A BAYESIAN APPROACH TO TOP-SCORING PAIRS CLASSIFICATION

Emre Arslan and Ulisses M. Braga-Neto

Texas A&M University
Department of Electrical and Computer Engineering
College Station, TX, USA

ABSTRACT

We extend the popular Top Scoring Pair (TSP) classification rule to a Bayesian setting, with the purpose of obtaining robust and effective classifiers for small-sample, high-dimensional data. We employ the Bradley-Terry model for rank data, and infer its parameters using a previously proposed Gibbs sampling algorithm. The parameters are then used to define a Bayesian TSP score, which is used to select the gene pairs to define the proposed Bayesian TSP classifiers. Accuracy of the proposed Bayesian classification rules is evaluated against those of the conventional TSP classifiers as well as other well-known machine learning methods, using a total of 12 gene-expression data sets. The results indicate that the Bayesian k-TSP classifier obtained the best overall average accuracy rate and the best accuracy rate over the majority of the individual data sets.

Index Terms— Gene expression classification, Top Scoring Classifier, Bayesian methods.

1. INTRODUCTION

Classification of high-dimensional gene expression data has been a topic of ongoing interest in Genomics, with applications in diagnosis and prognosis of cancer, infectious diseases, and more [1, 2, 3, 4, 5, 6, 7, 8]. When selecting a classification algorithm, a balance has to be struck between "black-box" accuracy and interpretability; many classification rules appear to produce accurate results on test data, but they produce complex decision boundaries that have little hope of interpretability and further validation by physicians and other biomedical experts. The Top-Scoring Pair (TSP) classifier was introduced by D. Geman and collaborators in [9], and further studied in [10, 11, 12]. It is a rank-based classifier that ignores the raw gene expression values and instead seeks simple rank changes across the two conditions. Its main virtue is to produce interpretable classifiers, which are yet powerful in terms of its classification accuracy rates.

In this paper, we extend the basic TSP and k-TSP classification rules to a Bayesian setting, with the purpose of obtaining robust and effective classifiers for small-sample, high-dimensional data. We employ the Bradley-Terry model for

rank data [13], and infer its parameters using the Gibbs sampling algorithm proposed by Caron and Doucet in [14]. The parameters are then used to define a Bayesian TSP score, which is used to select the gene pairs to define the proposed Bayesian TSP classifiers. Accuracy of the proposed Bayesian classification rules is evaluated against those of the conventional TSP classifiers as well as other well-known machine learning methods, using a total of 12 gene-expression data sets. The results indicate that the Bayesian k-TSP classifier obtained the best overall average accuracy rate and the best accuracy rate over the majority of the individual data sets.

2. BRADLEY TERRY MODEL

We begin by introducing the Bradley-Terry model. Let i and j be a pair of individuals randomly drawn from a population of size M . The Bradley-Terry model [13] stipulates that:

$$\pi_{ij} = P(i > j \mid \lambda_i, \lambda_j) = \frac{\lambda_i}{\lambda_i + \lambda_j}, \qquad (1)$$

where parameter $\lambda_i > 0$ can be interpreted as the "skill of player i," for $i = 1, \ldots, M$. For our purposes, λ_i denotes the propensity of the rank of gene i being larger than the rank of gene j. As expected, (1) implies that $\pi_{ij} + \pi_{ji} = 1$. The Bradley-Terry (BT) model can also be identified as a logistic model by a non-linear reparameterization of the parameters, $\lambda_i = e^{\beta_i}$, in which case

$$\pi_{ij} = \frac{1}{1 + e^{-(\beta_i - \beta_j)}} = \eta(\beta_i - \beta_j) \qquad (2)$$

where $\eta(\alpha) = 1/(1 + e^{-\alpha})$ is the inverse logit function [15]. There are numerous applications of the BT Model, such as ranking of Chess and Go players by their respective international federations, estimation of the influence of statistical journals, and more.

Given sample data, let w_{ij} denote the number of times $i > j$, w_i be the number of "wins" by i, and n_{ij} represent the number of comparisons between i and j, over the data. These statistics can be used to estimate the model parameters $\{\lambda_i\}$ by maximum likelihood (ML). If pairings are assumed

independent, the log-likelihood function for the BT Model is given by

$$
\begin{aligned}
\ell(\lambda) &= \sum_{1 \leq i \neq j \leq M} w_{ij} \log \lambda_i - w_{ij} \log(\lambda_i + \lambda_j) \\
&= \sum_{1 \leq i \neq j \leq M} w_i \log \lambda_i - \sum_{1 \leq i < j \leq M} n_{ij} \log(\lambda_i + \lambda_j).
\end{aligned}
\tag{3}
$$

The ML estimator can be found by an iterative procedure [16],

$$
\lambda_i^{(k+1)} = w_i \left(\sum_{i \neq j} \frac{n_{ij}}{\lambda_i^{(k)} + \lambda_j^{(k)}} \right)^{-1}, \quad i = 1, \ldots, M, \tag{4}
$$

which is repeated until convergence. The main drawback with the ML approach is the strong assumption that no player may win all games against another (i.e., $n_{ij} \neq 0$ for all i, j).

3. BAYESIAN INFERENCE FOR THE BRADLEY-TERRY MODEL

To overcome the difficulties associated with maximum-likelihood methods, Caron and Doucet [14] introduced a Bayesian approach to the inference of the parameters $\{\lambda_i\}$ in the BT model, which we briefly describe next.

The BT Model has a "Thurstonian" interpretation: for each pair $1 \leq i < j \leq K$ and for each pair comparison $k = 1, 2, ..., n_{ij}$, consider two independent random variables $Y_{ki} \sim \text{Exp}(\lambda_i)$ and $Y_{kj} \sim \text{Exp}(\lambda_j)$, where $\text{Exp}(\lambda)$ denotes an exponential distribution with rate parameter λ. Then a simple calculation reveals that

$$
P(Y_{ki} < Y_{kj}) = \frac{\lambda_i}{\lambda_i + \lambda_j}. \tag{5}
$$

In order to get a simpler complete log-likelihood, Caron and Doucet seek to introduce new latent variables Z_{ij} given by

$$
Z_{ij} = \sum_{k=1}^{n_{ij}} \min\{Y_{kj}, Y_{ki}\}. \tag{6}
$$

Owing to the facts that $U_k = \min\{Y_{kj}, Y_{ki}\} \sim \text{Exp}(\lambda_i + \lambda_j)$ and that U_k and U_l are independent for $k \neq l$, we conclude that $Z_{ij} \sim \text{Gamma}(n_{ij}, \lambda_i + \lambda_j)$, a Gamma distribution with shape parameter n_{ij} and rate parameter $\lambda_i + \lambda_j$. The resulting density of the variables $\mathbf{Z} = \{Z_{ij}\}$ given the data $\mathcal{D}$ and vector of parameters λ is

$$
p(\mathbf{z} \mid \mathcal{D}, \lambda) = \prod_{\substack{1 \leq i < j \leq M \\ \text{s.t. } n_{ij} > 0}} \text{Gamma}(z_{ij}; n_{ij}, \lambda_i + \lambda_j) \tag{7}
$$

and the resulting complete data log-likelihood is,

$$
\begin{aligned}
\ell_c(\lambda) &= \sum_{\substack{1 \leq i \neq j \leq M \\ \text{s.t. } w_{ij} > 0}} w_{ij} \log(\lambda_i) \\
&\quad - \sum_{\substack{1 \leq i < j \leq M \\ \text{s.t. } n_{ij} > 0}} (\lambda_i + \lambda_j) z_{ij} + (n_{ij} - 1) \log z_{ij} - \log \Gamma(n_{ij})
\end{aligned}
\tag{8}
$$

where Γ is the Gamma function.

The prior for λ is assigned as in [17, 18]:

$$
p(\lambda) = \prod_{i=1}^{M} \text{Gamma}(\lambda_i; a, b), \tag{9}
$$

where a and b are hyper parameters. At this point, one needs to sample from the posterior $p(\lambda, \mathbf{Z} \mid \mathcal{D})$. Canon and Doucet suggest the following Gibbs sampling scheme to accomplish that. First, update $\mathbf{Z}$ from the previous value of λ using (6):

$$
Z_{ij}^{t+1} \mid \mathcal{D}, \lambda^t \sim \text{Gamma}(n_{ij}, \lambda_i^t + \lambda_j^t); \tag{10}
$$

Next, sample the new value of λ from $p(\lambda \mid \mathbf{Z}, \mathcal{D}) \propto p(\lambda)\ell(\lambda, \mathbf{z})$, which has a Gamma distribution with known parameters:

$$
\begin{aligned}
&\lambda_i^{t+1} \mid \mathcal{D}, \mathbf{Z}^{t+1} \\
&\sim \text{Gamma}\left(a + w_i, \ b + \sum_{\substack{i < j \\ \text{s.t. } n_{ij} > 0}} z_{ij}^{t+1} + \sum_{\substack{i > j \\ \text{s.t. } n_{ij} > 0}} z_{ji}^{t+1} \right).
\end{aligned}
\tag{11}
$$

4. TOP SCORING PAIR CLASSIFIERS

Consider p genes where X_i represents the expression value of the ith gene. The quantity of interest is $p_{ij}(c) = P(X_i < X_j \mid C = c)$ where C represents a given class (in this paper, we will assume the two-class case, for the sake of simplicity). A TSP classifier seeks the pair which maximizes the TSP score $\widehat{\Delta}_{ij} = |\hat{p}_{ij}(0) - \hat{p}_{ij}(1)|$, where $\hat{p}_{i,j}$ is the sample estimate of $p_{i,j}$. After choosing the pair (i^*, j^*) with with the highest score, and assuming $\hat{p}_{i^* j^*}(0) > \hat{p}_{i^* j^*}(1)$, the TSP classifier is defined as;

$$
h_{TSP}(\mathbf{x}_{new}) = \begin{cases} C_0, & \text{if } \mathbf{x}_{new, i^*} < \mathbf{x}_{new, j^*}, \\ C_1, & otw. \end{cases} \tag{12}
$$

The classes C_0 and C_1 are flipped in the definition of the TSP classifier if $\hat{p}_{i^* j^*}(0) \leq \hat{p}_{i^* j^*}(1)$.

Many variations of TSP classifier have been proposed [12, 11, 19]. Among these, the most popular is the k-TSP classifier[10]. It is a generalization of TSP; instead of using one pair as in TSP, it selects the top k pairs according to $\widehat{\Delta}_{ij}$

(usually an odd number k of disjoint pairs are selected), and performs majority voting to classify the given data point:

$$h_{kTSP}(\mathbf{x}_{new}) = \underset{C \in \{C_0, C_1\}}{\arg\max} \sum_{r=1}^{k} I(h_r(\mathbf{x}_{new}) = C), \quad (13)$$

where $h_r(\cdot)$ is the TSP classifier based on pair r, for $r = 1, \ldots, k$.

5. BAYESIAN TOP SCORING PAIRS

In this section we describe the application of the Bradley-Terry model in the design of TSP classifiers. Let $\boldsymbol{\lambda}^0$ and $\boldsymbol{\lambda}^1$ be the class-specific "skill" parameters sampled from the posterior distributions for each class separately, as described in the previous section, and define

$$\pi_{ij}^0 = \frac{\lambda_i^0}{\lambda_i^0 + \lambda_j^0} \quad \text{and} \quad \pi_{ij}^1 = \frac{\lambda_i^1}{\lambda_i^1 + \lambda_j^1}. \quad (14)$$

The main goal is to find pairs that swap frequently beteen classes in terms of $\boldsymbol{\lambda}^0$ and $\boldsymbol{\lambda}^1$. For this purpose, we define the *Bayesian TSP score*

$$\Omega_{ij} = \left| \pi_{ji}^0 < \pi_{ji}^1 \right| = \left| \frac{\lambda_j^0}{\lambda_i^0 + \lambda_j^0} - \frac{\lambda_j^1}{\lambda_i^1 + \lambda_j^1} \right|, \quad (15)$$

and choose the best pair according to this score,

$$(i^*, j^*) = \underset{(i,j) \in S}{\arg\max} \, \Omega_{ij} \quad (16)$$

where S is whole feature space. If $\pi_{i^*j^*}^0 < \pi_{i^*j^*}^1$, the *Bayesian Top Scoring Pair* (*BTSP*) classifier is defined as

$$h_{BTSP}(\mathbf{x}_{new}) = \begin{cases} C_0, & \text{if } \mathbf{x}_{new,i^*} < \mathbf{x}_{new,j^*}, \\ C_1, & otw. \end{cases} \quad (17)$$

The classes C_0 and C_1 are flipped in the definition of the BTSP classifier if $\pi_{i^*j^*}^0 \le \pi_{i^*j^*}^1$.

As in the case of TSP, the BTSP classification rule can be extended by choosing more than one pair. We choose an odd number k of disjoint top scoring pairs according to Ω_{ij} and define a classifier as in (13), where $h_r(\cdot)$ this time denotes the BTSP classifier based on pair r, for $r = 1, \ldots, k$. We call this the *k-BTSP classifier*.

6. EXPERIMENTAL RESULTS

We used 12 genomic data sets (Table 1) to compare the proposed Bayesian TSP classification rules against the conventional TSP classifiers as well as other well-known machine learning methods. A variance of filter was employed to reduce the number of genes in all data sets to 2000. In order to simulate small sample size problem with given data, classifiers were trained on 20% of the data and tested on the remaining 80%. This procedure was repeated 50 times and the average accuracy was recorded. The Gibbs sampler was run for 1500 iterations with 300 burn-in iterations. We found that the choice of hyper parameters of the prior distribution in (9) does not change results dramatically: b is a scaling parameter and the likelihood is invariant to rescaling, so changing b does not have a big effect, while a may be fixed or can be calculated by a Metropolis-Hasting algorithm — in both cases, accuracies did not change much.

To determine the number of pairs to be used in k-TSP, we used the methods described in [10] and [12], leading to two classification rules, which we called k-TSP1 and k-TSP2, respectively. The switchBox R package was used for the k-TSP1 analysis [24] and the ktspair R package was used for the k-TSP2 analysis [25]. In each case, we limit the number of pairs to at most 9. Over all experiments, we observed that k-TSP1 uses an average of 4.02 genes, or about 2 pairs, while k-TSP2 uses an average of 17.8 genes, or nearly all 9 pairs. By contrast, we used a fixed number $k = 9$ of pairs for the k-BTSP classifier. The number of genes used by the SVM-RFE classification rule was fixed at 100, while Naive Bayes (NB) and the Decision Tree (DT) use all 2000 genes.

The accuracy results are displayed in Table 2. We observe that k-BTSP obtained the best overall average accuracy rate, and has the best accuracy rate over more individual data sets than all other classification rules. There doesn't seem to be a conclusive difference between the accuracy rates of the conventional and Bayesian TSP classifiers, but there is a significant improvement of the k-BTSP classifier over both conventional k-TSP classifiers (and of course the TSP classifiers).

k-BTSP is robust to monotone transformations and eliminates the variations among platforms or pre-processing techniques. Therefore, the results from k-BTSP are more reliable and robust compared to many well-known classifiers. k-BTSP is also a great candidate to define clear biological connections between genes and the studies.

7. REFERENCES

[1] Todd R Golub, Donna K Slonim, Pablo Tamayo, Christine Huard, Michelle Gaasenbeek, Jill P Mesirov, Hilary Coller, Mignon L Loh, James R Downing, Mark A Caligiuri, et al., "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *science*, vol. 286, no. 5439, pp. 531–537, 1999.

[2] Margaret A Shipp, Ken N Ross, Pablo Tamayo, Andrew P Weng, Jeffery L Kutok, Ricardo CT Aguiar, Michelle Gaasenbeek, Michael Angelo, Michael Reich, Geraldine S Pinkus, et al., "Diffuse large b-cell lym-

Study	Number of Genes	Class 1 Size	Class 2 Size	Reference
Colon	2000	22	40	Alon et al. (1999) [20]
Leukemia$_1$	7129	25	47	Golub et al. (1999) [1]
DLBCL	7129	58	19	Shipp et al. (2002) [2]
Lung	12,533	150	31	Gordon et al. (2002) [3]
Breast	22,283	62	42	Chowdary et al. (2006) [4]
Leukemia$_2$	12,564	24	24	Armstrong et al. (2002) [5]
Monocytes	26,496	49	47	Maouche et al. (2008) [21]
Squamous	12,625	22	22	Kuriakose, Chen et al. (2004) [6]
Sarcoma	43,931	37	31	Price et al. (2007) [7]
Huntington's	22,283	14	17	Borovecki et al. (2005) [8]
CNS	7129	25	9	Pomeroy et al. (2002) [22]
Myeloma	12,625	137	36	Tian et al. (2003) [23]

Table 1. Gene expression data sets used in the numerical experiment.

Study	TSP	k-TSP$_1$	k-TSP$_2$	BTSP	k-BTSP	DT	NB	SVM-RFE
Colon	0.7112	0.7326	0.7392	0.7304	**0.8132**	0.6504	0.6344	0.7268
Leukemia$_1$	0.8789	0.8893	0.9072	0.8720	**0.9168**	0.8091	0.855	0.8896
DLBCL	0.779	0.8216	0.8412	0.7777	0.8761	0.7408	0.7795	**0.877**
Lung	0.9529	0.9609	0.968	0.9246	**0.9773**	0.9284	0.9678	0.9734
Breast	0.9356	0.9412	0.9465	0.9322	**0.9672**	0.8726	0.8918	0.9134
Leukemia$_2$	0.8936	0.9178	0.9331	0.8878	**0.9573**	0.8436	0.8826	0.9486
Monocytes	0.981	0.9838	0.989	0.9896	**0.9901**	0.8866	0.9884	0.9888
Squamous	0.796	0.8062	0.824	0.7668	0.8388	0.77	0.6871	**0.8708**
Sarcoma	0.8258	0.8233	0.847	0.8289	**0.8640**	0.7564	0.8133	0.852
Huntington's	0.76	0.804	0.7728	0.6248	**0.816**	0.6824	0.6008	0.7728
CNS	0.7188	0.7222	0.7274	0.6540	**0.7540**	0.6044	0.7103	0.7418
Myeloma	0.7009	0.7228	0.7349	0.6755	0.7472	0.7053	**0.7865**	0.7689
Average	0.8278	0.8438	0.8525	0.8068	**0.8765**	0.7708	0.7997	0.8603

Table 2. Accuracy rates obtained in the numerical experiment.

phoma outcome prediction by gene-expression profiling and supervised machine learning," *Nature medicine*, vol. 8, no. 1, pp. 68–74, 2002.

[3] Gavin J Gordon, Roderick V Jensen, Li-Li Hsiao, Steven R Gullans, Joshua E Blumenstock, Sridhar Ramaswamy, William G Richards, David J Sugarbaker, and Raphael Bueno, "Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma," *Cancer research*, vol. 62, no. 17, pp. 4963–4967, 2002.

[4] Dondapati Chowdary, Jessica Lathrop, Joanne Skelton, Kathleen Curtin, Thomas Briggs, Yi Zhang, Jack Yu, Yixin Wang, and Abhijit Mazumder, "Prognostic gene expression signatures can be measured in tissues collected in rnalater preservative," *The journal of molecular diagnostics*, vol. 8, no. 1, pp. 31–39, 2006.

[5] Scott A Armstrong, Jane E Staunton, Lewis B Silverman, Rob Pieters, Monique L den Boer, Mark D Minden, Stephen E Sallan, Eric S Lander, Todd R Golub, and Stanley J Korsmeyer, "Mll translocations specify a distinct gene expression profile that distinguishes a unique leukemia," *Nature genetics*, vol. 30, no. 1, pp. 41–47, 2002.

[6] MA Kuriakose, WT Chen, ZM He, AG Sikora, P Zhang, ZY Zhang, WL Qiu, DF Hsu, C McMunn-Coffran, SM Brown, et al., "Selection and validation of differentially expressed genes in head and neck cancer," *Cellular and Molecular Life Sciences CMLS*, vol. 61, no. 11, pp. 1372–1383, 2004.

[7] Nathan D Price, Jonathan Trent, Adel K El-Naggar, David Cogdell, Ellen Taylor, Kelly K Hunt, Raphael E Pollock, Leroy Hood, Ilya Shmulevich, and Wei Zhang, "Highly accurate two-gene classifier for differentiating gastrointestinal stromal tumors and leiomyosarcomas," *Proceedings of the National Academy of Sciences*, vol. 104, no. 9, pp. 3414–3419, 2007.

[8] F Borovecki, L Lovrecic, Jessica Zhou, Hyun Jeong,

Florian Then, HD Rosas, SM Hersch, P Hogarth, Berengere Bouzou, RV Jensen, et al., "Genome-wide expression profiling of human blood reveals biomarkers for huntington's disease," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 31, pp. 11023–11028, 2005.

[9] Donald Geman, Christian d'Avignon, Daniel Q Naiman, and Raimond L Winslow, "Classifying gene expression profiles from pairwise mrna comparisons," *Statistical applications in genetics and molecular biology*, vol. 3, no. 1, pp. 1–19, 2004.

[10] Aik Choon Tan, Daniel Q Naiman, Lei Xu, Raimond L Winslow, and Donald Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, vol. 21, no. 20, pp. 3896–3904, 2005.

[11] Xue Lin, Bahman Afsari, Luigi Marchionni, Leslie Cope, Giovanni Parmigiani, Daniel Naiman, and Donald Geman, "The ordering of expression among a few genes can provide simple cancer biomarkers and signal brca1 mutations," *BMC bioinformatics*, vol. 10, no. 1, pp. 1, 2009.

[12] Bahman Afsari, Ulisses M Braga-Neto, Donald Geman, et al., "Rank discriminants for predicting phenotypes from rna expression," *The Annals of Applied Statistics*, vol. 8, no. 3, pp. 1469–1491, 2014.

[13] Ralph Allan Bradley and Milton E Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, 1952.

[14] Francois Caron and Arnaud Doucet, "Efficient bayesian inference for generalized bradley–terry models," *Journal of Computational and Graphical Statistics*, vol. 21, no. 1, pp. 174–196, 2012.

[15] Peter McCullagh and John A Nelder, *Generalized linear models*, vol. 37, CRC press, 1989.

[16] David R Hunter, "Mm algorithms for generalized bradley-terry models," *Annals of Statistics*, pp. 384–406, 2004.

[17] Isobel Claire Gormley and Thomas Brendan Murphy, "Exploring voting blocs within the irish electorate: A mixture modeling approach," *Journal of the American Statistical Association*, vol. 103, no. 483, pp. 1014–1027, 2008.

[18] John Guiver and Edward Snelson, "Bayesian inference for plackett-luce ranking models," in *proceedings of the 26th annual international conference on machine learning*. ACM, 2009, pp. 377–384.

[19] Haiyan Wang, Hongyan Zhang, Zhijun Dai, Ming-shun Chen, and Zheming Yuan, "Tsg: a new algorithm for binary and multi-class cancer classification and informative genes selection," *BMC medical genomics*, vol. 6, no. 1, pp. 1, 2013.

[20] Uri Alon, Naama Barkai, Daniel A Notterman, Kurt Gish, Suzanne Ybarra, Daniel Mack, and Arnold J Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proceedings of the National Academy of Sciences*, vol. 96, no. 12, pp. 6745–6750, 1999.

[21] Seraya Maouche, Odette Poirier, Tiphaine Godefroy, Robert Olaso, Ivo Gut, Jean-Phillipe Collet, Gilles Montalescot, and François Cambien, "Performance comparison of two microarray platforms to assess differential gene expression in human monocyte and macrophage cells," *BMC genomics*, vol. 9, no. 1, pp. 1, 2008.

[22] Scott L Pomeroy, Pablo Tamayo, Michelle Gaasenbeek, Lisa M Sturla, Michael Angelo, Margaret E McLaughlin, John YH Kim, Liliana C Goumnerova, Peter M Black, Ching Lau, et al., "Prediction of central nervous system embryonal tumour outcome based on gene expression," *Nature*, vol. 415, no. 6870, pp. 436–442, 2002.

[23] Erming Tian, Fenghuang Zhan, Ronald Walker, Erik Rasmussen, Yupo Ma, Bart Barlogie, and John D Shaughnessy Jr, "The role of the wnt-signaling antagonist dkk1 in the development of osteolytic lesions in multiple myeloma," *New England Journal of Medicine*, vol. 349, no. 26, pp. 2483–2494, 2003.

[24] Bahman Afsari, Elana J Fertig, Donald Geman, and Luigi Marchionni, "switchbox: an r package for k–top scoring pairs classifier development," *Bioinformatics*, p. btu622, 2014.

[25] J Damond, "ktspair: k-top scoring pairs for microarray classification, 2011," *R package version*, vol. 1.