: 978-1-60198-628-3

Foundations and Trends® in Machine Learning

Volume 5 Issues 2–3, 2012

Editorial Scope

Foundations and Trends® in Machine Learning will publish survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends® in Machine Learning, 2012, Volume 5, 4 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Foundations and Trends® in
Machine Learning
Vol. 5, Nos. 2–3 (2012) 123–286

DOI: 10.1561/2200000044

Determinantal Point Processes for Machine Learning

Alex Kulesza[1] **and Ben Taskar**[2]

[1] *University of Michigan, USA, kulesza@umich.edu*
[2] *University of Pennsylvania, USA, taskar@cis.upenn.edu*

Abstract

Determinantal point processes (DPPs) are elegant probabilistic models of repulsion that arise in quantum physics and random matrix theory. In contrast to traditional structured models like Markov random fields, which become intractable and hard to approximate in the presence of negative correlations, DPPs offer efficient and exact algorithms for sampling, marginalization, conditioning, and other inference tasks. We provide a gentle introduction to DPPs, focusing on the intuitions, algorithms, and extensions that are most relevant to the machine learning community, and show how DPPs can be applied to real-world applications like finding diverse sets of high-quality search results, building informative summaries by selecting diverse sentences from documents, modeling nonoverlapping human poses in images or video, and automatically building timelines of important news stories.

Contents

1

Introduction

Probabilistic modeling and learning techniques have become indispensable tools for analyzing data, discovering patterns, and making predictions in a variety of real-world settings. In recent years, the widespread availability of both data and processing capacity has led to new applications and methods involving more complex, structured output spaces, where the goal is to simultaneously make a large number of interrelated decisions. Unfortunately, the introduction of structure typically involves a combinatorial explosion of output possibilities, making inference computationally impractical without further assumptions.

A popular compromise is to employ graphical models, which are tractable when the graph encoding local interactions between variables is a tree. For loopy graphs, inference can often be approximated in the special case when the interactions between variables are positive and neighboring nodes tend to have the same labels. However, dealing with global, negative interactions in graphical models remain intractable, and heuristic methods often fail in practice.

Determinantal point processes (DPPs) offer a promising and complementary approach. Arising in quantum physics and random matrix theory, DPPs are elegant probabilistic models of global, negative correlations, and offer efficient algorithms for sampling, marginalization, conditioning, and other inference tasks. While they have been studied extensively by mathematicians, giving rise to a deep and beautiful theory, DPPs are relatively new in machine learning. We aim to provide a comprehensible introduction to DPPs, focusing on the intuitions, algorithms, and extensions that are most relevant to our community.

1.1 Diversity

A DPP is a distribution over subsets of a fixed ground set, for instance, sets of search results selected from a large database. Equivalently, a DPP over a ground set of N items can be seen as modeling a binary characteristic vector of length N. The essential characteristic of a DPP is that these binary variables are negatively correlated; that is, the inclusion of one item makes the inclusion of other items less likely. The strengths of these negative correlations are derived from a kernel matrix that defines a global measure of similarity between pairs of items, so that more similar items are less likely to co-occur. As a result, DPPs assign higher probability to sets of items that are *diverse*; for example, a DPP will prefer search results that cover multiple distinct aspects of a user's query, rather than focusing on the most popular or salient one.

This focus on diversity places DPPs alongside a number of recently developed techniques for working with diverse sets, particularly in the information retrieval community [23, 26, 121, 122, 140, 158, 159]. However, unlike these methods, DPPs are fully probabilistic, opening the door to a wider variety of potential applications, without compromising algorithmic tractability.

The general concept of diversity can take on a number of forms depending on context and application. Including multiple kinds of search results might be seen as *covering* or *summarizing* relevant interpretations of the query or its associated topics; see Figure 1.1. Alternatively, items inhabiting a continuous space may exhibit diversity as a result of *repulsion*, as in Figure 1.2. In fact, certain repulsive quantum particles are known to be precisely described by a DPP; however, a DPP can also serve as a model for general repulsive phenomena, such

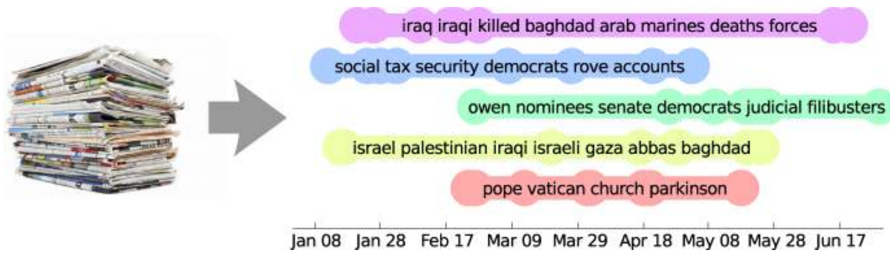

Fig. 1.1 Diversity is used to generate a set of summary timelines describing the most important events from a large news corpus.

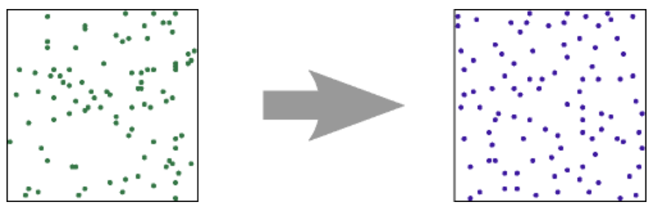

Fig. 1.2 On the left, points are sampled randomly; on the right, repulsion between points leads to the selection of a diverse set of locations.

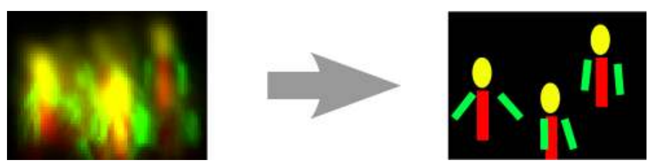

Fig. 1.3 On the left, the output of a human pose detector is noisy and uncertain; on the right, applying diversity as a filter leads to a clean, separated set of predictions.

as the locations of trees in a forest, which appear diverse due to physical and resource constraints. Finally, diversity can be used as a *filtering* prior when multiple selections must be based on a single detector or scoring metric. For instance, in Figure 1.3 a weak pose detector favors large clusters of poses that are nearly identical, but filtering through a DPP ensures that the final predictions are well separated.

Throughout this survey we demonstrate applications for DPPs in a variety of settings, including:

- The DUC 2003/2004 text summarization task, where we form extractive summaries of news articles by choosing diverse subsets of sentences (Section 4.2.1);

- An image search task, where we model human judgments of diversity for image sets returned by Google Image Search (Section 5.3),
- A multiple pose estimation task, where we improve the detection of human poses in images from television shows by incorporating a bias toward nonoverlapping predictions (Section 6.4), and
- A news threading task, where we automatically extract timelines of important news stories from a large corpus by balancing intra-timeline coherence with inter-timeline diversity (Section 6.6.4).

1.2 Outline

In this monograph we present general mathematical background on DPPs along with a range of modeling extensions, efficient algorithms, and theoretical results that aim to enable practical modeling and learning. The material is organized as follows.

Section 2: Determinantal Point Processes. We begin with an introduction to determinantal point processes tailored to the interests of the machine learning community. We focus on discrete DPPs, emphasizing intuitions and including new, simplified proofs for some theoretical results. We provide descriptions of known efficient inference algorithms and characterize their computational properties.

Section 3: Representation and Algorithms. We describe a decomposition of the DPP that makes explicit its fundamental trade-off between quality and diversity. We compare the expressive power of DPPs and MRFs, characterizing the trade-offs in terms of modeling power and computational efficiency. We also introduce a dual representation for DPPs, showing how it can be used to perform efficient inference over large ground sets. When the data are high-dimensional and dual inference is still too slow, we show that random projections can be used to maintain a provably close approximation to the original model while greatly reducing computational requirements.

Section 4: Learning. We derive an efficient algorithm for learning the parameters of a quality model when the diversity model is held fixed. We employ this learning algorithm to perform extractive summarization of news text.

Section 5: k-DPPs. We present an extension of DPPs that allows for explicit control over the number of items selected by the model. We show not only that this extension solves an important practical problem, but also that it increases expressive power: a k-DPP can capture distributions that a standard DPP cannot. The extension to k-DPPs necessitates new algorithms for efficient inference based on recursions for the elementary symmetric polynomials. We validate the new model experimentally on an image search task.

Section 6: Structured DPPs. We extend DPPs to model diverse sets of structured items, such as sequences or trees, where there are combinatorially many possible configurations. In this setting the number of possible subsets is doubly exponential, presenting a daunting computational challenge. However, we show that a factorization of the quality and diversity models together with the dual representation for DPPs makes efficient inference possible using second-order message passing. We demonstrate structured DPPs on a toy geographical paths problem, a still-image multiple pose estimation task, and two high-dimensional text threading tasks.

References

[1] A. Abdelbar and S. Hedetniemi, "Approximating maps for belief networks is NP-hard and other theorems," *Artificial Intelligence*, vol. 102, no. 1, pp. 21–38, 1998.

[2] J. Allan, R. Gupta, and V. Khandelwal, "Temporal Summaries of New Topics," in *Proceedings of the Annual Conference on Research and Development in Information Retrieval (SIGIR)*, 2001.

[3] A. J. Baddeley and M. N. M. Van Lieshout, "Area-interaction point processes," *Annals of the Institute of Statistical Mathematics*, vol. 47, no. 4, pp. 601–619, 1995.

[4] F. B. Baker and M. R. Harwell, "Computing elementary symmetric functions and their derivatives: A didactic," *Applied Psychological Measurement*, vol. 20, no. 2, p. 169, 1996.

[5] A. I. Barvinok, "Computational complexity of immanents and representations of the full linear group," *Functional Analysis and Its Applications*, vol. 24, no. 2, pp. 144–145, 1990.

[6] K. Berthelsen and J. Møller, "Bayesian analysis of Markov point processes," *Case Studies in Spatial Point Process Modeling*, pp. 85–97, 2006.

[7] D. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.

[8] J. Besag, "Some methods of statistical analysis for spatial data," *Bulletin of the International Statistical Institute*, vol. 47, no. 2, pp. 77–92, 1977.

[9] J. Besag and P. Green, "Spatial statistics and Bayesian computation," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 25–37, 1993.

[10] J. Besag, R. Milne, and S. Zachary, "Point process limits of lattice processes," *Journal of Applied Probability*, pp. 210–216, 1982.

[11] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 113–120, 2006.

[12] A. Borodin, "Determinantal point processes," URL http://arxiv.org/abs/0911.1153, 2009.

[13] A. Borodin, P. Diaconis, and J. Fulman, "On adding a list of numbers (and other one-dependent determinantal processes)," *American Mathematical Society*, vol. 47, no. 4, pp. 639–670, 2010.

[14] A. Borodin and G. Olshanski, "Distributions on partitions, point processes, and the hypergeometric kernel," *Communications in Mathematical Physics*, vol. 211, no. 2, pp. 335–358, 2000.

[15] A. Borodin and E. Rains, "Eynard-mehta theorem, schur process, and their pfaffian analogs," *Journal of Statistical Physics*, vol. 121, pp. 291–317, 2005. ISSN 0022-4715. 10.1007/s10955-005-7583-z.

[16] A. Borodin and A. Soshnikov, "Janossy densities. i. determinantal ensembles," *Journal of Statistical Physics*, vol. 113, no. 3, pp. 595–610, 2003.

[17] E. Boros and P. L. Hammer, "Pseudo-Boolean optimization," *Discrete Applied Mathematics*, vol. 123, no. 1-3, pp. 155–225, 2002.

[18] P. Bratley and B. Fox, "Algorithm 659: Implementing Sobol's quasirandom sequence generator," *ACM Transactions on Mathematical Software (TOMS)*, no. 1, pp. 88–100, 1988.

[19] J. L. Brylinski and R. Brylinski, "Complexity and completeness of immanants," Arxiv preprint cs/0301024, 2003.

[20] P. Bürgisser, "The computational complexity of immanants," *SIAM Journal on Computing*, vol. 30, p. 1023, 2000.

[21] R. Burton and R. Pemantle, "Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer-impedances," *The Annals of Probability*, pp. 1329–1371, 1993.

[22] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 679–698, 1986.

[23] J. Carbonell and J. Goldstein, "The use of MMR, diversity-based reranking for reordering documents and producing summaries," in *Proceedings of the Annual Conference on Research and Development in Information Retrieval (SIGIR)*, 1998.

[24] A. Cayley, "On the theory of determinants," *Transaction of the Cambridge Philosophical Society*, vol. 8, no. 1843, pp. 1–16, 1843.

[25] C. Chekuri, J. Vondrák, and R. Zenklusen, "Submodular function maximization via the multilinear relaxation and contention resolution schemes," Arxiv preprint arXiv:1105.4593, 2011.

[26] H. Chen and D. R. Karger, "Less is more: Probabilistic models for retrieving fewer relevant documents," in *Proceedings of the Annual Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 429–436, 2006.

[27] H. Chieu and Y. Lee, "Query based event extraction along a timeline," in *Proceedings of the Annual Conference on Research and Development in Information Retrieval (SIGIR)*, 2004.

[28] A. Çivril and M. Magdon-Ismail, "On selecting a maximum volume sub-matrix of a matrix and related problems," *Theoretical Computer Science*, vol. 410, no. 47–49, pp. 4801–4811, 2009.

[29] J. M. Conroy, J. Schlesinger, J. Goldstein, and D. P. Oleary, "Left-brain/right-brain multi-document summarization," in *Proceedings of the Document Understanding Conference (DUC)*, 2004.

[30] G. F. Cooper, "The computational complexity of probabilistic inference using Bayesian belief networks," *Artificial Intelligence*, vol. 42, no. 2–3, pp. 393–405, 1990.

[31] P. Dagum and M. Luby, "Approximating probabilistic inference in Bayesian belief networks is NP-hard," *Artificial Intelligence*, vol. 60, no. 1, pp. 141–153, 1993.

[32] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer, 2003.

[33] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: General Theory and Structure*, vol. 2. Springer Verlag, 2008.

[34] H. T. Dang, "Overview of DUC 2005," in *Procedings of the Document Understanding Conference (DUC)*, 2005.

[35] A. Deshpande and L. Rademacher, "Efficient volume sampling for row/column subset selection," in *2010 IEEE Annual Symposium on Foundations of Computer Science*, pp. 329–338, 2010.

[36] P. Diaconis, "Patterns in eigenvalues: The 70th Josiah Willard Gibbs lecture," *Bulletin-American Mathematical Society*, vol. 40, no. 2, pp. 155–178, 2003.

[37] P. Diaconis and S. N. Evans, "Immanants and finite point processes," *Journal of Combinatorial Theory, Series A*, vol. 91, no. 1-2, pp. 305–321, 2000.

[38] P. J. Diggle, T. Fiksel, P. Grabarnik, Y. Ogata, D. Stoyan, and M. Tanemura, "On parameter estimation for pairwise interaction point processes," *International Statistical Review/Revue Internationale de Statistique*, pp. 99–117, 1994.

[39] P. J. Diggle, D. J. Gates, and A. Stibbard, "A nonparametric estimator for pairwise-interaction point processes," *Biometrika*, vol. 74, no. 4, pp. 763–770, 1987.

[40] F. J. Dyson, "Statistical theory of the energy levels of complex systems. i," *Journal of Mathematical Physics*, vol. 3, pp. 140–156, 1962.

[41] F. J. Dyson, "Statistical theory of the energy levels of complex systems. ii," *Journal of Mathematical Physics*, vol. 3, no. 157–165, 1962.

[42] F. J. Dyson, "Statistical theory of the energy levels of complex systems. iii," *Journal of Mathematical Physics*, vol. 3, pp. 166–175, 1962.

[43] G. Erkan and D. R. Radev, "LexRank: Graph-based lexical centrality as salience in text summarization," *Journal of Artificial Intelligence Research*, no. 1, pp. 457–479, 2004. ISSN 1076-9757.

[44] S. N. Evans and A. Gottlieb, "Hyperdeterminantal point processes," *Metrika*, vol. 69, no. 2, pp. 85–99, 2009.

[45] J. Feder, "Random sequential adsorption," *Journal of Theoretical Biology*, vol. 87, no. 2, pp. 237–254, 1980.

[46] U. Feige, "A threshold of ln n for approximating set cover," *Journal of the ACM (JACM)*, vol. 45, no. 4, pp. 634–652, 1998.

[47] U. Feige, V. S. Mirrokni, and J. Vondrak, "Maximizing non-monotone submodular functions," in *Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pp. 461–471, 2007.

[48] M. Feldman, J. Naor, and R. Schwartz, "Nonmonotone submodular maximization via a structural continuous greedy algorithm," *Automata, Languages and Programming*, pp. 342–353, 2011.

[49] M. Feldman, J. S. Naor, and R. Schwartz, "A unified continuous greedy algorithm for submodular maximization," in *IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 570–579, 2011.

[50] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005. ISSN 0920-5691.

[51] L. Finegold and J. T. Donnell, "Maximum density of random placing of membrane particles," *Nature*, 1979.

[52] M. A. Fischler and R. A. Elschlager, "The representation and matching of pictorial structures," *IEEE Transactions on Computers*, vol. 100, no. 22, 1973.

[53] M. L. Fisher, G. L. Nemhauser, and L. A. Wolsey, "An analysis of approximations for maximizing submodular set functions — II," *Polyhedral Combinatorics*, pp. 73–87, 1978.

[54] I. M. Gel'fand, *Lectures on Linear Algebra*. Dover, 1989. ISBN 0486660826.

[55] P. E. Genest, G. Lapalme, and M. Yousfi-Monod, "Hextac: The creation of a manual extractive run," in *Proceedings of the Text Analysis Conference (TAC)*, Gaithersburg, Maryland, USA, 2010.

[56] S. O. Gharan and J. Vondrák, "Submodular maximization by simulated annealing," in *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1098–1116, 2011.

[57] J. Gillenwater, A. Kulesza, and B. Taskar, "Discovering diverse and salient threads in document collections," in *Proceedings of the 2012 Conference on Empirical Methods in Machine Learning*, 2012.

[58] J. Ginibre, "Statistical ensembles of complex, quaternion, and real matrices," *Journal of Mathematical Physics*, vol. 6, p. 440, 1965.

[59] V. Goel and W. Byrne, "Minimum Bayes-risk automatic speech recognition," *Computer Speech & Language*, vol. 14, no. 2, pp. 115–135, 2000.

[60] D. Graff and C. Cieri, "English Gigaword," 2009.

[61] G. R. Grimmett, "A theorem about random fields," *Bulletin of the London Mathematical Society*, vol. 5, no. 13, pp. 81–84, 1973.

[62] R. Grone and R. Merris, "An algorithm for the second immanant," *Mathematics of Comp*, vol. 43, pp. 589–591, 1984.

[63] O. Häggström, M. C. N. M. Van Lieshout, and J. Møller, "Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes," *Bernoulli*, vol. 5, no. 4, pp. 641–658, 1999.

[64] J. H. Halton, "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numerische Mathematik*, no. 1, pp. 84–90, 1960.

[65] W. Hartmann, “On the complexity of immanants,” *Linear and Multilinear Algebra*, vol. 18, no. 2, pp. 127–140, 1985.

[66] E. L. Hinrichsen, J. Feder, and T. Jøssang, “Geometry of random sequential adsorption,” *Journal of Statistical Physics*, vol. 44, no. 5, pp. 793–827, 1986.

[67] E. Hlawka, “Funktionen von beschränkter variatiou in der theorie der gleichverteilung,” *Annali di Matematica Pura ed Applicata*, vol. 54, no. 1, pp. 325–333, 1961.

[68] J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág, “Determinantal processes and independence,” *Probability Surveys*, vol. 3, pp. 206–229, 2006.

[69] M. L. Huber and R. L. Wolpert, “Likelihood-based inference for matérn type-iii repulsive point processes,” *Advances in Applied Probability*, vol. 41, no. 4, pp. 958–977, 2009.

[70] H. Ishikawa, “Exact optimization for Markov random fields with convex priors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1333–1336, 2003.

[71] J. L. Jensen and J. Moller, “Pseudolikelihood for exponential family models of spatial point processes,” *The Annals of Applied Probability*, vol. 1, no. 3, pp. 445–461, 1991.

[72] K. Johansson, “Non-intersecting paths, random tilings and random matrices,” *Probability Theory and Related Fields*, vol. 123, no. 2, pp. 225–280, 2002.

[73] K. Johansson, “Determinantal processes with number variance saturation,” *Communications in Mathematical Physics*, vol. 252, no. 1, pp. 111–148, 2004.

[74] K. Johansson, “The arctic circle boundary and the airy process,” *The Annals of Probability*, vol. 33, no. 1, pp. 1–30, 2005.

[75] K. Johansson, “Random matrices and determinantal processes,” Arxiv preprint math-ph/0510038, 2005.

[76] W. B. Johnson and J. Lindenstrauss, “Extensions of Lipschitz mappings into a Hilbert space,” *Contemporary Mathematics*, vol. 26, no. 189–206, pp. 1–1, 1984.

[77] C. W. Ko, J. Lee, and M. Queyranne, “An exact algorithm for maximum entropy sampling,” *Operations Research*, vol. 43, no. 4, pp. 684–691, 1995. ISSN 0030-364X.

[78] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

[79] V. Kolmogorov and R. Zabih, “What energy functions can be minimized via graph cuts?,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 147–159, 2004.

[80] A. Krause and C. Guestrin, “A note on the budgeted maximization of submodular functions,” Technical Report No. CMU-CALD, vol. 5, p. 103, 2005.

[81] A. Kulesza, J. Gillenwater, and B. Taskar, “Near-optimal map inference for determinantal point processes,” in *Proceedings of the Neural Information Processing Systems*, 2012.

[82] A. Kulesza and F. Pereira, “Structured learning with approximate inference,” *Advances in neural information processing systems*, vol. 20, pp. 785–792, 2008.

[83] A. Kulesza and B. Taskar, “Structured determinantal point processes,” in *Proceedings of the Neural Information Processing Systems*, 2010.

[84] A. Kulesza and B. Taskar, "k-DPPs: Fixed-size determinantal point processes," in *Proceedings of the International Conference on Machine Learning*, 2011.

[85] A. Kulesza and B. Taskar, "Learning determinantal point processes," in *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2011.

[86] S. Kumar and W. Byrne, "Minimum Bayes-risk word alignments of bilingual texts," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*-vol. 10, pp. 140–147, Association for Computational Linguistics, 2002.

[87] S. Kumar and W. Byrne, "Minimum Bayes-risk decoding for statistical machine translation," in *Proceedings of HLT-NAACL*, pp. 169–176, 2004.

[88] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the International Conference on Machine Learning*, pp. 282–289, Morgan Kaufmann Publishers Inc., 2001.

[89] S. L. Lauritzen and D. J. Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 157–224, 1988.

[90] J. Lee, V. S. Mirrokni, V. Nagarajan, and M. Sviridenko, "Non-monotone submodular maximization under matroid and knapsack constraints," in *Proceedings of the Annual ACM Symposium on Theory of Computing*, pp. 323–332, 2009.

[91] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD)*, 2009.

[92] Z. Li and J. Eisner, "First-and second-order expectation semirings with applications to minimum-risk training on translation forests," in *Proceedings of the Conference on Empirical Methods in National Language Processing (EMNLP)*, 2009.

[93] C. Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Proceedings of the Workshop on Text Summarization Branches out (WAS 2004)*, pp. 25–26, 2004.

[94] H. Lin and J. Bilmes, "Multi-document summarization via budgeted maximization of submodular functions," in *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics — Human Language Technologies (NAACL/HLT)*, 2010.

[95] H. Lin and J. Bilmes, "Learning mixtures of submodular shells with application to document summarization," in *Uncertainty in Artificial Intelligence (UAI)*, Catalina Island, USA, AUAI, July 2012.

[96] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1999.

[97] R. Lyons, "Determinantal probability measures," *Publications Mathématiques de l'IHÉS*, vol. 98, no. 1, pp. 167–212, 2003.

[98] O. Macchi, "The coincidence approach to stochastic point processes," *Advances in Applied Probability*, vol. 7, no. 1, pp. 83–122, 1975.

[99] A. Magen and A. Zouzias, "Near optimal dimensionality reductions that preserve volumes," *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pp. 523–534, 2008.

[100] B. Matérn, "Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations," *Meddelanden fran statens Skogsforskningsinstitut*, vol. 49, no. 5, 1960.

[101] B. Matérn, *Spatial Variation*. Springer-Verlag, 1986.

[102] A. McCallum, K. Nigam, J. Rennie, and K. Seymore, "Automating the construction of internet portals with machine learning," *Information Retrieval Journal*, vol. 3, pp. 127–163, 2000.

[103] P. McCullagh and J. Møller, "The permanental process," *Advances in Applied Probability*, pp. 873–888, 2006.

[104] M. L. Mehta and M. Gaudin, "On the density of eigenvalues of a random matrix," *Nuclear Physics*, vol. 18, no. 0, pp. 420–427, 1960. ISSN 0029-5582. doi: 10.1016/0029-5582(60)90414-4.

[105] W. Mei and C. Zhai, "Discovering evolutionary theme patterns from text: An exploration of temporal text mining," in *Proceedings of the SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD)*, 2005.

[106] J. Møller, M. L. Huber, and R. L. Wolpert, "Perfect simulation and moment properties for the matérn type III process," *Stochastic Processes and Their Applications*, vol. 120, no. 11, pp. 2142–2158, 2010.

[107] J. Møller and R. P. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes*, vol. 100. CRC Press, 2004.

[108] J. Møller and R. P. Waagepetersen, "Modern statistics for spatial point processes," *Scandinavian Journal of Statistics*, vol. 34, no. 4, pp. 643–684, 2007.

[109] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pp. 467–475, Morgan Kaufmann Publishers Inc., 1999.

[110] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions," *Mathematical Programming*, no. 1, pp. 265–294, 1978.

[111] A. Nenkova, L. Vanderwende, and K. McKeown, "A compositional context sensitive multi-document summarizer: Exploring the factors that influence summarization," in *Proceedings of the Annual Conference on Research and Development in Information Retrieval (SIGIR)*, 2006.

[112] H. Niederreiter, *Quasi-Monte Carlo Methods*. Wiley Online Library, 1992.

[113] J. Nocedal, "Updating quasi-Newton matrices with limited storage," *Mathematics of Computation*, vol. 35, no. 151, pp. 773–782, 1980.

[114] Y. Ogata and M. Tanemura, "Likelihood analysis of spatial point patterns," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 496–518, 1984.

[115] Y. Ogata and M. Tanemura, "Estimation of interaction potentials of marked spatial point patterns through the maximum likelihood method," *Biometrics*, pp. 421–433, 1985.

[116] A. Okounkov, "Infinite wedge and random partitions," *Selecta Mathematica, New Series*, vol. 7, no. 1, pp. 57–81, 2001.

[117] A. Okounkov and N. Reshetikhin, "Correlation function of Schur process with application to local geometry of a random 3-dimensional young diagram," *Journal of the American Mathematical Society*, vol. 16, no. 3, pp. 581–604, 2003.

[118] A. Oliva and A. Torralba, "Building the gist of a scene: The role of global image features in recognition," *Progress in Brain Research*, vol. 155, pp. 23–36, 2006. ISSN 0079-6123.

[119] J. Pearl, "Reverend bayes on inference engines: A distributed hierarchical approach," in *Proceedings of the AAAI National Conference on AI*, pp. 133–136, 1982.

[120] C. J. Preston, *Random Fields*. Springer-Verlag New York, 1976.

[121] F. Radlinski, R. Kleinberg, and T. Joachims, "Learning diverse rankings with multi-armed bandits," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2008.

[122] K. Raman, P. Shivaswamy, and T. Joachims, "Learning to diversify from implicit feedback," in *WSDM Workshop on Diversity in Document Retrieval*, 2012.

[123] J. J. Ramsden, "Review of new experimental techniques for investigating random sequential adsorption," *Journal of Statistical Physics*, vol. 73, no. 5, pp. 853–877, 1993.

[124] B. D. Ripley, *Statistical Inference for Spatial Processes*. Cambridge University Press, 1991.

[125] B. D. Ripley and F. P. Kelly, "Markov point processes," *Journal of the London Mathematical Society*, vol. 2, no. 1, p. 188, 1977.

[126] B. Sapp, C. Jordan, and B. Taskar, "Adaptive pose priors for pictorial structures," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'10)*, 2010.

[127] A. Schrijver, "A combinatorial algorithm minimizing submodular functions in strongly polynomial time," *Journal of Combinatorial Theory, Series B*, vol. 80, no. 2, pp. 346–355, 2000.

[128] D. Shahaf and C. Guestrin, "Connecting the dots between news articles," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD)*, 2010.

[129] D. Shahaf, C. Guestrin, and E. Horvitz, "Trains of thought: Generating information maps," in *Proceedings of the International Conference on World Wide Web*, 2012.

[130] S. E. Shimony, "Finding maps for belief networks is NP-hard," *Artificial Intelligence*, vol. 68, no. 2, pp. 399–410, 1994.

[131] T. Shirai and Y. Takahashi, "Fermion process and Fredholm determinant," in *Proceedings of the ISAAC Congress*, vol. 1, pp. 15–23, Kluwer Academic Publishers, 2000.

[132] T. Shirai and Y. Takahashi, "Random point fields associated with certain fredholm determinants ii: Fermion shifts and their ergodic and gibbs properties," *The Annals of Probability*, vol. 31, no. 3, pp. 1533–1564, 2003.

[133] T. Shirai and Y. Takahashi, "Random point fields associated with certain Fredholm determinants i: fermion, poisson and boson point processes," *Journal of Functional Analysis*, vol. 205, no. 2, pp. 414–463, 2003.

[134] I. M. Sobol, "On the distribution of points in a cube and the approximate evaluation of integrals," *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, vol. 7, no. 4, pp. 784–802, 1967.

[135] I. M. Sobol, "On quasi-Monte Carlo integrations," *Mathematics and Computers in Simulation*, vol. 47, no. 2, pp. 103–112, 1998.

[136] D. Sontag and T. Jaakkola, "New outer bounds on the marginal polytope," *Advances in Neural Information Processing Systems*, vol. 20, pp. 1393–1400, 2007.

[137] A. Soshnikov, "Determinantal random point fields," *Russian Mathematical Surveys*, vol. 55, p. 923, 2000.

[138] D. Stoyan and H. Stoyan, "On one of matérn's hard-core point process models," *Mathematische Nachrichten*, vol. 122, no. 1, pp. 205–214, 1985.

[139] D. Strauss, "A model for clustering," *Biometrika*, vol. 62, no. 2, pp. 467–475, 1975.

[140] A. Swaminathan, C. V. Mathew, and D. Kirovski, "Essential pages," in *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*, pp. 173–182, 2009.

[141] R. Swan and D. Jensen, "TimeMines: Constructing timelines with statistical models of word usage," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD)*, 2000.

[142] R. H. Swendsen, "Dynamics of random sequential adsorption," *Physical Review A*, vol. 24, no. 1, p. 504, 1981.

[143] M. Tanemura, "On random complete packing by discs," *Annals of the Institute of Statistical Mathematics*, vol. 31, no. 1, pp. 351–365, 1979.

[144] J. M. Tang and Y. Saad, "A probing method for computing the diagonal of a matrix inverse," *Numerical Linear Algebra with Applications*, 2011.

[145] T. Tao, "Determinantal processes," http://terrytao.wordpress.com/2009/08/23/determinantal-processes/, August 2009.

[146] B. Taskar, V. Chatalbashev, and D. Koller, "Learning associative Markov networks," in *Proceedings of the International Conference on Machine Learning*, p. 102, 2004.

[147] L. G. Valiant, "The complexity of computing the permanent," *Theoretical Computer Science*, vol. 8, no. 2, pp. 189–201, 1979.

[148] M. N. M. Van Lieshout, "Markov point processes and their applications," *Recherche*, vol. 67, p. 02, 2000.

[149] V. N. Vapnik, *The Nature of Statistical Learning Theory.* Springer Verlag, 2000.

[150] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," http://www.vlfeat.org/, 2008.

[151] S. S. Vempala, *The Random Projection Method*, vol. 65. American Mathematical Society, 2004.

[152] D. Vere-Jones, "Alpha-permanents and their applications to multivariate gamma, negative binomial and ordinary binomial distributions," *New Zealand Journal of Mathematics*, vol. 26, pp. 125–149, 1997.

[153] J. Vondrák, C. Chekuri, and R. Zenklusen, "Submodular function maximization via the multilinear relaxation and contention resolution schemes," in *Proceedings of the ACM Symposium on Theory of Computing (STOC)*, pp. 783–792, 2011.

[154] C. Wayne, "Multilingual topic detection and tracking: Successful research enabled by Corpora and evaluation," in *Proceedings of the International Language Resources and Evaluation (LREC)*, 2000.

[155] R. Yan, X. Wan, J. Otterbacher, L. Kong, X. Li, and Y. Zhang, "Evolutionary timeline summarization: A balanced optimization framewo rk via iterative substitution," in *Proceedings of the Annual Conference on Research and Development in Information Retrieval (SIGIR)*, 2011.

[156] C. Yanover, T. Meltzer, and Y. Weiss, "Linear programming relaxations and belief propagation — an empirical study," *The Journal of Machine Learning Research*, vol. 7, pp. 1887–1907, 2006.

[157] C. Yanover and Y. Weiss, "Approximate inference and protein folding," *Advances in Neural Information Processing Systems*, vol. 15, pp. 1457–1464, 2002.

[158] Y. Yue and T. Joachims, "Predicting diverse subsets using structural SVMs," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2008.

[159] C. X. Zhai, W. W. Cohen, and J. Lafferty, "Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval," in *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 10–17, 2003.