

A comparison study between one-class and two-class machine learning for MicroRNA target detection

Malik Yousef^{1,2,3*}, Naim Najami^{1,4}, Waleed Khalifa^{1*}

¹The Institute of Applied Research-the Galilee Society, Shefa Amr, Israel;

²Computer Science, The College of Sakhnin, Sakhnin, Israel;

³Al-Qasemi Academic College, Baqa Algharbiya, Israel;

⁴The Academic Arab College of Education, Haifa, Israel.

Email: yousef@gal-soc.org; khwahid@hotmail.com; najamina@gmail.com

Received 9 September 2009; revised 2 November 2009; accepted 5 November 2009.

ABSTRACT

The application of one-class machine learning is gaining attention in the computational biology community. Different studies have described the use of two-class machine learning to predict microRNAs (miRNAs) gene target. Most of these methods require the generation of an artificial negative class. However, designation of the negative class can be problematic and if it is not properly done can affect the performance of the classifier dramatically and/or yield a biased estimate of performance. We present a study using one-class machine learning for miRNA target discovery and compare one-class to two-class approaches. Of all the one-class methods tested, we found that most of them gave similar accuracy that range from 0.81 to 0.89 while the two-class naive Bayes gave 0.99 accuracy. One and two class methods can both give useful classification accuracies. The advantage of one class methods is that they don't require any additional effort for choosing the best way of generating the negative class. In these cases one-class methods can be superior to two-class methods when the features which are chosen as representative of that positive class are well defined.

Keywords: MicroRNA; One-Class; Two-Class; Machine Learning

1. INTRODUCTION

MicroRNAs (miRNAs) are single-stranded, non-coding RNAs averaging 21 nucleotides in length. The mature miRNA is cleaved from a 70-110 nucleotide (nt) "hairpin" precursor with a double-stranded region containing one or more single-stranded loops. MiRNAs target messenger RNAs (mRNAs) for cleavage, primarily by repressing translation and causing mRNA degradation [1]. Although recent findings [2] suggest microRNAs may

affect gene expression by binding to either 5' or 3' untranslated regions (UTRs) of mRNA, most studies have found that miRNA mark their target mRNAs for degradation or suppress their translation by binding to the 3'UTR and most target programs search there. These studies have suggested that the miRNA seed segment which includes 6-8 nt at the 5' end of the mature miRNA sequence is very important in the selection of the target site (see **Figure 1**).

Several computational approaches have been applied to miRNA gene prediction using methods based on sequence conservation and/or structural similarity [3,4,5,6,7]. Those methods that used machine learning were based on the two-class approaches, while our new reported results are based on the one-class approaches.

Several additional methods for the prediction of miRNA targets have been subsequently developed. These methods mainly use sequence complementarities, thermodynamic stability calculations, and evolutionary conservation among species to determine the likelihood of a productive miRNA: mRNA duplex formation [8,9]. John *et al.*, (2004) developed the miRanda [10] algorithm for miRNA target prediction. MiRanda uses dynamic programming to search for optimal sequence complementarities between a set of mature miRNAs and a given miRNA. Another algorithm RNAhybrid [8,9] is similar to a RNA secondary structure prediction algorithm like the Mfold program [11] but it determines the most favorable hybridization site between two sequences.

Lewis *et al.*, (2005) developed TargetScanS [12]. TargetScanS scores target sites based on the conservation of the target sequences between five genomes (human, mouse, rat, dog and chicken) as evolutionarily conserved target sequences are more likely to be true targets. In testing, TargetScanS was able to recover targets for all 5300 human genes known at the time to be targeted by miRNAs.

PicTar [13] is a computational method to detect common miRNA targets in vertebrates, nematodes (*C. ele-*

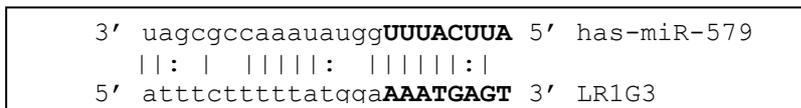


Figure 1. Duplex partitioned into two parts for miRNA hsa-miR-579 and its target LR1G3, the seed part and the out-seed part. The seed part appears in capital letters.

gans), and insects (*Drosophila melanogaster*). PicTar is based on a statistical method applied to eight vertebrate genome-wide alignments (multiple alignments of orthologous nucleotide sequences (3' UTRs). PicTar was able to recover validated miRNA targets at an estimated 30% false-positive rate. In a separate study PicTar was applied to target identification in *D. melanogaster* [14]. These studies suggest that one miRNA can target 54 genes on average and that known miRNAs are projected to regulate a large fraction of all *D. melanogaster* genes (15%). This is likely to be a conservative estimate due to the incomplete input data.

TargetBoost [15] is a machine learning algorithm for miRNA target prediction using only sequence information to create weighted sequence motifs that capture the binding characteristics between miRNAs and their targets. The authors suggest that TargetBoost is stable and identifies more of the already verified true targets than do other existing algorithms.

Sung-Kyu *et al.*, (2005), also reported the development of a machine learning algorithm using Support Vector Machine (SVM). The best reported results [16] were 0.921 sensitivity and 0.833 specificity. More recently Yan and others, used a machine learning approach that employs features extracted from both the seed and out-seed segments [17]. The best result obtained was an accuracy of 82.95% but it was generated using only 48 positive human and 16 negative examples, a relatively small training set to assess the algorithm.

In 2006, Thadani and Tammi [18] launched MicroTar, a novel statistical computational tool for prediction of miRNA targets from RNA duplexes which does not use sequence homology for prediction. MicroTar mainly relies on a quite novel approach to estimate the duplex energy. However, the reported sensitivity (60%) is significantly lower than that achieved using other published algorithms. At the same time, a miRNA pattern discovery method, RNA22 [19] was proposed to scan UTR sequences for targets. RNA22 does not rely upon cross-species conservation but was able to recover most of the known target sites with validation of some of its new predictions.

More recently, Yousef *et al.*, (2007) described a target prediction method, (NBmiRTar [20]) using instead machine learning by a Naïve Bayes classifier. NBmiRTar does not require sequence conservation but generates a model from sequence and miRNA:mRNA duplex information derived from validated target sequences and arti-

ficially generated negative examples. In this case, both the seed and “out-seed” segments of the miRNA:mRNA duplexes are used for target identification. NBmiRTar technique produces fewer false positive predictions and fewer target candidates to be tested than miRanda [10]. It exhibits higher sensitivity and specificity than algorithms that rely only on conserved genomic regions to decrease false positive predictions.

This paper describes a comparison study of using one-class and two-class approaches for miRNA target detection. The advantage of one class methods is that they don't require any additional effort for choosing the best way of generating the negative class while it is clear that the two class approaches performances are outperform the one-class methods.

2. METHODS

2.1. Designing Duplex Structure and Sequence Features

Machine learning enables one to generate automatic rules based on observation of the appropriate examples by the learning machine. However, the selection and design of the features that will be considered in order to represent each example for the learning process are very important and influence the classifier performance. We have followed [20] for feature design. We have partitioned the duplex into two parts, the seed (5' 8nt of the miRNA) and out-seed (3' remainder) as described in **Figure 1**. For each of these parts the following features are extracted to give 57 structural features: 1) the number of paired bases (bp), 2) The number of bulges (inserts on one strand between paired bases), 3) the number of loops (unpaired bases opposite each other between paired bases), 4) the number of asymmetric loops (loops with unequal numbers of unpaired bases on the two strands), 5) eight features, each representing the number of bulges of lengths 1-7 and those with lengths greater than 7. 6) Eight features, each representing the number of symmetric loops with lengths 1-7 and those with lengths greater than 7, 7) eight features each representing the number of asymmetric loops with lengths 1-7 and those with lengths greater than 7, 8) the distance from the start of the seed (the 3' end) to the first paired base of the 5' start of the out-seed part is an additional feature that is extracted. For the sequence features, we define “words” as sequences having lengths equal to or less than 3. The frequency of each word in the seed part

is extracted to form a representation in the vector space.

2.2. One-Class Methods

In general a binary learning (two-class) approach to miRNA discovery considers both positive (miRNA) and negative (non-miRNA) classes by providing examples for the two-classes to a learning algorithm in order to build a classifier that will attempt to discriminate between them. The most common term for this kind of learning is *supervised learning* where the labels of the two-classes are known before hand. One-class uses only the information for the target class (positive class) building a classifier which is able to recognize the examples belonging to its target and rejecting others as outliers.

Among the many classification algorithms available, we chose five one-class algorithms to compare for miRNA discovery. We give a brief description of each one-class classifier and we refer the references [21,22] for additional details including a description of parameters and thresholds. The LIBSVM library [23] was used as implementation of the SVM (one-class using the RBF kernel function) and the DDtools [24] for the other one-class methods. The WEKA software [25] was used as implementation of the two-class classifiers.

2.3. One-Class Support Vector Machines (OC-SVM)

Support Vector Machines (SVMs) are a learning machine developed as a two-class approach [26,27]. The use of one-class SVM was originally suggested by [22]. One-class SVM is an algorithmic method that produces a prediction function trained to “capture” most of the training data. For that purpose a kernel function is used to map the data into a feature space where the SVM is employed to find the hyper-plane with maximum margin from the origin of the feature space. In this use, the margin to be maximized between the two classes (in two-class SVM) becomes the distance between the origin and the support vectors which define the boundaries of the surrounding circle, (or hyper-sphere in high-dimensional space) which encloses the single class.

2.4. One-Class Gaussian (OC-Gaussian)

The Gaussian model is considered as a density estimation model. The assumption is that the target samples form a multivariate normal distribution, therefore for a given test sample z in n -dimensional space, the probability density function can be calculated as:

$$p(z) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{(-1/2(z-\mu)^T \Sigma^{-1} (z-\mu))} \quad (1)$$

where μ and Σ are the mean and covariance matrix of the target class estimated from the training samples.

2.5. One-Class Kmeans (OC-Kmeans)

Kmeans is a simple and well-known unsupervised machine learning algorithm used in order to partition the data into k clusters. Using the OC-Kmeans we describe the data as k clusters, or more specifically as k centroids, one derived from each cluster. For a new sample, z , the distance $d(z)$ is calculated as the minimum distance to each centroid. Then based on a user threshold, the classification decision is made. If $d(z)$ is less than the threshold the new sample belongs to the target class, otherwise it is rejected.

2.6. One-Class Principal Component Analysis (OC-PCA)

Principal component analysis (PCA) is a classical statistical method known as a linear transform that has been widely used in data analysis and compression. Mainly PCA is a projection method used for reducing dimensionality in a given dataset by capturing the most variance by a few orthogonal subspaces called principal components (PCs). For the one-class approach (OC-PCA) one needs to build the PCA model based on the training set and then for a given test example z the distance to the PCA(z) model is calculated and used as a decision factor for acceptance or rejection.

2.7. One-Class K-Nearest Neighbor (OC-KNN)

The one-class nearest neighbor classifier (OC-KNN) is a modification of the known two-class nearest neighbor classifier which learns from positive examples only. The algorithm operates by storing all the training examples as its model, then for a given test example z , the distance to its nearest neighbor $y(y=NN(z))$ is calculated as $d(z,y)$. The new sample belongs to the target class when:

$$\frac{d(z, y)}{d(y, NN(y))} < \delta \quad (2)$$

where $NN(y)$ is the nearest neighbor of y , in other words, it is the nearest neighbor of the nearest neighbor of z . The default value of δ is 1. The average distance of the k nearest neighbors is considered for the OC-KNN implementation.

3. TWO CLASS METHODS

3.1. Naïve Bayes

Naïve Bayes is a classification model obtained by applying a relatively simple method to a training dataset [28]. A Naïve Bayes classifier calculates the probability that a given instance (example) belongs to a certain class. It makes the simplifying assumption that the features

constituting the instance are conditionally independent given the class. Given an example X , described by its feature vector (x_1, \dots, x_n) , we are looking for a class C that maximizes the likelihood: $P(X | C) = P(x_1, \dots, x_n | C)$. The (naïve) assumption of conditional independence among the features, given the class, allows us to express this conditional probability $P(X | C)$ as a product of simpler probabilities: $P(X | C) = \prod_{i=1}^n P(x_i | C)$.

We used the Rainbow program [29] to train the naïve Bayes classifier. To combine the numeric features identified in the miRNA-target duplex with the sequence features (“words”) in the target candidate sequence, a dictionary of all the unique “words” was generated and the frequency of each “word” in the sequence is used.

3.2. Support Vector Machines (SVMs)

Support Vector Machines (SVMs) is a learning machine developed by Vapnik [27]. The performance of this algorithm, as compared to other algorithms, has proven to be particularly useful for the analysis of various classification problems, and has recently been widely used in the bioinformatics field [30,31,32]. Linear SVMs are usually defined as SVMs with linear kernel. The training data for linear SVMs could be linear non-separable and then soft-margin SVM could be applied. Linear SVM separates the two classes in the training data by producing the optimal separating hyper-plane with a maximal margin between the class 1 and class 2 samples. Given a training set of labeled examples (x_i, y_i) , $i = 1, \dots, \ell$ where $x_i \in R^d$ and $y_i \in \{+1, -1\}$, the support vector machines (SVMs) find the separating hyper-plane of the form $w \cdot x + b = 0$ $w \in R^d, b \in R$. Here, w is the “normal” of the hyper-plane. The constant b defines the position of the hyper-plane in the space. One could use the following formula as a predictor for a new instance: $f(x) = \text{sign}(w \cdot x + b)$ for more information see Vapnik [27].

3.3. Random Forest

Random forests are an ensemble of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [33]. The improvement in the classification accuracy is due to the growing of an ensemble of trees that vote for the most popular class. Random forests are becoming increasingly popular because their ability to deal with small sample size with high-dimensional space.

3.4. C4.5

C4.5 is a decision tree algorithm, developed by Quinlan (1993)[34]. A decision tree is a simple structure where

non-terminal nodes represent tests on one or more attributes and terminal nodes reflect decision outcomes.

4. RESULTS

4.1. Data

A collection of 326 confirmed MicroRNA targets (human, mouse, fruit fly worm and virus) were downloaded from the TarBase [35] (TarBase_V4, Tarbase flat file data as of 04/2007) web-site to serve as positive examples and 1,000 negative examples chosen at random from the negative class pool generated at the study of NBmiRTar [20].

To evaluate classification performance, we used the data generated from the positive class and 1,000 negative examples. The negative class is not used for training of the one-class classifiers, but merely for estimating the specificity performance

Each one-class algorithm was trained using 90% of the positive class and the remaining 10% was used for sensitivity evaluation. The randomly selected 1,000 negative examples were used for the evaluation of specificity. The whole process was repeated 100 times in order to evaluate the stability of the methods. Additionally, the Matthews Correlation Coefficient (MCC) [36] measurement is used to take into account both over-prediction and under-prediction in imbalanced data sets. It is defined as:

$$MCC = \frac{(TpTn - FpFn)}{\sqrt{(Tp + Fp)(Tp + Fn)(Tn + Fn)(Tn + Fp)}}$$

5. DISCUSSION

The one-class approach in machine learning has been receiving more attention particularly for solving problems where the negative class is not well defined [37,38,39,40]; moreover, the one class approach has been successfully applied in various fields including text mining [41], functional Magnetic Resonance Imaging (fMRI) [42], signature verification [43] and miRNA gene discovery [44].

This paper describes a comparison study of using one-class and two-class approaches for miRNA target detection. The advantage of one class methods is that they don't require any additional effort for choosing the best way of generating the negative class while it is clear that the two class approaches performances are outperform the one-class methods.

Table 1 shows the performance of five one-class classifiers while **Table 2** shows the performances of two-class methods. The results of the one-class approaches show a slight superiority for OC-Kmeans over the other one class methods based on the average of the MCC measurement. The MCC measurement with value of +1 represents a perfect prediction while 0 value indicates an average random prediction. However, accuracy is less

Table 1. One-Class results.

Method	TP	TN	ACC	MCC
OC-SVM	0.91	0.779	0.81	0.69
OC-Gaussian	0.86	0.89	0.89	0.75
OC-Kmeans	0.90	0.87	0.87	0.77
OC-PCA	0.77	0.77	0.77	0.55
OC-Knn	0.87	0.89	0.89	0.76

Table 2. Two class results.

Method	TP	TN	Acc	MCC
Naïve Bayes	0.93	0.99	0.99	0.93
SVM	0.98	0.9974	99.3	0.977
KNN4	0.858	0.952	92.88	0.813
C4.5	0.912	0.978	96.23	0.89
Random Forest	0.958	0.993	98.44	0.951

than the two-class approaches. During the training stage of the one-class classifier we have set the 10% of the positive data, whose likelihood is furthest from the true positive data based on the distribution, as “outliers” in order to produce a compact classifier. This factor might cause a loss of information about the target class which might also result in reducing performance compared to the two class approach.

6. CONCLUSIONS

The current results show that it is possible to build up a classifier based only on positive examples yielding a reasonable performance. Moreover, more efforts are required to figure out more biological features to be used in the design of the one-class classifier to improve the performance. However, we hypothesize that taken 10% of the training data as “outlier” is the cause of reducing the one-class performance.

7. ACKNOWLEDGEMENTS

This project is funded in part under a grant with the ESHKOL scholarship program of the ministry of science culture and sport to support KW.

REFERENCES

- [1] Bartel, D.P. (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*, **116**, 281-297.
- [2] Lytle, J.R., Yario, T.A. and Steitz, J.A. (2007) Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proceedings of the National Academy of Sciences*, **104**, 9667-9672.
- [3] Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B. and Bartel, D. P. (2003) Vertebrate MicroRNA Genes. *Science*, **299**, 1540.
- [4] Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B. and Bartel, D.P. (2003) The microRNAs of *Caenorhabditis elegans*. *Genes & Development*, **17**, 991-1008.
- [5] Weber, M.J. (2005) New human and mouse microRNA genes found by homology search. *FEBS Journal*, **272**, 59-73.
- [6] Lai, E., Tomancak, P., Williams, R. and Rubin, G. (2003) Computational identification of *Drosophila* microRNA genes. *Genome Biology*, **4**, R42.
- [7] Grad, Y., Aach, J., Hayes, G. D., Reinhart, B. J., Church, G.M., Ruvkun, G. and Kim, J. (2003) Computational and Experimental Identification of *C. elegans* microRNAs. *Molecular Cell*, **11**, 1253-1263.
- [8] Bartel, D. P. (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*, **116**, 281.
- [9] Lai, E. (2004) Predicting and validating microRNA targets. *Genome Biology*, **5**, 115.
- [10] John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) Human MicroRNA Targets. *PLoS Biology*, **2**, e363.
- [11] Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, **31** (13), 3406-3415.
- [12] Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. and Burge, C. B. (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787.
- [13] Krek, A. *et al.* (2005) Combinatorial microRNA target predictions. *Nature Genetics*, **37**, 495-500.
- [14] Grun, D., Wang, Y.L., Langenberger, D., Gunsalus, K.C. and Rajewsky, N. (2005) MicroRNA target predictions across seven drosophila species and comparison to mammalian targets. *PLoS Computational Biology*, **1**, e13.
- [15] SaeTrom, O.L.A., Snove, O.J. and SaeTrom, P.A.L. (2005) Weighted sequence motifs as an improved seeding step in microRNA target prediction algorithms. *RNA*, **11**, 995-1003.
- [16] Sung-Kyu, K., Jin-Wu, N., Wha-Jin, L. and Byoung-Tak, Z. (2005) A kernel method for microrna target prediction using sensible data and position-based features. In computational intelligence in bioinformatics and computational biology. *Proceedings of the 2005 IEEE Symposium on CIBCB*, 1-7.
- [17] Yan, X., *et al.* (2007) Improving the prediction of human microRNA target genes by using ensemble algorithm. *FEBS Letters*, **581**, 1587.
- [18] Thadani, R. and Tammi, M. (2006) MicroTar: Predicting microRNA targets from RNA duplexes. *BMC Bioinformatics*, **7**, S20.
- [19] Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.S., Tam, W.L., Thomson, A. M., Lim, B. and Rigoutsos, I. (2006) A pattern-based method for the identification of microrna binding sites and their corresponding. *Heteroduplexes*, **126**, 1203-1217.
- [20] Yousef, M., Jung, S., Kossenkov, A.V., Showe, L.C. and Showe, M.K. (2007) Naive Bayes for microRNA target predictions machine learning for microRNA targets. Oxford University Press, 2987-2992.
- [21] Tax, D.M.J. (2001) One-class classification; Concept-learning in the absence of counter-examples. Delft University of Technology ed.
- [22] Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R. C. (2001) Estimating the support of a

- high-dimensional distribution. *Neural Computation*, **13**, 1443-1471.
- [23] Chang, C.C. and Lin, C.J. (2001) LIBSVM: A library for support vector machines.
- [24] Tax, D.M.J. (2005) DDtools, the data description toolbox for matlab. Delft University of Technology ed.
- [25] Witten, I.H. and Frank, E. (2005) Data mining: Practical machine learning tools and techniques, Morgan Kaufmann, San Francisco.
- [26] Schölkopf, B., Burges, C.J.C. and Smola, A.J. (1999) Advances in kernel methods. MIT Press, Cambridge.
- [27] Vapnik, V. (1995) The Nature of Statistical Learning Theory, Springer.
- [28] Mitchell, T. (1997) Machine Learning, McGraw Hill.
- [29] McCallum, A.K. (1996) Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering.
- [30] Haussler, D. (1999) Convolution kernels on discrete structured, Technical Report UCSCCRL-99-10. Baskin School of Engineering, University of California, Santa Cruz.
- [31] Pavlidis, P., Weston, J., Cai, J. and Grundy, W.N. (2001) Gene functional classification from heterogeneous data. *Proceedings of the 5th Annual International Conference on Computational Biology*, ACM Press, Montreal, 249-255.
- [32] Donaldson, I. *et al.* (2003) PreBIND and Textomy-mining the biomedical literature for protein-protein interactions using a support vector machine. *BMC Bioinformatics*, **4**, 11.
- [33] Breiman, L. (2001) Random Forests. *Machine Learning* **45**, 5-32.
- [34] Quinlan, J.R. (1993) C4.5: Programs for machine learning Morgan Kaufmann Publishers Inc.
- [35] Sethupathy, P., Corda, B. and Hatzigeorgiou, A.G. (2006) TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA*, **12**, 192-197.
- [36] Matthews, B. (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta*, **405(2)**, 442-451.
- [37] Kowalczyk, A. and Raskutti, B. (2002) One class SVM for yeast regulation prediction. *SIGKDD Explorations*, **4**, 99-100.
- [38] Spinosa, E.J. and Carvalho, A.C.P.L.F.d. (2005) Support vector machines for novel class detection. *Bioinformatics Genetics and Molecular Research*, **4**, 608-615.
- [39] Crammer, K. and Chechik, G. (2004) A needle in a haystack: Local one-class optimization. *Proceedings of the 21st International Conference on Machine Learning*, Banff, 26.
- [40] Gupta, G. and Ghosh, J. (2005) Robust one-class clustering using hybrid global and local search. *Proceedings of the 22nd International Conference on Machine Learning*, ACM Press, Bonn, 273-280.
- [41] Manevitz, L.M. and Yousef, M. (2001) One-class SVMs for document classification. *Journal of Machine Learning Research*, 139-154.
- [42] Thirion, B. and Faugeras, O. (2004) Feature characterization in fMRI data: The information bottleneck approach. *Medical Image Analysis*, **8**, 403.
- [43] Koppel, M. and Schler, J. (2004) Authorship verification as a one-class classification problem. *Proceedings of the 21st International Conference on Machine Learning*, ACM Press, Banff, 62.
- [44] Yousef, M., Jung, S., Showe, L. and Showe, M. (2008) Learning from positive examples when the negative class is undetermined-microRNA gene identification. *Algorithms for Molecular Biology*, **3**, 2.