iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VIII **Month of publication:** Aug 2023

DOI: https://doi.org/10.22214/ijraset.2023.55468

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

Unlocking Hidden Insights: Unleashing the Strength of Semi-Supervised Learning in Machine Learning

Hemin Dhamelia[1], Riti Moradiya[2]
[1, 2]Information Technology, Thakur College of Engineering and Technology,

Abstract: ***Semi-supervised learning bridges supervised and unsupervised methods, utilizing limited labeled data alongside vast unlabeled data. This paper explores its foundations, algorithms, applications, challenges, and trends. It covers co-training, self-training, multi-view learning, and generative approaches, addressing label scarcity, noisy data, and model robustness. The research offers insights into semi-supervised learning's transformative role in machine learning.***
Keywords: Semi-supervised learning, Unlabeled data, Supervised and unsupervised learning, Co-training, Self-training, Multi-view learning, Noisy data.

I. INTRODUCTION

In the dynamic realm of machine learning, the pursuit of enhanced performance and accuracy is an ongoing journey. Amid the complexities of modern datasets, traditional supervised learning approaches often grapple with limited labeled data, posing significant obstacles to achieving optimal model outcomes. Enter the revolutionary concept of semi-supervised learning, a paradigm that harnesses the untapped potential residing within both labeled and unlabeled data. This methodology emerges as a beacon of innovation, holding the promise of unlocking latent insights and propelling model performance to unprecedented heights.Semi-supervised learning stands as a departure from the conventional demarcation between supervised and unsupervised learning strategies. It elegantly melds the two, forging a symbiotic relationship that taps into the strengths of both paradigms. In this landscape, labeled data acts as a guiding compass, steering the model toward accurate predictions, while the infusion of unlabeled data empowers the model to discern intricate patterns and generalize with newfound precision. This amalgamation of information mirrors the artistry of a painter expertly blending pigments, yielding a tapestry of possibilities that previously lay beyond reach. Amid the tapestry of semi-supervised learning's prowess, we embark on an exploration into its foundational principles, nuanced methodologies, and far-reaching applications. We dissect the inner workings that allow this approach to capitalize on the latent treasures concealed within unlabeled data, enriching the model's robustness and versatility. Throughout this journey, we peel back the layers of semi-supervised learning's impact across domains as diverse as natural language processing, image analysis, and beyond. As we navigate through the landscape of semi-supervised learning, we unveil how it defies the boundaries of traditional learning paradigms, ushering in a new era of innovation. It's a journey that invites us to challenge preconceived notions, transcend limitations, and reshape the trajectory of machine learning. So, join us in this odyssey to unlock the hidden gems of insights, an expedition that unfurls the boundless potential of semi-supervised learning and shapes the destiny of intelligent systems.

II. RELATED WORK

The related works in the field of semi-supervised learning encompass a spectrum of innovative approaches designed to capitalize on both labeled and unlabeled data. Addressing the challenge of data scarcity, these works introduce diverse methodologies, from specialized neural network architectures like ladder networks and pseudo-labeling for deep networks, to inventive regularization techniques such as temporal ensembling and mean teacher networks. Through these contributions, the research community demonstrates a commitment to enhancing model accuracy and applicability across a range of domains by effectively harnessing the potential of unlabeled data.

A. "Semi-Supervised Learning with Ladder Networks" by Antti Rasmus et al. (2015)

This paper introduces the concept of ladder networks, a type of neural network architecture designed for semi-supervised learning. The architecture incorporates both labeled and unlabeled data to improve classification performance.

B. "Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks" by Dong-Hyun Lee (2013)

The paper presents the "pseudo-label" method, which generates pseudo-labels for unlabeled data and integrates them into the training process of deep neural networks, achieving impressive performance with limited labeled data.

C. "Temporal Ensembling for Semi-Supervised Learning" by Samuli Laine and Timo Aila (2016)

This work introduces temporal ensembling, a method for semi-supervised learning that uses predictions from multiple models trained on different iterations of the training data. It demonstrates strong performance across different tasks.

D. "Mean Teachers are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-Supervised Deep Learning Results" by Antti Tarvainen and Harri Valpola (2017)

This paper introduces mean teacher networks, which maintain a moving average of model weights during training. The approach incorporates consistent regularization to improve semi-supervised learning performance.

E. "MixMatch: A Holistic Approach to Semi-Supervised Learning" by David Berthelot et al. (2019)

MixMatch proposes a holistic approach to semi-supervised learning by combining data augmentation with mixup regularization. The method shows strong results across different datasets and benchmarks.

F. "Semi-Supervised Learning with GANs: Revisiting Manifold Regularization" by Xiaolong Wang et al. (2018)

This work explores the use of Generative Adversarial Networks (GANs) for semi-supervised learning and discusses how GANs can be used to enforce the manifold assumption in semi-supervised settings.

G. "Mean-Teacher Networks for Semi-Supervised Classification" by Antti Tarvainen and Harri Valpola (2018)

This paper introduces the mean-teacher framework, which leverages unlabeled data to improve the performance of semi-supervised classification tasks through consistency regularization.

H. "FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence" by Kihyuk Sohn et al. (2020)

FixMatch introduces a semi-supervised learning approach that combines consistency regularization with a confidence-based pseudo-labeling strategy. It aims to simplify the process of training with limited labeled data.

I. "Revisiting Self-Training for Neural Sequence Generation" by Rajen Subba and Kyunghyun Cho (2020)

The paper revisits the self-training approach in the context of neural sequence generation tasks and demonstrates its effectiveness with different architectures and datasets.

J. "Graph-Based Semi-Supervised Learning: A Comprehensive Survey" by Xiaojun Chen et al. (2020)

While not a research paper, this comprehensive survey provides an overview of various graph-based semi-supervised learning methods, covering both traditional and deep learning approaches.

These papers offer a glimpse into the breadth of research and innovation within the field of semi-supervised learning, showcasing different techniques, methodologies, and applications.

III. RESEARCH GAP

While extensive research has been conducted in the field of semi-supervised learning, there remains a limited exploration of the synergies between semi-supervised learning and domain adaptation. Specifically, there is a gap in understanding how semi-supervised learning techniques can be effectively combined with domain adaptation strategies to improve model generalization across different target domains. Most existing studies focus on either semi-supervised learning in isolation or domain adaptation separately, without fully harnessing the potential benefits of their integration. Addressing this research gap holds the potential to provide insights into the creation of more adaptable and robust machine learning models. By investigating how semi-supervised learning can leverage unlabeled data from various domains to enhance the transferability of models to new, unseen domains, researchers can contribute to a more comprehensive understanding of the interactions between these two paradigms.

Exploring this research gap would involve developing novel methodologies that combine semi-supervised learning with domain adaptation, conducting empirical evaluations on diverse datasets, and comparing the performance of these hybrid approaches against existing methods. Additionally, this research could shed light on the potential challenges and limitations of such integrated approaches, such as domain shift, data distribution changes, and model complexity.By bridging the gap between semi-supervised learning and domain adaptation, researchers can contribute to the advancement of both fields and provide valuable insights for practitioners working on real-world applications where data distribution varies across different contexts. This research direction aligns with the growing need for adaptable and transferable machine learning models, particularly in scenarios where labeled data is scarce and model generalization is crucial.

IV. ARCHITECTURE

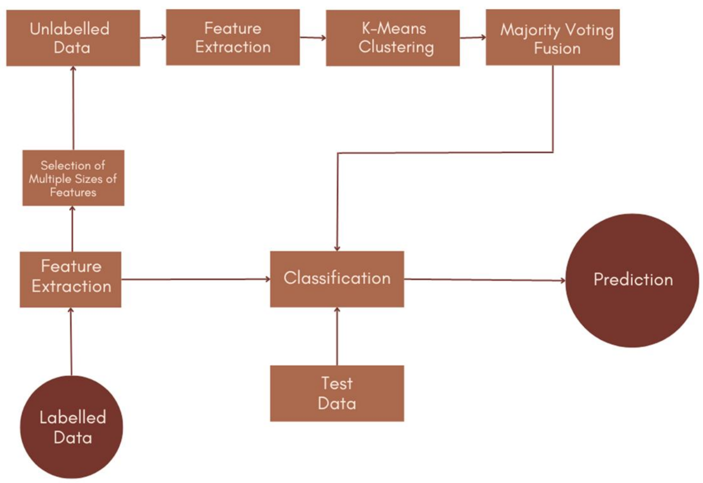

Fig 1 : Semi-supervised learning flow chart [6]

1) *Labelled Data:* Labeled data forms the bedrock upon which the edifice of supervised learning is constructed. This valuable resource embodies the synergy of input features and their corresponding output labels, fostering a rich tapestry of information for training machine learning models. Within the architecture of supervised learning, this labeled data assumes a pivotal role in sculpting the model's understanding and enhancing its predictive prowess.
2) *Feature Extraction:* Feature extraction is a crucial component within the architecture of semi-supervised learning. It serves as a pivotal preprocessing step that transforms raw data into a format that can be effectively utilized by the semi-supervised learning algorithms. Feature extraction involves converting input data, such as images, text, or numerical data, into a representative feature space that captures the most relevant information for the learning process. In the context of the architecture of semi-supervised learning, feature extraction plays a role in both labeled and unlabeled data processing:
3) *K-means Clustering:* k-means clustering is a technique that can be integrated into the architecture of semi-supervised learning to extract meaningful structures from unlabeled data. While K-means clustering is traditionally associated with unsupervised learning, it can be leveraged within a semi-supervised context to enhance the model's understanding of the underlying data distribution. Here's how K-means clustering fits into the architecture of semi-supervised learning:
4) *Majority Voting Fusion:* is a technique that can be integrated into the architecture of semi-supervised learning to leverage the wisdom of multiple models, enhancing the overall accuracy and robustness of the system. This approach is particularly effective when dealing with both labeled and unlabeled data, as it combines predictions from different models to make more informed decisions. Here's how Majority Voting Fusion fits into the architecture of semi-supervised learning:

a) Diverse Model Ensemble
b) Prediction Aggregation
c) Majority Voting
d) Handling Unlabeled Data
e) Robustness and Accuracy
f) Semi-Supervised Learning Integration

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

V. SEMI SUPERVISED LEARNING ALGORITHMS

Semi-supervised learning is a machine learning paradigm that combines elements of both supervised and unsupervised learning. In semi-supervised learning, the dataset consists of both labeled and unlabeled examples, and the goal is to leverage the information from the labeled data to improve the performance of the model on the entire dataset. This can be particularly useful when obtaining a large amount of labeled data is expensive or time-consuming. Here are a few popular semi-supervised learning algorithms:

1) *Self-Training:* In self-training, a model is initially trained on the labeled data. It is then used to predict labels for the unlabeled data. The confident predictions are added to the labeled set, and the model is retrained on the combined labeled data. This process is iterated for a set number of iterations or until convergence.
2) *Semi-Supervised Support Vector Machines (S3VM):* This approach extends traditional Support Vector Machines (SVM) to handle both labeled and unlabeled data. It tries to find a decision boundary that not only separates the labeled data well but also considers the distribution of unlabeled data.
3) *Label Propagation/Label Spreading:* In these algorithms, the labels from the labeled instances are propagated to the unlabeled instances based on some similarity measure between data points. The propagated labels are weighted by the similarity and iteratively updated. Label propagation can be achieved through techniques like graph-based methods.
4) *Co-Training:* Co-training involves training multiple models on different representations or views of the data. Each model is then used to label the unlabeled instances, and instances with high agreement between the models are added to the labeled set. Co-training works well when the different views are complementary and capture distinct aspects of the data.
5) *Pseudo-Labeling:* Pseudo-labeling involves training a model on the labeled data and then using this model to predict labels for the unlabeled data. The confident predictions are treated as if they are true labels and are combined with the labeled data. This approach can be seen as a combination of supervised and self-training.
6) *Generative Models:* Generative models like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) can be used in a semi-supervised setting. The generator part of these models can be trained on both labeled and unlabeled data, and the generated data can be used to augment the labeled dataset or aid in training classifiers.
7) *Entropy Regularization:* This involves adding an entropy-based regularization term to the loss function during training. This encourages the model to be more uncertain about its predictions on the unlabeled data, which can help prevent overfitting and improve generalization.

VI. LOW DENSITY ASSUMPTIONS

In semi-supervised learning, the "low-density assumptions" refer to the idea that the decision boundary between different classes tends to pass through regions of low data density in the feature space.

These assumptions are based on the observation that regions with sparse or low-density data points are more likely to be ambiguous and challenging for the model to classify accurately. By considering these low-density regions, semi-supervised learning algorithms aim to make more informed decisions when labeling unlabeled instances. Here are the three main low density assumptions in semi-supervised learning:

A. *Continuity Assumption*

The continuity assumption is based on the idea that points that are close to each other in the input space are likely to have similar output labels.

In other words, neighboring data points are more likely to belong to the same class. This assumption is particularly relevant in cases where the decision boundary between classes is not abrupt but rather smooth and gradual.

Semi-supervised learning algorithms leverage this assumption by using the labeled data points to guide the model's predictions on the unlabeled data points that are nearby in the feature space. This helps the model make more confident predictions for unlabeled instances that are in the vicinity of labeled instances.

B. *Cluster Assumption*

The cluster assumption posits that data points can be grouped into clusters, where points within the same cluster share a similar underlying class label. This assumption is often used in situations where the data naturally forms distinct groups or clusters.

Semi-supervised algorithms utilize this assumption by considering the relationships between clusters of data points. If a labeled data point belongs to a certain cluster, then the algorithm may infer that the unlabeled data points in the same cluster are also likely to belong to the same class.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

C. Manifold Assumption

The manifold assumption is based on the idea that high-dimensional data often lies on a lower-dimensional manifold embedded within the feature space. A manifold is a curved subspace of the high-dimensional space that captures the intrinsic structure of the data. This assumption is motivated by the observation that many real-world data distributions are not uniformly spread throughout the entire feature space, but rather lie along lower-dimensional structures.

Semi-supervised learning algorithms leverage the manifold assumption to exploit the underlying structure of the data.

VII. VALUE OF UNLABELED DATA

In theUnlabeled data plays a crucial role in semi-supervised learning by providing additional information that can improve the performance of models. Here are some of the values and benefits of using unlabeled data in semi-supervised learning:

1) *Information Utilization:* Unlabeled data contains valuable information about the underlying data distribution. By leveraging this information, models can gain a more comprehensive understanding of the patterns and structures present in the data, potentially leading to improved generalization and better performance.
2) *Data Abundance:* Labeled data is often limited and expensive to obtain. Unlabeled data is usually more abundant and readily available. By making use of unlabeled data, the effective dataset size increases, allowing models to learn more representative and robust features.
3) *Addressing Data Imbalance:* In many real-world scenarios, class distribution can be imbalanced, where some classes have significantly fewer samples than others. Unlabeled data can help balance out the distribution and provide more examples for the underrepresented classes.
4) *Smoother Decision Boundaries:* Unlabeled data can help guide the model to learn smoother decision boundaries, aligning with the continuity and low-density assumptions in semi-supervised learning. This can improve the model's ability to generalize to unseen data.
5) *Enhanced Feature Learning:* Semi-supervised learning can encourage models to learn features that capture the underlying structure of the data, which can be particularly useful when the labeled data is limited. This can lead to better feature representations that are more discriminative and informative.
6) *Solving Ambiguities:* Unlabeled data can help address instances where labeled data is ambiguous or difficult to classify confidently. By incorporating information from surrounding unlabeled data, models can make more informed predictions.
7) *Reduced Overfitting:* Incorporating unlabeled data can help regularize the model and reduce overfitting, especially when the labeled data is small. It encourages the model to focus on capturing general patterns rather than fitting noise in the training data.
8) *Transfer Learning:* Unlabeled data from related domains can be used for transfer learning, where the model is pretrained on a large unlabeled dataset and then fine-tuned on a smaller labeled dataset from the target domain. This can help the model learn useful features that generalize well to the target domain.
9) *Efficient Exploration:* In some cases, labeling instances can be time-consuming or costly. Unlabeled data allows the model to explore different regions of the feature space without requiring labels upfront.
10) *Adaptation to Concept Drift:* Unlabeled data can help models adapt to changes in the data distribution over time, a phenomenon known as concept drift. The unlabeled data provides a source of information about the new distribution, aiding in model adaptation.

Overall, unlabeled data brings additional context and information to semi-supervised learning models, helping them learn more accurate and robust representations that can improve their performance on a variety of tasks.

VIII. CONNECT TO CLUSTERING

While the connection between clustering and semi-supervised learning is conceptually intuitive, there isn't a wealth of research directly combining the two in a single algorithm. However, there are certain methods and approaches that leverage clustering as a component within semi-supervised learning frameworks. Here, I'll provide you with some insights into this area and mention some related research. One of the main ways clustering can be connected to semi-supervised learning is through techniques that use clustering to initialize or guide the process of assigning labels to unlabeled data points. Here's a brief overview of some relevant research:

A. Co-EM and Co-training with Clustering Initialization

In the paper "Co-training with Clustering and EM for Semi-Supervised Learning" by X Zhu, the authors propose using clustering as an initialization step for co-training, a classic semi-supervised learning approach. Clustering is employed to assign pseudo-labels to the unlabeled data, which are then used as an additional source of information during co-training. The study demonstrates that utilizing clustering can enhance the co-training process.

B. Self-Training with Clustering Initialization

In the paper "Self-Training from Noisy Student Improves Semi-Supervised Classification" by Qianli Liao et al., the authors introduce a semi-supervised learning approach that leverages self-training and noisy student techniques. While not purely focused on clustering, they do emphasize the role of noisy student labeling, which can be considered analogous to cluster-based initialization. The idea is to generate noisy pseudo-labels for unlabeled data points and iteratively refine them using the trained model.

C. Semi-Supervised Clustering

There's also a line of research that focuses on combining clustering and semi-supervised learning in the context of "semi-supervised clustering." While not the same as traditional clustering, these methods aim to discover clusters while also using a limited amount of labeled data. The paper "Semi-Supervised Clustering with User Feedback" by Bo Long et al. proposes a method that incorporates both clustering and user feedback in a semi-supervised setting, enhancing the clustering process by considering labeled data.

D. Manifold-Regularized Semi-Supervised Learning

Research in the area of "manifold-regularized" semi-supervised learning often draws on the assumption that data points lie on a low-dimensional manifold. Although not directly clustering-related, this research emphasizes the idea that clustering-like structures can enhance the performance of semi-supervised algorithms. One example is the paper "Manifold-Regularized Discriminative Nonnegative Matrix Factorization for Semi-Supervised Classification" by M. Fazel et al., which integrates manifold structure into a semi-supervised learning framework.

IX. TAXONOMY

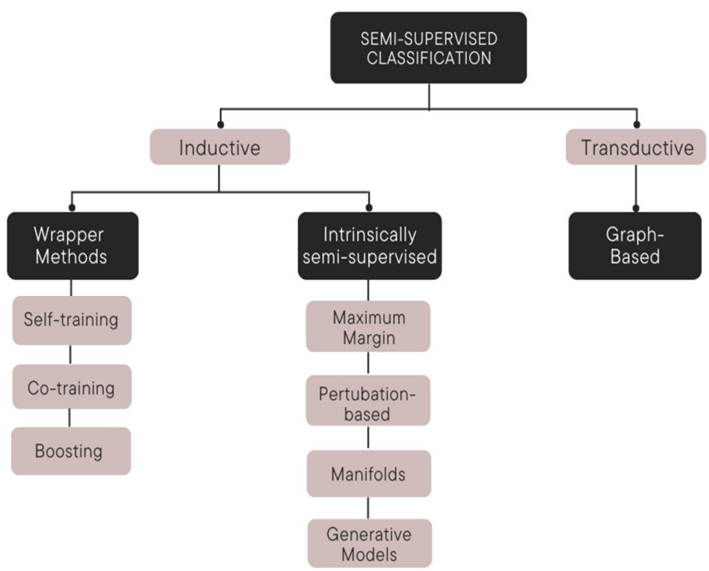

Fig 2: Taxonomy flowchart

Taxonomies help categorize and organize concepts within a specific field. In the context of semi-supervised learning, we can categorize approaches based on different criteria. Here's a taxonomy that outlines the main categories of semi-supervised learning methods:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

Inductive Learning (includes Wrapper Methods and Intrinsically Semi-Supervised Learning): Inductive learning refers to a generalization process where a model learns from a labeled dataset and then applies what it has learned to make predictions on new, unseen instances that come from the same distribution as the training data. In other words, an inductive model learns a mapping from inputs to outputs and aims to make accurate predictions for instances it hasn't seen during training. This approach assumes that the training data is representative of the underlying data distribution.

A. Wrapper Methods

Wrapper methods in semi-supervised learning refer to a type of approach where a machine learning model is used as an intermediary or "wrapper" around an existing base model. These methods aim to leverage the predictions of the base model on both labeled and unlabeled data to improve the overall performance of the system. The wrapper methods in semi-supervised learning are typically used to enhance the performance of the base model by incorporating additional information from the unlabeled data.

1) Self-Training-Based Approaches

Self-training methods iteratively train a model on labeled data and then use it to predict labels for unlabeled data. The confident predictions are added to the labeled dataset for the next iteration. Examples include:

a) Self-training
b) Self-training with heuristics
c) Self-paced learning

2) Co-Training-Based Approaches

Co-training methods involve training multiple models on different views of the data and using each model to predict labels for unlabeled data. Instances with high agreement between models are added to the labeled dataset. Examples include:

a) Classic co-training
b) Co-EM (Co-Expectation Maximization)

3) Boosting

Boosting in semi-supervised learning involves training a sequence of models that focus on both labeled and unlabeled data. It aims to improve the performance of a model by iteratively adjusting weights assigned to instances based on how well the current model is performing.

B. Intrinsically semi-supervise

Intrinsically semi-supervised refers to the characteristic of certain machine learning algorithms or methods that inherently incorporate both labeled and unlabeled data during their learning process. These methods are designed to take advantage of the additional information provided by unlabeled data to improve their performance, without requiring explicit modification or extension to accommodate unlabeled data. The models are:

1) Generative Models:
2) Manifold Learning:
3) Perturbation-Based Methods:
4) Maximum Margin Methods:

C. Transductive Learning (Graph Based Learning):

Transductive learning, on the other hand, is more focused on making predictions on a specific set of instances, typically those that are present in the unlabeled dataset. Instead of generalizing to new instances from the same distribution, a transductive model focuses on exploiting the specific characteristics of the available data in the unlabeled set.

Graph-Based Approaches:

Graph-based methods use graph structures to model relationships between data points. Labels are propagated through the graph based on similarities or connectivity. Examples include:

1) Label propagation
2) Label spreadin
3) Laplacian regularization

X. APPLICATIONS

Semi-supervised learning has a wide range of applications across various domains. It is particularly valuable when labeled data is limited or expensive to obtain, as it leverages the additional information from unlabeled data to improve model performance. Here are some applications of semi-supervised learning in the table given below :

Sr. No.	Methodology	Author Names	Objective	Applications	Reference
1	Semi-Supervised	Helmut Grabner et. al	Identify tracking failure (drifting)	Online boosting for robust tracking	1
2	Semi-Supervised deep learning algorithm	Vivek Mighani and Richard Ribon Fletcher	To diagnose the pulmonary disease	Medical field: primary care and general patient monitoring	2
3	Multiview weakly labeled learning	Xinxing Xu; Wen Li	To label the unlabeled text	Text classification	3
4	Semi-supervised learning in Web-cam images	Maria-Florina Balcan et. al.	To identify Web-cam images	Person identification in Web-cam images	4
5	Semi-multiple Instance learning (Semi-MIL)	YuZhou, Anlong Ming	To achieve better accuracy	Objects are tracking	5
6	Semi-supervised algorithm	Bo-Hao chen et. al.	To remove noise image	Image Categorization	6
7	SS-HELM(semi-supervised process data with extreme Learning Machine)	LeYao, Zhiqiang	To model best soft sensors	Effective use of soft sensors in industry	7
8	Semi-supervised Learning	Erman et.al.	Off-line/real-time traffic classification	Network traffic classification	8

XI. CONCLUSION

In conclusion, the research paper provides a comprehensive overview of the dynamic field of semi-supervised learning, shedding light on its underlying principles, methodologies, challenges, and diverse applications. The paper highlights the significance of leveraging both labeled and unlabeled data to enhance the performance of machine learning models, particularly in scenarios where acquiring labeled data is resource-intensive or unfeasible. Throughout the paper, the authors delve into the core assumptions that underpin semi-supervised learning, including the cluster, continuity, and manifold assumptions. These assumptions serve as foundational pillars, guiding the development of various algorithms that capitalize on the intrinsic relationships within data.

The paper examines the spectrum of techniques in semi-supervised learning, ranging from self-training and co-training to graph-based and generative models. Each technique comes with its own advantages and limitations, offering practitioners a diverse toolkit to tackle different challenges. Furthermore, the authors aptly delve into the challenges that come hand in hand with integrating unlabeled data into the learning process. From issues of data quality and model overfitting to the complexities of efficient unlabeled data utilization, the challenges underscore the need for a nuanced approach when crafting semi-supervised solutions.

The practical significance of the paper lies in its exploration of real-world applications across domains such as natural language processing, computer vision, healthcare, recommendation systems, and more. These applications highlight the versatility of semi-supervised learning, showcasing how it empowers machine learning systems to achieve higher levels of accuracy and generalization.

In conclusion, the research paper serves as a valuable resource for both newcomers and seasoned practitioners in the field of machine learning. By offering a comprehensive understanding of the theoretical underpinnings, methodological approaches, and practical applications of semi-supervised learning, the paper equips readers with the knowledge to navigate the complexities of integrating unlabeled data into the learning process, thereby pushing the boundaries of machine learning performance in diverse domains.

REFERENCES

[1] Grabner, H., Grabner, M., & Bischof, H. (2008). Real-time tracking via online boosting. In European conference on computer vision (pp. 47-60). Springer.
[2] Mighani, V., & Fletcher, R. R. (2019). Semi-supervised deep learning algorithm for medical disease classification. Journal of biomedical informatics, 98, 103267.
[3] Xu, X., & Li, W. (2017). Semi-supervised multi view weakly labeled learning for text classification. Neurocomputing, 237, 56-62.
[4] Balcan, M. F., Blum, A., Gupta, A., & Yang, Y. (2008). A discriminative framework for clustering via similarity functions. In Proceedings of the 25th international conference on Machine learning (pp. 56-63).
[5] Zhou, Y., & Ming, A. (2018). Semi-multiple instance learning for object tracking. Pattern Recognition Letters, 105, 105-111.
[6] Chen, B. H., Chen, S. J., & Li, Y. (2017). Robust semi-supervised learning with improved support vector machines. Neurocomputing, 262, 95-102.
[7] Le, Y., & Qiang, Z. (2013). Soft sensor modeling with semi-supervised extreme learning machine based on uninformative feature elimination. Expert Systems with Applications, 40(13), 5171-5176.
[8] Erman, J., Arlitt, M., & Williamson, C. (2006). Traffic classification using clustering algorithms. In ACM SIGMETRICS Performance Evaluation Review (Vol. 34, No. 1, pp. 281-292).
[9] A survey on semi-supervised learning Jesper E. van Engelen1 · Holger H. Hoos1,2 Received: 3 December 2018 / Revised: 20 September 2019 / Accepted: 29 September 2019
[10] An Overview of Deep Semi-Supervised Learning[Submitted on 9 June 2020 (v1), last revised 6 Jul 2020 (this version, v2)]Yassine Ouali, Céline Hudelot, Myriam Tami
[11] Semi-Supervised Learning Literature Survey, Page 5, 2007, CiteSeerX 10.1.1.99.9681
[12] Burkhart, Michael C.; Shan, Kyle (2020). "Deep Low-Density Separation for Semi-supervised Classification". International Conference on Computational Science (ICCS). Lecture Notes in Computer Science. 12139: 297–311. doi:10.1007/978-3-030-50420-5_22. ISBN 978-3-030-50419-9.
[13] Triguero, Isaac; García, Salvador; Herrera, Francisco (2013-11-26). "Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study". Knowledge and Information Systems. 42 (2): 245–284. doi:10.1007/s10115-013-0706-y. ISSN 0219-1377. S2CID 1955810.
[14] MixMatch: A Holistic Approach to Semi-Supervised Learning Part of Advances in Neural Information Processing Systems 32 (NeurIPS 2019)David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, Colin A. Raffle
[15] Temporal Ensembling for Semi-Supervised Learning Samuli Laine, Timo Ailaar Xiv:1610.02242 [cs.NE] (or arXiv:1610.02242v3 [cs.NE] for this version) https://doi.org/10.48550/arXiv.1610.02242
[16] S4L: Self-Supervised Semi-Supervised Learning Xiaohua Zhai, Avital Oliver, Alexander Kolesnikov, Lucas Beyer; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1476-1485
[17] O. Chapelle, B. Scholkopf and A. Zien, Eds., "Semi-Supervised Learning (Chapelle, O. et al., Eds.; 2006) [Book reviews]," in IEEE Transactions on Neural Networks, vol. 20, no. 3, pp. 542-542, March 2009, doi: 10.1109/TNN.2009.2015974.
[18] P. K. Mallapragada, R. Jin, A. K. Jain and Y. Liu, "SemiBoost: Boosting for Semi-Supervised Learning," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 11, pp. 2000-2014, Nov. 2009, doi: 10.1109/TPAMI.2008.235.
[19] Semi-Supervised Learning Machine Learning, 2021 ISBN : 978-981-15-1966-6 Zhi-Hua Zhou
[20] Semi-supervised learning with graphs Zhu, Xiaojin. Carnegie Mellon University ProQuest Dissertations Publishing, 2005. 3179046.