# UNSUPERVISED MACHINE TRANSLATION USING MONOLINGUAL CORPORA ONLY

Игошин Антон 141

# Основная идея

Построить общее латентное пространство между двумя языками (пространствами) и научиться переводить с учетом следующих принципов:

- Модель должна восстанавливать данное зашумленное предложение в выбранном языке (как в стандартных Denoising Auto-Encoders)
- Модель учится переводить зашумленное предложение из одного языка (source) по аналогичному зашумленному предложению из другого языка (target)

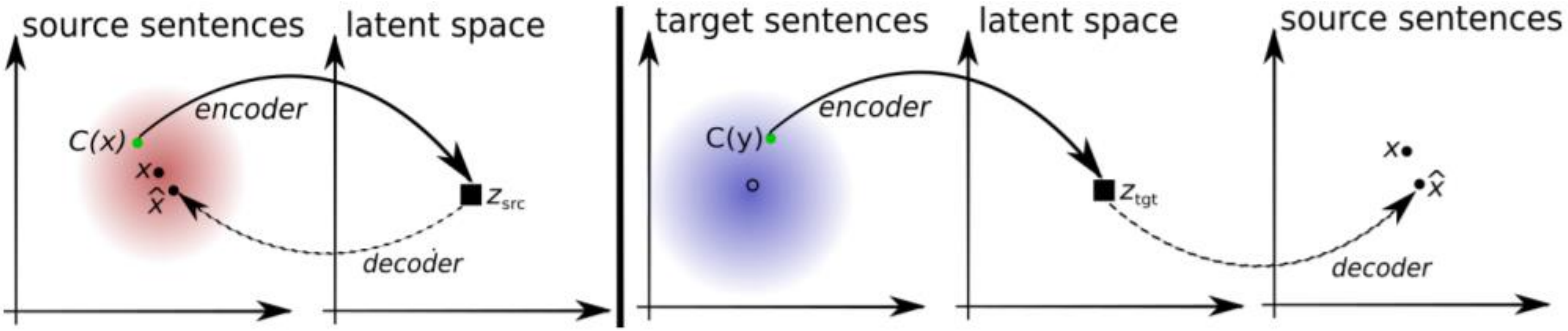


Figure 1: Toy illustration of the principles guiding the design of our objective function. Left (auto-encoding): the model is trained to reconstruct a sentence from a noisy version of it. $x$ is the target, $C(x)$ is the noisy input, $\hat{x}$ is the reconstruction. Right (translation): the model is trained to translate a sentence in the other domain. The input is a noisy translation (in this case, from source-to-target) produced by the model itself, $M$, at the previous iteration $(t)$, $y = M^{(t)}(x)$. The model is symmetric, and we repeat the same process in the other language.

$$\mathcal{L}_{auto}(\theta_{enc}, \theta_{dec}, \mathcal{Z}, \ell) = \mathbb{E}_{x \sim \mathcal{D}_\ell, \hat{x} \sim d(e(C(x),\ell),\ell)} \left[\Delta(\hat{x}, x)\right]$$

where $\hat{x} \sim d(e(C(x), \ell), \ell)$ means that $\hat{x}$ is a reconstruction of the corrupted version of $x$, with $x$ sampled from the monolingual dataset $\mathcal{D}_\ell$. In this equation, $\Delta$ is a measure of discrepancy between the two sequences, the sum of token-level cross-entropy losses in our case.

$$\mathcal{L}_{cd}(\theta_{enc}, \theta_{dec}, \mathcal{Z}, \ell_1, \ell_2) = \mathbb{E}_{x \sim \mathcal{D}_{\ell_1}, \hat{x} \sim d(e(C(y),\ell_2),\ell_1)} \left[\Delta(\hat{x}, x)\right]$$

where $y = M(x)$ and $\Delta$ is again the sum of token-level cross-entropy losses.

The discriminator operates on the output of the encoder, which is a sequence of latent vectors $(z_1, ..., z_m)$, with $z_i \in \mathbb{R}^n$, and produces a binary prediction about the language of the encoder input sentence: $p_D(l|z_1, ..., z_m) \propto \prod_{j=1}^{m} p_D(\ell|z_j)$, with $p_D : \mathbb{R}^n \rightarrow [0; 1]$, where 0 corresponds to the source domain, and 1 to the target domain.

$$\mathcal{L}_{adv}(\theta_{enc}, \mathcal{Z}|\theta_D) = -\mathbb{E}_{(x_i,\ell_i)}[\log p_D(\ell_j|e(x_i, \ell_i))]$$

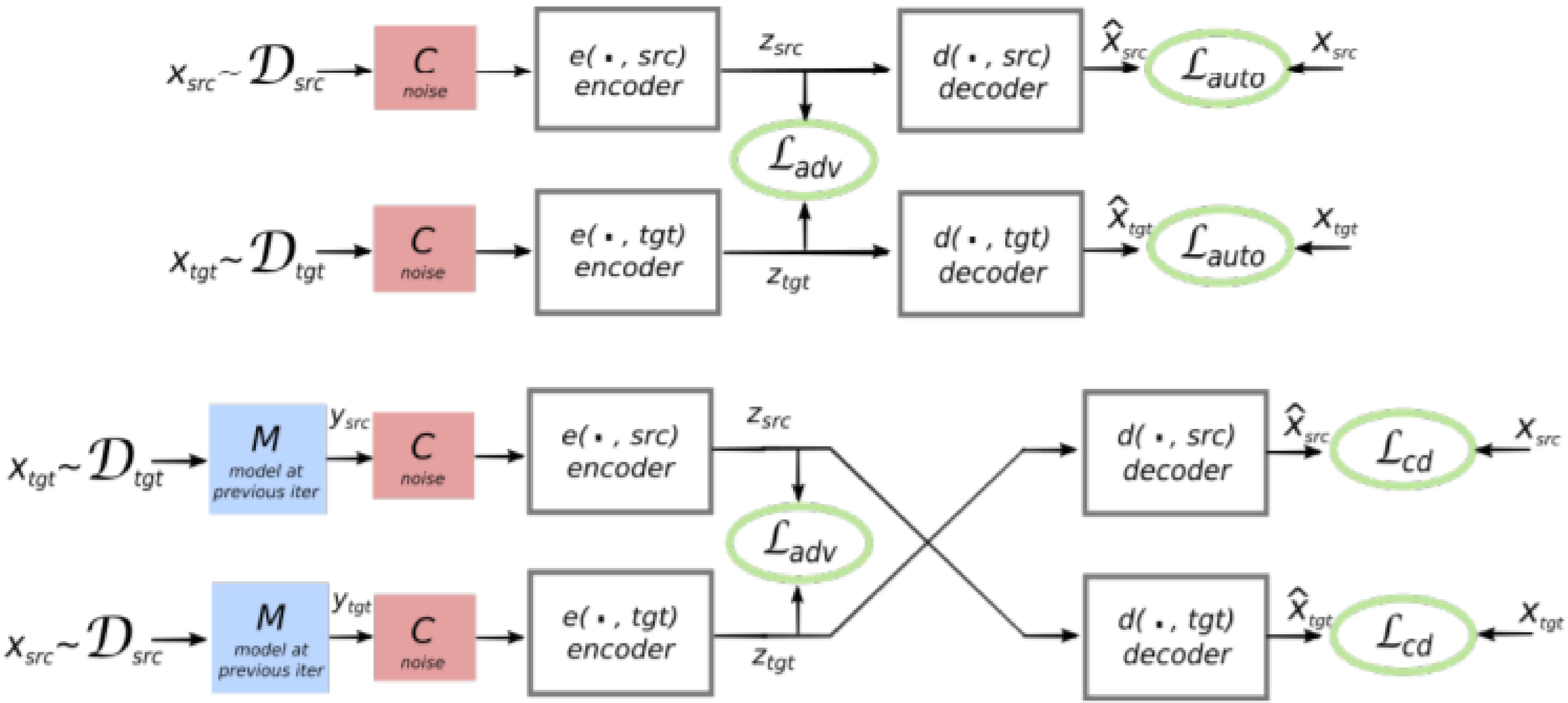


Figure 2: Illustration of the proposed architecture and training objectives. The architecture is a sequence to sequence model, with both encoder and decoder operating on two languages depending on an input language identifier that swaps lookup tables. Top (auto-encoding): the model learns to denoise sentences in each domain. Bottom (translation): like before, except that we encode from another language, using as input the translation produced by the model at the previous iteration (light blue box). The green ellipses indicate terms in the loss function.

$$
\begin{aligned}
\mathcal{L}(\theta_{enc}, \theta_{dec}, \mathcal{Z}) = & \lambda_{auto}[\mathcal{L}_{auto}(\theta_{enc}, \theta_{dec}, \mathcal{Z}, src) + \mathcal{L}_{auto}(\theta_{enc}, \theta_{dec}, \mathcal{Z}, tgt)] + \\
& \lambda_{cd}[\mathcal{L}_{cd}(\theta_{enc}, \theta_{dec}, \mathcal{Z}, src, tgt) + \mathcal{L}_{cd}(\theta_{enc}, \theta_{dec}, \mathcal{Z}, tgt, src)] + \\
& \lambda_{adv}\mathcal{L}_{adv}(\theta_{enc}, \mathcal{Z}|\theta_D)
\end{aligned}
$$

**Algorithm 1** Unsupervised Training for Machine Translation

**procedure** TRAINING($\mathcal{D}_{src}$, $\mathcal{D}_{tgt}$, $T$)
    Infer bilingual dictionary using monolingual data (Conneau et al., 2017)
    $M^{(1)} \leftarrow$ unsupervised word-by-word translation model using the inferred dictionary
    **for** $t = 1, T$ **do**
        using $M^{(t)}$, translate each monolingual dataset
        // discriminator training & model trainingl
        $\theta_{discr} \leftarrow \arg\min \mathcal{L}_D, \quad \theta_{enc}, \theta_{dec}, \mathcal{Z} \leftarrow \arg\min \mathcal{L}$
        $M^{(t+1)} \leftarrow e^{(t)} \circ d^{(t)}$ // update MT model
    **end for**
    return $M^{(t+1)}$
**end procedure**

| | |
|---|---|
| Source | un homme est debout près d' une série de jeux vidéo dans un bar . |
| Iteration 0 | a man is seated near a series of games video in a bar . |
| Iteration 1 | a man is standing near a closeup of other games in a bar . |
| Iteration 2 | a man is standing near a bunch of video video game in a bar . |
| Iteration 3 | a man is standing near a bunch of video games in a bar . |
| **Reference** | **a man is standing by a group of video games in a bar .** |
| Source | une femme aux cheveux roses habillée en noir parle à un homme . |
| Iteration 0 | a woman at hair roses dressed in black speaks to a man . |
| Iteration 1 | a woman at glasses dressed in black talking to a man . |
| Iteration 2 | a woman at pink hair dressed in black speaks to a man . |
| Iteration 3 | a woman with pink hair dressed in black is talking to a man . |
| **Reference** | **a woman with pink hair dressed in black talks to a man .** |
| Source | une photo d' une rue bondée en ville . |
| Iteration 0 | a photo a street crowded in city . |
| Iteration 1 | a picture of a street crowded in a city . |
| Iteration 2 | a picture of a crowded city street . |
| Iteration 3 | a picture of a crowded street in a city . |
| **Reference** | **a view of a crowded city street .** |

- Статья - https://arxiv.org/pdf/1711.00043.pdf