

Research Article

Open Access 

Predicting Protein Localization Sites Using an Ensemble Self-Labelled Framework

Emmanuel G Pintelas¹ and Panagiotis Pintelas^{*2}¹Department of Electrical & Computer Engineering, University of Patras, Greece²Department of Mathematics, University of Patras, GreeceReceived:  November 05, 2018; Published:  November 19, 2018***Corresponding author:** Panagiotis Pintelas, Department of Mathematics, University of Patras, Greece

Abstract

In recent years machine learning has been thoroughly used in the bioinformatics and biomedical field. The prediction of cellular localization of the proteins can be considered very significant task in bioinformatics since wrong localization site can cause various diseases and infections to humans. Ensemble learning algorithms and semi-supervised algorithms have been independently developed to build efficient and robust classification models. In this paper we focus on the prediction of protein localization site in *Escherichia Coli* and *Saccharomyces cerevisiae* organisms utilizing a semi-supervised self-labeled algorithm based on ensemble methodologies. The experimental results showed the efficiency of our proposed algorithm compared against state-of-the-art self-labeled techniques.

Introduction

Proteins are important molecules in our cells made up of long sequences of amino acid residues [1]. Each protein within the body has a specific function, while they work normally when they are in the correct localization site. The function of a protein in general can be affected by its cellular localization (the location a protein has in a cell) and contributes to many diseases like cardiovascular, metabolic, neurodegenerative diseases and cancer [2]. Also, it is of high interest in various research areas, like therapeutic target discovery, drug design and biological research [3]. Therefore, the prediction of cellular localization of the proteins can be considered very helpful and is a significant task in bioinformatics which has been studied a lot [4-6].

In general, a prediction tool can take as input some attributes of a protein such as its protein sequence of amino acids and predict the location where this protein resides in a cell, such as the nucleus and Endoplasmic reticulum. X-ray crystallography, electron crystallography and nuclear magnetic resonance are some traditionally biochemical experimental methods adopted [7] for predicting protein cellular location. These methods are accurate and precise in general, but they are inefficient and unpractical because they are expensive and time consuming. Therefore, in the last two decades computational methods especially using machine learning methods have been developed to make predictions [5,8-17]. *Escherichia Coli* (*E. coli*) and *Saccharomyces Cerevisiae* (Yeast)

are two well characterized unicellular organisms which have been exhaustively studied [18]. These two organisms have different proteins allocated in their cell where they must be at their accurate positions. A wrong localization site of these proteins in the cell can cause various diseases and infections to humans such as bloody diarrhea [19].

In the past, there have been significant efforts for predicting the localization sites of proteins [18-28]. Anastasiadis and Magoulas [18] investigated the performance of K nearest neighbours, feed-forward neural networks with and without cross-validation and ensemble-based techniques for the prediction of protein localization sites in *E. coli* and Yeast. Their results showed that the ensemble-based techniques had the highest average classification accuracy per class, achieving 91.7% and 66.2% for *E. coli* and Yeast respectively. Chen [22], implemented three different machine learning techniques: Decision tree, perceptrons, two-layer feed-forward neural network for predicting proteins' cellular localization on *E. coli* and Yeast datasets. From the results, a similar prediction accuracy was found for all three techniques and 65%~70% on *E. coli* dataset and 46%~50% on Yeast dataset. Sengur [23], investigated the performance of an artificial immune system based on fuzzy k-NN algorithm. The highest average classification accuracy was 97.29% for *E. coli* and 76.4% for Yeast. Bouziane et al. [21], utilized four supervised machine learning algorithms for the prediction of

cellular localization sites of proteins. For their experiments, they used Naïve Bayesian, k-Nearest Neighbour and feed-forward neural network classifiers.

The highest classification accuracy they managed to achieve was 95.8% for *E. coli* dataset and 73.4% for Yeast dataset. Very recently Priya and Chhabra [19], proposed a hybrid model of Support Vector Machine and the LogitBoost technique for the prediction of the protein localization site in *E. coli* bacteria. The maximum classification accuracy achieved was 95.23%. Motivated by previous work Satu et al. [20], utilized *E. coli* and Yeast datasets for the problem of protein localization prediction. For their experiments they used several data mining classification algorithms which were: lazy classifiers (kNN, KStar), meta classifiers (Iterative Classifies Optimizer, Logit boost, Random Committee, Rotation Forest), function classifiers (Logistics, Simple Logistics), tree classifier (LMT, Random Forest, Random Tree) and artificial neural networks, achieving 87.50% with Rotation Forest and 60.53% with Random Forest maximum classification accuracy for *E. coli* and Yeast respectively.

Nevertheless, the problem of prediction of protein localization sites is considered a challenging task since finding labeled data is often an expensive and time-consuming procedure [29], as it requires human efforts. To address this problem, Semi-Supervised Learning (SSL) algorithms utilize both labeled and unlabeled data since in general finding sufficient unlabeled data is significantly easier than finding labeled data [30-32]. The basic aim of SSL is to exploit the hidden information found in the unlabeled data in order to train classifiers more efficiently [33,34]. The most popular SSL algorithms are self-labeled algorithms. These algorithms make predictions on a large amount of unlabeled data aiming to enlarge a small amount of labeled data. Triguero et al. [35] made a taxonomy of self-labeled algorithms based on their main characteristics and conducted a comprehensive research of their classification efficacy on several datasets. Some of the most efficient and popular Self-labeled algorithms proposed in the literature are Self-training [30], Co-training [31], Tri-training [35], Democratic-Co learning [37], Co-Forest [38] and Co-Bagging [39].

In Self-training, one classifier following an iterative procedure is trained on a labeled dataset which is augmented by its most confident predictions on an unlabeled dataset. In Co-training, two classifiers are trained separately using two different views on a labeled dataset and then each classifier adds the most confident predictions on an unlabeled dataset to the training set of the other. Tri-training algorithm utilizes three classifiers which teach each other based on a majority voting strategy. Democratic-Co learning utilizes several classifiers following a majority voting and confidence measurement strategy for predicting the values of unlabeled examples. Co-Forest algorithm trains Random trees on bootstrap data from the dataset assigning few unlabeled examples to each tree, utilizing a majority voting. Co-Bagging algorithm trains multiple base classifiers on bootstrap data created by random resampling with replacement from the training set.

Ensemble Learning (EL) is a different approach, which has been developed in the last decades, for building more efficient composite

global model by the combination of several prediction models than using a single one [40]. Moreover, the combination of SSL and EL are beneficial to each other [41], leading to even better classification results by developing more accurate and robust classifiers [42-47] than utilizing EL and SSL independently. Recently, Livieris et al. [43,45] proposed some ensemble SSL algorithms which utilize the individual predictions of the most popular self-labeled methods i.e. Self-training, Co-training and Tri-training based on a combination of various voting strategies. Motivated by previous work, Livieris et al. [48] proposed a new semi-supervised learning algorithm which selects the most promising base learner from a number of classifiers utilizing a Self-training methodology.

In this work, we propose a semi-supervised self-labeled algorithm based on the ensemble approach for the prediction of protein localization sites on *E. coli* and Yeast organisms. The proposed algorithm constitutes a modification of the CST-Voting, utilizing each self-labeled algorithm with the base learner, which presents the highest accuracy. It is worth mentioning that we utilized only a 10%-50% ratio of the training set in our experiments in order to evaluate the efficiency of the SSL approach. Our experimental results reveal the efficiency of the proposed algorithm compared against state-of-the-art self-labeled methods. The remainder of this paper is organized as follows: Section 2 presents the proposed classification algorithm and a brief description of the data utilized in our study. Section 3 presents a series of experiments in order to evaluate the accuracy of the proposed algorithm against the most popular self-labeled classification algorithms. Finally, in Section 4 we present our concluding remarks.

Proposed Methodology

The main goal of the research described in this paper is the development of a prediction model for the classification of protein localization site in *Escherichia Coli* (*E. coli*) and *Saccharomyces Cerevisiae* (Yeast) organisms utilizing a semi-supervised self-labeled algorithm. For this purpose, we adopted a two-stages methodology, where the first stage deploys the self-labeled classification algorithm while the second one concerns dataset utilized in this study.

CST*-Voting Algorithm

In this section, we present a detailed description of the proposed SSL algorithm for the prediction of protein localization, which is based on an ensemble philosophy, entitled CST*-Voting. Recently, Livieris et al. [43], proposed the CST-Voting algorithm which combines the self-labeled framework along with ensemble learning techniques. In particular, this algorithm exploits the individual predictions of the most popular self-labeled algorithms namely, Co-training, Self-training and Tri-training utilizing simple majority voting. These self-labeled methods operate in a different way to take advantage of the hidden information found in the unlabeled data in order to enlarge a labeled dataset. The main difference between these self-labeled algorithms is the technique used to exploit the unlabeled data. More specifically, self-training and tri-training are single-view methods, while co-training is a multi-view method. Furthermore, it is worth mentioning that co-

training and tri-training are indeed ensemble methods, since they both make use of multiple classifiers.

Along with this line, we consider to improve the classification efficiency of the ensemble, by utilizing each self-labeled algorithm with the base learner, which presents the highest accuracy. To this end, Co-training utilizes Sequential Minimum Optimization (SMO) [49] as base learner, Self-training utilizes Multilayer perceptron (MLP) [50] and Tri-training utilizes C4.5 [51]. The motivation for this selection is based upon the fact that these algorithms were reported to present the best efficiency using these specific base learners [35,43]. A high-level description of the proposed CST*-Voting is presented in Algorithm 1. Initially, the classical semi-supervised algorithms, which constitute the ensemble, i.e., self-training (MLP), co-training (SMO) and tri-training (C4.5), are trained utilizing the same labeled and unlabeled U dataset. Subsequently, the final hypothesis on an unlabeled example of the test set combines the individual predictions of the self-labeled algorithms, hence utilizing a majority voting. Clearly, the ensemble output is the one made by more than half of them. An overview of proposed algorithm is depicted in Figure 1.

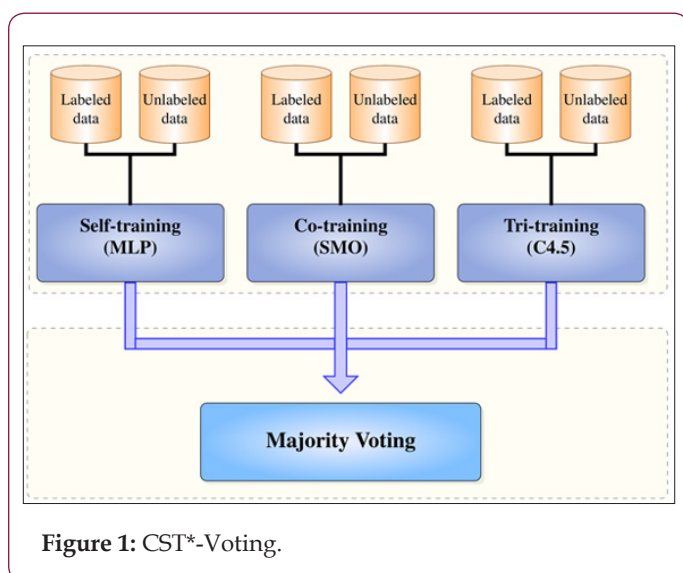


Figure 1: CST*-Voting.

Algorithm 1. CST*-Voting

Input: L- Set of labeled instances.

U- Set of unlabeled instances.

T- Training set.

Output: Trained ensemble classifier

(Phase I: Training)

- 1: Self-training(L, U) using MLP as base-learner.
- 2: Co-training(L, U) using SMO as base-learner.
- 3: Tri-training(L, U) using C4.5 as base-learner.

(Phase II: Voting)

Comment: The labels of instances in the testing set are placed in L through steps 4 - 7.

4: for each $x \in T$ **do**

5: Apply the trained classifiers on x.

6: Use majority vote to predict the label y^* of x.

7: end for

Datasets Description

In the experiments were used *E. coli* and Yeast datasets for the localization of proteins. Both are collected from UCI Machine learning data repository. *E. coli* dataset has 336 instances which are labeled into 8 classes. The attributes on this dataset are mcg, gvh, lip, chg, aac, alm1, alm2 which all are numeric. There are eight classes namely: CP (Cytoplasm), IM (Inner Membrane without signal sequence), PP(Periplasm), IMU (Inner Membrane with Uncleavable signal sequence), OM (Outer Membrane), OML (Outer Membrane Lipoprotein), IML (Inner Membrane Lipoprotein), IMS (Inner Membrane with cleavable signal sequence). Yeast dataset has 1462 instances which are labeled into 10 classes. The attributes on this dataset are mcg, gvh, alm, mit, erl, pox, vac, nuc which all are numeric. There are ten classes namely: Cytoplasmic (CYT), Nuclear (NUC), Vacuolar (VAC), Mitochondrial (MIT), Peroxisomal (POX), Extracellular (EXC), Endoplasmic Reticulum (ERL), Membrane proteins with a cleaved signal (ME1) membrane proteins with an uncleaved signal (ME2) and membrane proteins with no N-terminal signM (ME3).

Experimental Results

Next, we focus our interest on the experimental analysis for evaluating the classification performance of CST*-Voting against the most efficient and frequently utilized self-labeled methods, i.e. Self-training, Co-training, Tri-training. Notice that all self-labeled methods deployed base learners the SMO, the C4.5 and the MLP algorithm. These supervised classifiers probably constitute the most effective and popular machine learning algorithms for classification problems [50]. All self-labeled algorithms utilized the configuration parameter settings as in [44-48] and all base learners were used with their default parameter settings included in the WEKA 3.9 library [51] in order to minimize the effect of any expert bias, instead of attempting to tune any of the algorithms to the specific dataset. Furthermore, in order to study the influence of the amount of labeled data, five different ratios (R) of the training data were used, i.e. 10%, 20%, 30%, 40% and 50%. Tables 1 & 2 present the performance of all self-labeled methods on *E. coli* dataset and Yeast dataset, respectively. Notice that the highest classification performance for each labeled ratio and performance metric is highlighted in bold.

Table 1: Accuracy of the self-labeled algorithms on *E. coli* dataset.

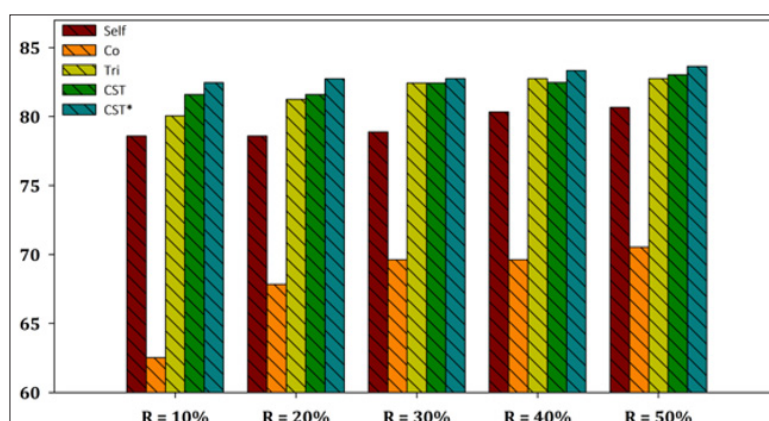
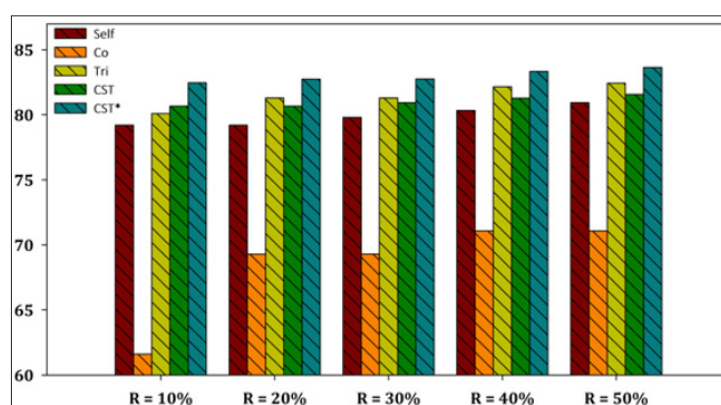
Ratio	C4.5				SMO				MLP				CST*
	Self	Co	Tri	CST	Self	Co	Tri	CST	Self	Co	Tri	CST	
10%	78,6%	62,5%	80,1%	81,6%	78,9%	65,5%	79,2%	80,7%	79,2%	61,6%	80,1%	80,7%	82,5%
20%	78,6%	67,8%	81,3%	81,6%	79,5%	66,7%	80,1%	81,8%	79,2%	69,3%	81,3%	80,7%	82,7%
30%	78,9%	69,6%	82,4%	82,4%	79,5%	67,8%	80,4%	81,8%	79,8%	69,3%	81,3%	80,9%	82,8%
40%	80,3%	69,6%	82,8%	82,5%	80,7%	71,7%	80,4%	82,4%	80,3%	71,1%	82,2%	81,3%	83,3%
50%	80,7%	70,5%	82,8%	83%	82,5%	71,7%	81,3%	83,3%	80,9%	71,1%	82,4%	81,6%	83,7%

Table 2: Accuracy of the self-labeled algorithms on Yeast dataset.

Ratio	C4.5				SMO				MLP				CST*
	Self	Co	Tri	CST	Self	Co	Tri	CST	Self	Co	Tri	CST	
10%	72,6%	71,2%	74,9%	74,9%	72,4%	72,7%	73,7%	74,1%	73,7%	72,1%	73,2%	74,1%	75,0%
20%	72,6%	71,4%	74,9%	75,1%	72,4%	73,1%	73,7%	74,4%	73,7%	72,3%	73,2%	74,1%	75,7%
30%	74,3%	72,4%	75,2%	75,3%	73,7%	74,1%	73,7%	74,4%	74,7%	73,2%	74,2%	74,6%	76,0%
40%	74,5%	72,4%	75,3%	76%	73,9%	74,1%	74,1%	74,8%	74,8%	73,3%	74,9%	75%	77,1%
50%	76,6%	73,3%	75,8%	76,1%	74,1%	74,9%	74,1%	74,8%	75,5%	73,3%	76,4%	76,6%	77,6%

The aggregated results showed that the CST*-Voting was by far the most efficient and robust method independent of the utilized ratio of labeled instances in the training set, performing better in all five ratio cases for both datasets. Furthermore, a more representative visualization of the classification accuracy of all compared SSL algorithms is presented in Figures 2-7 for *E. coli* and Yeast data-

sets respectively. In more detail, these figures present a performance comparison of CST*-Voting versus Self-training, Co-training, Tri-training, CST-Voting which utilize C4.5, SMO and MLP as base learners. Each box-plot presents the accuracy measure for each tested SSL algorithm according to the labeled ratio.

**Figure 2:** Performance evaluation of each SSL algorithm using C4.5 as base learner versus CST*-Voting for *E. coli* dataset.**Figure 3:** Performance evaluation of each SSL algorithm using MLP as base learner versus CST*-Voting for *E. coli* dataset.

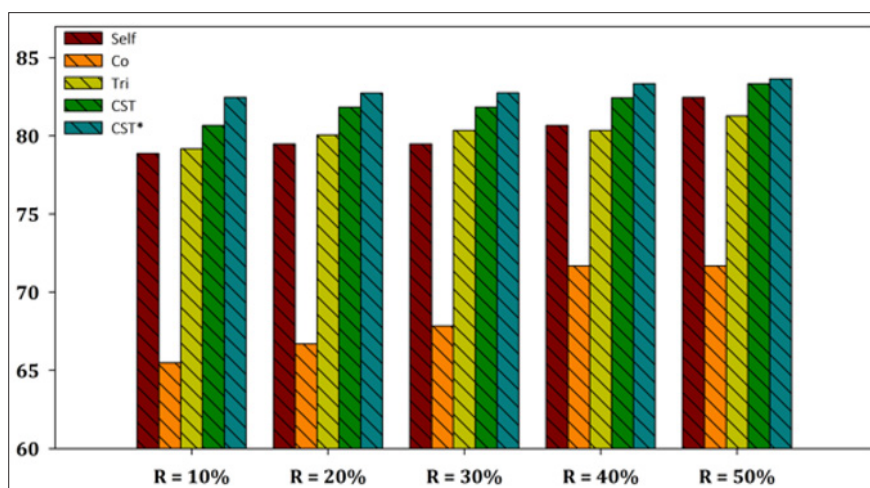


Figure 4: Performance evaluation of each SSL algorithm using SMO as base learner versus CST*-Voting for E. coli dataset.

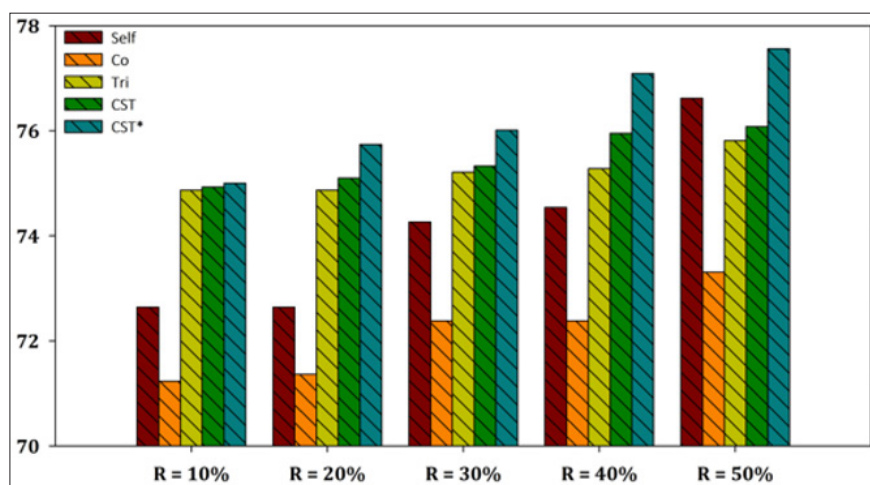


Figure 5: Performance evaluation of each SSL algorithm using C4.5 as base learner versus CST*-Voting for Yeast dataset.

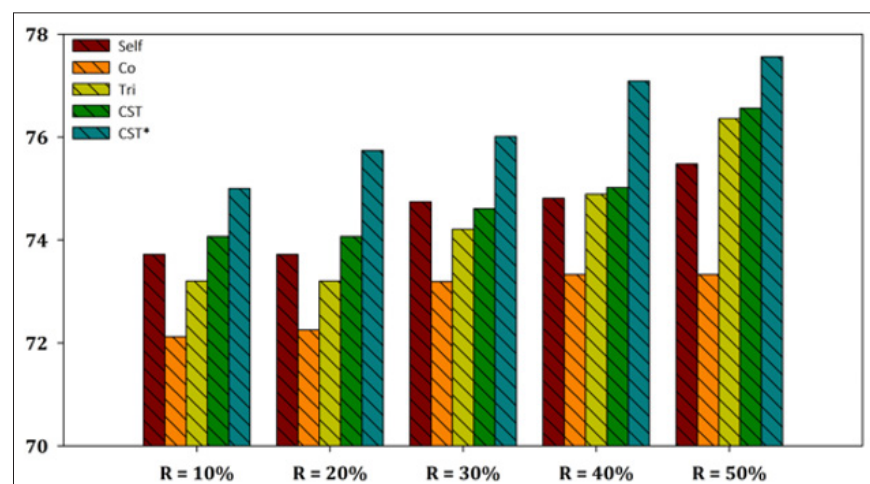


Figure 6: Performance evaluation of each SSL algorithm using MLP as base learner versus CST*-Voting for Yeast dataset.

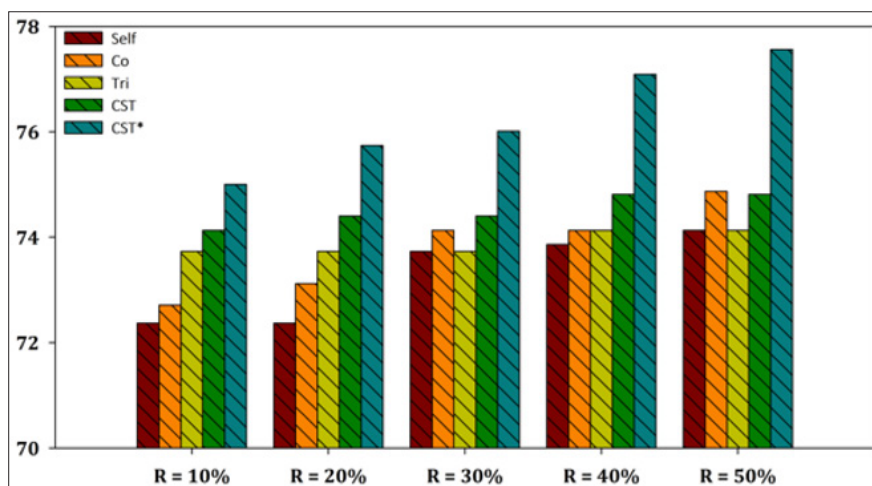


Figure 7: Performance evaluation of each SSL algorithm using SMO as base learner versus CST*-Voting for Yeast dataset.

Conclusion

In this work, we evaluated the performance of an ensemble-based self-labeled algorithm for protein localization sites, called CST*-Voting using two datasets (*E. coli* and Yeast). The proposed algorithm constitutes a modification of the CST-Voting, utilizing three self-labeled algorithms i.e. Self-training, Tri-training and Co-training, using the base learner which presented the highest accuracy in literature. A series of experiments were carried out in order to evaluate the classification performance of the proposed algorithm against the most efficient and frequently utilized self-labeled methods. To this end, we utilized only a 10%-50% ratios of the training set in our experiments, instead of the entire dataset, in order to evaluate the efficiency of the SSL approach. As our experimental results have shown, the efficiency of the proposed algorithm is better compared against state-of-the-art self-labeled methods. In our future work we intend to invest on extending our experiments of the proposed algorithm to several organism's cells for protein localization prediction and on improving the prediction accuracy of ensemble SSL utilizing more efficient and sophisticated self-labeled algorithms.

References

- Ahmed HR, Glasgow J (2014) A novel particle swarm-based approach for 3D motif matching and protein structure classification. Canadian Conference on Artificial Intelligence p. 1-12.
- Hung MC, Link W (2011) Protein localization in disease and therapy. J Cell Sci 124(20): 3381-3392.
- Eisenhaber F, Bork P (1998) Wanted: Subcellular localization of proteins based on sequence. Trends Cell Biol 8(4): 169-170.
- Emanuelsson, O, Brunak S, Von Heijne G, Nielsen H (2007) Locating proteins in the cell using targetp, signalp and related tools. Nat Protoc 2(4): 953-971.
- Imai K, Nakai K (2010) Prediction of subcellular locations of proteins: Where to proceed? Proteomics 10(22): 3970-3983.
- Wan S, Mak MW (2015) Machine learning for protein subcellular localization prediction. De Gruyter.
- Murphy RF, Boland MV, Velliste M (2000) Towards a systematics for protein subcellular location: quantitative description of protein localization patterns and automated analysis of fluorescence microscope images. Proc Int Conf Intell Syst Mol Biol 8: 251-259.
- Chou KC, Shen HB (2007) Recent progress in protein subcellular location prediction. Anal Biochem 370(1): 1-16.
- Chou KC (2013) Some remarks on predicting multi-label attributes in molecular biosystems. Mol Biosyst 9(6): 1092-1100.
- Du P, Xu C (2013) Predicting multisite protein subcellular locations: Progress and challenges. Expert Rev Proteomics 10(3): 227-237.
- Chou KC (2011) Some remarks on protein attribute prediction and pseudo amino acid composition. J Theor Biol 273(1): 236-247.
- Du P, Li T, Wang X (2011) Recent progress in predicting protein subcellular locations. Expert Rev Proteomics 8(3): 391-404.
- Chen Lin, Ying Zou, Ji Qin, Xiangrong Liu, Yi Jiang, et al. (2013) Hierarchical classification of protein folds using a novel ensemble classifier. PLoS One 8(2): e56499.
- Liu X, Li Z, Liu J, Liu L, Zeng X (2015) Implementation of arithmetic operations with time-free spiking neural P systems. IEEE Trans Nanobioscience 14(6): 617-624.
- Wu Y, Krishnan S (2011) Combining least-squares support vector machines for classification of biomedical signals: A case study with knee-joint vibroarthrographic signals. J Exp Theor Artif Intell 23(1): 63-77.
- Song L, Li D, Zeng X, Wu Y, Guo Li, et al. (2014) nDNA-prot: Identification of DNA-binding proteins based on unbalanced classification. BMC bioinformatics 15(1): 298.
- Wei L, Liao M, Gao Y, Ji R, He Z, et al. (2014) Improved and promising identification of human microRNAs by incorporating a high-quality negative set. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 11(1): 192-201.
- Anastasiadis AD, Magoulas GD (2006) Analysing the localisation sites of proteins through neural networks ensembles. Neural Computing & Applications 15(3-4): 277-288.
- Priya B, Chhabra A (2015) Prediction of protein cellular localization site by using data mining techniques. IEEE pp. 731-736.
- Satu MS, Akter T, Uddin MJ (2017) Performance analysis of classifying localization sites of protein using data mining techniques and artificial neural networks. IEEE p. 860-865.

21. Bouziane H, Messabih B, Chouarfia A (2011) A voting-based combination system for protein cellular localization sites prediction. In IEEE International Conference on Information and Computer Applications (ICICA) pp. 166-173.
22. Chen Y (2008) Predicting the cellular localization sites of proteins using decision tree and neural networks.
23. Sengur A (2009) Prediction of protein cellular localization sites using a hybrid method based on artificial immune system and fuzzy k-NN algorithm. Digital Signal Processing 19(5): 815-826.
24. Nakai K, Kanehisa M (1991) Expert system for predicting protein localization sites in gram-negative bacteria. Proteins 11(2): 95-110.
25. Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics 14(4): 897-911.
26. Horton P, Nakai K (1996) A probabilistic classification system for predicting the cellular localization sites of proteins. Proc Int Conf Intell Syst Mol Biol 4: 109-115.
27. Horton P, Nakai K (1997) Better prediction of protein cellular localization sites with the its k nearest neighbors classifier. Proc Int Conf Intell Syst Mol Biol 5: 147-152.
28. Cairns P, Huyck C, Mitchell I, Wu WX (2001) A comparison of categorisation algorithms for predicting the cellular localization sites of proteins. IEEE pp. 296-300.
29. Zhu, X, Ghahramani Z, Lafferty JD (2003) Semi-supervised learning using gaussian fields and harmonic functions. ICML pp. 912-919.
30. Yarowsky D (1995) Unsupervised word sense disambiguation rivaling supervised methods. Association for Computational Linguistics p. 189-196.
31. Blum A, Mitchell T (1998) Combining labeled and unlabeled data with co-training. ACM p. 92-100.
32. Duda, RO, Hart PE (1973) Pattern classification and scene analysis. A Wiley-Interscience Publication pp. 472.
33. Livieris I, Kanavos A, Tampakas V, Pintelas P (2018) An Ensemble SSL Algorithm for Efficient Chest X-Ray Image Classification. Journal of Imaging 4(7): 95.
34. Livieris IE (2019) A new ensemble self-labeled semi-supervised algorithm. Informatica p. 1-14.
35. Triguero I, Garcia S, Herrera F (2015) Self-labeled techniques for semi-supervised learning: Taxonomy, software and empirical study. Knowledge and Information Systems 42(2): 245-284.
36. Zhou ZH, Li M (2005) Tri-training: Exploiting unlabeled data using three classifiers. IEEE 17(11): 1529-1541.
37. Zhou Y, Goldman S (2004) Democratic co-learning. IEEE pp. 594-602.
38. Li, M, Zhou ZH (2007) Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. IEEE 37(6): 1088-1098.
39. Hady MFA, Schwenker F (2010) Combining committee-based semi-supervised learning and active learning. J Computer Sci Technol 25(4): 681-698.
40. Rokach, L (2010) Pattern classification using ensemble methods. World Scientific 75: 244.
41. Zhou ZH (2011) When semi-supervised learning meets ensemble learning. Frontiers of Electrical and Electronic Engineering in China 6(1): 6-16.
42. IE Livieris, E Pintelas, A Kanavos, P Pintelas (2018) Identification of blood cells subtypes from images using an improved SSL algorithm. Biomed J Sci Tech Res p. 7.
43. Livieris I, Kanavos A, Tampakas V, Pintelas P (2018) An Ensemble SSL Algorithm for Efficient Chest X-Ray Image Classification. Journal of Imaging 4(7): 95.
44. Livieris IE, Drakopoulou K, Mikropoulos TA, Tampakas V, Pintelas, P (2018) An ensemble-based semi-supervised approach for predicting students' performance. In Research on e-Learning and ICT in Education p. 25-42.
45. Livieris IE (2019) A new ensemble self-labeled semi-supervised algorithm. Informatica p. 1-14.
46. Livieris IE, Pintelas EG, Kanavos A, Pintelas P (2018) An improved self-labeled algorithm for cancer prediction. Advances in Experimental Medicine and Biology.
47. Livieris IE, Kiriakidou N, Kanavos A, Tampakas V, Pintelas P (2018) On ensemble SSL algorithms for credit scoring problem. Informatics.
48. IE Livieris, Kanavos A, Tampakas V, Pintelas P (2018) An Auto-Adjustable Semi-Supervised Self-Training Algorithm. Algorithms 11(9): 139.
49. Platt JC (1999) Using analytic QP and sparseness to speed training of support vector machines. In Advances in neural information processing systems pp. 557-563.
50. Rumelhart DE, McClelland JL (1986) Parallel distributed processing: explorations in the microstructure of cognition. volume 1. foundations.
51. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, et al. (2009) The WEKA data mining software: An update. ACM SIGKDD explorations newsletter 11(1): 10-18.

ISSN: 2574-1241

DOI: [10.26717/BJSTR.2018.11.002066](https://doi.org/10.26717/BJSTR.2018.11.002066)

Panagiotis Pintelas. Biomed J Sci & Tech Res



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>**Assets of Publishing with us**

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>