Optimizing AUC Related Performance Measures

Utkarsh Agarwal

1 Introduction

Extending the previous work done on CROC, concentrated ROC, a performance measure based on the ROC curve, we use bipartite and full ranking techniques to solve a bunch of more complex performance measures which are based on ROC curve and focus on early retrieval. We will look at a couple of performance measures one by one.

2 Rate Weighted AUC(rAUC)

2.1 Preliminaries

Consider a training sample $S = (S_+, S_-)$ consisting of m positive instances $S_+ = (x_1^+, x_2^+, ...x_m^+)$ and n negative instances $S_- = (x_1^-, x_2^-, ...x_n^-)$. Let y_i be the true labe of a point with label as 1 for a positive and 0 for a negative.
Let us the define the rate of a threshold t wrt a scoring function w as,

$$r_w(t) \;=\; \sum_{i=1}^{m+n} \mathbf{1}(w^\top x_i \geq t)$$

the accuracy of a scoring function w wrt a threshold t as

$$acc_w(t) \;=\; \sum_{i=1}^{m+n} \mathbf{1}(y_i(w^\top x_i - t) \geq 0) + \mathbf{1}((1-y_i)(w^\top x_i - t) \leq 0).$$

Let the fraction of positive points be π_+ and the fraction of negative points be π_-. Now

$$\pi_+ \;=\; \frac{1}{m+n}\sum_{i=1}^{m+n} \mathbf{1}(y_i = 1) \qquad \pi_- \;=\; \frac{1}{m+n}\sum_{i=1}^{m+n} \mathbf{1}(y_i = 0)$$

Now the minimum accuracy for a rate r, $acc_{min}(r)$ is given by

$$acc_{min}(r) = |\pi_- - r|$$

which can be seen easily. The minimum accuracy will be achieved when maximum negative points have been classified as positive. Take two scenarios, first when $\pi_- < r$. So the accuracy is the fraction of positiive points classified correctly as all negatives are classified as positive. In the second scenario the accuracy is just the fraction of negative points classified correctly as every positive is classified as negative. Now let us see the definition of $rAUC$[**?**, 2]

$$rAUC = \frac{1}{c}\int_0^1 \eta(r)(acc_w(r) - acc_{min}(r))dr$$

where c is the normalisation constant to make maximum value of $rAUC$ as 1.

2.2 Formulation of Optimisation Problem

We will be using full ranking matrix $\pi = [\pi_{ij}]$ to denote the relative ordering among m +n points. π_{ij} is 1 if ith point is ranked below jth point and 0 otherwise. We can frame the following optimisation problem:

$$\min_{w,\xi} \frac{1}{2}||w||^2 \;+\; C\xi$$

s.t. $\forall \pi \in \Pi_{m+n,m+n}$:

$$w^\top(\phi(S,\pi^*) - \phi(S,\pi)) \geq \triangle_{rAUC}(\pi^*,\pi) - \xi$$

where

$$\triangle_{rAUC}(\pi^*,\pi) = \sum_{r=1}^{m+n}\sum_{i=1}^{m+n} \eta(r)((1-y_i^*)\,\hat{y}_i^{w,r} + y_i^*(1-\hat{y}_i^{w,r}))$$

is the loss function for $rAUC$.
Note that y_i* is the true label and $\hat{y}_i^{w,r}$ is the labelling vector for $S = (S_+,\ S_-)$ according to w with exactly r points labelled as positive.
It can easily be seen that this loss function corresponds to $rAUC$ as we penalise each wrongly classified positive and negative with weight $\eta(r)$ for all possible rates ranging from 1 to $m+n$ where rate r is the number of points classified positive.
We define the joint feature map $\phi(S,\pi)$

$$\phi : (X^{m+n} \times X^{m+n}) \times \Pi_{m+n,m+n} \to \mathbb{R}^d \quad \text{as}$$

$$\phi(S,\pi) = \sum_{i=1}^{m+n}\sum_{j=1}^{m+n}(1-\pi_{ij})(x_i - x_j)$$

So $Q_w(\pi)$ the optimisation objective is

$$Q_w(\pi) \ = \ \triangle_{rAUC}(\pi^*,\pi) \ - \sum_{i=1}^{m+n}\sum_{j=1}^{m+n}(\pi_{ij}^* - \pi_{ij})w^\top(x_i - x_j)$$

Now

$$\begin{aligned}\triangle_{rAUC}(\pi^*,\pi) &= \sum_{i=1}^{m+n}\left[(1-y_i^*)\sum_{r=1}^{m+n}\eta(r)\,\hat{y}_i^{w,r} + y_i^*\sum_{r=1}^{m+n}\eta(r)(1-\hat{y}_i^{w,r})\right] \\ &= \sum_{i=1}^{m+n}\left[(1-y_i^*)\sum_{r=k_i}^{m+n}\eta(r) + \ y_i^*\sum_{r=1}^{k_i}\eta(r)\right]\end{aligned}$$

where k_i is the rank of the ith point ranked by scores according to w.

$$k_i = \sum_{j=1}^{m+n}\pi_{ij}$$

3 Robust Initial Enhancement(RIE)

3.1 Preliminaries

Consider a training sample $S = (S_+,\ S_-)$ consisting of m positive instances $S_+ = (x_1^+, x_2^+, ...x_m^+)$ and n negative instances $S_- = (x_1^-, x_2^-, ...x_n^-)$. Now let us see the definition of RIE[3].

$$RIE = \sum_{i=1}^{m} e^{-cr_i}$$

where c is a constant and r_i is the number of negatives ranked above the i^{th} positive.

3.2 Formulation of Optimisation Problem

We will be using bipartite ranking matrix $\pi = [\pi_{ij}]$ to denote the relative ordering among positives and negatives. π_{ij} is 1 if ith positive is ranked below jth negative and 0 otherwise. We can frame the following optimisation problem:

$$\min_{w,\xi} \frac{1}{2}||w||^2 \, + \, C\xi$$

s.t. $\forall \pi \in \Pi_{m,n}$:

$$w^\top(\phi(S,\pi^*) - \phi(S,\pi)) \geq \triangle_{RIE}(\pi^*,\pi) - \xi$$

where

$$\triangle_{RIE}(\pi^*,\pi) = \sum_{i=1}^{m} e^{r_i} = \sum_{i=1}^{m} e^{<\mathbf{1},\pi_i>}$$

is the loss function for RIE.
Note that π_i is the i^{th} row of the bipartite ranking matrix which contains its relative rank wrt the negatives and $\mathbf{1}$ is a n length vector with all entries as 1.
It can be seen that this loss function corresponds to RIE as maximizing e^{-x} is same as minimizing e^x . We define the joint feature map $\phi(S,\pi)$

$$\phi : (X^m \times X^n) \times \Pi_{m,n} \to \mathbb{R}^d \quad \text{as}$$

$$\phi(S,\pi) = \sum_{i=1}^{m}\sum_{j=1}^{n}(1-\pi_{ij})(x_i^+ - x_j^-)$$

So $Q_w(\pi)$, the optimisation objective is

$$Q_w(\pi) \; = \; \triangle_{RIE}(\pi^*,\pi) \; - \; \sum_{i=1}^{m}\sum_{j=1}^{n} \, (\pi^*_{ij} - \pi_{ij})w^\top(x_i^+ - x_j^-)$$

We have to find $\max\limits_{\pi} Q_w(\pi)$.

$$\begin{aligned}
\max_{\pi} Q_w(\pi) &= \max_{\pi}\left[\triangle_{RIE}\,(\pi^*,\pi)\,-\,\sum_{i=1}^{m}\sum_{j=1}^{n}(\pi^*_{ij}-\pi_{ij})w^\top(x_i^+-x_j^-)\right] \\
&= \max_{\pi}\left[\sum_{i=1}^{m}e^{<\mathbf{1},\pi_i>}\,-\,\sum_{i=1}^{m}\sum_{j=1}^{n}(0-\pi_{ij})w^\top(x_i^+-x_j^-)\right] \\
&= \max_{\pi}\left[\sum_{i=1}^{m}e^{<\mathbf{1},\pi_i>}\,+\,\sum_{i=1}^{m}\sum_{j=1}^{n}\pi_{ij}w^\top(x_j^- - x_i^+)\right] \\
&= \max_{\pi}\left[\sum_{i=1}^{m}e^{<\mathbf{1},\pi_i>}\,+\sigma(\pi,w)\right] \\
&= \max_{\pi}\left[\max_{\rho}[<\pi,\rho>-\Delta^*(\rho)]\,+\sigma(\pi,w)\right] \\
&= \max_{\rho}\left[\max_{\pi}[<\pi,\rho>-\Delta^*(\rho)]\,+\sigma(\pi,w)\right] \\
&= \max_{\rho}\left[\max_{\pi}\left[[<\pi,\rho>-\Delta^*(\rho)]\,+\sigma(\pi,w)\right]\right] \\
&= \max_{\rho}\left[\max_{\pi}\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\pi_{ij}[\rho_{ij}+w^\top(x_j^- - x_i^+)]\right]+\Delta^*(\rho)\right] \\
&= \max_{\rho}\left[\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\max[0,\rho_{ij}+w^\top(x_j^- - x_i^+)\right]+\Delta^*(\rho)\right]
\end{aligned}$$

Note that since e^x is a convex function, it can be written as a maximum of linear sum of convex functions using Fenchel Duality.

4 Area Under Accumalation Curve(AUAC)

4.1 Preliminaries

Consider a training sample $S = (S_+,\, S_-)$ consisting of m positive instances $S_+ = (x_1^+, x_2^+, ...x_m^+)$ and n negative instances $S_- = (x_1^-, x_2^-, ...x_n^-)$. Now let us see the definition of $AUAC$[1].

$$AUAC = 1 - \frac{1}{c}\sum_{i=1}^{m} r_i$$

where c is a normalisation constant and r_i is the absolute rank of the i^{th} positive.

4.2 Formulation of Optimisation Problem

We will be using bipartite ranking matrix $\pi = [\pi_{ij}]$ to denote the relative ordering among positives and negatives. π_{ij} is 1 if ith positive is ranked below jth negative and 0 otherwise. We can frame the following optimisation problem:

$$\min_{w,\xi}\frac{1}{2}||w||^2\,+\,C\xi$$

s.t. $\forall\pi \in \Pi_{m,n}$:

$$w^\top(\phi(S,\pi^*)-\phi(S,\pi)) \geq \triangle_{AUAC}(\pi^*,\pi)-\xi$$

where

$$\triangle_{AUAC}(\pi^*, \pi) = \sum_{i=1}^{m} r_i = \sum_{i=1}^{m} [r_i^+ + r_i^-] = \sum_{i=1}^{m} r_i^- + \sum_{i=1}^{m} r_i^+ = \sum_{i=1}^{m} r_i^- + \frac{m(m-1)}{2}$$

is the loss function for $AUAC$. We drop the constant term for convenience.

$$\triangle_{AUAC}(\pi^*, \pi) = \sum_{i=1}^{m} r_i^- = \sum_{i=1}^{m}\sum_{j=1}^{n} \pi_{ij}$$

It can be seen that this loss function corresponds to $AUAC$ as minimizing it directly maximizes the $AUAC$. We define the joint feature map $\phi(S, \pi)$

$$\phi : (X^m \times X^n) \times \Pi_{m,n} \to \mathbb{R}^d \quad \text{as}$$

$$\phi(S, \pi) = \sum_{i=1}^{m}\sum_{j=1}^{n}(1 - \pi_{ij})(x_i^+ - x_j^-)$$

So $Q_w(\pi)$, the optimisation objective is

$$Q_w(\pi) \;=\; \triangle_{AUAC}(\pi^*, \pi) \;-\; \sum_{i=1}^{m}\sum_{j=1}^{n} (\pi_{ij}^* - \pi_{ij}) w^\top (x_i^+ - x_j^-)$$

We have to find $\max\limits_{\pi} Q_w(\pi)$.

$$\begin{aligned}
\max_{\pi} Q_w(\pi) \; &= \max_{\pi} \left[\triangle_{AUAC}\,(\pi^*, \pi) \;-\; \sum_{i=1}^{m}\sum_{j=1}^{n} (\pi_{ij}^* - \pi_{ij}) w^\top (x_i^+ - x_j^-) \right] \\
&= \max_{\pi} \left[\sum_{i=1}^{m}\sum_{j=1}^{n} \pi_{ij} \;+\; \sum_{i=1}^{m}\sum_{j=1}^{n} \; \pi_{ij} w^\top (x_j^- - x_i^+) \right] \\
&= \max_{\pi} \left[\sum_{i=1}^{m}\sum_{j=1}^{n} \pi_{ij} \, [1 + \; w^\top (x_j^- - x_i^+)] \right]
\end{aligned}$$

5 Receiver Operating Characteristic Surface(ROCS)

5.1 Preliminaries

Consider a training sample $S = (S_+,\, S_-)$ consisting of m positive instances $S_+ = (x_1^+, x_2^+, ...x_m^+)$ and n negative instances $S_- = (x_1^-, x_2^-, ...x_n^-)$. Now let us see the definition of $AUAC$[4].

$$CROC = \int_{\delta=-\infty}^{\infty} (1 - FPR(\delta))TDR(\delta)dTPR(\delta)$$

where TDR is the True Discovery Rate, TPR is the True Positive Rate and FPR is the false positive rate.

5.2 Analysis of the Optimisation Objective

Let us define the labelling vector $\hat{y}$ as a length $m + n$ vector where y_i denotes the label alotted to the i^{th} point. Now let us rewrite the discretised form of the optimisation objective

$$\begin{aligned}
CROC &= \sum_{i=1}^{m+n} TP_i FP_i dTDR_i \\
&= \sum_{i=1}^{m+n} < a_i, \hat{y} > (i - < a_i, \hat{y} >)(< a_i - a_{i-1}, \hat{y} >)
\end{aligned}$$

This is a cubic objective function which is pseudo convex in nature, thus we are presently stuck trying to figure out the suitable optimisation technique.

References

[1] Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem Jean-François Truchon and and Christopher I. Bayly Journal of Chemical Information and Modeling 2007 47 (2), 488-508 DOI: 10.1021/ci600426e

[2] Rate-Constrained Ranking and the Rate-Weighted AUC. Millard, Louise A. C. and Flach, Peter A. and Higgins, Julian P. T. Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part II. DOI: 10.1007/978-3-662-44851-9_25

[3] Protocols for Bridging the Peptide to Nonpeptide Gap in Topological Similarity Searches Robert P. Sheridan,*, Suresh B. Singh,Eugene M. Fluder, and, and Simon K. Kearsley Journal of Chemical Information and Computer Sciences 2001 41 (5), 1395-1406 DOI: 10.1021/ci0100144

[4] Yu T (2012) ROCS: Receiver Operating Characteristic Surface for Class-Skewed High-Throughput Data. PLoS ONE 7(7): e40598. doi:10.1371/journal.pone.0040598