

Доменная адаптация для моделей обработки естественных языков

Виктория Чекалина

15.04.2019

Доменная адаптация (DA)

- Имеются два множества объектов: source и target. Множества состоят из похожих, но немного различающихся объектов. Такие множества называются доменами.

- *Source*: классические литературные текст
- *Target*: интернет комментарии
- *Source*: тексты на испанском
- *Target*: тексты на перуанском



- Объекты в source размечены. Объекты в target - нет.
- Возможно обучить модель для объектов из source, но невозможно - для объектов из target

Доменный сдвиг

(<http://www.cse.psu.edu/~duk17/papers/change.pdf>)

- Решение: обучить модель на объектах из source и применить её к объектам из target

Ожидание:

$$P_s(x, y) = P_t(x, y)$$

$$\text{Err}_s = \text{Err}_t$$

Ожидаемая ошибка на объектах домена

Реальность:

$$P_s(x, y) \neq P_t(x, y)$$

$$\text{Err}_s \neq \text{Err}_t$$

Распределение (объекта или части объектов) над множеством домена

- $\text{Err}_t \leq \text{Err}_s + 0.5 * H \Delta H(D_s, D_t)$ → Дивергенция между source и target доменами.

Доменная адаптация для задач естественных языков

1. В естественном языке размечать данные - очень затратно. Размеченных данных мало. Размеченных данных для узко-специализированных областей нет.
2. Почти все датасеты - для английского языка. Разметка для редких языков в основном отсутствует.

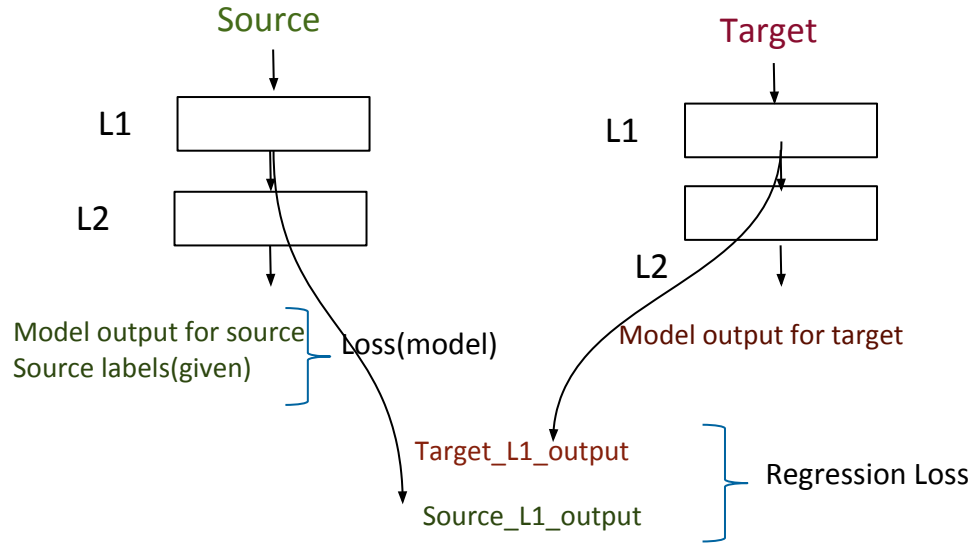
Как можно минимизировать доменный сдвиг

- Работать с объектами
 - Перевзвешивание
 - Self-training
 - Tri-training



- Работать с представлениями объектов для классификатора - делать распределение в пространстве дискриминирующих фич ближе для двух наборов объектов

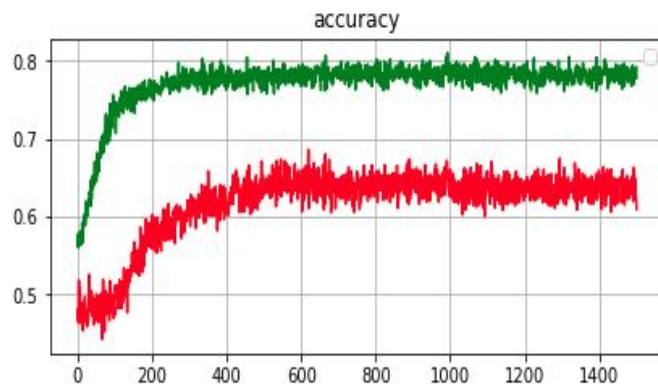
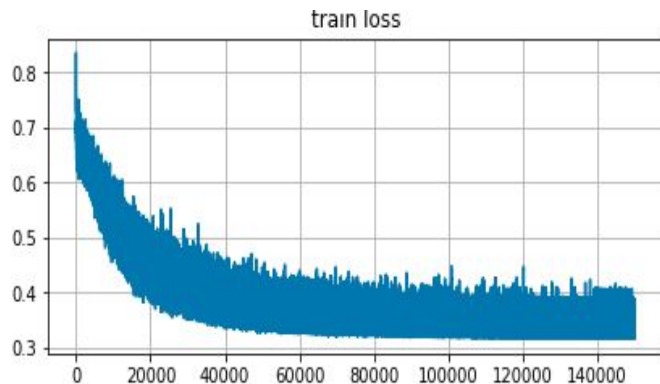
- Подход к задаче DA, изложенный в статье <https://arxiv.org/pdf/1505.07818.pdf>



На каждом шаге обучения:

- Обновляем $\text{Loss}(\text{model}) - \text{const} * \text{Regression loss}$
- Обновляем Regression loss

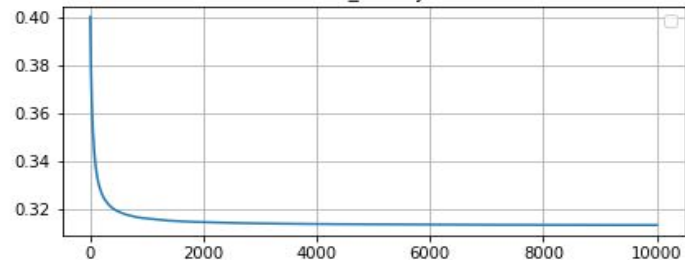
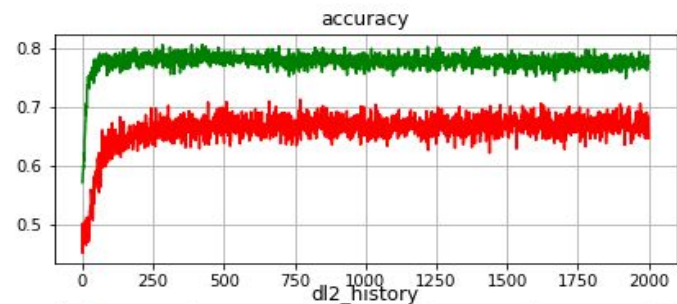
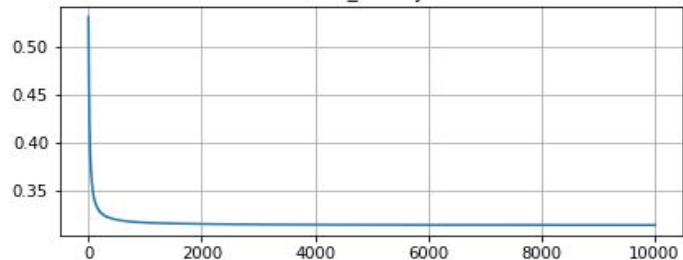
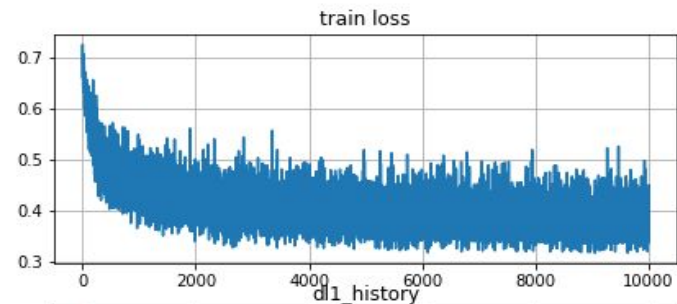
- Source: Amazon electronic reviews
- Target: Yelp restaurant reviews
- Task: Sentiment analys
- Metric: Accuracy
- Common models: 2 and 3-gram tokenization, number of features - 1000
- 2-layer NN
- Simple training (source 0.783, target 0.648)



Подход к задаче DA, изложенный в статье

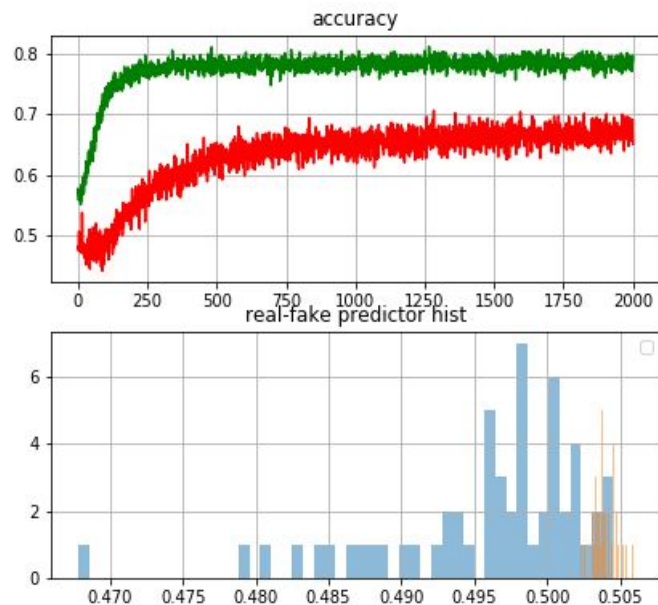
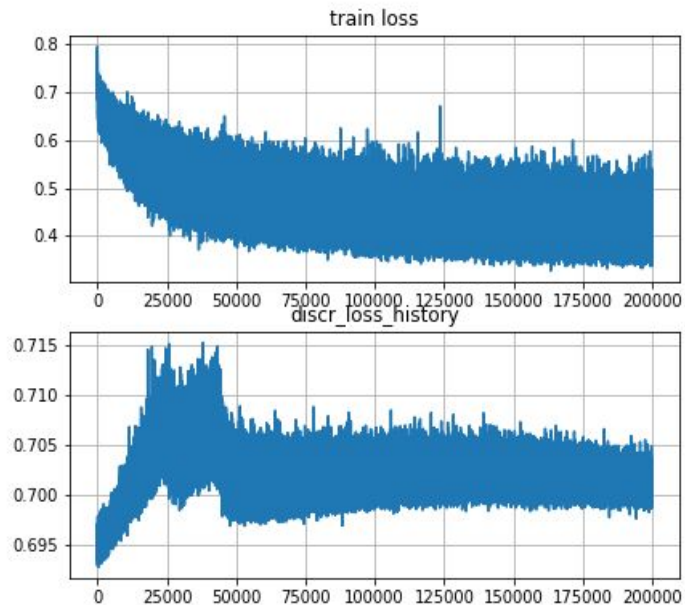
<https://arxiv.org/pdf/1505.07818.pdf>

- Regression: 1-layer NN

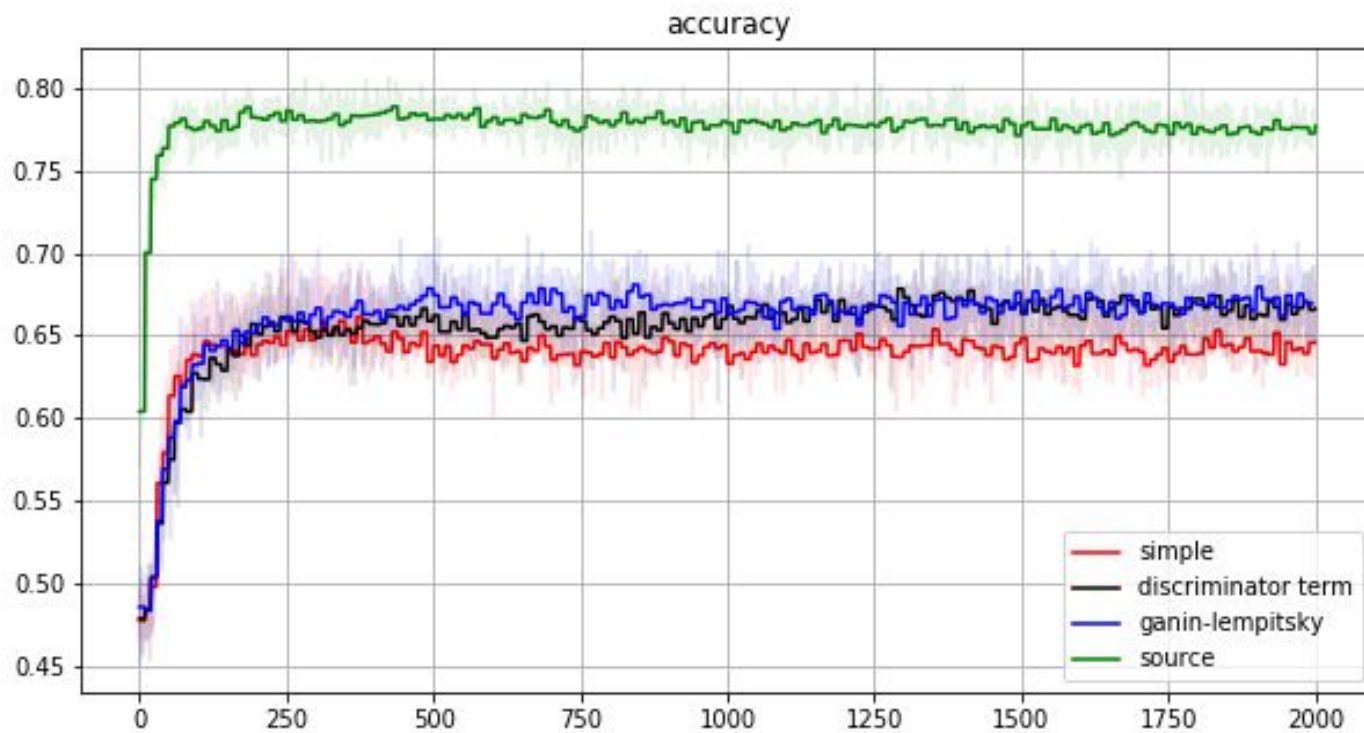


Если добавить вместо регрессии дискриминатор:

- Regression: 1-layer NN

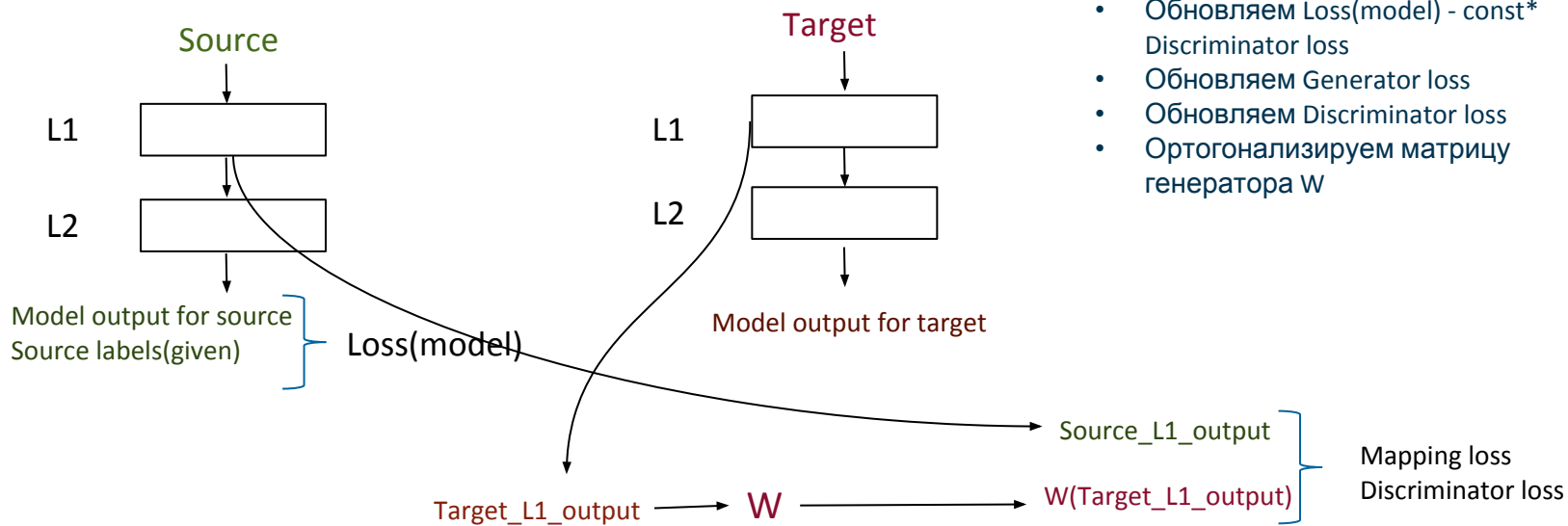


Общий результат

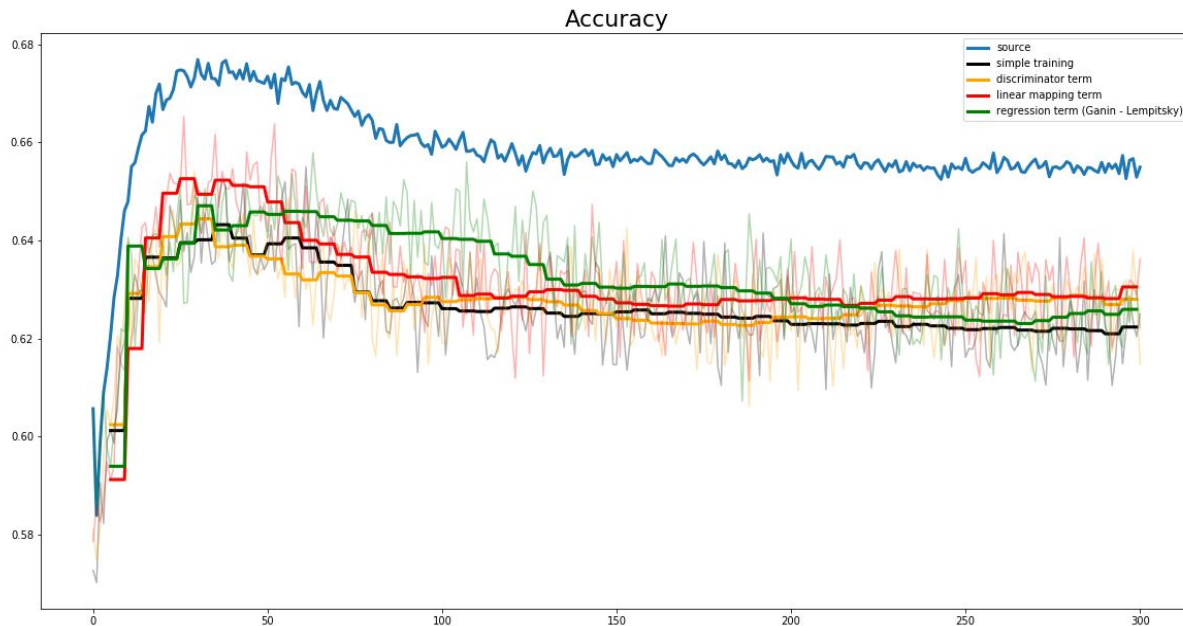


Если добавить генератор и дискриминатор

- В статье <https://arxiv.org/pdf/1805.11222.pdf> описан общий подход к задаче выравнивания распределений в многомерном пространстве
- В качестве генератора выступает матрица поворота, в качестве дискриминатора - Wasserstein Gan with Gradient Penalty(<https://arxiv.org/pdf/1704.00028.pdf>)



- Source: Amazon electronic reviews
- Target: Amazon book reviews



Сравнение алгоритмов на нескольких датасетах

- Sign test

Domain pair/ way of training	Simple training	Regression term	Discriminator term	Linear mapping
Electornic-books	0.621	0.629	0.633	0.636
Books-electronics	0.612	0.624	0.616	0.612
Kitchen-books	0.631	0.606	0.623	0.637
Books-kitchen	0.659	0.572	0.663	0.665
Kitchen-electronics	0.704	0.606	0.713	0.719
Electronics-kitchen	0.739	0.746	0.743	0.744
Electornics-cd	0.629	0.622	0.640	0.641
Cd - electornics	0.625	0.633	0.623	0.614
Books - cosmetics	0.667	0.661	0.657	0.674
Cosmetics-books	0.624	0.533	0.645	0.637

Дальнейшие шаги

- Source: English cinema reviews
Target: Spain cinema reviews
- FastText эмбединги
- Применить алгоритмы доменной адаптации на этих доменах.

Спасибо за внимание!