

# Combining Active Learning and Graph-Based Semi-Supervised Learning

Jhonatan de Paula Candao, Lilian Berton<sup>1</sup>

<sup>1</sup>Institute of Science and Technology – Federal University of São Paulo (UNIFESP)  
São José dos Campos – SP – Brazil

jhonatan.candao@unifesp.br, lberton@unifesp.br

**Abstract.** *The scarcity of labeled data is a common problem in many applications. Semi-supervised learning (SSL) aims to minimize the need for human annotation combining a small set of label data with a huge amount of unlabeled data. Similarly to SSL, Active Learning (AL) reduces the annotation efforts selecting the most informative points for annotation. Few works explore AL and graph-based SSL, in this work, we combine both strategies and explore different techniques: two graph-based SSL and two query strategy of AL in a pool-based scenario. Experimental results in artificial and real datasets indicate that our approach requires significantly less labeled instances to reach the same performance of random label selection.*

Machine learning. semi-supervised learning. active learning. label propagation.

## 1. Introduction

Machine Learning (ML) is a subarea of Artificial Intelligence (AI) that deals with the development and application of methods capable of predicting or describe situations from past experiences, learned through data sets. In one of his main works, [Mitchell 1997] defines ML as a research area that uses computer programs that can automatically improve its performance through experience.

In many practical applications of ML, such as classification or clustering, unlabeled examples are readily available, but labeled ones are relatively expensive to obtain since labeling training data often requires a physical experiment or a statistical trial, and human labor is usually required. Therefore, for many analysts, it is very expensive to fully label the entire training data set. Therefore, semi-supervised learning (SSL) algorithms have aroused considerable interest. Graph-based semi-supervised learning has become one of the most active areas of research in the SSL community [Chapelle et al. 2010, Zhu 2005, Vega-Oliveros et al. 2014, Berton et al. 2015, Berton et al. 2017, Berton et al. 2018].

However, SSL does not employ any process to select better label data. This way, we can employ active learning (AL) to select unlabeled instances to be labeled by a domain expert and use them in the SSL. Active learning aims to determine the best set of labeled data for classification performance by actively querying the labeled data [Settles 2010]. This approach gives better results than selecting label data randomly. The effectiveness of AL has been shown in typical classification tasks such as automatic speech recognition [Riccardi and Hakkani-Tur 2005], multimedia retrieval [Wang and Hua 2011], speech emotion recognition [Zhang et al. 2015], among others.

In this paper, we exploit active learning and graph-based SSL combining different SSL algorithms and AL query strategy. We considered the pool-scenario where every instance in the pool can be evaluated in terms of their informativeness and chosen for the training iteration. We employ two query strategies: certainty-based sampling and query-by-committee. Finally, the algorithms label propagation [Zhu and Ghahramani 2002] and local and global consistency [Zhou et al. 2003] were used for classifying the unlabeled points.

The remaining of this paper is organized as follows: Section 2 presents some related work that used SSL and active learning. Section 3 presents the materials and methods used in this work. Section 4 presents the experimental results and finally Section 5 presents the final remarks.

## 2. Related work

[Calma et al. 2018] combined active and semi-supervised learning for support vector machines (SVM) in classification tasks. Structure information is captured employing probabilistic models that are iteratively improved at runtime when label information becomes available. The probabilistic models are considered in a selection strategy based on distance, density, diversity, and distribution (4DS strategy) information for active learning in a kernel function for SVM.

Few works have combined active learning and graph-based SSL. [Zhu et al. 2003] proposed a framework based on Gaussian random fields and harmonic function to estimate the expected generalization error after querying a point, which leads to a better query selection criterion than naively selecting the point. [Chen and Wang 2017] developed a selective semi-supervised learning algorithm, called selective label spreading (SLS) by integrating the active learning model into the label spreading framework. SLS optimizes the process of selecting better-labeled data to improve classification performance.

Several applications have been reported: [Han et al. 2016] combined active learning and self-training intending to minimize the need for human annotation for sound classification model training. [Tomanek and Hahn 2009] employed semi-supervised self-training with active learning for sequence labeling and strongly reduces the number of tokens to be manually annotated. [Chellapriyadharshini et al. 2018] proposed an acoustic-model refinement using semi-supervised and active learning for a low resource Indian language.

## 3. Materials and methods

This section describes the algorithms employed to perform the semi-supervised and active learning experiments, and the datasets used.

### 3.1. Semi-supervised learning (SSL)

Semi-supervised learning (SSL) uses  $l$  labeled points  $L = \{(x_1, y_1), \dots, (x_l, y_l)\}$  and  $u$  unlabeled points  $U = (x_{l+1}, \dots, x_{l+u})$ ,  $l \ll u$  and  $n = l + u$ . A connected graph is generated  $G = (V, E, W)$  with  $V$  corresponding to the data points, and  $E$  corresponding to the edges and represented by a weight matrix  $W$  which usually is represented by the

radial basis function (RBF) that assigns large weights to nearby points in Euclidean space:

$$w_{ij} = \exp\left(-\frac{1}{\sigma^2} \sum_{d=1}^m (x_{id} - x_{jd})^2\right) \quad (1)$$

The Label Propagation (LP) [Zhu and Ghahramani 2002] algorithm presented in Algorithm 1 operates iteratively and uses the graph structure to gradually propagate the labels of the nodes that are labeled to the non-labeled. Thus, at each step, the node spreads its label to its neighboring nodes until it reaches convergence. The label of each unlabeled point will be the class of which it has received most information during the iteration process.

The probability that a vertex  $i$  has to pass its label to a vertex  $j$  is given by the probabilistic transition matrix, according to the Equation 2.

$$T_{ij} = P(j \rightarrow i) = \frac{W_{ij}}{\sum_k 1_{kj}^{l+u}} \quad (2)$$

---

**Algorithm 1:** Algorithm *Label Propagation* [Zhu and Ghahramani 2002]

---

Calculate the adjacency matrix  $W$  according to the Equation 1;  
 Calculate the matrix  $T$  according to the Equation 2;  
 Initialize the labels  $Y^{(0)} \leftarrow (y_1, y_2, \dots, y_l, 0, \dots, 0)$ ;  
 achieve convergence;  $Y^{(t+1)} \leftarrow TY^{(t)}$ ;  
 $Y_l^{(t+1)} = Y_l$ ; Label the sample  $x_i$ , according to the sign of  $Y_i$ ;

---

Another popular label propagation algorithm is the Local and Global Consistency (LGC) [Zhou et al. 2003] presented in Algorithm 2. The parameter  $\alpha$  specifies the relative amount of the information each point receives from its neighbors (first term) and its initial label information (second term). The label of each unlabeled point will be the class of which it has received most information during the iteration process.

---

**Algorithm 2:** Algorithm *Local and Global Consistency* [Zhou et al. 2003]

---

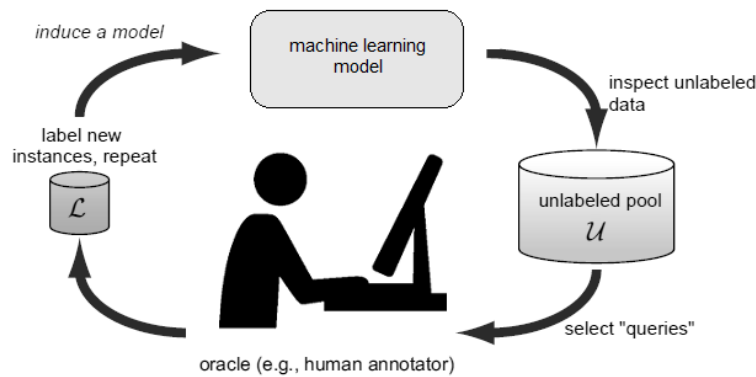
Calculate the adjacency matrix  $W$  according to the Equation 1;  
 Calculate the diagonal matrix  $D$ , using  $D_{ij} = \sum_{j=0}^n d_{ij}$ ;  
 Calculate the normalized Laplacian of the graph, where  $L = D^{\frac{1}{2}}WD^{\frac{1}{2}}$ ;  
 Choose a parameter  $\alpha \in [0,1)$ ;  
 Initialize  $Y^0 = (y_1, y_2, \dots, y_l, 0, 0, \dots, 0)$ ;  
**while** until  $f$  converge **do**  
 |  $f(t+1) \leftarrow \alpha Lf(t) + (1 - \alpha)Y$ ;  
**end**  
 Label the examples  $x_i$  with label  $c = \operatorname{argmax}_{1 \leq j \leq c} f_{ij}$ , where  $f_{ij}$  is the solution of the algorithm;

---

### 3.2. Active learning

Active learning (AL) aims to exploit the informativeness of unlabeled data by selecting those with high potential to improve the classification's algorithm performance for annotation. There are various strategies by which such informativeness can be processed.

The pool-based active learning usually starts with a large pool  $U$  of unlabeled samples and a small set of labeled samples  $L$  (with  $|L| \ll |U|$  and  $X = U \cup L$ ). Then, the number of labeled samples  $L$  is increased to train a classifier  $C$ . In each query round  $i$ , a query set  $S_i$  of unlabeled, informative samples is determined using a selection strategy  $Q$ , which takes into account the “knowledge” of the classifier. These samples are presented to an oracle (or a human expert) to be labeled. Then,  $S_i$  is added to  $L$ , removed from  $U$ , and  $C$  is updated. If a given stopping criterion is met, the process stops, otherwise the next query round  $i + 1$  starts. Figure 1 presents these cycle.



**Figure 1. The pool-based active learning cycle.**

There are various query strategies that process informativeness. Three categories are most common [Settles 2010]:

1. Certainty-based sampling: the active learner determines the certainty of the predictions on unlabeled data based on a previously trained model, and queries an annotator for the labeling of those with the least certain classification.
2. Query-by-committee: involves two or more classifiers and the selection of those instances about which the various models disagree the most are then delivered for human annotation.
3. Expected error reduction: is a decision-theoretic approach that aims to estimate how much the model’s generalization error is likely to be reduced. The instances estimated to have a high impact on the expected model’s error are selected for human annotation. In most cases, this method is the most computationally expensive.

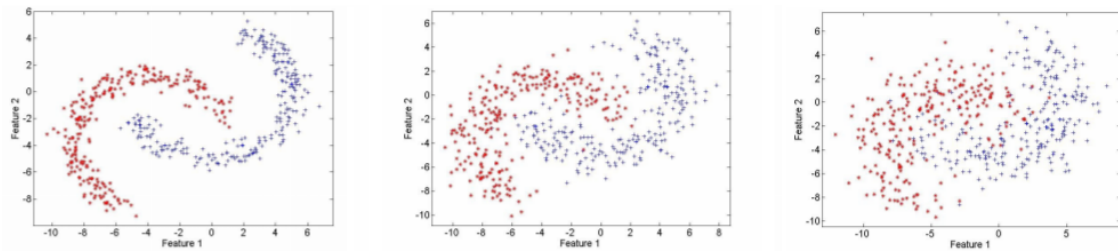
### 3.3. Datasets

Three artificial datasets and five real datasets: Iris, Balance scale, Ecoli, Banknote authentication and Contraceptive Method Choice datasets available from the UCI Machine Learning Repository<sup>1</sup> were employed. We used datasets with a different number of classes and class distribution to test the effectiveness of the algorithms.

The three datasets Two-moons are artificially generated in which the instances belong to two clusters with a banana shape, the difference among them is the overlapping

<sup>1</sup><https://archive.ics.uci.edu/ml/index.php>

among the classes as shown in Figure 2. Iris dataset contains numerical data that characterize the Iris flowers, where each may belong to one of three distinct classes: iris setosa, iris virginica and iris versicolor. The Balance scale dataset was generated to model psychological experimental results. Each example is classified as having the balance scale tilted to the right, tilted to the left or balanced. The Ecoli dataset contains information about Escherichia coli. It is a bacterium of the genus Escherichia that is commonly found in the small intestine of the warm-blooded organism and has eight classes. The dataset banknote authentication represents images taken from genuine and forged specimens of banknotes. The Contraceptive Method Choice is a dataset from Indonesia’s National Contraceptive Prevalence Survey realized in 1987. It is about the current choice of contraceptive methods (unused, long-term or short-term). Table 1 summarizes some information about the datasets.



**Figure 2. Data set Two-moons 1, 2 and 3.**

**Table 1. Dataset information.**

Dataset	# Examples	# Classes	Class distribution
Two-moons1	500	2	250,250
Two-moons2	500	2	250,250
Two-moons3	500	2	250,250
Iris	150	3	50,50,50
Balance Scale	625	3	288,49,288
Ecoli	336	8	2,2,35,5,143,20,77,52
Banknote authentication	1372	2	762,610
Contraceptive Method Choice	1473	3	629,333,511

## 4. Experiments

In this section, we describe the experiments conducted with the purpose of empirically investigating the effectiveness of combining AL and SSL learning methods.

### 4.1. Experimental setup

In the experiments a subset of instances is selected for “human labeling”, and the latter for machine labeling. We consider two approaches: 1) The label sampling  $L$  is randomly selected. 2) The label sampling  $L$  is selected employing active learning. In both cases, the label set  $L$  together with the unlabeled sampling  $U$  are employed in semi-supervised learning.

In the random label sampling, we select a different percentage of label data to perform semi-supervised learning (1, 5, 10, 15, 20 and 25 %). Except for the artificial

dataset we need to start with 10% of labels because few labels did not converge. Each execution of the experiment consists of the repetition of LP and LGC algorithms 30 times per dataset and per percentage of the label. At the end of the 30 runs, the arithmetic mean accuracy (Acc), the standard deviation (std) and the F1 score are obtained.

In the active learning label sampling, we start with a small set of random labeled instances  $L$  for training an initial classifier. With this classifier, we estimate the confidence scores  $S$  for the instances that are candidates for labeling. In the pool-based scenario, the entire pool of unlabeled instances  $U$  is estimated, and only those instances with confidence scores lower than a pre-defined threshold are selected for human annotation. We also select a different percentage of label data to perform semi-supervised learning (1, 5, 10 %), but in this case, a less percentage of the label already leads to high accuracy. For the Query-by-committee we execute 30 runs and present the arithmetic mean accuracy (Acc), the standard deviation (std) and the F1 score. In the Certainty-based-sampling, we do not need to execute many times since there is no randomness in the method. The arithmetic mean accuracy (Acc) and the F1 score are the results of one execution.

## 4.2. Results

Tables 2, 3, 4, 5, 6, 7, 8 and 9 present the classification results - Accuracy (Acc), Standard deviation (std) and F1 Score - considering random and active learning label selection.

In the artificial datasets as the overlapping among classes becomes high the accuracy decreases. This way, Two-moons3 achieved low accuracy than Two-moons2 and Two-moons1, respectively. In general, active learning needs fewer label data to achieve high accuracy, for example, in the Two-moons3 even with 25% of label data the accuracy of random sampling is small than active learning.

In the real datasets, the same occurs and we need a high percentage of labeled data in the random sampling (25%) to achieve an accuracy similar to active learning with less labeled data (5 or 10%). Only for the dataset Contraceptive Method Choice, it didn't get high accuracy with this percentage of the label since this dataset is hard to classify [Lim et al. 2000].

The results show the effectiveness of combine active learning with SSL in order to select better label data and achieve better classification results. In AL the number of examples to learn are often much lower than the number required in random case. In general, both LP and LGC benefit from the AL label sampling. The active learning Certainty-based learning achieved the high results in most datasets and it demands less execution since it selects those points where the current model is least certain, while the Query-by-committee train a variety of models and label those points for which the committee disagrees the most.

**Table 2. Classification results for dataset Two-moons1.**

% of labels	LP		LGC	
	Acc and std	F1 Score	Acc and std	F1 Score
Random				
10	0,96 ± 0,01	0,96	0,99 ± 0,00	0,99
15	0,99 ± 0,00	0,99	1,00 ± 0,00	1,00
20	0,99 ± 0,00	0,99	1,00 ± 0,00	1,00
25	0,99 ± 0,00	0,99	1,00 ± 0,00	1,00
Query-by-committe				
1	0,96 ± 0,05	0,96	0,96	0,96
5	0,99 ± 0,00	0,99	1,00	1,00
10	0,99 ± 0,00	0,99	1,00	1,00
Certainty-based sampling				
1	0,93	0,93	0,99	0,99
5	1,00	1,00	1,00	1,00

**Table 3. Classification results for dataset Two-moons2.**

% of labels	LP		LGC	
	Acc and std	F1 Score	Acc and std	F1 Score
Random				
10	0,97 ± 0,01	0,97	0,96 ± 0,01	0,96
15	0,97 ± 0,01	0,97	0,97 ± 0,01	0,97
20	0,97 ± 0,01	0,97	0,97 ± 0,01	0,97
25	0,98 ± 0,01	0,98	0,98 ± 0,01	0,98
Query-by-committe				
1	0,81 ± 0,07	0,81	0,81	0,81
5	0,98 ± 0,00	0,98	0,98	0,98
10	0,98 ± 0,00	0,98	0,98	0,98
Certainty-based sampling				
1	0,91	0,91	0,92	0,92
5	0,99	0,99	0,98	0,98
10	0,99	0,99	0,99	0,99

**Table 4. Classification results for dataset Two-moons3.**

% of labels	LP		LGC	
	Acc and std	F1 Score	Acc and std	F1 Score
Random				
10	0,88 ± 0,02	0,88	0,89 ± 0,03	0,88
15	0,89 ± 0,02	0,89	0,89 ± 0,01	0,89
20	0,90 ± 0,02	0,90	0,90 ± 0,02	0,90
25	0,90 ± 0,02	0,90	0,90 ± 0,01	0,90
Query-by-committe				
1	0,64 ± 0,08	0,64	0,66	0,66
5	0,89 ± 0,02	0,89	0,88	0,88
10	0,92 ± 0,01	0,92	0,93	0,93
Certainty-based sampling				
1	0,83	0,83	0,83	0,83
5	0,93	0,93	0,93	0,93
10	0,93	0,93	0,93	0,93

**Table 5. Classification results for dataset Iris.**

% of labels	LP		LGC	
	Acc and std	F1 Score	Acc and std	F1 Score
Random				
10	0,93 ± 0,07	0,89	0,93 ± 0,07	0,93
15	0,94 ± 0,03	0,91	0,95 ± 0,03	0,95
20	0,96 ± 0,02	0,96	0,97 ± 0,02	0,97
25	0,97 ± 0,02	0,97	0,98 ± 0,01	0,98
Query-by-committe				
1	0,47 ± 0,16	0,47	0,54	0,53
5	0,94 ± 0,02	0,93	0,93	0,93
10	0,97 ± 0,01	0,97	0,97	0,97
Certainty-based sampling				
1	0,67	0,67	0,67	0,67
5	0,97	0,97	0,94	0,94
10	0,99	0,99	0,99	0,99

**Table 6. Classification results for dataset Ecoli.**

% of labels	LP		LGC	
	Acc and std	F1 Score	Acc and std	F1 Score
Random				
10	0,66 ± 0,05	0,38	0,79 ± 0,04	0,40
15	0,74 ± 0,05	0,36	0,80 ± 0,02	0,44
20	0,76 ± 0,04	0,46	0,83 ± 0,02	0,44
25	0,78 ± 0,03	0,41	0,83 ± 0,02	0,63
Query-by-committe				
1	0,64 ± 0,08	0,25	0,68	0,28
5	0,81 ± 0,02	0,56	0,83	0,57
10	0,85 ± 0,02	0,66	0,87	0,63
Certainty-based sampling				
1	0,73	0,37	0,73	0,37
5	0,77	0,39	0,86	0,64
10	0,87	0,66	0,88	0,66

**Table 7. Classification results for dataset Banknote authentication.**

% of labels	LP		LGC	
	Acc and std	F1 Score	Acc and std	F1 Score
Random				
10	0,99 ± 0,01	0,99	0,99 ± 0,02	0,99
15	0,99 ± 0,00	0,99	0,99 ± 0,02	0,99
20	0,99 ± 0,00	0,99	0,99 ± 0,02	0,99
25	0,99 ± 0,00	0,99	0,99 ± 0,02	0,99
Query-by-committe				
1	0,96 ± 0,01	0,96	0,97	0,97
5	1,00 ± 0,00	1,00	1,00	1,00
Certainty-based sampling				
1	0,97	0,97	0,93	0,92
5	1,00	1,00	1,00	1,00



**Table 8. Classification results for dataset Balance Scale.**

	LP		LGC	
Random				
% of labels	Acc and std	F1 Score	Acc and std	F1 Score
10	0,82 ± 0,02	0,65	0,81 ± 0,02	0,63
15	0,84 ± 0,02	0,66	0,84 ± 0,01	0,71
20	0,86 ± 0,02	0,67	0,85 ± 0,01	0,68
25	0,87 ± 0,02	0,71	0,86 ± 0,01	0,71
Query-by-committe				
1	0,75 ± 0,04	0,54	0,75	0,54
5	0,82 ± 0,02	0,62	0,82	0,52
10	0,85 ± 0,01	0,69	0,85	0,69
15	0,87 ± 0,01	0,70	0,75	0,70
Certainty-based sampling				
1	0,67	0,48	0,67	0,48
5	0,74	0,61	0,74	0,61
10	0,85	0,74	0,85	0,74
15	0,87	0,83	0,87	0,83

**Table 9. Classification results for dataset Contraceptive Method Choice.**

	LP		LGC	
% of labels	Acc and std	F1 Score	Acc and std	F1 Score
Random				
10	0,50 ± 0,01	0,47	0,48 ± 0,01	0,46
15	0,54 ± 0,01	0,51	0,52 ± 0,01	0,49
20	0,56 ± 0,01	0,53	0,55 ± 0,01	0,54
25	0,59 ± 0,01	0,55	0,59 ± 0,02	0,57
Query-by-committe				
1	0,45 ± 0,02	0,42	0,45	0,42
5	0,53 ± 0,01	0,52	0,52	0,51
10	0,60 ± 0,01	0,58	0,60	0,58
15	0,65 ± 0,01	0,64	0,65	0,63
Certainty-based sampling				
1	0,38	0,37	0,39	0,38
5	0,46	0,45	0,45	0,45
10	0,54	0,52	0,50	0,48
15	0,57	0,56	0,54	0,52

## 5. Conclusion

In this paper, we demonstrated the effectiveness of combining graph-based semi-supervised and active learning in order to reduce the need for human annotations. Two query strategies of AL and two SSL label propagation algorithms were used with different percentage of labeled points (1, 5, 10, 15, 20 e 25 %).

We show that AL achieved a high accuracy with less labeled data compared to random label selection. Certainty-based sampling leads to better accuracy with the same label portion compared to Query-by-committee and random.

As future work more AL strategies can be employed as well more datasets of specific applications such as text or image classification.

## 6. Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and São Paulo Research Foundation (FAPESP) grant #2018/01722-3.

## References

- Berton, L., de Andrade Lopes, A., and Vega-Oliveros, D. A. (2018). A comparison of graph construction methods for semi-supervised learning. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.
- Berton, L., de Paulo Faleiros, T., Valejo, A., Valverde-Rebaza, J., and de Andrade Lopes, A. (2017). Rgcli: Robust graph that considers labeled instances for semi-supervised learning. *Neurocomputing*, 226:238 – 248.
- Berton, L., Valverde-Rebaza, J., and de Andrade Lopes, A. (2015). Link prediction in graph construction for supervised and semi-supervised learning. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.
- Calma, A., Reitmaier, T., and Sick, B. (2018). Semi-supervised active learning for support vector machines: A novel approach that exploits structure information in data. *Information Sciences*, 456:13–33.
- Chapelle, O., Schlkopf, B., and Zien, A. (2010). *Semi-Supervised Learning*. The MIT Press, 1st edition.
- Chellapriyadharshini, M., Toffy, A., Raghavan K. M., S., and Ramasubramanian, V. (2018). Semi-supervised and active-learning scenarios: Efficient acoustic model refinement for a low resource indian language. pages 1041–1045.
- Chen, X. and Wang, T. (2017). Combining active learning and semi-supervised learning by using selective label spreading. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 850–857.
- Han, W., Coutinho, E., Ruan, H., Li, H., Schuller, B., Yu, X., and Zhu, X. (2016). Semi-supervised active learning for sound classification in hybrid learning environments. *PLOS ONE*, 11(9):1–23.
- Lim, T.-S., Loh, W.-Y., and Shih, Y.-S. (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning*, 40(3):203–228.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edition.
- Riccardi, G. and Hakkani-Tur, D. (2005). Active learning: theory and applications to automatic speech recognition. *IEEE Transactions on Speech and Audio Processing*, 13(4):504–511.
- Settles, B. (2010). Active learning literature survey. Technical report, University of Wisconsin–Madison.
- Tomanek, K. and Hahn, U. (2009). Semi-supervised active learning for sequence labeling. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the*

- 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ACL '09, pages 1039–1047, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Vega-Oliveros, D. A., Berton, L., Eberle, A. M., de Andrade Lopes, A., and Zhao, L. (2014). Regular graph construction for semi-supervised learning. *Journal of Physics: Conference Series*, 490:012022.
- Wang, M. and Hua, X.-S. (2011). Active learning in multimedia annotation and retrieval: A survey. *ACM Trans. Intell. Syst. Technol.*, 2(2):10:1–10:21.
- Zhang, Y., Coutinho, E., Zhang, Z., Quan, C., and Schuller, B. (2015). Dynamic active learning based on agreement and applied to emotion recognition in spoken interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, pages 275–278, New York, NY, USA. ACM.
- Zhou, D., Bousquet, O., Lal, T. N., Weston, J., and Schölkopf, B. (2003). Learning with local and global consistency. In *Proceedings of the 16th International Conference on Neural Information Processing Systems*, NIPS'03, pages 321–328, Cambridge, MA, USA. MIT Press.
- Zhu, X. (2005). Semi-supervised learning literature survey. Technical Report 1530, Computer Sciences, University of Wisconsin-Madison.
- Zhu, X. and Ghahramani, Z. (2002). Learning from labeled and unlabeled data with label propagation. Technical Report CMU-CALD-02-107, Carnegie Mellon University.
- Zhu, X., Lafferty, J., and Ghahramani, Z. (2003). Combining active learning and semi-supervised learning using gaussian fields and harmonic functions. In *ICML 2003 workshop on The Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining*, pages 58–65.