

# Обучение машинного перевода без параллельных текстов

Саттаров Тагир

Московский физико-технический институт

Курс: Численные методы обучения по прецедентам  
(практика, В.В. Стрижов)/Группа 674, весна 2019

## Цель работы

Улучшить качество перевода на низкоресурсных парах языков

## Проблема

При переводе некоторых пар языков необходимо учитывать морфологическое строение слов, что существенно усложняет задачу машинного перевода

## Решение

Предлагается использовать T2T модель кодировщика-декодировщика вместо используемой seq2seq

# Постановка задачи

## Дано

Пусть заданы  $\mathcal{D}_{src}$  и  $\mathcal{D}_{tgt}$ . Пусть также задана  $\mathcal{D}_{valid}$  для проверки качества метода.

## Вводим

Модели декодировщика  $g$  и кодировщика  $f$

# Требования к кодировщику и декодировщику

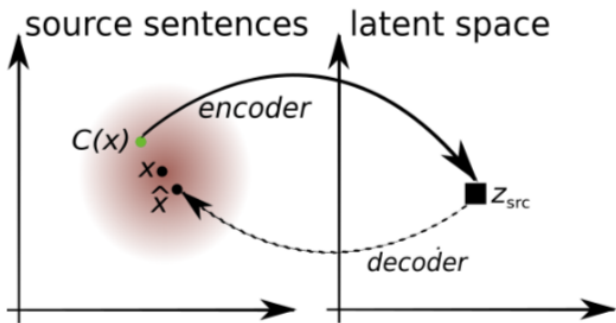


Рис.: Auto-encode

# Требования к кодировщику и декодировщику

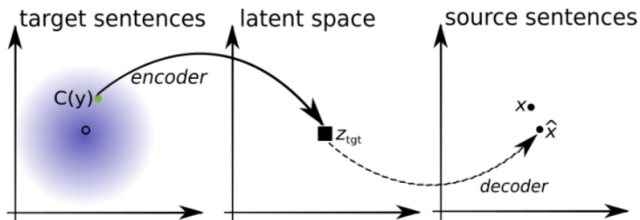


Рис.: Cross-domain

# Вычислительный эксперимент

