

Конференция ММРО-15. Конкурс по анализу данных. Задача предсказания отклика клиентов ОТП Банка.

ВВЕДЕНИЕ

В рамках всероссийской конференции «Математические методы распознавания образов» ММРО-15 проводится конкурс по решению прикладной задачи из области банковского маркетинга.

Один из способов повышения эффективности взаимодействия банка с клиентами заключается в том, чтобы отправлять предложение о новой услуге не всем клиентам банка, а только некоторой части, выбираемой по принципу наибольшей склонности к отклику на данное предложение.

Конкурсное задание заключается в том, чтобы предложить алгоритм, который будет выдавать оценку склонности клиента к положительному отклику по его признаковому описанию. Эта оценка может (хотя и не обязана) интерпретироваться как вероятность положительного отклика. Предполагается, что, получив такие оценки для некоторого множества клиентов, банк обратится с предложением только к тем клиентам, у которых значение оценки выше некоторого порога.

Данные предоставлены ОТП Банком (<http://www.otpbank.ru/>), который входит в число 50 крупнейших банков России, занимает 2-е место на рынке потребительского кредитования, 3-е место на рынке беззалогового кредитования, и 4-е место на рынке кредитных карт.

ДААННЫЕ

Исходная выборка («выборка А») содержит записи о 15 223 клиентах, классифицированных на два класса: 1 — отклик был (1812 клиентов), 0 — отклика не было (13411 клиентов).

Ещё 14910 записей отложены в качестве тестовых («выборка Б») — ответы по ним известны только банку. Тестовые данные будут использоваться для определения победителя конкурса.

Записи (признаковые описания) клиентов состоят из 50 признаков, в состав которых входит, в частности, возраст, пол, социальный статус относительно работы, социальный статус относительно пенсии, количество детей, количество иждивенцев, образование, семейное положение, отрасль работы.

Данные доступны по следующим адресам:

- выборка А: http://www.machinelearning.ru/wiki/images/2/26/Contest_MMRO15_OTP.rar
- выборка Б: [http://www.machinelearning.ru/wiki/images/5/52/Contest_MMRO15_OTP_\(validation\).rar](http://www.machinelearning.ru/wiki/images/5/52/Contest_MMRO15_OTP_(validation).rar)

ОЦЕНИВАНИЕ

Качество работы алгоритма будет оцениваться, согласно сложившейся в банковском маркетинге практике, при помощи показателя AUC (area under curve) — площади под ROC-кривой (<http://www.machinelearning.ru/wiki/index.php?title=ROC-кривая>), вычисляемой по контрольным данным («выборка Б»).

Для оценивания качества алгоритма на «Выборке Б» участник предоставляет csv-файл, строки которого в точности соответствуют строкам csv-файла «Выборки Б», а единственный столбец содержит вещественные оценки принадлежности клиентов классу 1. Все оценки должны находиться в отрезке $[0,1]$ и могут интерпретироваться как вероятности положительного отклика.

ПРОЦЕДУРА ПРОВЕДЕНИЯ КОНКУРСА И СРОКИ

Конкурс проходит в два этапа.

Этап первый: настройка алгоритмов

Этап проходит с момента объявления конкурса и до **2 сентября 2011 года, 23:59** по московскому времени. В течение этого времени участники могут настраивать свои алгоритмы на обучающей части данных конкурсной задачи («выборка А»).

Для проверки своего алгоритма и сравнения с алгоритмами других участников конкурса можно воспользоваться системой «Полигон» (<http://polygon.machinelearning.ru>), в которой уже загружены исходные данные ([выборка А](#)), имеется несколько стандартных алгоритмов, реализовано вычисление AUC и сравнение алгоритмов по методике скользящего контроля. Методика тестирования в «Полигоне» заключается в следующем. Исходная выборка (выборка А) 50 раз разбивается на две подвыборки, обучающую и контрольную, по методике 10-кратного случайного разбиения на 5 блоков (10 x 5-fold cross-validation) со стратификацией классов (т. е. сохраняя равные пропорции классов в обеих подвыборках). В отчетах «Полигона» указывается оценка AUC, усредненная по всем 50 разбиениям, строятся доверительные интервал и графики. Предварительный рейтинг алгоритмов строится по убыванию среднего AUC.

Для подключения своего алгоритма к системе «Полигон» необходимо реализовать модуль обмена данными. Техническая документация и необходимые шаблоны предоставлены на странице [Полигон алгоритмов классификации](#) вики-ресурса www.MachineLearning.ru. Участники конкурса могут регистрировать неограниченное количество алгоритмов в системе «Полигон» и тестировать их в рамках конкурсной задачи.

Доступ к конкурсному тестированию алгоритмов будет открыт на сайте системы «Полигон» **15 августа в 12:00**. Начиная с этого момента будут приниматься результаты классификации «Выборки Б», также через систему «Полигон». При этом участник конкурса может пройти конкурсное тестирование при помощи алгоритма, подключенного к Полигону или самостоятельно загрузить файл с результатами на «Выборке Б». В зачет конкурса пойдет только один результат от каждого участника — тот, который был загружен последним.

Этап второй: сравнение результатов и объявление победителя

После завершения первого этапа заявки на новые алгоритмы для участия в конкурсе не принимаются. Результаты тестирования передаются банку, который вычисляет AUC по «Выборке Б», и таким образом определяется победитель конкурса.

НАГРАДЫ

Победитель конкурса будет объявлен 16 сентября, в день закрытия конференции ММРО-15.

Призовой фонд конкурса – **100 000 рублей** (без учета налогов).

Денежный приз предоставлен партнером проведения конкурса – ОТП Банком.