

Overview

- Hypothesis: natural languages are semantically similar
 - Homologous words in different languages should occur in homologous contexts
- Word embedding distributions should be nearly isomorphic
- Learn to match word embedding distributions
 - Using Adversarial Autoencoders
 - Without using any parallel resource

Architecture

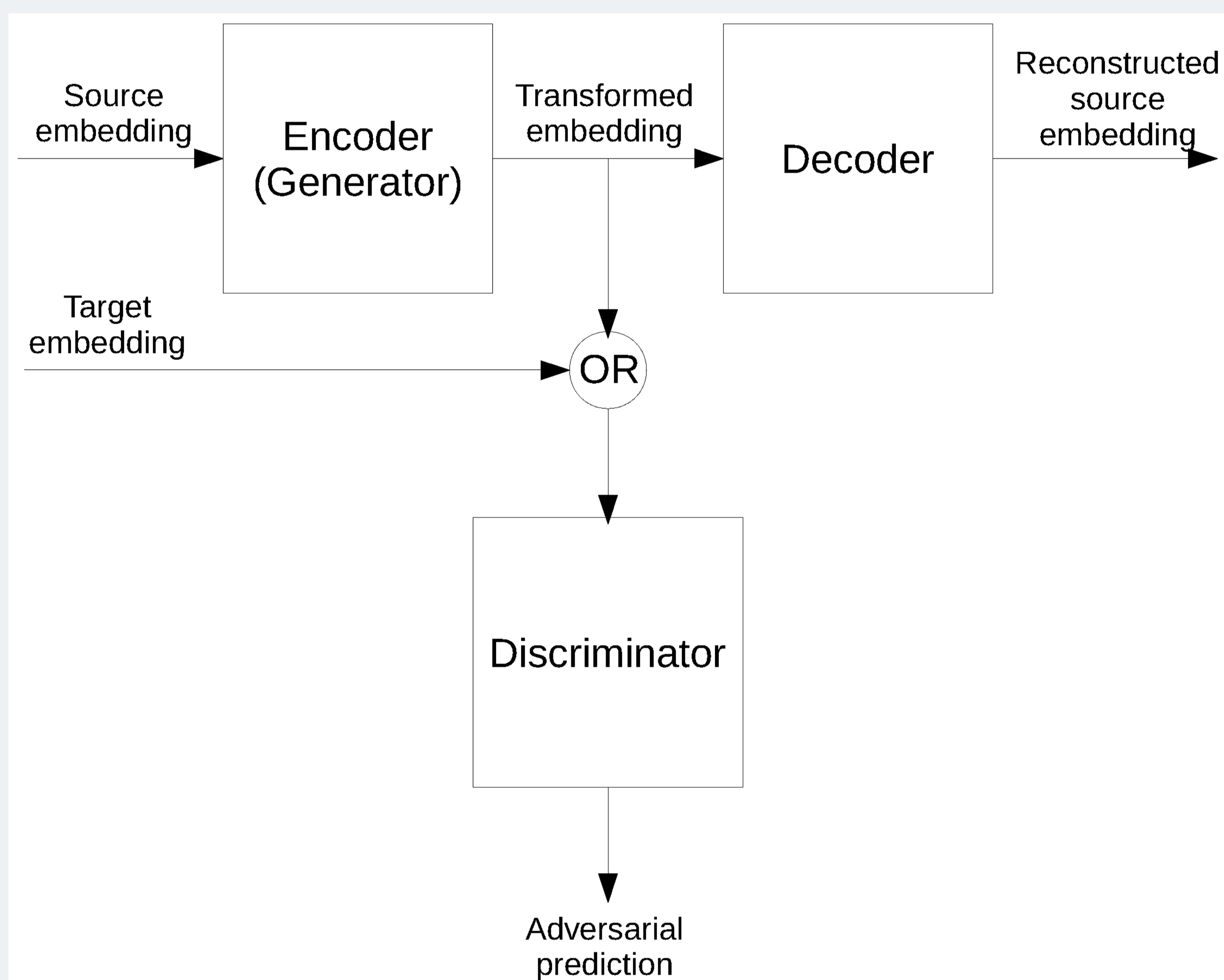
- Autoencoder: linear model with tied matrices
- Reconstruction loss: cosine dissimilarity
- Discriminator: Deep Residual Network (He et al., 2015) with leaky ReLU, batch normalization and dropout.
- Optimization: Adam (Kingma & Ba, 2014)

Links



Code: <https://github.com/Avmb/clweadv>

Adversarial Autoencoder (AAE)



- Map source embeddings to latent representations and back
- Force the latent and target distribution to match
- The discriminator tries to distinguish between true and mapped target embeddings
- The encoder and the decoder try to fool the discriminator and get good reconstructions
- Extends Makhzani et al., 2015

Experiments

- Qualitative evaluation
 - English and Italian word2vec (Mikolov et al., 2013) embeddings trained on Wikipedia
- Document classification (Klementiev et al., 2012)
 - English and German word2vec embeddings trained on Reuters + News Commentary corpora embeddings.
 - Train document classifier on German embeddings, then evaluate it on embeddings mapped from English

Qualitative results

- Correct mappings: 'computer' (en) → 'computer' (it); 'man' (en) → 'Mann' (de)
- Close mappings: 'rain' (en) → 'gelo', 'gela', 'intensissimo', 'galleggiava', 'rigidissimo', 'arida', 'semi-desertico', 'fortunale', 'gelata', 'piovosa' (it 10-best)
- Bad mappings: 'France' (en) → 'Radiomobile', 'Cartubi', 'Freniatria', 'UNUCI', 'Cornhole', 'Internazione', 'CSCE', 'Folklorica', 'UECI', 'Rientro' (it 10-best)

Document classification results

Small improvement on the smallest training set size.

Conclusions

- Adversarial training on point-mass distributions
 - Difficult, lots of tricks are needed (e.g. ResNet)
 - Standard GANs (Goodfellow et al. 2014) fail by mapping everything to a single embedding
 - The decoder preserves diversity
- Embedding transfer without parallel resources
 - Feasible to some extent, but not yet competitive to methods that use parallel resources
 - May be an issue of training difficulty or too strong language similarity hypothesis
 - The proposed method can be augmented with parallel resources and extended to sentence embeddings