

# Generalized Multi-Instance Learning: Problems, Algorithms and Data Sets

Min-Ling Zhang  
College of Computer and Information Engineering  
Hohai University  
Nanjing, China  
Email: zhangml@hhu.edu.cn

## Abstract

*In multi-instance learning, each example is represented by a bag of instances while associated with a binary label. Under standard multi-instance learning settings, one example is labeled as a positive bag if at least one of its instances is positive. Otherwise, it is labeled as a negative bag. Although based on the above assumption, standard multi-instance learning has achieved much success in solving diverse learning tasks, there are still many real-world problems where this assumption may not necessarily hold. Therefore, researchers aimed to expand the underlying assumption of standard multi-instance learning where two frameworks of generalized multi-instance learning have been proposed. In this paper, the problem definition, learning algorithms and also experimental data sets related to either generalized multi-instance learning framework are briefly reviewed.*

## 1. Introduction

In their investigation of drug activity prediction problem, Dietterich et al. [7] initialized the framework of multi-instance learning. In this learning framework, each example is represented by a *bag* comprising multiple instances and associated with one binary label. The task of multi-instance learning is trying to correctly predict the labels of unseen bags through learning from training bags with known labels. In standard settings of multi-instance learning, it is assumed that one bag is labeled positively if and only if *at least* one instance contained in the bag is positive. Otherwise, the bag is labeled negatively.

Many real-world problems can be appropriately modeled under multi-instance learning. For example, in drug activity prediction [7], each molecule may exhibit a number of low-energy (stable) conformations each can be represented by an instance; In content-based image categorization [5], [16] or retrieval [28], [32], each image usually contains several naturally disjoint regions each can be regarded as an instance; In text applications [2], [34], each document or web page generally encompasses some paragraphs or links each can be expressed as an instance.

Although multi-instance learning has achieved considerable success in solving a wide range of learning tasks,

the basic assumption that the qualification of a bag being positive totally relies on the existence of one *key* (i.e. positive) instance may not hold in many real-world cases. For instance, an image being labeled as *Arctic* is not attributed to the *sole* existence of either a bear patch or an ice patch, but actually attributed to *joint* effects of both image patches. Therefore, researchers have proposed two different generalized multi-instance learning frameworks [20], [26] respectively to accommodate more sophisticated interactions among instances in the bag. In this paper, related works on generalized multi-instance learning are briefly reviewed.

The rest of this paper is organized as follows. In Section 2, the formal definition and the state-of-the-art of multi-instance learning are given. In Sections 3 and 4, the formal definition, learning algorithms and benchmark data sets of either generalized multi-instance learning framework are reviewed respectively. Finally in Section 5, the paper is concluded and some future works are raised.

## 2. Multi-Instance Learning

Let  $\mathcal{X}$  be the input space and  $\mathcal{Y} = \{0, 1\}$  be the binary output space. In traditional supervised learning, the task is to learn a function  $f : \mathcal{X} \rightarrow \mathcal{Y}$  from a set of training instances  $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$ , where  $\mathbf{x}_i \in \mathcal{X}$  is a single instance and  $y_i \in \mathcal{Y}$  is the label associated with  $\mathbf{x}_i$ . In multi-instance learning [7] however, the training set is composed of  $m$  labeled bags  $\{(X_1, y_1), (X_2, y_2), \dots, (X_m, y_m)\}$ , where  $X_i = \{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{in_i}\}$  is a bag of  $n_i$  instances and  $y_i$  is the binary label associated with  $X_i$ . The task of multi-instance learning is to learn a function  $f_{\text{MIL}} : 2^{\mathcal{X}} \rightarrow \mathcal{Y}$  from the given multi-instance training set.

Following the seminal work of Dietterich et al. [7], multi-instance learning has been extensively studied by the machine learning community. Theoretically, the PAC-learnability of multi-instance learning algorithms, especially the APR (Axis-Parallel Rectangles) approach proposed by Dietterich et al. [7], are thoroughly investigated [3], [4], [14]. Researchers have also proposed numerous multi-instance learning algorithms by adapting traditional supervised learning approaches, e.g. Diverse Density [15], Citation- $k$ NN

[25], Relic [19], EM-DD [31], ID3-MI [6], RBF-MIP [29], MI-SVM [2], and MIBoosting [27], etc.

Multi-instance learning techniques have been successfully applied to a number of real-world problems such as CBIR [10], [28], [32], scene analysis [5], [16], computer security [19], and web mining [34], etc. Besides standard multi-instance classification problems, multi-instance learning has also been investigated under various learning scenarios, such as multi-instance regression [8], [18], semi-supervised multi-instance learning [17], multi-instance ensembles [27], [35], multi-instance learning aided relational learning [1], and multi-instance multi-label learning [36], etc. Comprehensive reviews on multi-instance learning can be found in [33].

In the above-mentioned standard multi-instance learning, the basic assumption is that one example, i.e. a bag of instances, is regarded to be positive if and only if it contains *at least* one positive instance. In other words, there is always an *instance-level* target concept  $c : \mathcal{X} \rightarrow \{0, 1\}$  governing any standard multi-instance learning problem. Firstly, the multi-instance learner usually attempts to induce  $c$  by learning a classifier  $c'$  from  $\{(X_1, y_1), (X_2, y_2), \dots, (X_m, y_m)\}$ . After that, the label for unseen bag  $X$  is predicted as  $f_{\text{MIL}}(X) = \arg \max_{\mathbf{x} \in X} c'(\mathbf{x})$ . As shown in Section 1, the basic multi-instance notion based on the existence of *key* instance may not necessarily hold in some real-world learning problems. In the next section, we will briefly review related works on *generalized* multi-instance learning frameworks where more complex and sophisticated underlying assumptions are adopted for multi-instance learning.

### 3. Generalized Multi-Instance Learning: the First Scenario

#### 3.1. Problem Definition

The first type of generalized multi-instance learning was proposed by Weidmann et al. [26]. Specifically, they defined three different kinds of generalized multi-instance concepts, i.e. *presence-based* multi-instance learning, *threshold-based* multi-instance learning, and *count-based* multi-instance learning. In any of the generalized multi-instance problems, it is assumed that a set of  $r$  *instance-level* concepts  $C = \{c_1, c_2, \dots, c_r\}$  ( $c_i : \mathcal{X} \rightarrow \{0, 1\}$ ) will govern the labeling process of multi-instance bags. The essential differences among those generalized multi-instance problems arise in how the set of underlying concepts in  $C$  are combined to determine the label of each example.

Let  $\Delta(X, c)$  be the number of instances in bag  $X$  which correspond to concept  $c$ . For presence-based multi-instance learning, the learned multi-instance classifier  $f_{\text{PB\_MIL}} : 2^{\mathcal{X}} \rightarrow \{0, 1\}$  predicts the label of unseen bag  $X$  as follows:

$$f_{\text{PB\_MIL}}(X) = \arg \min_{1 \leq i \leq r} [\Delta(X, c_i) \geq 1] \quad (1)$$

Here,  $[\pi]$  returns 1 if predicate  $\pi$  holds, and 0 otherwise. In presence-based multi-instance learning,  $X$  is labeled to be positive if it contains at least one instance for each concept. For instance, an image labeled as *Africa* rather than *Australia* may be attributed to the co-existence of image patches describing grassland, lions and hutch.

For threshold-based multi-instance learning, the learned multi-instance classifier  $f_{\text{TB\_MIL}} : 2^{\mathcal{X}} \rightarrow \{0, 1\}$  predicts the label of unseen bag  $X$  as follows:

$$f_{\text{TB\_MIL}}(X) = \arg \min_{1 \leq i \leq r} [\Delta(X, c_i) \geq t_i] \quad (2)$$

Here  $t_i \in \mathbb{N}$  represents the *lower bound* for concept  $c_i$ . In threshold-based multi-instance learning,  $X$  is labeled to be positive if it contains at least a certain number of instances for each concept. For instance, a video snapshot labeled as *sporting match* rather than *training session* may be attributed to the existence of *some* athletes, *several* referees and also *a group of* spectators.

For count-based multi-instance learning, the learned multi-instance classifier  $f_{\text{CB\_MIL}} : 2^{\mathcal{X}} \rightarrow \{0, 1\}$  predicts the label of unseen bag  $X$  as follows:

$$f_{\text{CB\_MIL}}(X) = \arg \min_{1 \leq i \leq r} [t_i \leq \Delta(X, c_i) \leq z_i] \quad (3)$$

Here  $z_i \in \mathbb{N}$  represents the *upper bound* for concept  $c_i$ . In count-based multi-instance learning,  $X$  is labeled to be positive if it contains at least a minimum number of instances as well as at most a maximum number of instances for each concept. For instance, a molecule labeled as *ethanol* rather than *carbinol* may be attributed to the existence of *two* carbon atoms, *one* oxygen atom and also *six* hydrogen atoms (assuming that  $t_i = z_i$  for all  $c_i \in C$ ).

#### 3.2. Learning Algorithms

Weidmann et al. [26] proposed the two-level classification approach (named as TLC in this paper) to deal with their proposed generalized multi-instance learning problems. The basic idea of TLC is trying to identify the underlying *instance-level* concepts in the first level, where each bag of instances is re-represented by a *meta-instance* whose features encode the relationships between instances in the bag and the identified underlying concepts. After that, those newly created meta-instances (along with labels of the generating bags) are fed to a second-level classifier in hope of inducing a function capturing the interactions between meta-instance and bag's label.

Concretely, given the multi-instance training set  $\{(X_1, y_1), (X_2, y_2), \dots, (X_m, y_m)\}$ , let  $N = \sum_{i=1}^m |X_i|$  denote the total number of instances contained in all bags. In the first level of TLC, each instance in  $X_i$  is assigned with label  $y_i$  and weight  $\frac{1}{|X_i|} \cdot \frac{N}{m}$  such that bags of different size share the same weight of  $\frac{N}{m}$ . Then a decision tree is built on all the labeled (weighted) instances using

splitting criterion of *information gain*. Each leaf node of the induced decision tree is supposed to be related to a certain underlying concept. Therefore, the bag of instances  $X$  can be re-represented by a *meta-instance* each of its features encodes how many instances in  $X$  have fallen into a specific leaf node. In the second level, TLC employs the Logit-boosted decision stump [9] to learn from those meta-instances to discover the mappings from feature values to bag’s class label, i.e. how the bag’s label can be inferred from all the underlying concepts (*presence-*, *threshold-*, or *count-based*).

### 3.3. Data Sets

Weidmann et al. [26] have generated artificial data sets for each kind of their proposed generalized multi-instance learning problems to test the effectiveness of TLC. For any of the three different problems,  $|C|$  underlying concepts were considered. The input space  $\mathcal{X} = \{0, 1\}^l$  corresponds to bit-strings of length  $l$ . Each underlying concept corresponds to a matching template with first  $l_r$  bits fixed, i.e. an instance is a member of some concept  $c^*$  if and only if its first  $l_r$  binary attributes are identical to the specified matching template for  $c^*$ .

For presence-based multi-instance learning, a positive bag is formed by generating a number of instances (varying from 1 to 10) of each concept  $c_i$  and also a number of random instances (varying from  $10|C|$  to  $10|C| + 10$ ). The negative bag is formed by replacing all instances of some concept  $c^*$  by random instances. The minimal and maximal bag size in presence-based data set are  $|C| + 10|C|$  and  $20|C| + 10$  respectively. For threshold-based multi-instance learning, a positive bag is formed by generating a number of instances (varying from  $t_i$  to 10) of each concept  $t_i$ . The negative bag is formed by replacing at least  $\Delta(X, c^*) - t^* + 1$  instances of some concept  $c^*$  by random instances. The minimal and maximal bag size in threshold-based data set are  $\sum_i t_i + 10|C|$  and  $20|C| + 10$  respectively. For count-based multi-instance learning, the lower bound  $t_i$  and upper bound  $z_i$  are set to be equal and a positive bag is formed by generating a specific number of instances (i.e.  $z_i$ ) of each concept  $c_i$ . The negative bag is formed by either increasing or decreasing the required number  $z^*$  of some concept  $c^*$  ( $10 - z^*$  increment or  $z_i$  decrement at most). The minimal and maximal bag size in count-based data set are  $\sum_i z_i + 10|C|$  and  $\sum_i z_i + 10|C| + 10$  respectively.

## 4. Generalized Multi-Instance Learning: the Second Scenario

### 4.1. Problem Definition

In addition to the work of Weidmann et al. [26], Scott et al. [20] also studied the problem of generalizing standard

multi-instance learning scenario. They defined another type of generalized multi-instance learning framework where the underlying concepts are related to *two sets of* points. Concretely, there are a set of  $k$  *attraction* points  $P = \{p_1, p_2, \dots, p_k\}$  as well as a set of  $k'$  *repulsion* points  $Q = \{q_1, q_2, \dots, q_{k'}\}$  in the input space.

They assume that for a bag  $X$  to be labeled positively it is equivalent to find a subset of  $r$  points  $C_X \subseteq P \cup Q$  with the following properties: each attraction point  $p \in C_X$  is near *some instance* in  $X$  and each repulsion point  $q \in C_X$  is far from *any instance* of  $X$ . Therefore, the attraction and repulsion points actually act as the underlying concepts which are desirable or undesirable to be present<sup>1</sup>. For instance, an image labeled as *forests* rather than *desert* may be attributed to the existence of *trees*, *flowers* or *birds* as well as the non-existence of *rocks*, *sands* and *camel*.

### 4.2. Learning Algorithms

A series of learning algorithms have been proposed to handling the above generalized multi-instance learning problem, such as GMIL-1 [20], GMIL-2 [21],  $k_\wedge$  [22] and  $k_{\min}$  [23]. GMIL-1 [20] assumes a discretized  $d$ -dimensional input space  $\mathcal{X} = \{0, 1, \dots, s\}^d$ . All the possible axis-parallel boxes in  $\mathcal{X}$  (including degenerated ones) are enumerated (denoted as  $B_{\mathcal{X}}$  with  $|B_{\mathcal{X}}| = \binom{s+2}{2}^d$ ), each of which characterizes a possible underlying concept which is required to be present or not. Based on those boxes, a bag  $X$  can be mapped into a  $2|B_{\mathcal{X}}|$ -dimensional boolean vector  $\phi(X)$ . Specifically, for each box  $b \in B_{\mathcal{X}}$ , two boolean features  $a_b$  and  $\bar{a}_b$  are created for  $\phi(X)$  with  $a_b = \arg \max_{x \in X} [b \text{ covers } x]$  and  $\bar{a}_b = 1 - a_b$ .

With regard to the definition given in Subsection 4.1, the target function of Scott et al. [20]’s generalized multi-instance learning can therefore be expressed as an  $r$ -of- $(k + k')$  threshold function over features  $a_b$  (with box  $b$  characterizing certain concept in  $P$ ) together with features  $\bar{a}_{b'}$  (with box  $b'$  characterizing certain concept in  $Q$ ). Fortunately, the task of finding such an  $r$ -of- $(k + k')$  threshold function in  $2|B_{\mathcal{X}}|$ -dimensional boolean space can be easily fulfilled by the online-style algorithm Winnow [13]. Afterwards, Tao and Scott [21] proposed GMIL-2 which utilized some heuristics to ameliorate the computational complexity of GMIL-1 which is exponential in both  $\log s$  and  $d$ .

To cope with the rather demanding computational requirements of both GMIL-1 and GMIL-2, Tao et al. [22] proposed a kernel reformulation of previous algorithms. Firstly, given two bags  $X$  and  $Z$ , they observed that calculating the dot-product between their mapped vectors  $\phi(X)$  and  $\phi(Z)$  is equivalent to count the number of boxes in  $B(X \wedge Z)$ , where

1. Note that although neither type of the generalized multi-instance learning frameworks proposed in [26] and [20] is more expressive than the other one, they do overlap under certain circumstances [22], [30].

each box in  $B(X \wedge Z)$  covers at least one point from both  $X$  and  $Z$ . They further proved that the above box counting problem is  $\#P$ -complete and thus can't be *precisely* solved in polynomial time. Therefore, they instead *approximately* solved the box counting problem based on techniques from Karp et al. [11], which aims to estimate (at arbitrarily specified precision) the size of set  $B$  from the sizes of  $u$  sets  $\{B_1, B_2, \dots, B_u\}$  with  $B = \bigcup_{i=1}^u B_i$ . In their case,  $B(X \wedge Z) = \bigcup_{\mathbf{x} \in X, \mathbf{z} \in Z} B(\{\mathbf{x}\} \wedge \{\mathbf{z}\})$ . With the dot-product  $\langle \phi(X), \phi(Z) \rangle$  approximated by the above techniques, Tao et al. [22] then implemented a kernel reformulation named  $k_\wedge$  for the generalized multi-instance problem proposed by Scott et al. [20]. Later, this kernel is further revised to  $k_{\min}$  in light of Weidmann et al. [26]'s *count-based* generalized multi-instance problems.

### 4.3. Data Sets

Scott et al. [20] and Tao et al. [21], [22], [23] tested their algorithms on three different data sets, i.e. the image data, the protein superfamily identification data<sup>2</sup> and the multi-site drug binding affinity data. The image data corresponds to two CBIR tasks: one is a single concept task of distinguishing images containing *sunsets* from those not containing sunsets, another is a conjunctive concept task of distinguishing images containing *a field with no sky* from those containing both field and sky or containing no field. Images are represented as bags of 5-dimensional instances based on Zhang et al. [32]'s image bag generator. For single concept task, 30 different data sets are created each containing 120 positive bags and 600 negative bags; For conjunctive concept task, around 25 positive bags as well as 30 negative bags comprise the training set and around 50 positive bags as well as 670 negative bags comprise the test set.

The protein superfamily identification data was initially used by Wang et al. [24] in their investigation of identifying new Trx-fold (Thioredoxin-fold) proteins based on each protein's primary sequence. Given the primary sequence of each protein, all the motifs in the sequence (typically of the form CxxC known to exist in Trx-fold protein) are identified. Eight numeric properties [12] extracted from the enclosing window of size 204 around each motif (20 residues upstream, 180 downstream) are served as instances in the bag. The resultant data set contains 180 bags with 20 labeled as positive and 160 labeled as negative.

The multi-site drug binding affinity data is an extension of the synthetic standard multi-instance data studied by Dooly et al. [8]. Here, a molecule is regarded to be positive if it could bind to multiple sub-target receptors simultaneously. Artificially, a set of sub-target points  $T$  are generated and a bag  $X$  is labeled positively if for each sub-target point  $t \in T$ ,

there is at least one instance  $x_t$  in  $X$  whose normalized binding energy [8], [30] with  $t$  is greater than 1/2. Scott et al. [20] generated ten 5-dimensional data sets each with 4 sub-targets and 200 bags in either training or test set. Tao et al. [22] further generated ten 10 and 20-dimensional data sets each with 5 sub-targets and 200 bags in both training and test sets.

## 5. Conclusion

This paper aims to succinctly while informatively review one of the important progress in multi-instance learning, i.e. two formalizations of generalized multi-instance learning. The problem definition, learning algorithms as well as experimental data sets of either framework are briefly introduced. Designing algorithms able to deal with both generalized multi-instance learning problems and applying them to more real-world tasks would be interesting future works.

## Acknowledgements

This work was supported by the Natural Science Foundation of China (60805022), 863 High-Tech Program of China (2007AA01Z178), and the Startup Foundation for Excellent New Faculties of Hohai University.

## References

- [1] E. Alphonse and S. Matwin. Filtering multi-instance problems to reduce dimensionality in relational learning. *Journal of Intelligent Information Systems*, 22(1):23–40, 2004.
- [2] S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 561–568. MIT Press, Cambridge, MA, 2003.
- [3] P. Auer, P. M. Long, and A. Srinivasan. Approximating hyper-rectangles: learning and pseudo-random sets. *Journal of Computer and System Sciences*, 57(3):376–388, 1998.
- [4] A. Blum and A. Kalai. A note on learning from multiple-instance examples. *Machine Learning*, 30(1):23–29, 1998.
- [5] Y. Chen and J. Z. Wang. Image categorization by learning and reasoning with regions. *Journal of Machine Learning Research*, 5(Aug):913–939, 2004.
- [6] Y. Chevaleyre and J.-D. Zucker. Solving multiple-instance and multiple-part learning problems with decision trees and decision rules. Application to the mutagenesis problem. In E. Stroulia and S. Matwin, editors, *Lecture Notes in Artificial Intelligence 2056*, pages 204–214. Springer, Berlin, 2001.
- [7] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez. Solving the multiple-instance problem with axis-parallel rectangles. *Artificial Intelligence*, 39(1-2):31–71, 1997.
- [8] D. R. Dooly, Q. Zhang, S. A. Goldman, and R. A. Amar. Multiple-instance learning of real-valued data. *Journal of Machine Learning Research*, 3(Dec):651–678, 2002.

2. [http://cse.unl.edu/~qtao/datasets/mil\\_dataset\\_\\_Trx\\_protein.html](http://cse.unl.edu/~qtao/datasets/mil_dataset__Trx_protein.html).

- [9] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28(2):307–337, 2000.
- [10] X. Huang, S.-C. Chen, M.-L. Shyu, and C. Zhang. Mining high-level user concepts with multiple instance learning and relevance feedback for content-based image retrieval. In O. R. Zaïane, S. J. Simoff, and C. Djeraba, editors, *Lecture Notes in Artificial Intelligence 2797*, pages 50–67. Springer, Berlin, 2002.
- [11] R. Karp, M. Luby, and N. Madras. Monte-carlo approximation algorithms for enumeration problems. *Journal of Algorithms*, 10(3):429–448, 1989.
- [12] J. Kim, E. N. Moriyama, C. G. Warr, P. J. Clyne, and J. R. Carlson. Identification of novel multi-transmembrane proteins from genomic databases using quasi-periodic structural properties. *Bioinformatics*, 16(9):767–775, 2000.
- [13] N. Littlestone. Redundant noisy attributes, attribute errors, and linear threshold learning using winnow. In *Proceedings of the 4th Annual Workshop on Computational Learning Theory*, pages 147–156, Santa Cruz, CA, 1991.
- [14] P. M. Long and L. Tan. PAC learning axis-aligned rectangles with respect to product distribution from multiple-instance examples. *Machine Learning*, 20(1):7–21, 1998.
- [15] O. Maron and T. Lozano-Pérez. A framework for multiple-instance learning. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors, *Advances in Neural Information Processing Systems 10*, pages 570–576. MIT Press, Cambridge, MA, 1998.
- [16] O. Maron and A. L. Ratan. Multiple-instance learning for natural scene classification. In *Proceedings of the 15th International Conference on Machine Learning*, pages 341–349, Madison, WI, 1998.
- [17] R. Rahmani and S. A. Goldman. Missl: multiple-instance semi-supervised learning. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 705–712, Pittsburgh, PE, 2006.
- [18] S. Ray and D. Page. Multiple instance regression. In *Proceedings of the 18th International Conference on Machine Learning*, pages 425–432, Williamstown, MA, 2001.
- [19] G. Ruffo. *Learning single and multiple decision trees for security applications*. PhD thesis, Department of Computer Science, University of Turin, Italy, 2000.
- [20] S. Scott, J. Zhang, and J. Brown. On generalized multiple-instance learning. Technical Report UNL-CSE-2003-5, Department of Computer Science and Engineering, University of Nebraska, Lincoln, NE, 2003.
- [21] Q. Tao and S. Scott. A faster algorithm for generalized multiple-instance learning. In *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference*, pages 550–555, Miami Beach, FL, 2004.
- [22] Q. Tao, S. Scott, N. V. Vinodchandran, and T. T. Osugi. Svm-based generalized multiple-instance learning via approximate box counting. In *Proceedings of the 21st International Conference on Machine Learning*, pages 799–806, Banff, Canada, 2004.
- [23] Q. Tao, S. Scott, N. V. Vinodchandran, T. T. Osugi, and B. Mueller. An extended kernel for generalized multiple-instance learning. In *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence*, pages 272–277, Boca Raton, FL, 2004.
- [24] C. Wang, S. D. Scott, J. Zhang, Q. Tao, D. Fomenko, and V. Gladyshev. A study in modeling low-conservation protein superfamilies. Technical Report UNL-CSE-2004-0003, Department of Computer Science and Engineering, University of Nebraska, Lincoln, NE, 2004.
- [25] J. Wang and J.-D. Zucker. Solving the multiple-instance problem: a lazy learning approach. In *Proceedings of the 17th International Conference on Machine Learning*, pages 1119–1125, Stanford, CA, 2000.
- [26] N. Weidmann, E. Frank, and B. Pfahringer. A two-level learning method for generalized multi-instance problem. In N. Lavrač, D. Gamberger, H. Blockeel, and L. Todorovski, editors, *Lecture Notes in Artificial Intelligence 2837*, pages 468–579. Springer, Berlin, 2003.
- [27] X. Xu and E. Frank. Logistic regression and boosting for labeled bags of instances. In H. Dai, R. Srikant, and C.-Q. Zhang, editors, *Lecture Notes in Computer Science 3056*, pages 272–281. Springer, Berlin, 2004.
- [28] C. Yang and T. Lozano-Pérez. Image database retrieval with multiple-instance learning techniques. In *Proceedings of the 16th International Conference on Data Engineering*, pages 233–243, San Diego, CA, 2000.
- [29] M.-L. Zhang and Z.-H. Zhou. Adapting RBF neural networks to multi-instance learning. *Neural Processing Letters*, 23(1):1–26, 2006.
- [30] M.-L. Zhang and Z.-H. Zhou. Multi-instance clustering with applications to multi-instance prediction. *Applied Intelligence*, in press.
- [31] Q. Zhang and S. A. Goldman. EM-DD: an improved multiple-instance learning technique. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, pages 1073–1080. MIT Press, Cambridge, MA, 2002.
- [32] Q. Zhang, W. Yu, S. A. Goldman, and J. E. Fritts. Content-based image retrieval using multiple-instance learning. In *Proceedings of the 19th International Conference on Machine Learning*, pages 682–689, Sydney, Australia, 2002.
- [33] Z.-H. Zhou. Multi-instance learning: a survey. Technical report, AI Lab, Department of Computer Science & Technology, Nanjing University, Nanjing, China, 2004.
- [34] Z.-H. Zhou, K. Jiang, and M. Li. Multi-instance learning based web mining. *Applied Intelligence*, 22(2):135–147, 2005.
- [35] Z.-H. Zhou and M.-L. Zhang. Ensembles of multi-instance learners. In N. Lavrač, D. Gamberger, H. Blockeel, and L. Todorovski, editors, *Lecture Notes in Artificial Intelligence 2837*, pages 492–502. Springer, Berlin, 2003.
- [36] Z.-H. Zhou and M.-L. Zhang. Multi-instance multi-label learning with application to scene classification. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 1609–1616. MIT Press, Cambridge, MA, 2007.