

Децентрализованная доменная адаптация

Шокоров Вячеслав Александрович

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Научный руководитель д.ф.-м.н. В. В. Стрижов

Москва
2021 г

Общая проблематика

Решается задача доменной адаптации. Суть этой адаптации заключается в обучении модели на данных из домена-источника так, чтобы она показывала сравнимое качество на целевом домене.

Предлагаемое решение

Данную задачу предлагается решить через построение функции преобразования одного домена в другой. Мотивация подхода заключается в том, что после того, как мы построим такую функцию, т.е. получим возможность переводить распределение одного домена в другой, и умея решать некую задачу на первом домене, мы получим сравнимое качество для второго домена на данной задаче.

Исследуемая проблема

Исследуется проблема построения и анализа вероятностного пространства параметров этого преобразования. Проблема усложнена тем, что домены могут принадлежать непересекающихся или слабо пересекающихся пространствах.

Метод исследования

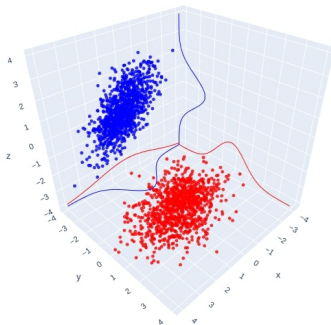
Байесовский подход к нахождению параметров модели и анализ апостериорных распределений.

Определение

Каждый *домен* \mathcal{D} описывается парой $\{G, \mathcal{I}\}$, где \mathcal{I} - множество индексов, G - область пространства признакового описания, причем $G \subset \mathbb{R}^{\mathcal{I}}$.

Определение

Функцией преобразования домена $\{G_1, \mathcal{I}_1\}$ в домен $\{G_2, \mathcal{I}_2\}$ назовем функцию $f : \mathbb{R}^{\mathcal{I}_1} \rightarrow \mathbb{R}^{\mathcal{I}_2}$. На параметры этого преобразования может задаваться какая-то априорная информация.



Определение

Оптимальной функцией преобразования домена \mathcal{D}_1 в домен \mathcal{D}_2 относительно функции сходства g назовем функцию \hat{f} , на которой достигается максимальное сходство распределений второго домена и образа функции f , т.е.:

$$\hat{f} = \arg \max_f g(p(x, x \in \mathcal{D}_2), p(f(x), x \in \mathcal{D}_1))$$

Определение

Назовем функцией сходства s – *score* пары распределений $g_1 : \mathbb{R}^n \rightarrow \mathbb{R}^+$, $g_2 : \mathbb{R}^n \rightarrow \mathbb{R}^+$, определенных на одном пространстве, функцию вида

$$s_0(g_1, g_2) = \frac{\int g_1(\mathbf{w})g_2(\mathbf{w})d\mathbf{w}}{\max_{\mathbf{b} \in \mathbb{R}^n} \int g_1(\mathbf{v})g_2(\mathbf{v} - \mathbf{b})d\mathbf{v}}$$

В работе будут также использоваться функции сходства построенные на дивергенции Кульбака-Лейблера, метрика Васерштейна.

Параметрический вид функции преобразования

В качестве функции преобразования будем брать обобщенно линейную функцию, вида

$$f_{\mathbf{v}_1, \mathbf{V}_2, \mathbf{b}_2, \mathbf{b}_1} = (\mathbf{v}_1)\sigma(\mathbf{V}_2\mathbf{X} + \mathbf{b}_2) + \mathbf{b}_1,$$

где $\mathbf{v}_1, \mathbf{V}_2, \mathbf{b}_2, \mathbf{b}_1$ - параметры преобразования.

Задача минимизации

Будем искать оптимальные параметры функции относительно функции сходства g через задачу максимизации:

$$\mathbf{v}_1, \mathbf{V}_2, \mathbf{b}_2, \mathbf{b}_1 = \arg \max_{\mathbf{v}_1, \mathbf{V}_2, \mathbf{b}_2, \mathbf{b}_1} g\left(p(x, x \in \mathcal{D}_2), p(f_{\mathbf{v}_1, \mathbf{V}_2, \mathbf{b}_2, \mathbf{b}_1}(x), x \in \mathcal{D}_1)\right)$$

Архитектура

Зададим классификатор как нейронную сеть и построим состязательную процедуру обучения. Задача функции преобразования сделать так, чтобы классификатор не мог угадать из какого домена взяты объекты, а задача классификатора научиться угадывать из какого домена взят объект.

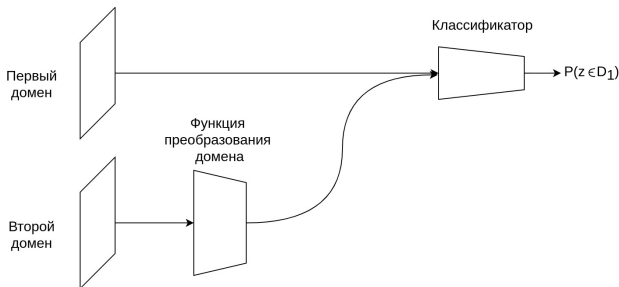


Рис.: Предлагаемая архитектура модели для решения задачи.

Задача минимизации

Для упрощения записи зададим C, f функцию классификатора и функцию преобразования, тогда итоговый вид задачи:

$$L(C, V) = \mathbb{E}_{x \sim \mathcal{D}_1} [\log C(x)] + \mathbb{E}_{x \sim \mathcal{D}_2} [\log(1 - C(f(x)))]$$

Описание данных

Будем проводить вычислительный эксперимент на базе данных отзывов с сайта Amazon, разный домен соответствует категории товара на который оставлен отзыв, например, домены могут соответствовать отзывам про Amazon Fashion, Electronics, Musical Instruments и тд. Для построения признакового описания используется подход TF-IDF.

В качестве разных доменов используются бинаризованные картинки с изображенными на них кругами и квадратами, для каждого домена решается задача регрессии - нахождение размера фигуры на изображении.

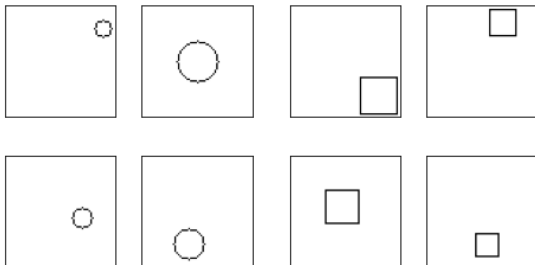






Рис.: Пример изображений из различных доменов

Хотим:

- 1 Статистически проверить гипотезу о равенстве весов в задачах регрессии модели строятся для разных доменов и для разных функций преобразования.
- 2 Проверить качество функций преобразования построенных относительно различных функций сходства с помощью точности решения задачи классификации доменов.
- 3 Для архитектуры с классификатором проверить незначимость каждого признака через априорную матрица корреляции $\mathbf{A} = \text{diag}(\boldsymbol{\alpha})$, которая обладает следующим свойством, если $\alpha_j \rightarrow \infty$ это значит, что этот признак не важен. Цель функции преобразования сделать все признаки шумовыми для классификатора. Поэтому хочется проверить, что $\min_i \alpha_i^C \rightarrow \infty$.

-  А.А. Адуенко. Выбор мультимodelей в задачах классификации, 2017
-  Jing Jiang, A Literature Survey on Domain Adaptation of Statistical Classifiers, 2008
-  А. В. Грабовой, В. В. Стрижов, Анализ выбора априорного распределения для смеси экспертов
-  Guo, Jiang and Shah, Darsh J and Barzilay, Regina, Multi-Source Domain Adaptation with Mixture of Experts, 2018, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, <http://aclweb.org/anthology/D18-1498>