



Таблица 2. Результаты экспериментов

Количество слоёв	Доля верных ответов на наборе для тестирования
1	0,9797
2	0,9903
3	0,9859

Литература

1. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов. [Текст]/ В.Н. Вапник, А.Я. Червоненкис: – М.: Наука, 1974.
2. Обучение с подкреплением // Википедия. [2008—2019]. Дата обновления: 31.03.2019. URL: <https://ru.wikipedia.org/?oldid=98962763> (дата обращения: 01.05.2019).
3. Эволюционные алгоритмы // Википедия. [2008—2018]. Дата обновления: 24.10.2018. URL: <https://ru.wikipedia.org/?oldid=95788219> (дата обращения: 01.05.2019).
4. Hanxiao Liu, Karen Simonyan, Yiming Yang DARTS: Differentiable Architecture Search [Электронный ресурс] // arXiv.org. 2001. Дата обновления: 23.04.2019. URL: <https://arxiv.org/abs/1806.09055> (дата обращения: 01.05.2019)
5. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, Backpropagation Applied to Handwritten Zip Code Recognition, 1989
6. Linear programming relaxation // Википедия. [2006—2019]. Дата обновления: 09.02.2019. URL: https://en.wikipedia.org/w/index.php?title=Linear_programming_relaxation&oldid=882528878 (дата обращения: 19.05.2019).

Н.М. Барсков, И.М. Куликовских

ПРОГНОЗИРОВАНИЕ ВЗАИМОДЕЙСТВИЙ ПОЛЬЗОВАТЕЛЕЙ С РЕКЛАМНЫМ СОДЕРЖИМЫМ ИНТЕРНЕТ СТРАНИЦ

(Самарский университет)

В рамках данной работы был разработан и реализован алгоритм для прогнозирования взаимодействия пользователей с рекламными элементами интернет-страниц. Дана обзорная структура предложенного метода, оценки точности прогнозируемых величин.

Предложенный алгоритм представляет собой комбинацию метода наивного сезонного прогнозирования, модели прогнозирования Prophet и метода калибровки выборки.



В качестве исходных данных был использован набор данных, представляющий собой обфусцированную историю взаимодействия пользователей с рекламными элементами интернет-страниц.

Предложенный метод быть использован при ограничении вычислительных возможностей рабочей станции или в случае малых выборок.

Описание задачи

Целью данного исследования является построения алгоритма для прогнозирования взаимодействий пользователей с рекламными элементами интернет-страниц.

Основным источником данных о взаимодействиях пользователей с рекламными элементами элементов интернет-страниц (информация о пользовательском поведении) является протокол работы системы реализации рекламного содержимого. Протокол представляет собой файл или совокупность файлов, в котором содержатся хронологический порядок взаимодействий пользователей с рекламными элементами интернет-страниц и описание данных взаимодействий на основе информации, собранной интернет-обозревателем, информации, полученной из внешних источников, и набора технических данных, сгенерированных в процессе розыгрыша рекламного места. Пример формата данных, содержащихся в данном протоколе, описан в таблице 1.

Таблица 1 — Описание исходных данных протокола работы рекламной системы

Название поля	Описание	Множество допустимых значений
Идентификатор пользователя	Уникальный идентификатор пользователя	Строка, генерируемая системой алгоритмической рекламы
Идентификатор рекламного элемента	Уникальный идентификатор рекламного элемента, с которым пользователь взаимодействовал	Целое число, генерируемое системой алгоритмической рекламы
Время взаимодействия	Время взаимодействия пользователя с рекламным элементом	Целое число

Представим каждое взаимодействие пользователя с рекламными элементами в виде вектора U_i , элементами U_i которого являются признаки взаимодействия (данные, содержащиеся в журнале работы рекламной системы, например, как в таблице 1. Тогда следующее выражение содержит формализованное описание каждого взаимодействия



$$U_i = (t_k, u_1, \dots, u_n), i = \overline{1, M}, n = \overline{1, N}, k = \overline{1, K}, u_n \in D_n.$$

Описание алгоритма

Для решения поставленной задачи был разработан алгоритм на основе комбинации метода прогнозирования Naive Seasonal Approach (наивное сезонное прогнозирование) и метода калибровки выборки.

Пусть дана выборка пользователей $\{U\}$ за период времени $[T_0, T_1]$. Пусть u_i элемент вектора взаимодействия содержит значение идентификатора пользователя, совершившего взаимодействие, а u_j — идентификатор рекламного элемента, с которым взаимодействие было совершено, t_k — временной отсчет события взаимодействия

$$\text{Imps}(U, A, t_k, \{U\}) = \begin{cases} 1, & u_i = U \wedge u_j = A \\ 0, & u_i \neq U \vee u_j \neq A \end{cases}, \forall t_k \in [T_0, T_1]$$

определяет функцию взаимодействия пользователя U с рекламным элементом A в момент времени t_k .

$$\text{Distinct}(A, T_0, T_1, \{U\}) = \begin{cases} |\{u_i\}|, & u_j = A \\ 0, & u_j \neq A \end{cases}, \forall t_k \in [T_0, T_1]$$

определяет метрику количества уникальных пользователей, взаимодействующих с рекламным элементом.

На основе исходных наблюдений можно построить временные прогнозы метрик количества наблюдений и количества уникальных пользователей при помощи модели Prophet.

При построении многомерного временного прогноза используется метод наивного сезонного прогнозирования

$$\hat{U}_{T+h|T} = U_{T+h-(k+1)m}.$$

Решение задачи калибровки происходит при помощи решения следующей системы линейных алгебраических уравнений

$$\begin{pmatrix} \text{MIMPS}_{00} & \dots & \text{MIMPS}_{0L} \\ \vdots & \ddots & \vdots \\ \text{MIMPS}_{M0} & \dots & \text{MIMPS}_{ML} \\ \text{MDSTN}_{00} & \dots & \text{MDSTN}_{0L} \\ \vdots & \ddots & \vdots \\ \text{MDSTN}_{M0} & \dots & \text{MDSTN}_{ML} \end{pmatrix} \times \begin{pmatrix} \text{us}_0 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \text{us}_L \end{pmatrix} = \begin{pmatrix} \text{IMPS}_0 \\ \vdots \\ \text{IMPS}_M \\ \text{DSNT}_0 \\ \vdots \\ \text{DSNT}_M \end{pmatrix}$$



$$\text{MIMPS} = \text{Imps} \left(U_i, A_j, T_k, T_{k+1}, \{\hat{U}\} \right), \quad \text{IMPS} = \text{Imps} \left(A_j, T'_k, T'_{k+1}, \{\hat{U}\} \right),$$

где

$$\text{MDSTN} = \begin{cases} 1, & \text{Imps} (U_i, A_j, T_k, T_{k+1}, \{\hat{U}\}) \neq 0 \\ 0, & \text{Imps} (U_i, A_j, T_k, T_{k+1}, \{\hat{U}\}) = 0 \end{cases} \quad \text{DSTN} = \text{Distinct} \left(A_j, T'_k, T'_{k+1}, \{\hat{U}\} \right).$$

Выводы

Проведена разработка алгоритма прогнозирования взаимодействий пользователей с рекламными элементами интернет-страниц, формализована математическая модель.

В качестве дальнейшей работы автором будет произведена детальная оценка качества предложенного алгоритма при различных случаях распределения значений временных рядов, производится улучшение метода построения одномерного прогноза, скорость его выполнения и его точность.

М.А. Борисов, Н.Г. Крупец

АЛГОРИТМЫ ПОСТРОЕНИЯ ДЕРЕВА ПРИНЯТИЯ РЕШЕНИЙ В ЗАДАЧАХ КЛАССИФИКАЦИИ

(Самарский университет)

В представленной работе рассматриваются методы построения правил классификации объектов, характеризуемых вектором признаков, измеренных в различных шкалах измерения.

Целью работы является исследование точностных характеристик алгоритмов составления дерева принятия решений, позволяющего решать задачи классификации на моделях кластеров обучающих и тестирующих последовательностей объектов.

Дерево принятия решений (также может называться деревом классификации или регрессионным деревом) - средство поддержки принятия решений, используемое в машинном обучении, анализе данных и статистике [1]. Структура дерева представляет собой «листья» и «ветки». На рёбрах («ветках») дерева решения записаны атрибуты, от которых зависит целевая функция (функция классификации объектов), в «листьях» записаны значения целевой функции, в остальных узлах — атрибуты, по которым различаются случаи. Чтобы классифицировать новый случай, надо спуститься по дереву до листа и выдать соответствующее значение.

Подобные деревья решений широко используются в интеллектуальном анализе данных. Цель состоит в том, чтобы создать модель, которая предсказывает значение целевой переменной на основе нескольких переменных на входе.