

УДК 004.89

Д. Дармопук, Р. Жаровський канд. техн. наук

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

АНАЛІЗ УСПІШНОСТІ СТУДЕНТІВ НА ОСНОВІ ТЕХНОЛОГІЇ GRITNET

UDC 004.89

D. Darmopuk, R. Zharovskyu Ph.D.

STUDENT PERFORMANCE ANALYSIS BASED ON GRITNET TECHNOLOGY

Проблема прогнозування успішності студентів частково досліджувалась спільнотою аналітиків та фахівцями з аналізу освітніх даних у формі прогнозування імовірності відрахування (або завершення навчання) студента. Ця задача є важливою проблемою підкласу прогнозування успішності здобувачів освіти. Більшість таких робіт можна поділити на дві групи, в залежності від використовуваних підходів.

Перший класичний підхід в основному ґрунтується на узагальнених лінійних моделях, які включають логістичну регресію, лінійний SVM та аналіз «виживання». Кожна модель враховує різні типи поведінкових і прогнозних особливостей, добутих з різних неопрацьованих записів активності студентів (наприклад, потоку кліків, оцінок, форуму).

Другий, новий підхід, включає дослідження успішності з використанням нейронних мереж. Дослідники використовували моделі глибоких нейронних мереж (DNN), рекурентні та згорткові нейронні мережі. Проте всі ці нові моделі, поки що, показали примітивну продуктивність. Це пов'язано головним чином із тим, що моделі все ще покладаються на розробку функцій для зменшення вхідних розмірів даних, що, здається, обмежує розробку більших (тобто кращих) моделей нейронних мереж. Записи діяльності студентів, зібрані з різних курсів, часто мають різну довжину, формат і зміст, тому функції, ефективні в одному курсі, можуть бути не такими в іншому. Навіть ретельно розроблені розміри об'єкта зазвичай мають бути малими.

Обидва підходи формують вхідні дані, які поки що занадто обмежені, щоб використати всі переваги моделей послідовного глибокого навчання. Щоб уникнути недоліків попередніх робіт, GritNet сприймає навчальну діяльність здобувачів освіти у часі як вихідні дані і (неявно) шукає частини послідовності вбудовування подій, які є найбільш дискримінаційними для прогнозування успішності студента без необхідності розробляти ці частини як (явні) вхідні ознаки.

Задачу прогнозування успішності студентів можна виразити у вигляді задачі послідовного прогнозування подій: враховуючи попередню (історичну) послідовність $o \triangleq (o_1, o_2, \dots, o_T)$ у яких бере участь здобувач освіти потрібно оцінити імовірність майбутньої послідовності подій $y \triangleq (y_{T+D}, \dots, y_{T'})$, де $D \in Z_+$.

У формі онлайн-занять кожна подія представляється у вигляді деякої діяльності студента, яка пов'язана з відповідною міткою часу. Іншими словами, o_t визначається як кортеж (a_t, d_t) .

Кожна дія a_t може представляти собою, наприклад, «переглянуто відео лекції», «відповідь на тестове запитання правильне/неправильне» або «проект виконано та здано/не здано», а d_t вказує відповідну (zareєстровану) мітку часу.

Таким чином, метою GritNet є обчислення індивідуального журналу логарифмічної правдоподібності $p(y_i | v; \theta)$, і ці оцінені бали можна просто додати для оцінки довгострокових результатів успішності студента.