Learning to Construct Novel Structures

Dino Oglic [†] **Roman Garnett** [†] **Thomas Gärtner**[†] [‡]
[†]University of Bonn and [‡]Fraunhofer IAIS, Sankt Augustin, Germany
{dino.oglic, roman.garnett, thomas.gaertner}@uni-bonn.de

Abstract

We investigate the problem of constructing novel representatives of a fixed but unknown concept over a structured, discrete domain, such as the set of unlabeled graphs. In particular, we consider an online scenario in which the learning algorithm immediately observes the true label of each proposed structure. This problem is highly relevant in many application domains and has so far been tackled mostly by special-purpose algorithms. We propose a general algorithm based on iteratively sampling novel structures from the domain and updating a conditional probability model of the concept based on the observed labels. The conditional probability model is used by the sampling procedure to focus on representatives of the desired concept. An empirical study of our approach demonstrates its effectiveness on a variety of problems.

1 Introduction

We consider the class of problems in which an algorithm is trying to design novel representatives of a fixed but unknown concept over a structured, discrete domain, such as the set of unlabelled graphs. A simple example of such a problem is '*de novo*' drug design: how can we find a molecule that binds to a particular protein binding site? Whereas a standard approach consists of screening a large database of molecules, in reality the set of synthesizable molecules is much larger than what can be stored in any database or enumerated in any reasonable time. Similar problems are also considered in the recently growing field of computational creativity (e.g., see McCormack and d'Inverno (2012)). The goal of computational creativity is to develop systems exhibiting behaviour considered creative by people. These systems either support people in a creative process or are supposed to function as autonomous artists, musicians, writers, designers, engineers, or scientists Colton (2012).

In this paper, we investigate an iterative process for *de novo* construction of structured concept representatives, focusing on the construction of graphs with desirable properties. In each step, we first sample a candidate from the posterior distribution $p(x \mid y)$ of instances x given the desired label y and obtain its true label from the oracle. We then update the conditional distribution $p(y \mid x)$ by fitting a probability model to the current set of examples. The posterior distribution is initially equal to some given instrumental distribution and is in further iterations updated by the learned conditional probability model. The procedure to obtain samples from the updated posterior resembles a Metropolis-Hastings chain in which the instrumental distribution serves as the proposal distribution and the acceptance probability is computed from the conditional probability model. A simple argument shows that the stationary distribution of this chain is exactly the posterior $p(x \mid y)$. To be able to discover concept representatives, the support of the instrumental distribution must contain the desired concept. The uniform distribution is generally a safe choice, as it guarantees that the entire space will eventually be discovered. In the absence of any other information about the desired concept, we advocate the uniform distribution from a maximum-entropy standpoint. However, when domain-specific knowledge is available, the instrumental distribution can be modified appropriately. This can be especially useful when the concept class is relatively small compared to the entire search space, in which case a uniform sampler might need very many samples to see a positive candidate.

To evaluate the efficiency and effectiveness of our approach without relying on subjective judgements or extensive wet lab experiments we investigate how well our approach is able to construct novel representatives of structured concepts that can be evaluated algorithmically. In particular, we investigate how well it can pick up concepts from oracle feedback that corresponds to planar, acyclic, and Hamiltonian graphs, respectively. Our empirical evaluation is promising and indicates the proposed approach is able to construct concept representatives from scratch.

2 Related Work

Perhaps the most-related problem setting to ours is computer-based *de novo* drug design (see, e.g., Schneider and Fechner (2005)), which can be seen as a special case of our approach. The setting defines a process in which a molecule with desired properties is constructed from scratch. Similar to our setting, the search space is enormous and exhaustive search is infeasible. From various methods implementing the *de novo* design process, the most similar to ours is a Monte Carlo search that can be combined with a Metropolis criterion. For instance, Pearlman and Murcko (1993) and Nishibata and Itai (1991) designed novel molecules that bind to a target protein by sampling atoms and making random connections based on interaction properties of atoms and the binding site. The proposed candidate is then accepted or rejected using a fixed energy-based scoring function. This process can be seen as sampling from a fixed, expert-designed distribution. Our method, in contrast, learns the concept in tandem with the construction process. The *de novo* process has also been implemented with other structure sampling methods: linking, growing, lattice-based sampling, random structure mutation, transitions driven by molecular dynamics simulations, and graph-based sampling methods (Schneider and Fechner, 2005).

Related to our work are also methods for interactive learning and optimization (ILO). These methods implement a two-step iterative process in which an agent interacts with a user until a satisfactory solution is obtained. Some of the well-known ILO methods are designed for problems in information retrieval (see, e.g., Shivaswamy and Joachims (2012); Yue and Joachims (2009)) and reinforcement learning (see, e.g., Jain et al. (2013); Wilson et al. (2012)). In contrast to ILO methods, we are not necessarily interested in approximating the single 'best' point from the instance space. For instance, we could be searching for as many positive instances as possible that are at the same time as diverse as possible. This makes our problem setting more general and more difficult. In addition to this, ILO methods are designed mainly to construct structures over the domain of real-valued vectors and can not be directly applied to highly structured domains.

3 Algorithm

In this section, we first give a pseudo-code description of the algorithm for *de novo* construction of concept representatives given examples. We then review exponential family models and bound the mixing time of a Metropolis–Hastings (MH) chain for sampling from these models.

We propose to construct structured entities by sampling from an iteratively refined posterior distribution. The sampling process resembles a Metropolis-Hastings chain with an instrumental proposal distribution $\mathcal{Q}\colon \mathcal{X} \to \mathbb{R}$ which is independent of the current state of the chain. To refine the posterior, the algorithm is provided with an oracle $\mathcal{O}\colon \mathcal{X} \to \mathcal{Y}$ to evaluate the proposed candidates. Usually, the evaluation is expensive and a budget B will be given to the algorithm. In our experiments, we use a kernelized exponential family model $p(y \mid x, \theta)$ to learn the concept from labelled examples; this is discussed below. For this choice of model, we require a positive definite kernel function on $\mathcal{X} \times \mathcal{Y}$. We note here that the proposed approach is not limited to instrumental distributions independent of past states of the chain. Instead, a transition kernel that satisfies the detailed balance condition can be used to design novel entities by modifying the previously designed ones (e.g., a Markov chain procedure with the prior distribution as its stationary distribution).

The algorithm first samples n entities from the instrumental distribution $\mathcal{Q}$ and adds the entities, along with their labels, into the training sample S. Then, the algorithm starts iterating until we either obtain a satisfactory number of desired candidates or deplete the oracle budget (see line 2). In the first step of each iteration, the algorithm is trained on the previously seen examples S (see line 3). This is achieved by finding the maximum *a posteriori* (MAP) estimate of the parameter vector θ. In the next step (see lines 4–10), the Metropolis–Hastings algorithm (see, e.g., Andrieu et al. (2003))

Algorithm 1 DE-NOVO-MACHINE

Input: target concept class $y \in \mathcal{Y}$, parameterized model $p(y \mid x, \theta)$ and parameter prior $p(\theta)$, instrumental distribution $\mathcal{Q}$, oracle $\mathcal{O}\colon \mathcal{X} \to \mathcal{Y}$ with budget B, and initial training set size n.
Output: list of potential concept representatives $x_1, x_2, \ldots, x_B$

1: $S \leftarrow \{(\hat{x}, \hat{y}) \mid \hat{x} \sim \mathcal{Q}, \hat{y} \leftarrow \mathcal{O}(\hat{x})\}$ and $|S| = n$
2: **while** $t = 1, 2, \ldots, B$ **do**
3: $\quad \theta_t \leftarrow \arg\max_\theta \log p(S \mid \theta) + \log p(\theta)$
4: $\quad x_t \sim \mathcal{Q}$
5: $\quad$ **repeat**
6: $\quad\quad x \sim \mathcal{Q}$ and $u \sim \mathcal{U}[0, 1]$
7: $\quad\quad$ **if** $u < p(y \mid x)/p(y \mid x_t)$ **then**
8: $\quad\quad\quad x_t \leftarrow x$
9: $\quad\quad$ **end if**
10: $\quad$ **until** CHAIN MIXED
11: $\quad$ **output** x_t
12: $\quad y_t \leftarrow \mathcal{O}(x_t)$
13: $\quad S \leftarrow S \cup \{(x_t, y_t)\}$
14: **end while**

is used to sample the posterior $p(x \mid y) = \frac{p(y|x)p(x)}{p(y)}$. As the Markov chain samples the next state independently from the instrumental distribution $\mathcal{Q}$, the acceptance threshold is equal to

$$\frac{p(y \mid x_c)p(x_c)\mathcal{Q}(x_t)}{p(y \mid x_t)p(x_t)\mathcal{Q}(x_c)} = \frac{p(y \mid x_c)}{p(y \mid x_t)}, \tag{1}$$

where x_c is the proposed candidate, x_t is the last accepted state, and $p(x) = \mathcal{Q}(x)$ $(\forall x \in \mathcal{X})$. After sampling an instance x_t from the posterior, the algorithm outputs it as a potential concept representative (line 11) and presents it to an oracle. The oracle evaluates it providing feedback y_t to the algorithm (see line 12). Finally, the labelled pair (x_t, y_t) is added to the training sample S (see line 13) for the next iteration (see line 3).

As we have presented the algorithm, we do not specify the nature of the model $p(y \mid x, \theta)$. In this paper we advocate the conditional exponential family model, because it is flexible and we can give theoretical bounds on the mixing time of the Markov chain. Let $\mathcal{X}$ be a set of structured entities (e.g., graphs, binary vectors, etc.) and $\mathcal{Y}$ be a set of concepts (e.g., planar graphs). The conditional exponential family models the conditional distribution of an output $y \in \mathcal{Y}$ given $x \in \mathcal{X}$ as $p(y \mid x; \theta) = \exp\big(\langle \phi(x, y), \theta \rangle\big) - g(\theta \mid x)$, where θ is a parameter vector, $\langle \cdot, \cdot \rangle$ is a dot product in a Hilbert space, $\phi(x, y)$ are the sufficient statistics of $y \mid x$ and $g(\theta \mid x) = \log \int_{\mathcal{Y}} \exp\big(\langle \phi(x, y), \theta \rangle\big)\, \mathrm{d}y$ is the log-partition function. Throughout the paper, we assume that the ℓ_2-norm of the sufficient statistics and the parameters are bounded, i.e., $\|\phi(x, y)\| \leq r$ and $\|\theta\| \leq R$. Here, we note that we can take $r = 1$ by normalizing the covariance function such that $\big\langle \phi(x, y), \phi(x, y) \big\rangle = 1$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

To bound the mixing time of a Metropolis–Hastings chain for sampling from the posterior $p(x \mid y)$ we provide a simple *coupling* analysis (see also Vembu et al. (2009)). In a nutshell, $|\mathcal{X}|$ parallel chains are started such that each chain starts from a different state $x \in \mathcal{X}$. Then, an identical random bit sequence is used to simulate all the chains. After two chains move to a common state, all the future transitions of the two chains are the same. Thus, from that point on it is sufficient to track only one of the chains and such chain coupling is called a *coalescence* (Huber, 1998). Propp and Wilson (1996) have showed that if all the chains coalesce to a single chain at step T then samples from this resulting chain are samples from the stationary distribution.

For a finite concept space $\mathcal{Y}$, from Eq. (1) it follows

$$p(x \to x') = \frac{\exp\big(\langle \phi(x', y), \theta \rangle - g(\theta \mid x')\big)}{\exp\big(\langle \phi(x, y), \theta \rangle - g(\theta \mid x)\big)} = \frac{\sum_{\overline{y} \in \mathcal{Y}} \exp\big(\langle \phi\,(x', y) + \phi\,(x, \overline{y})\,, \theta \rangle\big)}{\sum_{\overline{y} \in \mathcal{Y}} \exp\big(\langle \phi\,(x, y) + \phi\,(x', \overline{y})\,, \theta \rangle\big)}$$
$$\geq \frac{|\mathcal{Y}| \exp\big(2 \cdot \langle \phi\,(x_\downarrow, y_\downarrow)\,, \theta \rangle\big)}{|\mathcal{Y}| \exp\big(2 \cdot \langle \phi\,(x_\uparrow, y_\uparrow)\,, \theta \rangle\big)} \geq \exp\big(-2 \cdot \big|\big\langle \phi(x_\downarrow, y_\downarrow) - \phi(x_\uparrow, y_\uparrow), \theta \big\rangle\big|\big),$$

where $\langle \phi(x_\downarrow, y_\downarrow), \theta \rangle$ and $\langle \phi(x_\uparrow, y_\uparrow), \theta \rangle$ are the minimum and maximum values of the dot products appearing in the numerator and denominator of $p(x \to x')$, respectively.

Now, using the Cauchy–Schwarz inequality we get

$$p(x \to x') \geq \exp\big(-2 \cdot \big\|\phi(x_\downarrow, y_\downarrow) - \phi(x_\uparrow, y_\uparrow)\big\| \|\theta\|\big) \geq \exp(-4R). \tag{2}$$

Hence, the expected number of steps until coalescence is upper bounded with

$$\mathbb{E}[T] \leq \sum_{k=1}^{\infty} k \exp(-4R)\big(1 - \exp(-4R)\big)^{k-1} = \exp(4R).$$

4 Application-Specific Details

In this section, we describe instrumental distributions and covariance functions used in Section 5. For graphs we follow the terminology and notation of standard textbooks (Bondy and Murty, 2008).

As the set of graphs is a complicated, combinatorial object, it can be difficult to design an efficient proposal distribution. In general, to sample a random unlabelled graph it is common to use the Erdős–Rényi model with $p = 1/2$. This approach, however, does not yield unlabelled graphs uniformly at random (e.g., the probability of sampling an unlabelled path with n vertices is $\frac{n!}{2}$ times higher than the probability of sampling the complete graph with the same number of vertices).

Let $\mathcal{G}_n$ denote the set of all canonically labelled graphs with n vertices and $\mathcal{G} = \cup_{n \geq 0} \mathcal{G}_n$. A *left action* of a group S on a set X is a function $\mu : S \times X \to X$ with the following two properties: $(i)(\forall x \in X)(\forall s, t \in S) : \mu(t, \mu(s, x)) = \mu(ts, x)$; $(ii)(\forall x \in X) : \mu(e, x) = x$ (where e is the identity element of the group S). If no confusion arises we write $\mu(s, x) = sx$. A group action defines the equivalence relation $\sim$ on a set X, i.e., $a \sim b \Leftrightarrow sa = b$ for some $s \in S$ and $a, b \in X$. The equivalence classes determined by the relation $\sim$ are called *orbits* of S in X and the number of orbits can be computed using the Frobenius–Burnside theorem (see, e.g., Cameron (1998)).

Theorem 1 (Frobenius–Burnside)**.** *Let X be a finite non-empty set and S be a finite group. If X is an S-set, then the number of orbits of S in X is equal to $\frac{1}{|S|} \sum_{s \in S} |\operatorname{Fix}(s)|$, where $\operatorname{Fix}(s) = \{x \in X \mid sx = x\}$.*

To sample unlabelled graphs uniformly at random, Wormald (1987) proposed a rejection sampling method based on Theorem 1. The idea is to consider the action of a symmetric group S_n over the set $\mathcal{G}_n$. Then, the orbits of S_n in the set $\mathcal{G}_n$ are non-isomorphic unlabelled graphs and to sample unlabelled graphs uniformly it suffices to uniformly sample the orbits (see, e.g., Dixon and Wilf (1983)). Moreover, it is possible to show (see, e.g., Dixon and Wilf (1983); Wormald (1987)) that the uniform orbit sampling is equivalent to uniform sampling of an element from the set $\Gamma = \big\{(\pi, g) \mid \pi g = g; \pi \in S_n, g \in \mathcal{G}_n\big\}$.

According to Theorem 1, an element $(\pi, g) \in \Gamma$ can be sampled u.a.r. by choosing a permutation π with probability proportional to $|\operatorname{Fix}(\pi)|$ and then choosing $g \in \operatorname{Fix}(\pi)$ u.a.r. Dixon and Wilf (1983) propose a more efficient sampling algorithm by partitioning the symmetric group into conjugacy classes $[\pi_i]$ $(i = \overline{1, l})$ and sampling: (i) $[\pi_i] \sim {|[\pi_i]||\operatorname{Fix}(\pi_i)|}/{o_n |S_n|}$, (ii) $g \in \operatorname{Fix}(\pi_i)$ u.a.r.; where o_n denotes the number of non-isomorphic unlabelled graphs with n vertices and π_i is a class representative for the class $[\pi_i]$. As it holds $|\operatorname{Fix}(\pi)| = |\operatorname{Fix}(\pi')|$ and $|\operatorname{Fix}(\pi) \cap [g]| = |\operatorname{Fix}(\pi') \cap [g]|$ for $\pi, \pi' \in [\pi_i]$ then

$$P\big([g]\big) = \sum_{i=1}^{l} P\big([\pi_i]\big) P\big([g] \mid [\pi_i]\big) = \sum_{i=1}^{l} \frac{\big|[\pi_i]\big|\big|\operatorname{Fix}(\pi_i)\big|}{o_n |S_n|} \frac{\big|\operatorname{Fix}(\pi_i) \cap [g]\big|}{\big|\operatorname{Fix}(\pi_i)\big|} = \frac{1}{o_n}$$

(see, e.g., Wormald (1987) for a more detailed proof)

The problem with the approach is the fact that we need to know the exact number of non-isomorphic graphs with n vertices o_n to apply the algorithm and this number is not computable in polynomial time. To overcome this, Wormald (1987) partitions the elements of the group S_n into classes $[c_k] = \{\pi \in S_n \mid \operatorname{support}(\pi) = k\}$, $k = \overline{0, n}$, and upper bounds $\big|[c_i]\big|\big|\operatorname{Fix}(\pi_i)\big| \leq B_i$. The algorithm then samples an unlabelled graph u.a.r. as follows: (i) $[c_i] \sim {B_i}/{\sum_j B_j}$, (ii) $\pi_i \in [c_i]$ u.a.r., (iii)

$g \in \mathrm{Fix}(\pi_i)$ u.a.r., (*iv*) accept the sampled graph g with probability $B_i^{-1} \, |[c_i]| \, |\mathrm{Fix}(\pi_i)|$; otherwise, restart. On average, the method generates an unlabelled graph in time that is polynomial in the number of vertices (Wormald, 1987).

For covariance functions of conditional exponential family, we assume we can write the inner product of sufficient statistics as $\langle \phi(x, y), \phi(x', y') \rangle = k_{\mathcal{X}}(x, x') k_{\mathcal{Y}}(y, y')$, where $k_{\mathcal{X}}$ and $k_{\mathcal{Y}}$ are covariance functions over spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. To learn graph concepts we use the random walk kernel as the instance space kernel $k_{\mathcal{X}}$ (see, e.g., Borgwardt (2007); Gärtner et al. (2003)). The kernel performs random walks on both graphs and counts the number of matching walks. The concept space kernel function is defined by the cardinality of the intersection set of concepts $y_i, y_j \in \mathcal{Y}$, i.e., $k_{\mathcal{Y}}(y_i, y_j) = |y_i \cap y_j|$.

5 Experiments

In this section, we investigate performances of Algorithm 1 over different concepts: the construction of acyclic, Hamiltonian, and planar graphs.

When constructing graphs we restrict ourselves to generation of concept representatives from sets of (non-isomorphic) unlabelled graphs with 7 and 10 vertices. Here we note that $|\mathcal{G}_7| = 1\,044$ and $|\mathcal{G}_{10}| = 12\,005\,168$. Each graph concept we investigate partitions the space of graphs into two disjoint classes, e.g., planar and non-planar graphs. Thus, the concept classes are orthogonal and the concept space kernel matrix is the identity matrix. In construction of Hamiltonian and planar graphs, we propose graphs using the uniform instrumental distribution described above. For the construction of acyclic graphs, we use the Erdős–Rényi sampler with $p \sim \mathcal{U}[0, 1]$ (for $\mathcal{G}_{10}$, already, we rarely see acyclic graphs with the uniform sampler).

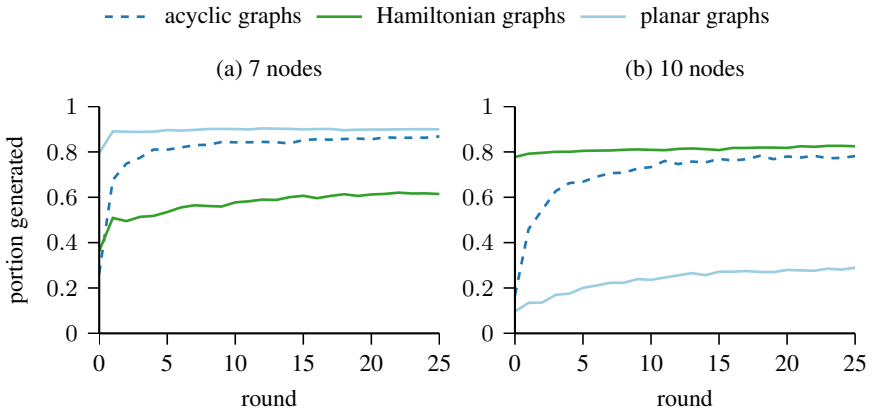

Figure 1: Learning to construct novel concept representatives on small unlabeled graphs. The concepts are acyclic, Hamiltonian, and planar graphs. We implement 25 rounds of our algorithm for graphs with seven (a) and ten nodes (b) and report the portion of generated samples reflecting the desired concept.

All the reported results were obtained by averaging over 5 runs of the algorithm. Metropolis–Hastings sampling was performed with a burn-in sample of 50 000 proposals, and sampling was done in a batch by taking 10 samples with a skip size of 1 000 proposals. We used the exponential family regularizer parameter $\sigma = 0.001$ and random walk kernel parameter $\lambda = (n^2 - n + 1)^{-1}$, where $\mathcal{G}_n$ denotes the search space of graphs with n vertices. This choice of λ guarantees that the kernel converges for any pair of graphs $g, g' \in \mathcal{G}_n$. We measure our progress by tracking the fraction of target samples within the accepted states of a Metropolis–Hastings chain throughout the rounds. Also, at the first step of each run we take 50 000 samples from the prior and report the fraction of target concepts within the sample (step-0 statistics in Figure 1).

Figure 1 demonstrates the performance of Algorithm 1 on graph concepts. Results indicate we are learning in all cases. As seen from the experiments, some concepts are easier from the perspective of the prior as it is biased toward that concept class. To stress the importance of a prior, we note that for

acyclic graphs, we had to use the Erdős–Rényi sampler as with the uniform sampler we rarely see an acyclic graphs. Finally, we note that some concepts are difficult to discriminate with the kernel we use (e.g., planar graphs); improvements in this direction are beyond the scope of this paper.

6 Conclusion

We have presented an algorithm for the *de novo* construction of novel structured concept representatives. The effectiveness of the approach was tested on the domain of graphs with a fixed number of vertices. Empirical results are promising and indicate our algorithm is capable of constructing structured concept representatives from scratch. For future work, we plan to extend the domain of applications to the space of molecular graphs and derive a theoretical lower-bound on the number of constructed target representatives.

Acknowledgment: Part of this work was supported by the German Science Foundation (DFG) under the reference number 'GA 1615/1-1'.

References

Christophe Andrieu, Nando de Freitas, Arnaud Doucet, and Michael I. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1-2):5–43, 2003.

Adrian J. Bondy and Uppaluri S. R. Murty. *Graph Theory*. Springer, 2008.

Karsten Michael Borgwardt. *Graph kernels*. PhD thesis, Ludwig Maximilians University Munich, 2007.

Peter J. Cameron. *Introduction to Algebra*. Oxford University Press, 1998.

Simon Colton. The Painting Fool: Stories from Building an Automated Painter. In Jon McCormack and Mark d'Inverno, editors, *Computers and Creativity*, pages 3–38. Springer, 2012.

John D. Dixon and Herbert S. Wilf. The Random Selection of Unlabeled Graphs. *Journal of Algorithms*, 4(3): 205–213, 1983.

Thomas Gärtner, Peter A. Flach, and Stefan Wrobel. On graph kernels: Hardness results and efficient alternatives. In *Proceedings of the 16th Annual Conference on Computational Learning Theory (COLT 2003) and the 7th Kernel Workshop*, 2003.

Mark Huber. Exact Sampling and Approximate Counting Techniques. In *Proceedings of the 30th Annual ACM Symposium on the Theory of Computing*, 1998.

Ashesh Jain, Brian Wojcik, Thorsten Joachims, and Ashutosh Saxena. Learning Trajectory Preferences for Manipulators via Iterative Improvement. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, 2013.

Jon McCormack and Mark d'Inverno, editors. *Computers and Creativity*. Springer, 2012.

Yoshihiko Nishibata and Akiko Itai. Automatic creation of drug candidate structures based on receptor structure. Starting point for artificial lead generation. *Tetrahedron*, 47(43):8985–8990, 1991.

David A. Pearlman and Mark A. Murcko. Concepts: New dynamic algorithm for de novo drug suggestion. *Journal of Computational Chemistry*, 14(10):1184–1193, 1993.

James G. Propp and David B. Wilson. Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics, 1996.

Gisbert Schneider and Uli Fechner. Computer-based de novo design of drug-like molecules. *Nature Reviews Drug Discovery*, 4(8):649–63, 2005.

Pannaga Shivaswamy and Thorsten Joachims. Online Structured Prediction via Coactive Learning. In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, 2012.

Shankar Vembu, Thomas Gärtner, and Mario Boley. Probabilistic Structured Predictors. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI 2009)*, 2009.

Aaron Wilson, Alan Fern, and Prasad Tadepalli. A Bayesian Approach for Policy Learning from Trajectory Preference Queries. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2012.

Nicholas C. Wormald. Generating Random Unlabelled Graphs. *SIAM Journal on Computing*, 16(4):717–727, 1987.

Yisong Yue and Thorsten Joachims. Interactively Optimizing Information Retrieval Systems As a Dueling Bandits Problem. In *Proceedings of the 26th International Conference on Machine Learning (ICML 2009)*, 2009.