Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence

Distribution Kernel Methods for Multiple-Instance Learning

Gary Doran
Department of Electrical Engineering and Computer Science
Case Western Reserve University
Cleveland, OH 44106, USA
gary.doran@case.edu

Abstract

I propose to investigate learning in the multiple-instance (MI) framework as a problem of learning from distributions. In many MI applications, bags of instances can be thought of as samples from bag-generating distributions. Recent kernel approaches for learning from distributions have the potential to be successfully applied to these domains and other MI learning problems. Understanding when distribution-based techniques work for MI learning will lead to new theoretical insights, improved algorithms, and more accurate solutions for real-world problems.

Motivating Application

The 3-dimensional Quantitative Structure-Activity Relationship (3D-QSAR) problem entails predicting whether a molecule is "active" or "inactive," that is, whether it will bind to a target protein or not. Although this problem sounds simple, designing a new drug (from initial research through testing) can cost nearly $1 billion (Adams and Brantner 2006). Accurate computational approaches to solving this problem can greatly reduce these costs. For example, the potential to bind to a protein largely depends on a molecule's structure, which can be represented as a feature vector and used in conjunction with *supervised learning* techniques, such as feature selection and linear regression, to predict activity.

However, the problem is complicated by the fact that flexible bonds in molecules allow each molecule to exist in multiple shapes, called *conformations*, when dissolved in solution. Figure 1 shows schematically how a molecule activates a receptor if and only if some conformation binds to the target. The multiple-instance (MI) learning (MIL) framework is an extension of supervised learning motivated by the above problem (Dietterich, Lathrop, and Lozano-Pérez 1997), and encodes this relationship between an observed label and a set instances responsible for that label. In this case, each conformation represented as a feature vector is an *instance*, and a molecule corresponds to a set of instances (conformations) called a *bag*. If a bag is labeled positive, then at least one instance in the bag is positive. However, if a bag is negative, then every instance in the bag is negative.

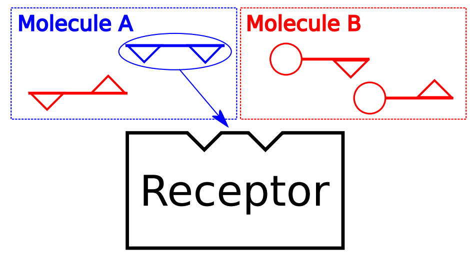

Figure 1: Molecule A activates the receptor since one of its conformations binds. On the other hand, no conformation of Molecule B activates the receptor, so Molecule B is inactive.

The learner for an MI classification problem has to produce a classifier that can accurately label new bags or instances. MI learning is an important problem with applications beyond 3D-QSAR in content-based image retrieval, remote sensing, sub-goal discovery for reinforcement learning, and robotics.

Key Idea and Thesis Statement

The key insight of my research is that in many MIL applications, bags of instances can be thought of as samples from underlying *bag generating distributions*. In the 3D-QSAR problem described above, conformations smoothly transform from one to another, existing in a dynamic equilibrium probabilistically governed by their Gibbs free energy. The bags observed are samples drawn from the distributions corresponding to the energy functions over all conformations, which naturally prefer low-energy conformations. Similar arguments for bag generating distributions can be made for other MI problem domains as well. I propose to investigate learning in the MI framework as a problem of *learning from distributions*. Understanding when distribution-based techniques work for MI learning will lead to new theoretical insights, improved algorithms, and more accurate solutions for real-world problems.

Research Topics

Recent techniques for learning from distributions with kernels (Smola et al. 2007; Muandet et al. 2012) have the potential to be successfully applied to these domains and other

MI learning problems. In particular, the kernel mean map is an injective mapping of distributions into a feature space that compactly represents distributions while preserving information relevant to the learning process. A related MI kernel is the *average-normalized set kernel* (Gärtner et al. 2002), which is equivalent to an "empirical" mean map computed using samples from distributions. The properties of these techniques raise several interesting theoretical questions related to *when* distribution-based learning can be applied successfully to MIL. The distribution-based learning framework also suggests several novel methods and applications for MIL, described below.

Theory

A key focus of my research is understanding which techniques for embedding distributions into feature spaces are best suited for MIL. Set kernels lack a property I call *soundness*, which means that they allow zero-loss solutions that are inconsistent with the MI assumption that at least one instance in each bag is positive. This property is unique to MIL, since supervised learning algorithms are sound by construction. In other words, a support vector machine (SVM) using the set kernel might "hallucinate" a solution that separates positive and negative *bags* in its feature space when no solution exists that separates positive and negative *instances* in the related instance feature space. Because this contradicts the logical relationship between instance and bag labels, it remains unclear why set kernels perform well for MIL.

It is also important to understand when *instance* labels can be learned from a *bag* (distribution) classifier. Some prior work on learning theory for MIL assumes that each bag is generated by drawing instances independently from an underlying instance distribution (Blum and Kalai 1998). However, as argued above, it often makes more sense to consider a hierarchical generative process in which each bag is drawn from a distribution over bags, then each instance in a bag is sampled according to a bag-specific distribution over instances. Other existing work describes when instance or bag concepts are learnable in the probably approximately correct (PAC) framework from MI data using instance classifiers (Auer, Long, and Srinivasan 1997; Blum and Kalai 1998; Sabato and Tishby 2012). However, it remains unclear when learnability of instance concepts is possible using algorithms that learn using bag-level distribution information. I will explore these learning theory questions into next year.

Methods

Distribution kernels might be used to solve other MI problems such as MI regression efficiently. The 3D-QSAR problem described above can also be formulated as an MI regression problem in which each molecule is labeled with a real-valued activity level rather than a binary active/inactive label. My current research suggests that set kernels can outperform state-of-the-art baselines on this task.

Techniques like the empirical mean map suggest other set kernels that would be more well-suited for particular MIL problems. For example, if bag sizes are small, then noisy instances might affect the value of the empirical mean map of bag distributions. By computing the median rather than the mean, a distribution might be embedded into a feature space in a way that is more robust to noise. I have implemented this "kernel median map" for the MI classification and regression problems, but more work is needed to theoretically characterize the behavior of this approach.

Applications

I plan to integrate the MI regression approach described above into a system that will evaluate the performance of set kernels on real 3D-QSAR datasets, which will provide a comparison to other techniques developed for this particular application. I am also exploring the application of MI regression to problems in climate science such as the remote sensing of aerosols using satellite measurements.

Conclusion

Understanding *when* distribution-based techniques work for MI learning will lead to improved algorithms and more accurate solutions for real-world problems. Furthermore, my research will more generally provide insight into the power of distribution-based learning techniques for learning from structured objects. These insights might illustrate when these approaches can be successfully applied to the related problems of learning from trees, graphs, and relational data.

References

Adams, C., and Brantner, V. 2006. Estimating the cost of new drug development: is it really $802 million? *Health Affairs* 25(2):420–428.

Auer, P.; Long, P.; and Srinivasan, A. 1997. Approximating hyper-rectangles: Learning and pseudo-random sets. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computation*, 314–323. ACM.

Blum, A., and Kalai, A. 1998. A note on learning from multiple-instance examples. *Machine Learning Journal* 30:23–29.

Dietterich, T. G.; Lathrop, R. H.; and Lozano-Pérez, T. 1997. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence* 89(1–2):31–71.

Gärtner, T.; Flach, P.; Kowalczyk, A.; and Smola, A. 2002. Multi-instance kernels. In *Proceedings of the International Conference on Machine Learning*, 179–186.

Muandet, K.; Fukumizu, K.; Dinuzzo, F.; and Schölkopf, B. 2012. Learning from distributions via support measure machines. In *Advances in Neural Information Processing Systems*. 10–18.

Sabato, S., and Tishby, N. 2012. Multi-instance learning with any hypothesis class. *Journal of Machine Learning Research* 13:2999–3039.

Smola, A.; Gretton, A.; Song, L.; and Schölkopf, B. 2007. A hilbert space embedding for distributions. In *Algorithmic Learning Theory*, 13–31. Springer.