

Why Most Published Research Findings Are False

НИС машинное обучение и приложения, 2016/2017

Введем обозначения: s - число всех опубликованных работ в данной области, α, β - вероятности ошибок первого и второго рода в публикации, R - отношение числа работ, исследовавших существующие закономерности ($TP + FN$), к числу работ, исследовавших данные без закономерности ($FP + TN$).

1. Выразить через R априорную вероятность наличия закономерности в данных.
2. Оценить число работ в области, которые ошибочно отвергают закономерность в данных.
3. Как оценивается точность (PPV) работ данной области?
4. **Феномен Протея** описывает ситуацию, при которой ученые в новой популярной области стремятся опубликовать результаты, опровергающие предшественников. Предположим, что вследствие феномена Протея $TP + FP = FN + TN$. Например, для каждой найденной закономерности было опубликовано опровержение. Рассмотрим модель со смещением. Как в этом случае параметр u ¹ зависит от α, β, R ?
5. Согласно модели с множеством независимых исследователей, к чему будут стремиться точность и полнота области исследования по мере увеличения n ?

¹В оригинальной статье u был независимым параметром, определявшим смещенность результатов в области.