

Классификация товарных предложений

Лимит времени: 30 минут

Лимит памяти: 1 Гб

Лимит дискового пространства: 2Гб

[Яндекс.Маркет](#) – это сервис выбора товара и места его покупки. Это также сервис привлечения клиентов для продавцов. Для Яндекса – это один из основных сервисов, приносящий существенную часть доходов компании.

Правильная классификация товарных предложений позволяет продавцам конкурировать на более таргетированном множестве предложений, что влечет за собой возрастание средней ставки на предложение и, соответственно, доходов, а так же, учитывая отношения между категориями, предлагать дополнительные аксессуары и сопутствующие товары своим клиентам.

Дана размеченная обучающая выборка (архивный файл learn.in.bz2 74Мб, около 1,9 млн. примеров), где в каждой строке дан обучающий пример:

Title\tDescription\tShopCategoryName\tPrice\tCategoryId

- o Title – название товарного предложения (например, Nokia n95)
- o Description – неформальное описание, в котором могут указываться характеристики (цвет, размеры), а может содержаться информация рекламного характера (например, лучший телефон на свете)
- o ShopCategoryName имеет следующую структуру:

Host_name:CatnameLevel1\CatnameLevel2...

Например: nokia-shop.ru:Все товары\Телефоны\Nokia

- o Price – цена в рублях
- o CategoryId – id категории из дерева категорий (см. ниже)

А так же дерево категорий (файл categories.txt):

CategoryId\tParentId\tName

У корневой вершины ParentId = -1

Используя данные обучающей выборки, иерархию и названия категорий нужно разработать программу, которая будет классифицировать новые товарные предложения по товарному дереву.

В исходных данных символ последовательность символов \\n — означает перенос строки, \\ - означает \

Решения всех участников будут оцениваться следующим образом:

Для каждого тестового примера будет считаться расстояние по дереву (минимальное число ребер от одной вершины до другой).

Очки за тестовый пример будут начисляться по формуле:

$$score = \left(\frac{1}{4}\right)^{dist}, \text{ dist} - \text{расстояние по дереву, когда dist} = 0 \text{ score} = 1$$

Среднее количество очков по всем тестам является результатом команды.

Замечания:

1. В обучающей выборке ЕСТЬ ошибки. Она составлена в полуавтоматическом режиме операторами. Число ошибок там не превышает 5%. Могут встречаться категории, где все примеры ошибочны, но таких категорий мало, примеров касательно этих категорий так же мало и они не должны оказать существенное влияние на итоговую оценку
2. В Яндекс.Маркете нет особых формальных требований к формату полей, организации магазинного дерева категорий и прочего. Тем не менее использование структуры и специфика магазина могут быть полезной информацией
3. Программа может использовать внешние данные (файлы)

Входные данные:

На вход подается несколько товарных предложений следующего формата

Title\tDescription\tShopCategoryName\tPrice

Всего будет дано около сотни тысяч тестовых примеров

Выходные данные:

Для каждого тестового предложения вывести соответствующий ему id категории

Пример:

Вход	Выход
Nokia n95 Хороший телефон shop-nokia.ru:Товары 14500	91491
Стул KUADRA 1271 Каркас-сатин,сиденье и спинка-пластик www.newbar.ru:Стулья\Стулья на металлокаркасе7560	90673