

Рецензия на работу:  
**“Learning co-evolution information with natural language processing  
for protein folding problem”** Зверева Егора

Рассматривается задача классификации белковых складок и методы ее решения, основанные на ко-эволюционной информации между белками. Проводится анализ предобученных языковых моделей, как альтернативных к методу MSA подходов для решения данной задачи.

Ошибки, недочеты, замечания:

В разделе 2.3 рекомендуется привести более подробные определения используемых понятий, а также пронумеровать и выделить формулы через \$\$ ... \$\$.

В разделе 2.4 следует добавить более подробное описание.

В разделе 3.2 и 3.3 требуется добавить более подробное описание используемой архитектуры. Также в работе не приводится мотивация того, почему было решено использовать именно такую архитектуру модели.

В разделе 4.2 графики очень мелкие. Необходимо либо увеличить масштаб, либо вставить эти графики отдельно.

Также нет комментариев про обобщение проведенных экспериментов на задачу классификации всех 1517 типов белков.

Итог:

В работе присутствует четкое целеполагание, однако далее теряется нить повествования из-за недостаточно подробного содержания разделов 2, 3, 4. Кроме того, не сделан общий вывод по проведенным экспериментам, что придает незавершенность разделу 4 с описанием вычислительного эксперимента. Также в списке используемой литературы отсутствует ссылка на исходный код. Данная работа посвящена очень актуальной проблеме классификации, однако необходимо более подробно описать используемую архитектуру модели и пояснить, почему была выбрана именно эта архитектура, почему не рассматривались другие предобученные архитектуры, а также дать комментарий о возможностях обобщения данного подхода с классификации на 2 и 10 классов (эксперименты в данной работе), на классификацию на 1517 классов.

Рецензент:

Толмачев А. Д.