KNOWLEDGE TRANSFER FOR UTTERANCE CLASSIFICATION IN LOW-RESOURCE LANGUAGES

Andrei Smirnov and Valentin Mendelev

CONTENTS

STC innovations

INTRODUCTION

MOTIVATION

- Labelled data: expensive to produce, limited in quantity
- Even more true for low-resource languages

GOAL

- Build a text classifier without labelled data in a target language

STC innovations

INTRODUCTION

DATA

- Target language – Kazakh, source language – Russian
- 6000 users’ requests, 40 classes
- Average request length 7.6 words

«I have a umm ... sort of ... non-technical question. Can I, like, suspend services during my vacation»

Service suspension

CONTENTS

APPROACH

Target utterance

↓ Word embeddings

Vector in a target language space

↓ Linear mapping

Vector in a source language space

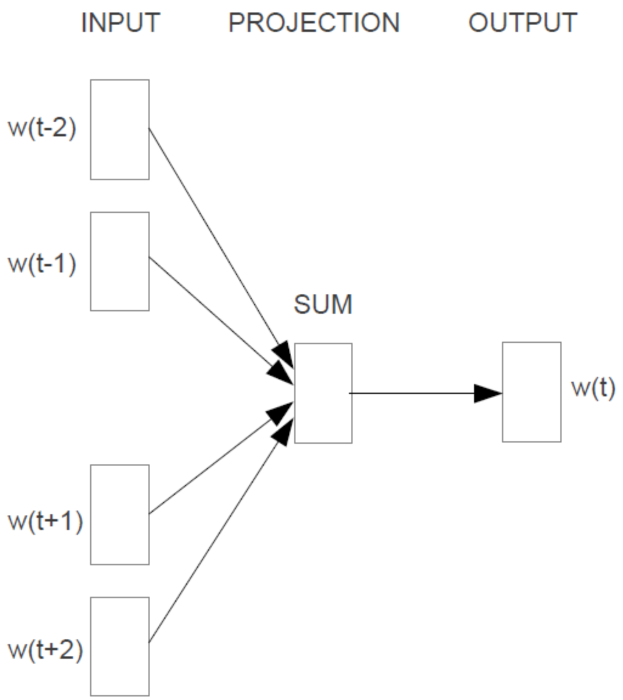
↓ Source language classifier

Predicted class

WORD EMBEDDINGS

DETAILS

- Training set for Russian: ~200m tokens
  (Conversations, books, news articles)
- Training set for Kazakh: ~30m tokens
  (Kazakh Wikipedia and news articles)
- Embeddings dimension is
  100 for Russian and 500 for Kazakh



CBOW

The CBOW architecture predicts the current word based on the context

Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013)

KNOWLEDGE TRANSFER

VECTOR SPACE TRANSFORMATION APPROACH*

- Translate a set of words or phrases from the target to the source language
- Train a linear model by minimizing L_2 distance

$$\min_A \sum_{i=0}^{N} ||Av_i^{tar} - v_i^{src}||^2$$

v_i^{tar} - embedding in the target space

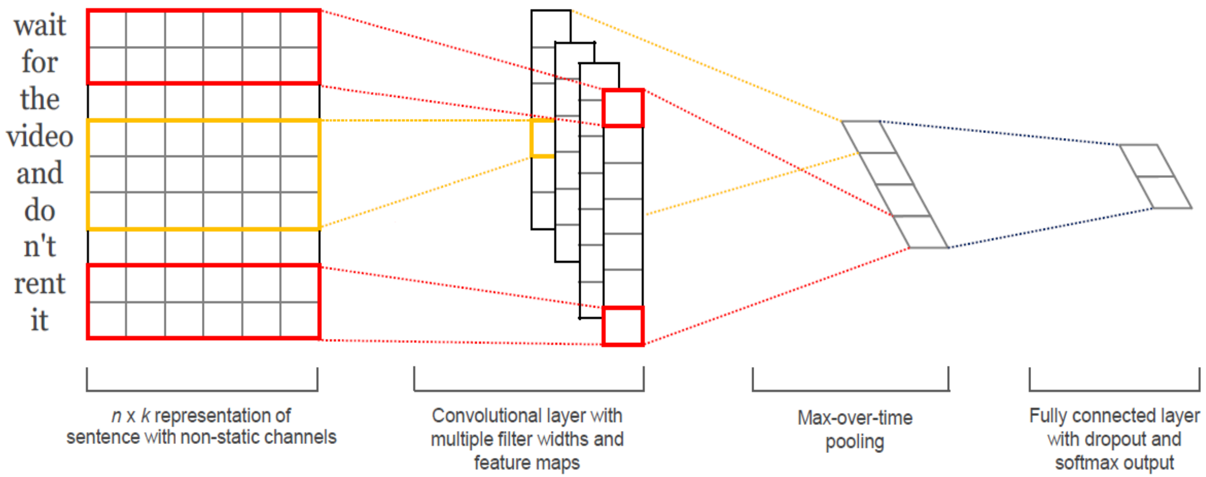
v_i^{src} - embedding in the source space

*Mikolov, Tomas, Quoc V. Le, and Ilya Sutskever. "Exploiting similarities among languages for machine translation." *arXiv preprint arXiv:1309.4168*(2013).

SOURCE LANGUAGE CLASSIFIER

DETAILS

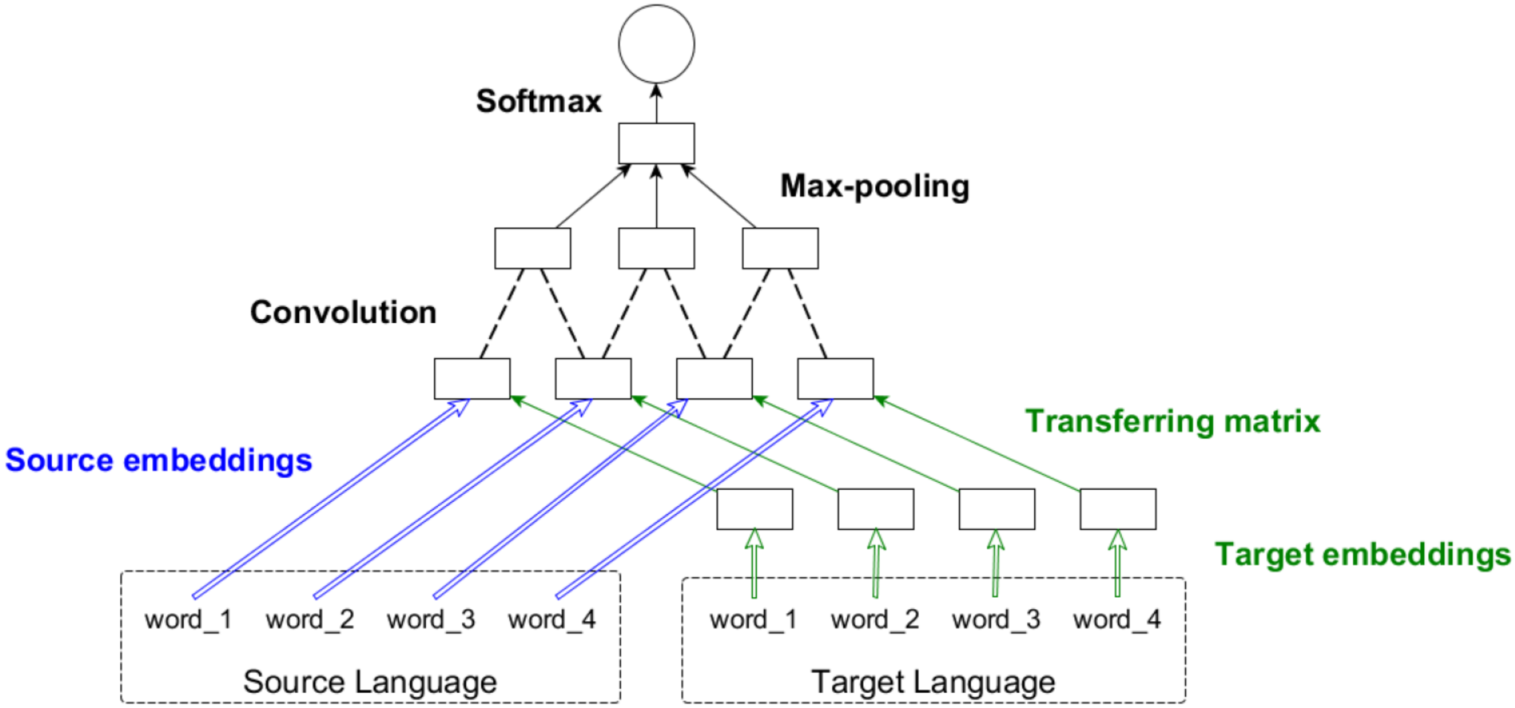
- 192 filters of filter size 2
- ReLU activation function
- Dropout rate 0.5



CNN architecture for an example sentence

Kim, Y.: Convolutional neural networks for sentence classification, EMNLP, 2014

APPROACH



Model architecture. Data flow for the source language is shown in blue, for the target language in green

STC innovations

CONTENTS

RESULTS

DATASETS

- 6000 requests, 40 classes;

  4200 requests, 10 classes; 1980 requests, 2 classes
- Train/Development/Test   80%/10%/10%

DATA FOR THE TRANSFERRING MATRIX

- 5000 most frequent words from the telecom-related text corpus
- Translated by Google Translate

DEPENDENCE ON THE TYPE OF TRANSLATION MECHANISM

# of classes	Manual	Google Translate	Transferring matrix (GT-dict)
40	84.3	61.4	37.6
10	89.1	71.9	51.8
2	97.8	95	84.4

DEPENDENCE ON THE DATA FOR THE TRANSFERRING MATRIX

Training data	**Accuracy**
GT-dict	51.8
GT-dict + Train≤*2* (415)	60.2
GT-dict + Train≤*4* (1440)	67.7
GT-dict + Train≤*∞* (3140)	73.3

Train≤*N* – all requests from train data consisting of less than *N* words

STC innovations

RESULTS

DEPENDENCE ON THE CLASSIFIER

# of classes	Manual			Transferring matrix (GT-dict)		
	k-NN	CNN	Δ	k-NN	CNN	Δ
40	72.8	84.3	**7.5**	32.8	37.6	**4.8**
10	78.4	89.1	**11.3**	46.7	51.8	**5.1**
2	90.5	97.8	**7.3**	76.4	84.4	**8.0**

CONTENTS

CONCLUSION

- Knowledge transfer allows to achieve reasonable classification accuracy for low-resource tasks
- To boost the performance add task-specific phrases to the training data for transferring matrix
- We want to do better

Thank you for your attention