COMMENTARY

CrossMark

Reply to the comment made by Šicho, Voršilák and Svozil on 'The Power metric: a new statistically robust enrichment-type metric for virtual screening applications with early recovery capability'

Hans De Winter[*] and Julio Cesar Dias Lopes

The authors of the comment [1] raised an interesting remark about the relation between the power metric (*PM*) [2] and the precision metric (*PR*), also known as the positive predictive value (*PPV*).

In fact, this relation was noted before by the authors of the article that introduced the power metric [2]. Actually, this relationship is shared by all enrichment-type metrics, like the enrichment factor (*EF*) and *ROC* enrichment (*ROCE*), as can be noted by these equations:

$$EF = \frac{PPV}{R_a} \quad (1)$$

$$ROCE = \frac{PPV \cdot R_i}{(1 - PPV) \cdot R_a} \quad (2)$$

$$PM = \frac{PPV \cdot R_i}{PPV \cdot R_i + (1 - PPV) \cdot R_a} \quad (3)$$

in which R_i and R_a being the proportion of active and inactive instances in the whole dataset with N instances:

$$R_a = \frac{n_a}{N} \text{ and } R_i = \frac{n_i}{N} \quad (4 \text{ and } 5)$$

with n_a and n_i the number of active and inactive instances in the dataset.

This relationship was one of the reasons to classify the power metric as another enrichment-type metric. In fact, all enrichment-type metrics can be expressed by the same representation:

$$\text{'Enrichment-type metric'} = \frac{TPR}{x} \quad (6)$$

in which the threshold χ will be the cutoff that defines the hitlist of selected compounds. It can be expressed differently for each particular metric:

(a) in *EF*, χ is the fraction of compounds selected ($\chi = N_s/N$), related to the number of true and false positives (*TP* and *FP*):

$$\chi = \frac{TP + FP}{N} \quad (7)$$

(b) in *ROCE*, χ can be related to the fraction of inactive instances wrongly classified as positives:

$$\chi = FPR = \frac{FP}{n_i} \quad (8)$$

(c) in *PM*, χ can be related to the sum of the true and false positive rates:

$$\chi = TPR + FPR = \frac{TP}{n_a} + \frac{FP}{n_i} \quad (9)$$

Due to these characteristics all these metrics are interconvertible.

A second remark made by Šicho, Voršilák and Svozil [1] is that the power metric 'should be accompanied by a metric taking negative classification into account'. We do not entirely agree with this statement as one can estimate all other metrics from the 2-by-2 contingency (confusion) matrix using only the power metric value and the

*Correspondence: hans.dewinter@uantwerpen.be
Laboratory of Medicinal Chemistry, University of Antwerp,
Universiteitsplein 1, 2610 Wilrijk, Belgium

Springer Open

user-defined threshold χ. Combining Eqs. (6) and (9), we can redefine *PM* as a function of χ and *FPR*:

$$PM = \frac{TPR}{TPR + FPR} \tag{10}$$

and derive:

$$TPR = PM \cdot \chi \tag{11}$$

$$FPR = \chi - TPR = \chi - PM \cdot \chi = \chi \cdot (1 - PM) \tag{12}$$

$$TNR = 1 - FPR \tag{13}$$

$$FNR = 1 = TPR \tag{14}$$

In addition, using the number of actives and inactives, all values of *TP*, *FP*, *TN* (true negatives) and *FN* (false negatives) can be calculated, and from these values any metric can be derived.

The fact that all these metrics are functionally related to the precision metric do not invalidated them as being useful metrics ('not suitable for performance assessment', as stated by the authors of the comment). All these metrics have their scopes, strengths and weaknesses. Each one has its meaning and can be used by the user depending on the desired aims. For example, the precision or *EF* metrics might be more appropriate if the user is more concerned about false positives, while in applications with more emphasis on true positive rates the *PM* or *ROCE* metrics would be recommended instead.

In order to have a better understanding on the interpretation of the power metric, lets investigate the dependency of *PM* on threshold χ. In case of a 'perfect' screening method in which *FPR* approaches zero, the *PM* tends to approach one (Eq. 10) and *TPR* tends to become equal to χ (Eq. 9). Thus, in this case the maximum value of the *TPR* is limited by the user-defined threshold value χ:

$$TPR_{\max} = \chi \tag{15}$$

and the PM could be expressed as:

$$PM = \frac{TPR}{TPR_{\max}} \tag{16}$$

This leads us to the interpretation of the *PM* as the fraction of active compounds that are correctly predicted in relation to the maximum fraction of active compounds that could be recovered at the chosen threshold χ, or, in other words, *PM* express the probability of an active compound to be correctly classified.

> This comment refers to the article available at https://doi.org/10.1186/s13321-018-0267-x.

Authors' contributions
HDW and JCDL wrote, reviewed and edited the manuscript. Both authors read and approved the final manuscript.

References
1. Svozil D, Šícho M, Voršilák M (2018) Comment on "The power metric: a new statistically robust enrichment-type metric for virtual screening applications with early recovery capability". J Cheminf. https://doi.org/10.1186/s13321-018-0267-x
2. Lopes JCD, Dos Santos FM, Martins-José A, Augustyns K, De Winter H (2017) The power metric: a new statistically robust enrichment-type metric for virtual screening applications with early recovery capability. J Cheminform 9:7