POSITIVE AND UNLABELED LEARNING ALGORITHMS AND APPLICATIONS: A SURVEY

Kristen Jaskie and Andreas Spanias

SenSIP Center, School of ECEE, Arizona State University, USA

ABSTRACT

This paper will address the Positive and Unlabeled learning problem (PU learning) and its importance in the growing field of semi-supervised learning. In most real-world classification applications, well labeled data is expensive or impossible to obtain. We can often label a small subset of data as belonging to the class of interest. It is frequently impractical to manually label all data we are not interested in. We are left with a small set of positive labeled items of interest and a large set of unknown and unlabeled data. Learning a model for this is the PU learning problem.

In this paper, we explore several applications for PU learning including examples in biological/medical, business, security, and signal processing. We then survey the literature for new and existing solutions to the PU learning problem.

Index Terms—PU learning, positive unlabeled learning, machine learning, artificial intelligence, classification
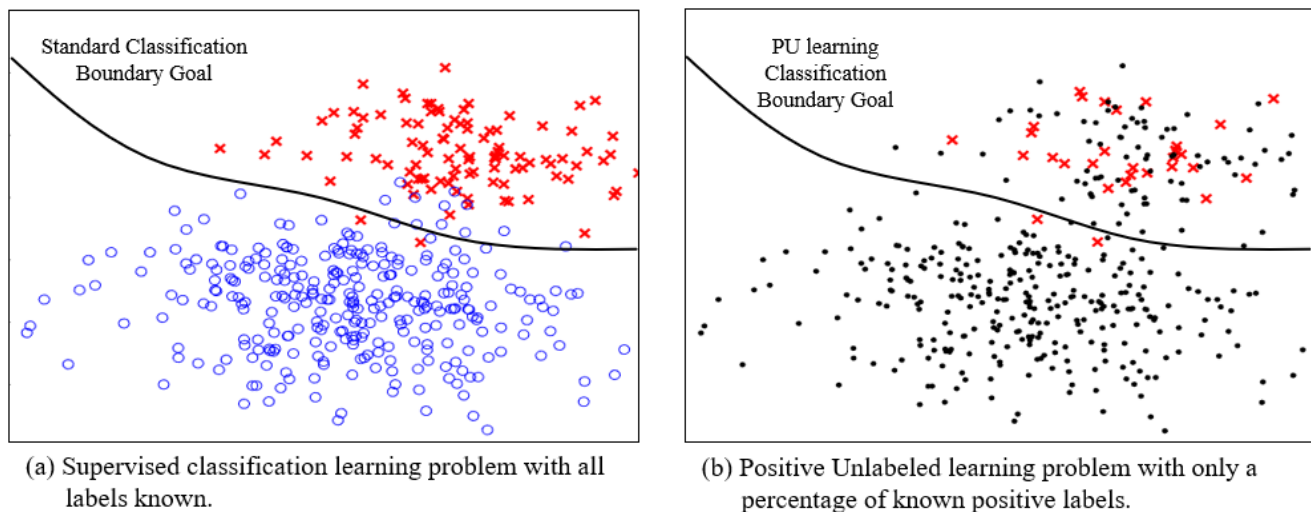
1. INTRODUCTION

In the modern world, big data and machine learning provide many opportunities for knowledge extraction. Data accumulation is exploding. By some estimates, 2.5 quintillion bytes of data are generated daily at our current pace, and some 90% of the data in the world was generated in the last two years [1]. The Internet of Things (IoT) is only accelerating this growth [2], [3]. Machine learning algorithms use this data to identify patterns, make connections, and learn representative models.

The most frequently used machine learning algorithms are supervised or unsupervised. Supervised learning is used when the "ground truth" or labels of the data are known. Given the data and the known labels, a model can be created to predict the unknown label of a new, unfamiliar data sample. When no labels are available for training, unsupervised learning methods are used [4]–[6]. See section 5.1 for additional information and resources on general Machine Learning.

A common task in machine learning is supervised binary classification. Given data sample and binary label pairs (x, y), where y is typically considered either positive ($y = 1$) or negative ($y = 0$), a classification algorithm learns a model $f(x)$ from the features of these labeled samples. Given a new data sample with no label, this model then places that unlabeled sample into either the positive or the negative class. Several algorithms work well with this problem, including logistic regression, support vector machines (SVMs), and artificial neural networks (ANNs) among others.

While enormous quantities of data are accumulated in the modern world, labeling these can be prohibitively expensive and time-consuming, if even possible, making supervised algorithms unusable. Semi-supervised learning addresses the problem where labels are not known for all training samples. Many real-world classification problems inherently have a small number of known positive data, no known negative data, and a large quantity of unlabeled data. This is known as the Positive and Unlabeled learning problem (PU learning). The difference between traditional binary classification and PU classification is illustrated in Figure 1.

(a) Supervised classification learning problem with all labels known.

(b) Positive Unlabeled learning problem with only a percentage of known positive labels.

Figure 1: Illustration of a Positive Unlabeled binary classification problem.

One application of PU learning involves remote sensing - including identifying objects in satellite images. In a database of satellite images, some number of images may be labeled as containing the item of interest, but there are too many images for all to be labeled. We would like to classify all the images in the database as either containing the item of interest or not. This and many more PU learning applications will be described in more detail in Section 2.

This paper will address the PU learning problem as follows. Section 2 will describe several applications for PU learning. Section 3 will discuss some of the assumptions that are fundamental to several of the problem solutions. Section 4 surveys the existing methods of solving the PU learning problem. Section 5 concludes and provides a brief overview into Machine Learning along with suggested reading and a discussion of future work and opportunities.

2. POSITIVE AND UNLABELED APPLICATIONS

A surprisingly large number of real-world classification problems naturally fall into the Positive and Unlabeled learning scenario (PU learning). This section provides a brief survey of those topics found in the literature and ideas for potential future applications. Current uses emphasize medical/biological and business applications. Signal processing and security applications are underrepresented in the literature and are discussed in section 2.3.

2.1 Medical and Biological Applications

Many medical situations are a natural fit for the PU learning problem. One example is identifying [7]–[9] or priority ranking [10]–[12] genes or gene combinations that influence disease incidence. Many diseases including cancer, Alzheimer's disease, cystic fibrosis, sickle cell anemia, and even anxiety disorders are influenced by genetics. A small set of genes are known to cause or influence the specified diseases - these compose the positive set, and very little is known about how all other genes influence the disease of interest - these are unknown. In addition to genetic analysis, diseases such as cancer may be detectable by examining general patient records including blood test results and patient history [13]–[15].

Virtual screening of drug compounds – identifying which drug compounds could be useful in treating a given disease is another application for PU learning [11]. Drugs and chemicals known to be effective against a given disease make up the positive set. All other compounds in a drug or chemical database make up the unknown set. [16] uses PU learning to identify transport proteins from a general protein database. [17] reconstructs gene regulatory networks from gene expression data, identifying which gene pairs are most likely to interact. [18] identifies small non-coding RNA (ncRNA) genes from intergenic sequences. A related, though more difficult learning problem – that of learning only from the proportions of the positive and unlabeled sets – is being used for embryo selection in assisted reproduction [19]. In this scenario, only the most viable embryos should be implanted, and a variant of PU learning can be used to do this. Growing medical fields concerning gut microbiome analysis and epigenetics could also benefit from this type of learning.

PU learning for ecological [20] and environmental monitoring [21], [22] are scenarios that have been lightly touched upon in the literature. [20] discusses using PU learning to identify species presence. Determining species absence in a region is difficult and expensive. Geographical regions with reported sightings of the species of interest make up the positive set, while all other areas remain in the unknown set.

2.2 Business Applications

Some of the best studied applications for PU learning involve text classification or document classification [10], [23], [32]–[37], [24]–[31]. This can include categorizing the subject of a paper, webpage, or email. One common application is email spam identification. Users identify some emails as spam, and these make up the positive class. All other emails are considered unknown. Search topic identification/classification [10] and web page text retrieval and classification [11], [15], [30]–[32] are other important text applications.

Recommender systems can be considered as PU learning opportunities. For web page recommenders, a user's browsing history or bookmarks make up the positive set. All other webpages constitute the unknown set. From this, web pages of interest can be recommended [13], [33], [38]. This could also be useful in recommending movies/tv shows or social media contacts or posts that a user would enjoy. Every show, post, or group that is "liked" by a user would make up the positive set, while all others would be unknown. [39] attempts to predict which subjects would interest a politician by looking at their past work. Recommender systems can also suffer from deceptive reviews. [40] and [41] use PU learning to identify these.

Reject inference for loan approval and other tasks learns from the applications of both accepted and rejected individuals. The behavior of rejected individuals is unknown and thus fits within the Positive and Unlabeled framework. [42] applies PU learning to reject inference problems which can include epidemiology, econometrics, and clinical trial evaluation along with the more standard financial applications.

Direct, or targeted, marketing allows a business to save a significant amount of money and is a natural match for PU learning. Known customer profiles compose the positive set, and large unknown databases of customer information make up the negative set [13], [38], [43], [44]. Fraud detection is another application for which PU learning could be useful [45], [46]. The positive set could be composed of known fraudulent transactions, while all others would be unlabeled.

2.3 Security and Signal Processing Applications

Security and signal processing applications are severely underrepresented in the Positive and Unlabeled learning domain. Initial forays into image classification [10], [15], [22], [47]–[50], have explored satellite image land-type classification and facial authentication. Classifying satellite images, radar images, and others is a natural fit for PU learning. A subset of objects from the desired class are manually labeled, and all others are classified using PU learning. This could be used to identify man-made objects in satellite imagery such as unknown archeological sites or new military installations and build-ups.

Activity or emotion classification could be identified from video streams. Illegal or dangerous activities such as an unexpected package being left in a populated area could result in a proactive police presence for increased safety. Nervousness or stress could be recognized and assist in determining persons to subject to more strenuous security checks at airports or large events. Network intrusion detection is another area that can benefit from PU learning [11], [21]. Online cyberattacks and security attacks on IoT (Internet of Things) platforms could also be identified using PU learning [2], [3]. Existing attack information could make up the positive class while all other data is used to populate the unlabeled class.

Sound classification and authentication may also be good PU learning applications. The positive set would be composed of a target sound or sounds. General audio from the target location would make up the unlabeled set. This could be used for diverse applications including song bird identification, helicopter approach, or human presence detection. [45], [46], [51]–[55] apply PU learning on data streams and signals.

Fault detection, related to novelty or outlier detection [53]-[56] is an important signal processing application. Many areas of manufacturing are concerned with identifying faulty products before shipment. Identifying microscopic faults in airline wings using sensors for flex analysis could improve airline safety. Another growing area is photovoltaic (PV) solar array fault detection [2], [57]–[61]. Could PU learning be used to classify such faults using real-time sensor data?

Other potential security applications include threat detection or submarine detection using an array of multi-int domain sensors. It may also be possible to learn to isolate a signal from a very noisy channel using PU learning.

Many of these security and signal applications have not yet been explored in the Positive and Unlabeled learning domain, though they fit the profile well – containing small amounts of positive data and large amounts of unlabeled data.
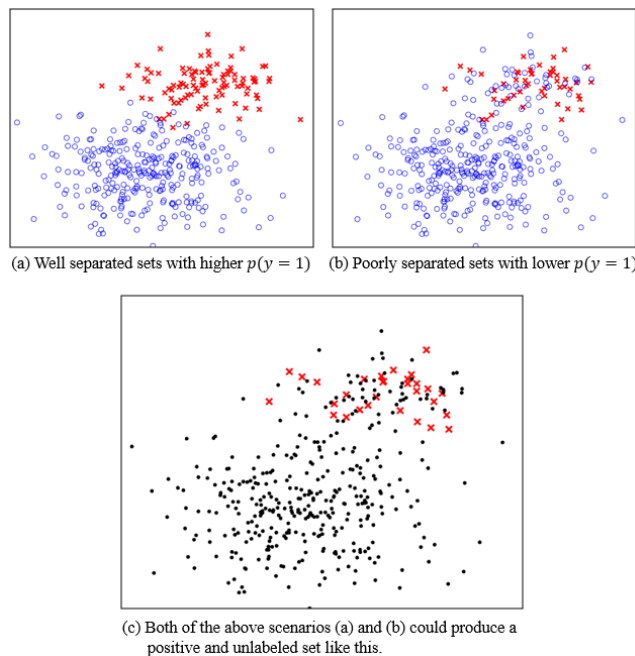
3. PU LEARNING ASSUMPTIONS

Most solutions to the PU learning problem assume that at least a portion of the positive class is separable from the negative class. This is illustrated in Figure 2 and is sometimes referred to as a positive subdomain or anchor set [62]. Another very common assumption is called the SCAR assumption where SCAR stands for "Selected Completely At Random". The SCAR assumption was first described in [63]. The rationale behind the SCAR assumption is that any bias in the labeled sample selection process will translate to bias in the model. This is a strong assumption and not always realistic in practice. Some more recent papers such as [50], [64]–[66] are trying to weaken these assumptions and will be discussed in Section 4.4 below.

4. PU LEARNING ALGORITHMS

There have been many papers published on PU learning in the past twenty years. Solutions to this problem generally fall into one of a couple of categories. We will outline these methods here and then describe them in more detail in the following subsections.

- Approach 1: The first, and oldest, approach is a two-step algorithm that first finds probable negative samples from

(a) Well separated sets with higher $p(y = 1)$ (b) Poorly separated sets with lower $p(y = 1)$

(c) Both of the above scenarios (a) and (b) could produce a positive and unlabeled set like this.

Figure 2: Separability Assumptions.

the unlabeled set, and then learns a traditional classifier from them. This approach has mostly fallen out of favor and is described in Section 4.1.

- Approach 2: The second type of algorithm assigns weights to the unlabeled sample and then learns a classifier from these weighted samples. The first approach above can be thought of as a special case of the second, where all weights are either 0 or 1. This approach is described in more detail in Section 4.2.

- Approach 3: A third type of solution treats unlabeled data as noisy or uncertain negatives. See Section 4.3 for more details.

- Approach 4: Several newer algorithms work to deal with and compensate for selection bias by weakening the SCAR assumption described in Section 3. These algorithms are described in Section 4.4 below.

- Other methods: recent algorithms continue to use the strategies described above, but also include methods to improve scalability, minimize overfitting, and stabilize the algorithms including bagging, convex risk/loss reformulations, and deep learning techniques such as generative adversarial networks (GANs). Other unusual approaches are described here as well. See Section 4.5 below for more on these.

4.1 The two-step approach to PU learning

In the first—generally older—method, the first step identifies probable negative samples from the unlabeled data using various techniques. [28] suggests a method called S-EM that uses a "Spy" technique to identify likely negatives. [27] use a Rocchio algorithm in their Roc-SVM paper. [32] describes PEBL, an algorithm that uses a technique called 1-DNF. [18] uses a

distance metric and SVM to identify negatives in PSoL. [15] creates an algorithm called SVMC that uses the margin maximization property of an SVM. [67] proposes NPSVM, a Nonparallel hyperplane SVM to improve performance. A more recent algorithm described in [30], builds on [32]'s 1-DNF strategy while introducing an iterative classification and a voting, or bagging, method for final classification (discussed further below). [68] proposes A-EM which adds additional unlabeled samples that are expected to be mostly negative to better separate the classes.

Step two of this approach involves applying a standard supervised learning algorithm, possibly iterated, to these estimated negatives and known positives. The S-EM algorithm uses an EM (Expectation Maximization) algorithm with a Naive Bayes classifier. PEBL, Roc-SVM, PSol, and SVMC use SVMs. [43] provides an analysis and comparison of several of the algorithms in this approach. The overall percentage of positive samples is usually considered unknown or is estimated using domain knowledge. As such, there is no way to independently verify the final classification results of the algorithm and convergence is usually used to cease iteration between labeling and learning.

4.2 The weighted approach to PU learning

The second approach to positive and unlabeled classification assigns real valued weights to each unlabeled sample. These weights represent the likelihood, or conditional probability, that an unlabeled sample x belongs to the positive set or negative set [20], [33], [69], or both [13], [63]. A standard learning algorithm then uses the weighted, unlabeled samples as either constantly weighted negative samples [33], variably weighted negative samples [20], [69], or as variably weighted positive and negative samples concurrently [13], [63] to learn a classifier, $f(x) = p(y = 1|x)$.

Various methods are used to estimate these likelihoods such as generalized linear models [33], logistic regression [63], boosted trees nonlinear logit model [20], soft-margin SVMs with linear kernels and Platt scaling [63], minimum distance to the positive set [69], and validation set experimentation and PrTFIDF [13].

Some algorithms, such as [43], use a more general weighting scheme where weights represent the cost of mislabeling an unlabeled positive sample as negative and vice versa, rather than directly weighting each sample's likelihood. These cost weights vary according to $p(y = 1)$ and are determined experimentally, using a F-score performance measure [43] to determine which values result in the highest performing algorithm. This method avoids the need to estimate the likelihood, $p(y = 1|x)$, or $p(y = 1)$ directly but is quite slow.

A hybrid method called PUDI is used by [8], combining aspects of the two-step and weighted learning approaches to identify likely disease genes. Instead of using weights, bins are used to partition the unlabeled set into four subsets: reliably negative samples, likely negative samples, weak negative samples, and likely positive samples. Samples are placed into each of these bins based on the Euclidean distance between the feature vectors of the unlabeled sample and a "positive representative vector" determined by averaging the genes in the known positive set. Weighted SVMs are then trained on these sets to create a final classifier.

A new variant of the weighted approach to PU learning, created by the authors of this paper and described in depth elsewhere [70], uses a modified logistic regression to build on [63] with an adaptive upper bound in place of a standard logistic regression. This produces improved results more than 87% of the time using an F-Score performance metric.

4.3 The noisy negative approach to PU learning

The third approach to the Positive and Unlabeled learning problem involves treating the unlabeled set as noisy negatives. Occasionally the positive set is assumed to be noisy as well, though the levels of noise are usually class-conditional. A classifier for this noisy data is then learned using techniques developed to deal with such scenarios [71]. [33] discusses this problem and develop a classifier using a weighted logistic regression algorithm. The authors of [43] offer a classifier they call the Biased-SVM with a modified cost function to account for the noisy data. [72] demonstrate that surrogate loss functions with importance reweighting allow any traditional classifier to be modified to work on noisy data. [21] uses a tree augmented naïve Bayes algorithm called UPTAN to deal with the uncertainty of the noisy data by creating a Bayesian network using dependence information among uncertain attributes to create a classifier. [73] uses mixture proportion estimation (MPE) to deal with noisy labels. [74] assumes that most of the unlabeled data are negative and employs a Laplacian unit-hyperplane classifier to deal with the resultant noisy data.

4.4 Weakening SCAR to deal with Selection Bias

One of the obstacles in PU learning is dealing with selection bias. The SCAR assumption described in Section 3 assumes that there is no selection bias which is often found to be unrealistic in practice. Papers such as [47], [48], [66], [75], and [76], attempt to compensate for this selection bias.

Rather than SCAR, [66] and [77] propose a weaker assumption, Selected at Random (SAR), which no longer assumes that positive regions are sampled with consistent frequency. They introduce an algorithm, SAR-EM, that uses a propensity score as a function of the attributes and the expectation maximization (EM) algorithm to solve the PU learning problem.

Other authors such as [50] attempt to weaken the SCAR assumption and deal with selection bias by minimizing the classification risk function using what they call the invariance of order. That is, they assume that the order of the probability of a sample being labeled positive is the same as the probability of it *being* positive. They also use a "partial identification" technique to extract some useful information from a function, without attempting to identify the entire function.

The authors in [64] also modify the classification risk function to deal with selection bias. They claim that unbiased risk estimators will produce negative empirical risks if the model being used is very flexible. To deal with this, they introduce a non-negative risk estimator that they argue produces better results in these situations and allows for hyperflexible deep learning solutions.

4.5 Other techniques

In recent years, heuristic optimizations for the Positive and Unlabeled learning such as bagging have gained traction. Bootstrap aggregating, or bagging, has become a popular approach to increase classifier performance [10], [54]. [10] proposes that "by nature, PU learning problems have a particular structure that leads to instability of classifiers, which can be advantageously exploited by a bagging-like procedure". Bagging involves repeatedly generating random subsets of the training data to classify. Each classifier then votes on the correct classification of a new sample. [10] and [11] both use bagging SVMs, while [30] adds bagging to the 1-DNF approach described in [32].

Traditional PU learning methods have not usually considered scalability concerns such as computational efficiency and overfitting. [78] propose an iterative algorithm USMO (Unlabeled data in Sequential Minimal Optimization) using Gram matrices and breaking the problem down into smaller subproblems that they argue is both computationally efficient and theoretically optimal. [79] also focuses on scalability by introducing a double hinge convex loss function. Improved computational efficiency is shown experimentally. In addition to softening the SCAR assumption and selection bias as described above, [64] also handles overfitting by introducing a non-negative risk estimator that allows the use of flexible, deep learning algorithms without the propensity to overfitting that many methods have. [80] proposes a solution that deals with large-scale datasets by introducing a closed-form classifier under certain conditions. [81] analyzes some of the theory behind PU learning and discusses how different convex representations may be more effective than others due to bias.

A new and extremely promising method uses a generative approach to PU learning using a deep learning Generative Adversarial Network (GAN) to identify both positive and negative data distributions [65]. Their algorithm, GenPU, uses two generative agents, one to generate positive samples and one for negative samples, and three discriminator agents – one each for the positive, unlabeled, and negative classes. The authors provide a theoretical analysis that claims GenPU can recover both positive and negative data distributions at equilibrium.

5. CONCLUDING REMARKS

This paper illustrates the importance of both the Positive and Unlabeled learning problem and its applications. Surprisingly many important scenarios naturally contain a small amount of positively labeled data and a large amount of unlabeled data. Some of these applications have been extensively studied using PU learning such as text classification (see section 2.2 above). Other problems such as disease gene, molecular classification, and drug identification are still in their infancy (see section 2.1). In addition, we have included several problems that have never or only rarely been investigated using PU learning such as image, sound, and video classification and various security applications (see section 2.3).

In addition to applications, we have surveyed a variety of old and new algorithms for solving the PU Learning problem. We have attempted to provide a broad overview of methods in the literature.

There is significant opportunity for further research in this field. A useful next step would be a comparative evaluation and benchmarking study to apply the existing algorithms to a variety of known and labeled datasets for comparative purposes.

5.1 Additional Reading on Machine Learning

Machine Learning (ML) is the science of finding patterns in big data. Many algorithms exist today, but most can be generally grouped in to one of three learning paradigms: supervised learning, unsupervised learning, and reinforcement learning.

As mentioned briefly in the introduction, the appropriate type of learning algorithm depends on the data available for learning. If the training data includes the desired outputs to be learned (labels), then a supervised learning algorithm is used. If the desired outputs or labels are unknown, the problem falls into the unsupervised paradigm. Reinforcement learning is different in that it is generally agent based using a stochastic exploratory/reward structure to learn a behavior or path in such a way as to maximize a reward.

Many different learning tasks can be accomplished within these general ML archetypes. Supervised learning algorithms usually provide real-valued predictions (called regression) or discrete class identification (classification). For each task, various algorithms exist to solve it. For example, regression problems for supervised, real-valued predictions can be learned using linear regression, simple artificial neural networks (ANNs), and various deep learning techniques such as recurrent neural networks (RNNs). Classification problems can be solved using support vector machines (SVMs), logistic regression, and various ANNs and deep learning algorithms such as convolutional neural networks (CNNs).

Unlike supervised learning algorithms, unsupervised learning is often used to find hidden or unknown patterns in data. This can be used for dimensionality reduction such as in principal component analysis (PCA) or through using complex deep learning neural networks to create autoencoders to more effectively encode learning data. Unsupervised algorithms such as k-means, spectral clustering, ANNs, and deep learning algorithms can also be used for clustering and other pattern identification purposes.

Reinforcement learning algorithms can be used for non-convex optimization problems using search algorithms such as genetic algorithms and simulated annealing, and for learning behavior in areas such as robotics. ANNs and deep learning techniques can be used for reinforcement learning, along with more recent algorithms such as generative adversarial networks (GANs) [65]. GANs also fit well in both supervised and unsupervised scenarios.

Combinations of many of these supervised, unsupervised, and even reinforcement learning algorithms are used for solving the semi-supervised Positive and Unlabeled learning problem described in this paper. Further approaches and combinations of methods and algorithms will continue to improve solutions to the PU learning problem. To learn more about Machine Learning, we recommend looking into these books [4], [5], [82], [83] and papers [6], [84]–[87] on the subject. Additional papers on ML and the PU problem can be found in the references [88]–[107].

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] B. Marr, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read," *Forbes*, May-2018.

[2] J. Lee, M. Stanley, A. Spanias, and C. Tepedelenlioglu, "Integrating machine learning in embedded sensor systems for Internet-of-Things applications," in *ISSPIT*, Limassol, Cyprus, IEEE, pp. 290–294, Dec. 2017.

[3] M. Stanley and J. M. Lee, *Sensors for IoT Applications*. Morgan and Claypool Publishers, Mar. 2018.

[4] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2011.

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, vol. 26. MIT Press, 2016.

[6] U. S. Shanthamallu, A. Spanias, C. Tepedelenlioglu, and M. Stanley, "A brief survey of machine learning methods and their sensor and IoT applications," in *IISA*, Larnaca, Cyprus, IEEE, Aug. 2017.

[7] P. Yang, X. Li, H. N. Chua, C. K. Kwoh, and S. K. Ng, "Ensemble positive unlabeled learning for disease gene identification," *PLoS One*, vol. 9, no. 5, p. e97079, May 2014.

[8] P. Yang, X. L. Li, J. P. Mei, C. K. Kwoh, and S. K. Ng, "Positive-unlabeled learning for disease gene identification," *Bioinformatics*, vol. 28, no. 20, pp. 2640–2647, Oct. 2012.

[9] P. Geurts, "Learning from positive and unlabeled examples by enforcing statistical significance," *JMLR*, vol. 15, pp. 305–314, 2011.

[10] F. Mordelet and J. P. Vert, "A bagging SVM to learn from positive and unlabeled examples," *Pattern Recognit. Lett.*, vol. 36, pp. 201–209, 2014.

[11] M. Claesen, F. De Smet, J. A. K. Suykens, and B. De Moor, "A robust ensemble approach to learn from positive and unlabeled data using SVM base models," *Neurocomputing*, vol. 160, pp. 73–84, Jul. 2015.

[12] F. Mordelet and J. P. Vert, "ProDiGe : PRioritization Of Disease Genes with multitask machine learning from positive and unlabeled examples," *BMC Bioinformatics*, pp. 1–15, Oct. 2011.

[13] D. Zhang and W. S. Lee, "A Simple Probabilistic Approach to Learning from Positive and Unlabeled Examples," in *UKCI*, London, England, Springer, pp. 83–87, Sep. 2005.

[14] F. Denis, R. Gilleron, and F. Letouzey, "Learning from positive and unlabeled examples," *Theor. Comput. Sci.*, vol. 348, no. 1, pp. 70–83, Dec. 2005.

[15] H. Yu, "Single-class classification with mapping convergence," *Mach. Learn.*, vol. 61, no. 1–3, pp. 49–69, Nov. 2005.

[16] S. Das, M. H. Saier, and C. Elkan, "Finding Transport Proteins in a General Protein Database," in *ECML PKDD*, Warsaw, Poland, Springer, vol. 4702, no. Lecture Notes in Computer Science, pp. 54–66, Sep. 2007.

[17] L. Cerulo, C. Elkan, and M. Ceccarelli, "Learning gene regulatory networks from only positive and unlabeled data," *BMC Bioinformatics*, p. 11:228, 2010.

[18] C. Wang, C. Ding, R. F. Meraz, and S. R. Holbrook, "PSoL: A positive sample only learning algorithm for finding non-coding RNA genes," *Bioinformatics*, vol. 22, no. 21, pp. 2590–2596, Nov. 2006.

[19] J. Hernández-González, I. Inza, and J. A. Lozano, "Learning from Proportions of Positive and Unlabeled Examples," *Int. J. Intell. Syst.*, vol. 32, no. 2, pp. 109–33, Feb. 2017.

[20] G. Ward, T. Hastie, S. Barry, J. Elith, and J. R. Leathwick, "Presence-only data and the em algorithm," *Biometrics*, vol. 65, no. 2, pp. 554–63, May 2009.

[21] H. Gan, Y. Zhang, and Q. Song, "Bayesian belief network for positive unlabeled learning with uncertainty," *Pattern Recognit. Lett.*, vol. 90, pp. 28–35, Apr. 2017.

[22] W. Li, Q. Guo, and C. Elkan, "A positive and unlabeled learning algorithm for one-class classification of remote-sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 717–725, Feb. 2011.

[23] F. Denis, R. Gilleron, and M. Tommasi, "Text classification from positive and unlabeled examples," in *IPMU*, Annecy, France, Jul. 2002.

[24] Q. Liu, B. Zhang, H. Sun, Y. Guan, and L. Zhao, "A novel K-means clustering algorithm based on positive examples and careful seeding," in *ICCIS*, Chengdu, China, IEEE, pp. 767–770, Dec. 2010.

[25] T. Joachims, "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization," *ICML*, pp. 143–151, Jul. 1997.

[26] M. Li, S. Pan, Y. Zhang, and X. Cai, "Classifying networked text data with positive and unlabeled examples," *Pattern Recognit. Lett.*, vol. 77, pp. 1–7, Jul. 2016.

[27] X. Li and B. Liu, "Learning to classify texts using positive and unlabeled data," in *IJCAI*, Acapulco, Mexico, Morgan Kaufmann, p. Pages 587-592, Aug. 2003.

[28] B. Liu, W. S. Lee, P. S. Yu, and X. Li, "Partially Supervised Classification of Text Documents," in *ICML*, Sydney, Australia, Morgan Kaufmann, pp. 387–394, Jul. 2002.

[29] G. P. C. Fung, J. X. Yu, H. Lu, and P. S. Yu, "Text classification without negative examples revisit," *TKDE*, vol. 18, no. 1, pp. 6–20, Jan. 2006.

[30] J. Han, W. Zuo, L. Liu, Y. Xu, and T. Peng, "Building text classifiers using positive, unlabeled and 'outdated' examples," *Concurr. Comput. Pract. Exp.*, vol. 28, no. 13, pp. 3691–3706, Sep. 2016.

[31] B. Z. Zhang and W. L. Zuo, "Co-EM support vector machine based text classification from positive and unlabeled examples," in *ICINIS*, Wuhan, China, IEEE, pp. 745–748, Nov. 2008.

[32] H. Yu, J. Man, and K. C. C. Chang, "PEBL: Web Page Classification without Negative Examples," *TKDE*, vol. 16, no. 1, pp. 70–81, Jan. 2004.

[33] W. S. Lee and B. Liu, "Learning with Positive and Unlabeled Examples Using Weighted Logistic Regression," in *ICML*, Washington, D.C., AAAI Press, pp. 448–455, Aug. 2003.

[34] H. Yu, C. Zhai, and J. Han, "Text classification from positive and unlabeled documents," in *CIKM*, New Orleans, USA, ACM, pp. 232–239, Nov. 2003.

[35] F. Denis, A. Laurent, R. Gilleron, and M. Tommasi, "Text Classification and Co-training from Positive and Unlabeled Examples," in *ICML*, Washington, D.C., ICML, pp. 80–87, Aug. 2003.

[36] S. Yu and C. Li, "PE-PUC: A Graph Based PU-Learning Approach for Text Classification," *MLDM*, vol. 4571, pp. 574–584, Jul. 2007.

[37] K. Ren *et al.*, "A Robust AUC Maximization Framework With Simultaneous Outlier Detection and Feature Selection for Positive-Unlabeled Classification," *IEEE Transactions on Neural Networks and Learning Systems*, IEEE, pp. 1–12, Oct-2018.

[38] S. Sellamanickam, P. Garg, and S. K. Selvaraj, "A pairwise ranking based approach to learning with positive and unlabeled examples," in *ACM CIKM*, Glasgow, Scottland, ACM, pp. 663–672, Oct. 2011.

[39] L. M. de Campos, J. M. Fernández-Luna, J. F. Huete, and L. Redondo-Expósito, "Positive unlabeled learning for building recommender systems in a parliamentary setting," *Inf. Sci. (Ny).*, vol. 433–434, pp. 221–32, Apr. 2018.

[40] Y. Ren, D. Ji, and H. Zhang, "Positive unlabeled learning for deceptive reviews detection," in *EMNLP*, Doha, Qatar, Association for Computational Linguistics (ACL), pp. 488–498, Oct. 2014.

[41] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, "Spotting Fake Reviews via Collective Positive-Unlabeled Learning," in *ICDM*, Washington, D.C., IEEE, pp. 899–904, Dec. 2014.

[42] A. Smith and C. Elkan, "A Bayesian network framework for reject inference," in *SIGKDD*, Seattle, Washington, ACM, pp. 286–295, Aug. 2004.

[43] B. Liu, Y. Dai, X. Li, W. S. Lee, and P. S. Yu, "Building text classifiers using positive and unlabeled examples," in *ICDM*, Melbourne, Florida, IEEE, pp. 179–186, Dec. 2003.

[44] H. Fei, Y. Kim, S. Sahu, M. Naphade, S. K. Mamidipallis, and J. Hutchinson, "Heat pump detection from coarse grained smart meter data with positive and unlabeled learning," in *SIGKDD*, Chicago, USA, Association for Computing Machinery, vol. Part F1288, pp. 1330–1338, Aug. 2013.

[45] X. Qin, Y. Zhang, C. Li, and X. Li, "Learning from data streams with only positive and unlabeled data," *JIIS*, vol. 40, no. 3, pp. 405–430, Jun. 2013.

[46] C. Liang, Y. Zhang, P. Shi, and Z. Hu, "Learning very fast decision tree from uncertain data streams with positive and unlabeled samples," *Inf. Sci. (Ny).*, vol. 213, pp. 50–67, Dec. 2012.

[47] Y. Xu, C. Xu, C. Xu, and D. Tao, "Multi-positive and unlabeled learning," in *IJCAI*, Melbourne, Australia, IJCAI, pp. 3182–88, Aug. 2017.

[48] F. Chiaroni, M. C. Rahal, N. Hueber, and F. Dufaux, "Learning with a generative adversarial network from a positive unlabeled dataset for image classification," in *ICIP*, Athens, Greece, IEEE, pp. 1368–1372, Oct. 2018.

[49] C. Gong, H. Shi, J. Yang, J. Yang, and J. Yanga, "Multi-Manifold Positive and Unlabeled Learning for Visual Analysis," *TCSVT*, pp. 1–14, Mar. 2019.

[50] M. Kato, T. Teshima, and J. Honda, "Learning from Positive and Unlabeled Data with a Selection Bias," *ICLR*, May 2019.

[51] L. de Carvalho Pagliosa and R. F. de Mello, "Semi-supervised time series classification on positive and unlabeled problems using cross-recurrence quantification analysis," *Pattern Recognit.*, vol. 80, pp. 53–63, Aug. 2018.

[52] M. N. Nguyen, X. L. Li, and S. K. Ng, "Positive unlabeled learning for time series classification," *IJCAI*, pp. 1421–1426, 2011.

[53] X.-L. Li, P. S. Yu, B. Liu, and S.-K. Ng, "Positive Unlabeled Learning for Data Stream Classification," in *SDM*, Sparks, SIAM, pp. 259–270, Apr. 2009.

[54] J. Zhang, Z. Wang, J. Meng, Y. P. Tan, and J. Yuan, "Boosting positive and unlabeled learning for anomaly detection with multi-features," *IEEE Trans. Multimed.*, vol. 21, no. 5, pp. 1332–1344, May 2019.

[55] A. Kumar and B. Raj, "Audio event detection using weakly labeled data," in *ACM MM*, New York, New York, USA, ACM Press, pp. 1038–1047, 2016.

[56] G. Blanchard, G. Lee, and C. Scott, "Semi-supervised novelty detection," *JMLR*, vol. 11, pp. 2973–3009, Nov. 2010.

[57] E. Pedersen *et al.*, "PV Array Fault Detection using Radial Basis Networks," in *IISA*, Patras, Greece, IEEE, Jul. 2019.

[58] S. Rao, A. Spanias, and C. Tepedelenlioglu, "Solar Array Fault Detection using Neural Networks," in *ICPS*, Taipei, Taiwan, Institute of Electrical and Electronics Engineers (IEEE), pp. 196–200, May 2019.

[59] A. S. Spanias, "Solar energy management as an Internet of Things (IoT) application," in *IISA*, Larnaca, Cyprus, IEEE, vol. 2018-Janua, pp. 1–4, Aug. 2017.

[60] V. S. Narayanaswamy, R. Ayyanar, A. Spanias, C. Tepedelenlioglu, and D. Srinivasan, "Connection Topology Optimization in Photovoltaic Arrays using Neural Networks," in *ICPS*, Taipei, Taiwan, IEEE, pp. 167–172, May 2019.

[61] R. Ramakrishna, A. Scaglione, A. Spanias, and C. Tepedelenlioglu, "Distributed Bayesian Estimation with Low-rank Data: Application to Solar Array Processing," in *ICASSP*, Brighton, UK, IEEE, vol. 2019-May, pp. 4440–4444, 2019.

[62] J. Bekker and J. Davis, "Learning From Positive and Unlabeled Data: A Survey," *ArXiv*, Nov. 2018.

[63] C. Elkan and K. Noto, "Learning classifiers from only positive and unlabeled data," in *SIGKDD*, Las Vegas, ACM, pp. 213–20, Aug. 2008.

[64] R. Kiryo, G. Niu, M. C. du Plessis, and M. Sugiyama, "Positive-Unlabeled Learning with Non-Negative Risk Estimator," in *NIPS*, Long Beach, Curran Assoc. Inc., pp. 1674–84, Dec. 2017.

[65] M. Hou, B. Chaib-Draa, C. Li, and Q. Zhao, "Generative Adversarial Positive-Unlabelled Learning," in *IJCAI*, Stockholm, Sweden, International Joint Conferences on Artificial Intelligence, Jul. 2018.

[66] J. Bekker and J. Davis, "Learning from Positive and Unlabeled Data under the Selected At Random Assumption," in *Proc. Second Int. Work. Learn. with Imbalanced Domains Theory Appl.*, Dublin, Ireland, PMLR, pp. 94:8–22, Sep. 2018.

[67] Y. Zhang, X. C. Ju, and Y. J. Tian, "Nonparallel hyperplane support vector machine for PU learning," in *ICNC*, Xiamen, China, IEEE, pp. 703–708, Aug. 2014.

[68] X.-L. Li and B. Liu, "Learning from Positive and Unlabeled Examples with Different Data Distributions," *ECML PKDD*, pp. 218–229, Jan. 2005.

[69] Z. Liu, W. Shi, D. Li, and Q. Qin, "Partially Supervised Classification: Based on Weighted Unlabeled Samples Support Vector Machine," *IJDWM*, vol. 2, no. 3, pp. 42–56, 2006.

[70] K. Jaskie, C. Elkan, and A. S. Spanias, "A Modified Logistic

[71] Regression for Positive and Unlabeled Learning," in *ACSSC*, Pacific Grove, California, IEEE, Nov. 2019.

[71] A. K. Menon, B. Van Rooyen, C. S. Oong, and R. C. Williamson, "Learning from Corrupted Binary Labels via Class-Probability Estimation," in *ICML*, Lille, France, JMLR, vol. 37, pp. 125–34, Jul. 2015.

[72] T. Liu and D. Tao, "Classification with Noisy Labels by Importance Reweighting," *TPAMI*, vol. 38, no. 3, pp. 447–461, Mar. 2016.

[73] C. Scott, "A rate of convergence for mixture proportion estimation, with application to learning from noisy labels," in *AISTATS*, San Diego, JMLR and Microtome, vol. 38, pp. 838–846, May 2015.

[74] Y. H. Shao, W. J. Chen, L. M. Liu, and N. Y. Deng, "Laplacian unit-hyperplane learning from positive and unlabeled examples," *Inf. Sci. (Ny).*, vol. 314, no. 1, pp. 152–168, Sep. 2015.

[75] A. T. Smith and C. Elkan, "Making generative classifiers robust to selection bias," in *SIGKDD*, San Jose, ACM, pp. 657–666, Aug. 2007.

[76] F. He, G. I. Webb, T. Liu, and D. Tao, "Instance-Dependent PU Learning by Bayesian Optimal Relabeling," *ArXiv*, Aug. 2018.

[77] J. Bekker and J. Davis, "Beyond the Selected Completely At Random Assumption for Learning from Positive and Unlabeled Data," *ArXiv*, Jun. 2018.

[78] E. Sansone, F. G. B. De Natale, and Z. H. Zhou, "Efficient Training for Positive Unlabeled Learning," *TPAMI*, Jul. 2018.

[79] M. C. du Plessis, G. Niu, and M. Sugiyama, "Convex Formulation for Learning from Positive and Unlabeled Data," in *ACML*, Hong Kong, China, Springer, pp. 221–236, Nov. 2015.

[80] Y. Kwon, W. Kim, M. Sugiyama, and M. C. Paik, "An analytic formulation for positive-unlabeled learning via weighted integral probability metric," *ArXiv*, 2019.

[81] M. C. du Plessis, G. Niu, and M. Sugiyama, "Analysis of Learning from Positive and Unlabeled Data," *NIPS*, pp. 703–711, Dec. 2014.

[82] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*. Elsevier Ltd, Mar. 2015.

[83] M. Virvou, E. Alepis, G. A. Tsihrintzis, and L. C. Jain, Eds., *Machine Learning Paradigms, Advances in Learning Analytics*. Springer, May 2020.

[84] H. Song, J. Thiagarajan, K. Ramamurthy, A. S. Spanias, and P. Turaga, "Iterative Kernel Fusion for Image Classification," in *ICASSP*, Shanghai, China, IEEE, Mar. 2016.

[85] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3. pp. 199–222, Aug-2004.

[86] H. Song, J. J. Thiagarajan, P. Sattigeri, and A. Spanias, "Optimizing kernel machines using deep learning," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 11, pp. 5528–5540, Feb. 2018.

[87] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic Differentiation in Machine Learning: a Survey," *JMLR*, vol. 18, pp. 1–43, Apr. 2018.

[88] M. C. Du Plessis, G. Niu, and M. Sugiyama, "Class-prior estimation for learning from positive and unlabeled data," in *ACML*, Hong Kong, China, Asian Conference on Machine Learning, pp. 221–236, Nov. 2015.

[89] J. He, Y. Zhang, X. Li, and Y. Wang, "Naive Bayes Classifier for Positive Unlabeled Learning with Uncertainty *," in *SDM*, Columbus, SIAM, p. 12, Apr. 2010.

[90] A. Blum and T. Mitchell, "Combining Labeled and Unlabeled Data with Co-Training," *Proc. 11th Annu. Conf. Comput. Learn. Theory*, pp. 92–100, Jul. 1998.

[91] K. Pelckmans and J. A. K. Suykens, "Transductively learning from positive examples only," in *ESANN*, Bruges, Belgium, ESANN, vol. 08, no. April 2009, pp. 2007–2011, Apr. 2009.

[92] S. Dan and N. S. Cardell, "Estimating logistic regression models when the dependent variable has no variance," *Commun. Stat. - Theory and Methods*, vol. 21, no. 2, pp. 423–450, Jun. 2007.

[93] S. Jain, M. White, and P. Radivojac, "Estimating the class prior and posterior from noisy positives and unlabeled data," *NIPS*, pp. 2693–2701, Dec. 2016.

[94] J. Bekker and J. Davis, "Estimating the Class Prior in Positive and Unlabeled Data Through Decision Tree Induction," in *AAAI*, New Orleans, Louisiana, AAAI Press, pp. 2712–2719, Feb. 2018.

[95] B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C.

Williamson, "Estimating the Support of a High-Dimensional Distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.

[96] M. C. Du Plessis and M. Sugiyama, "Class Prior Estimation from Positive and Unlabeled Data," *IEICE Trans. Inf. Syst.*, vol. E96-D, no. 5, pp. 1358–1362, 2014.

[97] B. Zhang and W. Zuo, "Learning from positive and unlabeled examples: A survey," in *ISIP WMWA*, Moscow, Russia, IEEE, May 2008.

[98] H. G. Ramaswamy, C. Scott, and A. Tewari, "Mixture Proportion Estimation via Kernel Embedding of Distributions," in *ICML*, New York, JMLR, vol. 48, pp. 2052–60, Jun. 2016.

[99] J. T. Zhou, Q. Mao, I. W. Tsang, and S. J. Pan, "Multi-view positive and unlabeled learning," in *ACML*, Singapore, Singapore, JMLR, vol. 25, pp. 555–570, Nov. 2012.

[100] S. Jain, M. White, M. W. Trosset, and P. Radivojac, "Nonparametric semi-supervised learning of class proportions," Jan. 2016.

[101] S. S. Khan and M. G. Madden, "One-class classification: Taxonomy of study and review of techniques," *Knowl. Eng. Rev.*, vol. 29, no. 3, pp. 345–374, Jun. 2014.

[102] D. Ienco and R. G. Pensa, "Positive and unlabeled learning in categorical data," *Neurocomputing*, vol. 196, pp. 113–124, Jul. 2016.

[103] P. Yang, W. Liu, and J. Yang, "Positive unlabeled learning via wrapper-based adaptive sampling," in *IJCAI*, Melbourne, Australia, IJCAI, pp. 3273–79, Aug. 2017.

[104] C. Hsieh, N. Natarajan, and I. S. Dhillon, "PU Learning for Matrix Completion," in *ICML*, Lille, France, ICML, vol. 37, pp. 2445–53, Jul. 2015.

[105] M. C. Du Plessis and M. Sugiyama, "Semi-supervised learning of class balance under class-prior change by distribution matching," *Neural Networks*, vol. 50, pp. 110–119, Feb. 2014.

[106] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Second Edi. John Wiley & Sons, Incorporated, Sep. 2002.

[107] G. Niu, M. C. du Plessis, T. Sakai, Y. Ma, and M. Sugiyama, "Theoretical Comparisons of Positive-Unlabeled Learning against Positive-Negative Learning," in *NIPS*, Barcelona, Spain, ACM, pp. 1207–15, Dec. 2016.