Increasing Weak Classifiers Diversity by Omics Networks

Vladimír Kunc, Jiří Kléma, and Michael Anděl

Department of Computer Science,
Czech Technical University in Prague,
Technická 2, 166 27 Prague, Czech Republic,
{kuncvlad,klema,andelmi2}@fel.cvut.cz
http://ida.felk.cvut.cz/

Abstract. The common problems in machine learning from omics data are the scarcity of samples, the high number of features and their complex interaction structure. The models built solely from measured data often suffer from overfitting. One of possible methods dealing with overfitting is to use prior knowledge for regularization. This work analyzes contribution of feature interaction networks in regularization of ensemble classifiers representing another approach to overfitting reduction. We study how utilization of feature interaction networks influences diversity of weak classifiers and thus accuracy of the resulting ensemble model. The network and its random walks are used to control the feature randomization during construction of weak classifiers, which makes them more diverse than in the well-known random forest. We experiment with different types of weak classifiers (trees, logistic regression, naïve Bayes) and different random walk lengths and demonstrate that diversity of weak classifiers grows with increasing network locality of weak classifiers.

Keywords: ensemble learning, random forest, prior knowledge, diversity, gene expression

1 Introduction

In recent years, the fields of genomics, proteomics and metabolomics (collectively omics) have been strongly influenced by progress in high-throughput technologies which led to a boost of generated data. These *omics* data are thoroughly analyzed to learn about the mechanisms of shaded diseases, to predict the disease onset and progression, or to set a proper treatment protocol. One of the common problems in learning from omics data is the scarcity of available samples, namely when compared with their large dimensionality. This inconvenient ratio, altogether with common noisiness, leads to *overfitting* as the number of possible hypotheses immensely exceeds the number of training examples. In the field of machine learning, the overfitting is commonly addressed by the means of *regularization*, when a prior hypothesis is imposed during the learning process. Another approach to deal with overfitting, namely the one caused by the noise,

is *ensemble learning*. Ensemble methods train multiple classifiers and use these classifiers to create an aggregated classifier for a single task. These ensembles usually outperform each of the base classifiers, which they are composed from, in most classification tasks [7,8]. The key assumption of ensemble models is that the underlying classifiers are diverse, i.e., that they make different errors, and thus they can together achieve higher predictive performance than could have been obtained from any of the constituent classifiers [7,23].

In this paper, we study how regularization by domain knowledge may contribute to the diversity of ensemble classifier and consequently boost its performance. This paper follows the work of [2,1] where random forests (RFs) [5] get enriched with domain knowledge. The knowledge is here defined as prior known interaction between omics features. Since the interacting features are considered correlated, the base classifiers built on different sets of interacting features are assumed to be decorrelated and therefore more accurate in common. The goal is to show how ensemble classifiers can increase their diversity due to prior knowledge and be further used for better predictions of the onset and progression of heterogeneous multifactorial diseases such as *myelodysplastic syndrome* that serves as a case study in the experimental part.

2 Motivation and Related Work

The key feature of an ensemble is the diversity between its base classifiers. The ensemble provides higher accuracy, only if the ensemble members disagree about some inputs [6,15,25]. So far, there have been several attempts to increase this diversity and potentially its ability of generalization.

A straightforward approach to deliver the ensemble diversity is *manipulating training samples*. This method is applied to unstable classifiers such as neural networks (NNs) and decision trees (DTs) [20]. Most known examples are *bagging* [4] and *boosting* [22]. Similarly, one may *manipulate with the feature set*. The random subspace (RS) method (also called attribute bagging) creates random subspaces of the feature space, each base classifier is trained using one of these subspaces [20,21,11]. This method is recommended especially when the dimension of the feature space is very high and most other classification methods suffer from the *curse of dimensionality* [11]. An orthogonal way to achieve diversity is by *manipulating the algorithm* that creates base classifiers. The learning algorithm can change its parameters, for example the topology of NN [12], the pruning factor of DTs [20] or the starting point and the way of traversal in the hypothesis search space [6]. According to Rokach [20], there are two methods of manipulating the space traversal — *random-based strategy* and *collective performance strategy* [20]. The random-based strategy uses randomness to gain higher diversity; one of the most common examples is a random forest [5], in which a weak classifier does not select the best feature in each its node but only the best feature from a certain feature subset. A different forest randomization strategy was used in [8], where, for each tree, "the 20 best candidate splits are computed, and then one of these is chosen uniformly at random". Contrarily, the *collective*

performance based strategy creates the ensemble as a whole while trying to increase its accuracy by various means. The base classifiers might cooperate with each other in order to specialize, i.e., be diverse from others [20].

The most impacting of collective performance strategy is the family of *penalty methods*. They add a *penalty term* to the *error function* of an ensemble to encourage diversity among base classifiers [6,20,21]. Several penalty methods were proposed and analysed in the literature — e.g., *negative correlation learning* [20,6] or *root-quartic negative correlation learning* [6].

Besides ensembles, another approach to address overfitting is *regularization*. It applies especially when the sample size n is much smaller than the number of measured features p, $n \ll p$. As mentioned in Sect. 1, regularization means imposing certain hypothesis during the learning process, while this need not to be admitted if there is not enough training examples in its favour. This is very suitable in the case of domain knowledge [1,19,16]. The hypothesis may be defined as a set of feature interactions, which shows valid (or not) in a certain domain context. Nonetheless, the hypothesis imposed may be defined uninformedly too, merely by restraining the hypothesis space geometrically as in the case of margin classifiers, or purely as restrained hypothesis space as in the case of DT pruning.

As mentioned above, one of the ways to make base classifiers diverse, is to manipulate their learning hypotheses. Here we can see how regularization meets ensemble learning, that regularizing each of base classifiers should make them diverse and potentially more powerful altogether. This was a motivation for our recent ensemble method [1], where the base trees are induced only from the genes lying close to each other in omics, namely protein-interaction, network. In the other words, as genes whose corresponding proteins bind or interact are assumed to be correlated in their mRNA expression, we assume that trees built on the interacting features shall be decorrelated in their prediction. Disagreement between the base prediction is fundamental for the ensemble diversity, being the key quantity in its measuring [15,6]. In this paper we investigate possibilities to gain this diversity through the domain knowledge about protein interactions.

3 Methods

In this chapter we briefly describe the measures commonly used for quantifying the ensemble diversity. Next, we discuss our recently proposed ensemble method, which we here further generalize and investigate as to its diversity induced by domain knowledge. There are two main approaches of measuring the diversity, *pairwise* and *non-pairwise* [15,6]. The *non-pairwise* measures mostly compare the output of base classifiers with the averaged output of the whole ensemble or are based on the idea of entropy. The *pairwise* measures calculate the average of a particular measure of all possible pairings of ensemble members [6]. For this research, four diversity measures were chosen — *entropy* [15] and *Kohavi-Wolpert measure* (KW) [15,13] represent non-pairwise measures, *average Q statistics* (Q_{ave}) [15,28] and *double-fault measure* (DF) [15,10] fall into pairwise measures.

3.1 Network-constrained Forest (NCF)

NCF algorithm [1] combines two approaches to solving the $n \ll p$ problem common in the omics field, it utilises prior knowledge for creation of an ensemble of decision trees. Unlike RF, the NCF biases "the feature sampling process towards the genes and loci in general, which have been previously reported as candidates for causing the phenomenon being studied (...) and consequently the omics features which directly or indirectly interact with those candidate genes" [1]. This sampling process is driven by a random walk on the biological interaction network integrating both mRNA and miRNA prior knowledge, the process starts from the candidate causal genes called *seeds*. When candidates causal genes are unknown, seeds are randomly sampled from the entire set and the probability of a gene being sampled as a seed is proportional to its out degree in the network. Further implementation details and pseudo-code are available in [1,2].

The crucial assumption behind the NCF is that gene that are close in the biological feature network are also correlated in their expression, therefore it is suitable to create weak classifiers grouping these features because it leads to decorrelating the individual weak classifiers and therefore better ensemble diversity. The biological background behind this method is discussed in [1] and the conclusion presented is that the weak "DTs may vaguely correspond to the individual disease factors and their network-local manifestations" [1]. The individual trees are constructed using the features in the network neighbourhood of a particular seed gene that was chosen for the tree. The neighbourhood is represented by a distribution function using which the feature set is sampled. This distribution is defined as a random walk of length k from the seed gene — it is more dense when closer to the seed gene and also it is not possible to reach genes that are further in the network than k. Therefore, the NCF is parametrized by the walk length k whose optimal value may be different for different tasks as it strongly influences the feature sampling [1]. A heuristic based on *incidence of underfitted trees* for setting the parameter k was proposed in [2]. The influence of k on the accuracy and diversity of weak learners and the overall accuracy of the ensemble is further analyzed in Sect. 4.

3.2 The Novel Method of Network-constrained Random Subspaces

In this section, we propose a generalization of the NCF algorithm called *network-constrained random subspaces* (NCRS) which applies the idea of biased sampling of the feature set to the general ensemble *random subspace* method (see Sect. 2). The idea of NCF is not strictly related to ensembles of DTs and it is easily extensible to ensembles of other weak learners. Even though DTs as weak classifiers of forest have many advantages as, for example, direct interpretability and possible use of such forest for feature selection, other classifiers such as *logistic regression* (LR) or *naïve Bayes* (NB) might be used as well. Moreover, turning RF and NCF into a tree independent ensemble method is allowed by a simple modification of the algorithm, feature sampling performed independently in tree nodes changes into sampling performed before construction of a whole weak learner.

The relationship between RF, RS, NCF, and NCRS is depicted in Fig. 1 — the RF and the NCF both sample the feature space in each node of each tree, however, the RS and the NCRS both sample the feature space only for each weak learner. The sampling in the NCF and the NCRS is network-constrained, i.e., its sampling procedure generates samples using random walks over the interaction network, while the sampling procedure in the RF and the RS is random.

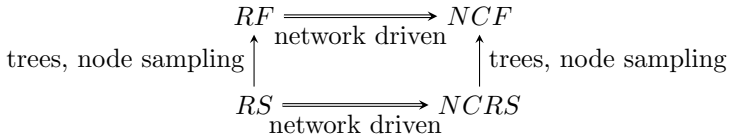


Fig. 1: The relationship between RF, RS, NCF and NCRS.

4 Experiments

Further described experiments had several objectives. First of all, the goal was to verify the accuracy of the newly proposed NCRS method, namely the generalization from DTs used in [2,1] towards LR and NB as weak classifiers. The second objective was to analyze the impact of different values of the parameter k defining the length of a random walk on the accuracy of both the whole ensemble and also of the individual weak classifiers. Moreover, [2,1] implies that the diversity of weak classifiers should be strongly influenced by the parameter k and in most cases, a longer walk should lead to smaller diversity among the weak classifiers in the ensemble as they become less specialized. The parameter k was analysed for similar values as used in [2,1] but also for more extreme values — e.g., for a random walk of length 100.

Another objective was to experimentally validate the convergence of both NCRS and NCF as $k \to \infty$. The NCF does not converge to RF, rather it converges to the stationary distribution of random walk $\pi^{\infty}(v) = \frac{\deg(v)}{|\mathcal{I}|}$, where $\mathcal{I}$ is the the set of edges in the biological network [1]. However, the NCF converges to the stationary distribution only if there are no miRNA interactions present because such interactions are handled in a special way — when encountering the miRNA node in the walk, the walk always ends there, details are again available in [1]. The convergence was not experimentally validated in [1].

4.1 Domain and Data

Data related to *myelodysplastic syndrome* (MDS) were used for most of the experiments. It is the same data that was used in the original experiment with NCF [2,1]. The data were provided by a collaborative laboratory at the Institute of Hematology and Blood Transfusion in Prague. The data were obtained for analysis of lenaledomide treatment of patients with myelodysplastic syndrome.

The data consist of two datasets — mRNA with 16,666 attributes measuring the gene expression level and miRNA with 1,146 attributes measuring the

expression level of particular miRNAs [1]. The samples were obtained from bone marrow (BM) CD34+ progenitor cells and from peripheral blood (PB) CD14+ monocytes and were obtained either before the treatment (BE) or during the treatment (DU). Moreover, the data can be further categorized by the partial deletion of the chromosome 5 (5q or non-5q). Using these categories, the data consisting of 75 samples were divided into 10 related datasets.

Again, for the coherence of experiments with [2,1], the same prior knowledge in the form of gene networks and candidates causal genes was used. The prior knowledge is publicly available — in vitro validated miRNA-mRNA interactions are from TarBase 6.0 [24], in silico predicted interactions are from miRWalk database [9], experimentally validated protein-protein interactions are from Human Protein Reference Database [18], predicted protein-protein interactions are from [3] and MDS causal genes are from [27], according to [1].

4.2 Experimental Protocol

The NCRS ensemble classifier was implemented in Python 3 as a modification of both the original NCF [1] and general *Bagging classifier* from machine learning library Scikit-learn [17] version 0.16.1. 10 times repeated stratified m-fold cross-validation was used for MDS experiments, where $m := \min\{10, c\}$, where c is the number of samples in the smallest class. This setting of m maximizes the number of stratified folds in tasks with small sample sets and keeps the common number of folds for the rest of tasks. All ensembles were built from 1000 weak classifiers using the RS method, each weak classifier accessed 100 features. Both the parameters were set in advance with no tuning. The number of weak classifiers was strongly limited by computational costs of both learning period and calculating pairwise diversity measures. The number of accessed features roughly followed the rule of thumb $\sqrt{p}$ implemented in [17].

Matthew's correlation coefficient (MCC) was chosen as a measure of classification quality insensitive to classes with different sizes. The MCC was calculated for predictions for the whole dataset, not for individual folds, and then averaged over repetitions — in contrast to [1], where median was used instead of averaging. The random walk length k was set to $k \in \{1, 2, \ldots, 15\}$ for most experiments, for the rest it is explicitly noted which set of k was used.

5 Results

The results split into several parts — the comparison of NCRS with the unbiased RS method, the analysis of diversity, and the analysis of convergence of NCRS.

NCRS and Unbiased Random Subspace Method. In the original study [2], the NCF was compared to the random subspace forest of DTs, however, our NCRS generalization allows the use of different weak classifiers in the ensemble. For this part of experiment, we have used NCRS with DTs (CART), LR and NB classifiers. In most tasks, the NCRS was better in terms of MCC for some

values of k than the unbiased RS with the same type of weak classifiers. For each datasets, there are three possible results — NCRS better for some vales of the parameter k (*win*), NCRS with exactly the same performance as RS for some values of k and worse for the rest (*tie*) and NCRS worse than RS for all k (*loss*).

Tab. 1 displays results for different types of weak classifiers in the NCRS compared with the unbiased RS. However, this comparison is optimistically biased because NCRS was considered to be the winner if it was better for any value of k — in real case scenario, the parameter k could be either determined using internal cross-validation or by heuristic proposed in [2]. On the other hand, the NCRS was better in terms of MCC for any $k \in \{1, 2, \ldots, 15\}$ for many tasks — the k independent results are displayed in column *Pessimistic* Tab. 1 — therefore, the optimistic bias is not present in those experiments as this table only contains results that hold for any value of $k \in \{1, 2, \ldots, 15\}$.

Table 1: Performance of three different types of weak classifiers in terms of wins, ties, and losses, which were consistent for any $k \in \{1, 2, \ldots, 15\}$

Classifier Type	**k-dependent**			**k-independent**		
	wins	**ties**	**losses**	**wins**	**ties**	**losses**
Decision Tree	8	1	1	5	1	1
Logistic Regression	5	4	1	3	1	1
Naïve Bayes	7	1	2	6	1	2

The results displayed in Tab. 1 compare whether the NCRS with the particular type of weak classifiers is better than RS with the same type of weak classifiers, they do not compare the suitability of used weak classifiers for the task as they do not show the absolute accuracy over the datasets. From this point of view, the NCRS with logistic regression performs best as depicted in Tab. 2. However, the original NCRS with decision trees is also very close to the NCRS LR — both in the rank and the average MCC.

Table 2: Comparison of performance of different types of weak classifiers for both NCRS and RS ensembles. The MCC values are taken as the maximum MCC for $k \in \{1, 2, \ldots, 15\}$ for given classifier

Task	#samples	NCRS DT	NCRS LR	NCRS NB	RS DT	RS LR	RS NB
BMBE DU5q	16	**0.76**	0.36	0.38	0.34	0.46	0.10
BMH ABE5q	21	1.00	1.00	1.00	1.00	1.00	1.00
BMH ABEnon-5q	16	**1.00**	**1.00**	**1.00**	0.87	0.90	0.87
BMH ADU5q	15	0.72	0.79	**0.81**	0.75	0.66	0.71
BMnon-5q 5qBT	17	**1.00**	**1.00**	0.75	0.66	0.79	0.62
PBBE DU5q	22	0.57	**0.79**	0.33	0.57	0.62	0.14
PBH ABE5q	19	0.99	**1.00**	0.82	**1.00**	**1.00**	0.84
PBH ABEnon-5q	14	0.83	**1.00**	0.84	0.81	**1.00**	0.65
PBH ADU5q	23	**1.00**	0.93	0.56	0.92	**1.00**	0.66
PBnon-5q 5qBE	13	0.96	**1.00**	0.82	0.86	**1.00**	0.64
Average MCC		0.88	**0.89**	0.73	0.79	0.84	0.62
Average rank		2.85	**2.20**	3.85	4.00	2.70	5.40

However, these experiments were biased because the best value of k based on the performance on the *test* set was chosen, better approach would be use internal cross-validation for determining the optimal value of k and then use this value on the *test* set. However, the datasets are very small, from 13 to 23 samples, and internal cross-validation would reduce the training or testing set even further. Even though it would still be possible, for example using leave-one-out cross-validation, the experiments would be computationally costly, moreover, the k is to be set using the heuristic proposed in [2,1], therefore the cross-validation would not simulate the real use of the method. The heuristic also cannot be used for comparison as it is tree specific, modification of the heuristic for other learners is part of possible future work. Furthermore, the purpose of this experiment was to show that other weak classifiers are also suitable alternative to DT.

The conclusion arising from this part of experiments is clear — the NCRS method is suitable also for other types of weak classifiers than just the DT. The NCRS method outperformed the RS method in most tasks for any of the three tested types of weak classifiers. In terms of absolute performance, the NCRS with LR outperformed other ensemble classifiers both in ranks and average MCCs.

Analysis of Diversity. The analysis of the relationship between the walk length k and the diversity among classifiers in the ensemble is difficult because there are two main characteristics that are dependent on the parameter k — *diversity* and *weak classifiers accuracy* — and they cannot be analysed individually. For this reason, four different diversity measures were chosen to understand the dependency between diversity and accuracy in more depth.

As proposed in [2,1], the diversity indeed seems to decrease with the length of random walk k as the weak classifiers become less and less specialized. On the other hand, the average MCC of weak classifiers is increasing with the length k in most cases. Therefore, the overall MCC of the ensemble is based on the proportion between the diversity growth and the weak classifiers accuracy growth.

The overall MCC of the ensemble is the result of proportion of its weak classifiers accuracy and diversity. This is nicely shown in Fig. 2 where the ensemble

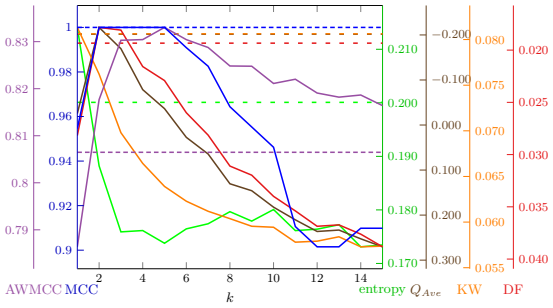


Fig. 2: The trade-off between the weak classifiers' diversity and the accuracy. The graph represents task BMH_ABE5q classified using NCRS with Logistic Regression weak classifiers.

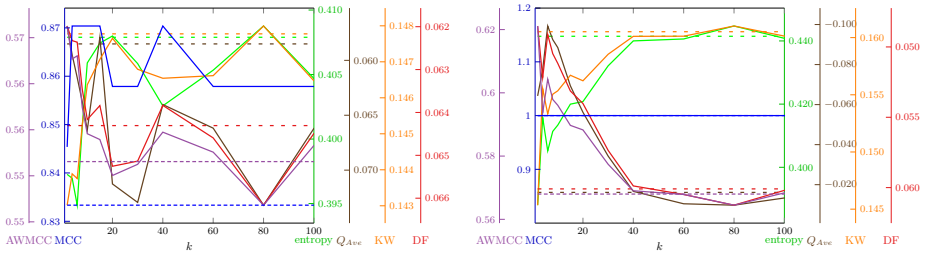
starts with diverse weak classifiers with lower accuracy for $k = 1$, then the diver-

sity is decreasing, however, the weak classifiers' accuracy is steeply increasing, therefore the overall MCC of the ensemble reaches 1.0 and holds there while the diversity is still decreasing and the weak classifiers' accuracy slowly increasing. However, even though the weak classifiers accuracy is increasing, they tend more and more to have correlated errors — these errors have bigger influence than the increasing accuracy and DF measure starts to decrease. At some point the accuracy of weak classifiers begins to slowly decrease but since the ensemble diversity is very low at this point, the ensemble MCC plummets — the decrease in MCC is not proportional to the decrease in the *average MCC of weak classifiers* (AWMCC) — roughly 1.5 % for the AWMCC while about 9 % for the MCC.

As a whole, the NCRS algorithm manages the diversity nicely, in most cases, it starts with specialized and diverse weak classifiers and with increasing value of the parameter k, the diversity usually decreases and the average accuracy of weak classifiers increases. Tuning the random walk length k may allow to find the optimal trade-off between the diversity and the AWMCC resulting in high MCC of the whole ensemble. Only in several cases, the NCRS ends up with unexpected distribution of weak classifiers with higher AWMCC than the overall MCC. This phenomenon requires further analysis. However, it occurs only for particular combination of the dataset and the type of weak classifier, moreover it appears for particular values of k only.

Analysis of The Convergence of NCRS. As described in Sect. 4, the NCRS converges to a stationary distribution of a random walk for $k \to \infty$ where no miRNA nodes are present in the network. The goal of this experiment was to empirically validate the convergence, therefore this experiments utilises only candidate causal genes and mRNA interactions as prior knowledge. Parameter k was chosen from $\{2, 4, 6, 8, 10, 15, 20, 30, 40, 60, 80, 100, 150, 200\}$. The NCRS algorithm was also modified to sample the features with a probability $\pi^{\infty}(v) = \frac{\deg(v)}{|\mathcal{I}|}$ — i.e., the probability of a feature being sampled is proportional to its degree in the biological network. Some results of this experiment are depicted in Fig. 3, where the dotted lines represent the values of measures for the k independent degree proportional sampling NCRS, while the full lines represent the k dependent random walk sampling NCRS. In contrast to other plots, the scale of y-axes is very important in the convergence analysis — e.g., seemingly unconverging lines might be just caused by small fluctuations caused by the stochastic nature of the classifier as, for example, in Fig. 3a, where the values seemingly do not converge for increasing values of k, however, the scales of axes are very small, therefore the observed chaotic behavior is just small fluctuations around the desirable values.

On the other hand, the convergence is ideally depicted in Fig. 3b, where all the measures nicely converge to the values obtained by the modified NCRS for higher values of k. The convergence manifests in other tasks too, albeit not as nicely. It seems that the values converge to different values in several tasks or that the convergence is biased a bit for some reason. Besides the bias, there are also other two possible explanations for such phenomenon. Firstly, it might be just a fluctuation of the stochastic-based original NCRS. Secondly and more im-



(a) Seemingly chaotic behavior. (b) The ideal convergence example.

Fig. 3: Empirical validation of proposed convergence

portantly, it might be caused by the stochastic nature of the modified NCRS as well. When there are changes due to the stochastic nature of the original NCRS, random fluctuation are expected as we fit the classifiers for different values of k and these fluctuations show in the smoothness of the measured points, however when dealing with stochastic nature of the modified NCRS, only one value is obtained and in spite of 10 repeated n-fold cross-validation, the obtained averaged values might still be significantly different from the hidden true expected values of the modified NCRS.

This experiment strongly suggests that the proposed convergence of NCRS (NCF) holds, even though there are still several tasks which would need further analysis as the values seem to converge to a slightly biased point.

6 Conclusion

Ensemble methods have been widely applied to problems showing high dimensionality, small sample size and complex structure. The omics high-throughput data often have all the above-mentioned characteristics and ensemble methods represent a popular alternative in their classification [26]. In this paper, we focus on the omics data where the complex structure is partially known or assumed. In particular, we suppose that a feature interaction network exists and the interactions imply feature dependencies. Note that the dependencies can hardly be reliably identified from the data itself for the sake of small sample size. By contrast, the interaction network can be composed from the relationships and regulation formerly described in literature.

We stem from our recently proposed NCF that modifies the well-known random forest for domains with known interaction networks. We have proposed its further simple generalization called *network-constrained random subspace* method which goes beyond DTs used in the original NCF. NCRS was empirically validated using the same datasets as in the original study [2,1]. It was conclusively shown that NCRS is suitable for different types of weak classifiers. To exemplify, the NCRS with logistic regression weak classifiers proved to outperform the originally proposed NCF (NCRS DT) on most MDS datasets. Importantly, both nave Bayes and logistic regression classifiers provide insight into the problem as they allow to easily analyze feature importance.

Furthermore, the role of diversity in NCRS was studied using popular diversity measures. As the feature sampling process in NCRS is parametrized by the length of random walk k, we have analysed its influence on the diversity and accuracy of the ensemble. We have empirically shown that the diversity usually decreases with increasing the length k as was hinted, but not tested, in [2,1]. The last experiments for the MDS datasets were validating the convergence NCRS for $k \to \infty$ proposed in [1]. After that, we have tested the behavior of NCRS on benchmark datasets from [14], which, in contrast with MDS datasets, do not have miRNA data and candidate causal genes. The NCRS performed similarly as RS on most of these datasets but it slightly outperformed the RS on several datasets and was outperformed by the RS method on only one datasets.

There is a lot of future work. First, with assumed increasing availability of omics data, the experiments will be replicated with more data. We expect increased statistical relevance and the possibility to analyse the influence of the size of the training set on the performance of NCRS compared to the unbiased RS. With more data, the prior knowledge is expected to become less important, however, the sample sizes where prior knowledge is superfluous are not realistic yet. Second, we plan to integrate other types of data and prior knowledge into NCRS (e.g.,DNA methylation arrays). The datasets with a large scale of measurements (GE, miRNA, DNA methylation) are still rare and small-sized, but their importance will increase. Third, a modified heuristic for finding the optimal length of random walk k that applies for ensembles of general weak classifiers, not just the NCF, shall be proposed. Last but not least, biology is not the only domain where the prior knowledge in the form of networks is available. The other tasks could be, e.g., document topic prediction or click prediction.

Acknowledgments. This work was supported by the grants NT14539 and NT14377 of the Ministry of Health of the Czech Republic.

References

1. Anděl, M., Kléma, J., Krejčík, Z.: Network-constrained forest for regularized classification of omics data. Methods 83, 88–97 (Jul 2015)
2. Anděl, M., Kléma, J., Krejčík, Z.: Network-constrained forest for regularized omics data classification. In: 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 410–417 (Nov 2014)
3. Bossi, A., Lehner, B.: Tissue specificity and the human protein interaction network. Mol Syst Biol 5 (Apr 2009)
4. Breiman, L.: Bagging predictors. Machine Learning 24(2), 123–140 (1996)
5. Breiman, L.: Randoms forests. Machine Learning 45(1), 5–32 (2001)
6. Brown, G., Wyatt, J., Harris, R., Yao, X.: Diversity creation methods: a survey and categorisation. Information Fusion 6(1), 5–20 (Mar 2005)
7. Dietterich, T.G.: Ensemble methods in machine learning. Lecture Notes in Computer Science 1857, 1–15 (2000)
8. Dietterich, T.G.: An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Mach Learn 40(2), 139–157 (Aug 2000)

9. Dweep, H., Sticht, C., Pandey, P., Gretz, N.: miRWalk – database: Prediction of possible miRNA binding sites by "walking" the genes of three genomes. Journal of Biomedical Informatics 44(5), 839–847 (Oct 2011)
10. Giacinto, G., Roli, F.: Design of effective neural network ensembles for image classification purposes. Image and Vision Computing 19(9-10), 699–707 (Aug 2001)
11. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Machine Intell. 20(8), 832–844 (1998)
12. Islam, M., Yao, X., Murase, K.: A constructive algorithm for training cooperative neural network ensembles. IEEE Trans Neural Netw 14(4), 820–834 (Jul 2003)
13. Kohavi, R., Wolpert, D.H.: Bias plus variance decomposition for zero-one loss functions. In: Proc. of the 13th Int. Conf. on Machine Learning. pp. 275–283. Morgan Kaufmann Publishers (1996)
14. Krejník, M., Kléma, J.: Empirical evidence of the applicability of functional clustering through gene expression classification. IEEE/ACM Trans. Comput. Biol. and Bioinf. 9(3), 788–798 (May 2012)
15. Kuncheva, L., Whitaker, C.: Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Mach Learn 51(2), 181–207 (May 2003)
16. Li, C., Li, H.: Network-constrained regularization and variable selection for analysis of genomic data. Bioinformatics 24(9), 1175–1182 (Mar 2008)
17. Pedregosa, F., Varoquaux, G., Gramfort, A., et al.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830 (2011)
18. Prasad, T.S.K., Goel, R., Kandasamy, K., et al.: Human protein reference database–2009 update. Nucleic Acids Res 37(Database), D767–D772 (Jan 2009)
19. Rapaport, F., Zinovyev, A., Dutreix, M., et al.: Classification of microarray data using gene networks. BMC Bioinformatics 8(1), 35 (2007)
20. Rokach, L.: Taxonomy for characterizing ensemble methods in classification tasks: A review and annotated bibliography. Computational Statistics & Data Analysis 53(12), 4046–4072 (Oct 2009)
21. Rokach, L.: Ensemble-based classifiers. Artif. Intell. Rev. 33(1-2), 1–39 (Feb 2010)
22. Schapire, R.E.: The boosting approach to machine learning: An overview. In: Nonlinear estimation and classification, pp. 149–171. Springer (2003)
23. Scherbart, A., Nattkemper, T.W.: The diversity of regression ensembles combining bagging and random subspace method. In: Advances in Neuro-Information Processing, pp. 911–918. Springer Science + Business Media (2009)
24. Vergoulis, T., Vlachos, I.S., Alexiou, P., et al.: TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. Nucleic Acids Res 40(D1), D222–D229 (Dec 2011)
25. Woniak, M., Graña, M., Corchado, E.: A survey of multiple classifier systems as hybrid systems. Inf. Fusion 16, 3–17 (Mar 2014)
26. Yang, P., Yang, Y.H., Zhou, B.B., Zomaya, A.Y.: A review of ensemble methods in bioinformatics. CBIO 5(4), 296–308 (Dec 2010)
27. Yu, W., Clyne, M., Khoury, M.J., Gwinn, M.: Phenopedia and genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. Bioinformatics 26(1), 145–146 (Oct 2009)
28. Yule, G.U.: On the association of attributes in statistics: With illustrations from the material of the childhood society, &c. Phil. Trans. R. Soc. A 194(252-261), 257–319 (Jan 1900)