

Синтез продукційної діагностичної моделі за допомогою негативного відбору з маскуванням

С.О. Зайцев¹

Анотація – Запропоновано модель негативного відбору і метод її навчання з використанням маскування детекторів, що дозволило підвищити швидкість роботи моделі та покращити інтерпретабельність отримуваних результатів.

Ключові слова – Продукційна модель, негативний відбір, маскування.

I. ВСТУП

Продукційні моделі знаходять своє застосування у технічному та медичному діагностуванні, оскільки дозволяють встановити взаємозв'язки між певними значеннями ознак екземпляра та його станом. Часто продукційні моделі синтезують за допомогою дерев прийняття рішень [1]. Головним їх недоліком є надмірна складність донавчання дерева з надходженням нових даних.

У випадку використання дискретних ознак доцільним є використання моделі негативного відбору, яка може виправити цей недолік, а також має ряд переваг – здатна навчатись на екземплярах одного класу та працювати з бінарним представленням даних. Проте результати роботи існуючих моделей негативного відбору [2] складно інтерпретувати, до того ж сформовані продукційні правила не завжди є оптимальними.

Мета роботи полягає у створенні такої моделі негативного відбору, яка дозволяла б синтезувати продукційні моделі на основі бінарного представлення ознак.

II. НЕГАТИВНИЙ ВІДБІР З МАСКУВАННЯМ

Зазвичай моделі негативного відбору навчаються за допомогою цензурування. Ця процедура полягає у видаленні таких детекторів, які активуються при співставленні зі “своїми” (бездефектними). Таким чином, модель містить у собі виключно такі детектори, які потенційно можуть реагувати на будь-який “чужий” (дефектний) екземпляр.

Бінарне представлення даних передбачає, що детектори також являють собою бітові рядки, і значення кожного біта детектора можна розглядати як окрему складову частину продукційного правила. У випадку високої розмірності задачі такі правила будуть містити занадто велику кількість умов в антицеденті правила.

З метою спрощення правил, авторами пропонується використання детекторів з маскуванням. Для цього до

алфавіту необхідно додати ще один символ – Z . Цей символ означатиме замаскований біт детектора. Таким чином, антицедент правил міститиме меншу кількість умов (будуть включатися лише незамасковані ознаки).

В ході навчання моделі генерація кожного окремого детектора проходить наступні етапи.

1. Згенерувати повністю замаскований детектор Z , де n – розмірність задачі.

2. Якщо $\exists s \in S, \forall i = \overline{1, n}, d_i = s_i \vee d_i = Z$, то перейти до етапу 3, в іншому випадку — перейти до етапу 4.

3. Обрати такий $i = \overline{1, n}, d_i = Z$. Встановити $d_i = \neg s_i$.

Перейти до етапу 2.

4. Зупин.

Для донавчання моделі необхідно виконати цензурування існуючих детекторів. Завдяки тому, що детектори є автономними, ця операція є досить простою.

Програмна реалізація моделі на мові Python тестувалася на задачі медичного діагностування [3], при цьому точність роботи моделі становила близько 95%.

III. ВИСНОВОК

Наукова новизна роботи полягає в тому, що розроблено новий метод навчання моделі негативного відбору, що дозволило завдяки маскуванню детекторів скоротити їх кількість та спростити продукційні правила, а також дозволяє проводити донавчання моделі.

Практична цінність полягає в тому, що було розроблене математичне та програмне забезпечення, яке спрощує синтез продукційних моделей.

СПИСОК ПОСИЛАНЬ

- [1] J. Ross Quinlan "C4.5: Programs for Machine learning" Morgan Kaufmann Publishers, 1993, P. 324.
- [2] Ji Z., Dasgupta D. "Revisiting negative selection algorithms", *Evol. Comput.*, vol. 15. pp. 223–251, 2007.
- [3] Герасимчук Т.С., Зайцев С.А., Субботин С.А. "Использование ИИС для прогнозирования риска развития рекуррентных респираторных инфекций у детей раннего возраста", матеріали конф. "Діагностика та лікування інфекційно опосередкованих соматичних захворювань у дітей", Донецьк, Україна, 2011, С. 27–29.

¹ Запорізький національний технічний університет, вул. Жуковського, 64, м. Запоріжжя, 69063, УКРАЇНА, E-mail: zaitsev.serge@gmail.com