

# T\_REX - ASR RESCORER-EXTENDER



PolEval 2020 Task 1 Submission

# Task description

## Task 1: Post-editing and rescoring of automatic speech recognition results

Input: **the output of an ASR system**

Given as:

- 1-best - the result of the original ASR
- N-best list - top-100 variants of the recognition
- Lattice - a graph containing all variants considered by ASR

Output: **(hopefully) a better transcription**

# Task description

- System can be adjusted to the original ASR
  - correct specific errors!
- *no punctuation, no capitalization and no digits, symbols or abbreviations*
- *Input data:*
  - Clarin-PL training set
    - WER 9.59%, lattice oracle WER 3.75%
  - Clarin-PL test set
    - WER 12.08%, lattice oracle WER 4.72%
  - Polish parliament corpus
    - WER 45.57%, lattice oracle **WER 30.71%**
- *The evaluation data was supposed to be quite similar to Polish Sejm Corpora*

Note: Quite big WER and oracle WER

## Data sample (hard example)

**ASR:** mamy gangów ruszy **wczoraj że wasz** sen **może liczyć** wielkiego akcji w naszego **klubu** rekord zasadę **ale także** d o w **słowach** ruszyliśmy garbarczyk **platformy obywatelskiej** każdego z fal

**REFERENCE:** mam pytanie do wnioskodawców otóż **wczoraj** dowiedzieliśmy się **że wasz** cenny projekt **może liczyć** nie tylko na poparcie waszego klubu tutaj kolegów z eselde **ale także** po tych ciepłych **słowach** które tutaj usłyszeliśmy na poparcie **platformy obywatelskiej** i polskiego stronnictwa ludowego

**Words in bold are recognized correctly**

# Data sample (easier example, ASR result)

przygotowany przez ministerstwo skarbu państwa

pragną przedstawi

stanowisko klubu palikota

ustawy dróg go dziewięćdziesiąt dwa

mamy do czynienia z rozsądne mi zmianami

w ustawie o działaniu tak zwanych w narodowych funduszach inwestycyjnych

począć zmiany nazwy

przywileje podatkowe

z tytułu i w ustawie z dziewięćdziesiątego drugiego roku ...

# Data sample (easier example, reference)

przygotowany przez ministerstwo skarbu państwa

pragnę przedstawić

stanowisko klubu ruchu palikota

apropo ustawy druk sto dziewięćdziesiąt dwa

mamy do czynienia z rozsądnymi zmianami

w ustawie o działaniu tak zwanych narodowych funduszków inwestycyjnych

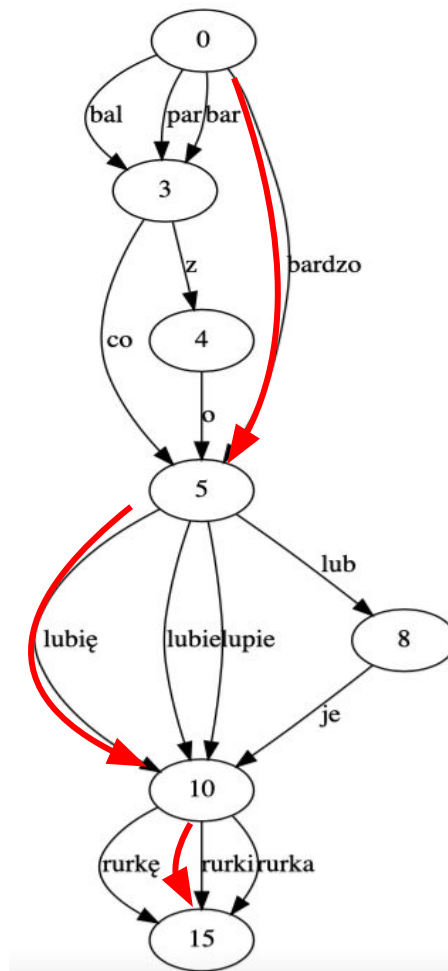
począwszy od zmiany nazwy

po przywileje podatkowe

z tytułu pit w ustawie z dziewięćdziesiątego drugiego roku...

# Kaldi lattice

- Graph
  - Edges - words
  - Nodes - costs (acoustic and language)
- Example lattice of sentence “Bardzo lubię rurki”



# Our approach

- Use lattices
- Replace original LM score by a score from a better language model
- Add some extra nodes and edges to the lattice (new words consisting of similar phonemes as the original words)
- Use beam-search decoder to find top-K best paths
- Rescore the most promising candidates using better (slower) LM



# Lattice expansion

- For each word from the lattice we find words which are close
  - We use BK-tree, Levenshtein distance and NKJP corpus
  - We increase search range for longer words and ignore very short ones

$x$  = length of phonemes form of a word

$$\text{search radius } (x) = \begin{cases} 0, & \text{if } x < 4 \\ 1, & \text{if } 4 \leq x < 8 \\ 2, & \text{if } x \geq 8 \end{cases}$$

- E.g. there are 180 words at distance 1 from “ala” and only 17 at distance smaller or equal to 4 from “rozentuzjzmowany”

# Lattice expansion - PageRank

1. Extract all the words
2. Create a directed graph
  - a. Nodes represents words
  - b. Nodes are connected by an edge if it is natural that they can appear in the same utterance
  - c. Edge between two words exists if results of formula is bigger than 1

$$\left[ v * \text{cosine\_distance}(\text{parent}, \text{child}) + b * \log_{10} \frac{\text{bi-gram}[\text{parent}, \text{child}]}{\text{uni-gram}[\text{parent}]} \right]^{-1}$$

3. Perform a PageRank algorithm on the resulting graph
4. Keep 15 best possible extensions

This procedure is loosely inspired by TextRank algorithm (for summarization)

# Lattice expansion - results

- Oracle WER on our evaluation set: 15.4%
- Oracle WER after lat. extension: 9.8% (30 times bigger than base)
- Oracle WER after lat. reduction: 11.3% (10 times bigger than base)

We used expanded-reduced lattice version

# Replacing Language Model

Every edge in the lattice has two costs: Acoustic cost and Graph (HMM) cost

Graph cost = State-transition-cost + LM Cost

We don't know the State-transition cost

To perform LM-rescoring we use the new LM log-likelihood in place of the Graph Cost.

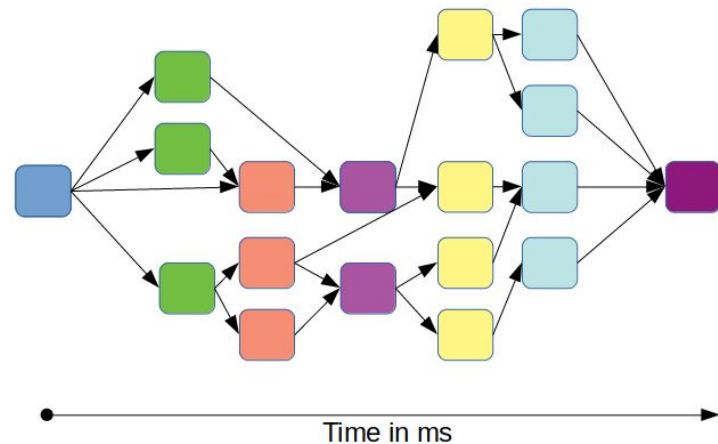
# Beam-search decoder

Time-aware beam search:

Each lattice node  $\mathbf{n}$  has an approximate time  
(time=number of ticks on the longest path from start to  $\mathbf{n}$ )

Path competes in the beam only with paths of similar length

Moreover, better GPU utilization by feeding neural network with extensions of words of similar durations



# Used Language Models

## 1. N-gram models

- trained on a large part of the Polish Parliamentary Corpus
- beams of size 5 000 and 10 000 with KenLM

## 2. Flair language model (Character-Level Bi-LSTM)

- pre-trained model pl-opus primarily based on the Wikipedia and OPUS corpus
- The model was fine-tuned on the Polish Parliamentary Corpus and its subcorpus of transcriptions provided in training data
- training objective == language modeling on separate forward and backward models

# Used Language Models (cont.)

## 3. Transformer model

- PolBert - Bert model trained on Polish corpus constructed from:
  - OpenSubtitles
  - ParaCrawl
  - Polish Wikipedia (February 2020)
  - Polish Parliamentary Corpus
- Used only for rescoring candidate utterances extracted from beam-search decoder
- Evaluation on 2 x Nvidia Tesla M40 24GB GPU



# Sample corrections

- tREX:

“(...) natomiast początkiem rozwiązywania **konfliktu** powinna być deklaracja **władz iranu** dotycząca **powstrzymywania** się od użycia **sił** oraz uwolnienia **uwięzionych** działaczy opozycji (...)”

- 1best:

“(...) natomiast początkiem rozwiązywania **konfliktów** powinna być deklaracja **właz iran** dotycząca **powstrzymywanie** się od użycia **siły** oraz uwolnienia **uwięziony** działaczy opozycji (...)”



# Results

System Name	Affiliation	WER %	CHG %
1best	PolEval 2020	27.6	0.0
KRS + spaces	UJ, AGH	25.9	3.6
Polbert	<a href="https://skok.ai">https://skok.ai</a>	26.9	2.1
BiLSTM-CRF edit ops tagger	Adam Mickiewicz University	24.7	6.2
base-4g-rr	Samsung R&D Institute Poland	27.7	2.0
PJA_CLARIN_20k	Polish-Japanese Academy of IT	31.8	9.9
MLM+bert_base_polish	-----	73.9	2.1
<b><i>t-REx_fbs</i></b>	<b><i>University of Wrocław</i></b>	<b>24.31</b>	<b>17.2</b>
<b>Out-Of-Competition</b>			
t-REx_fbs (with 2 from nbest)	University of Wrocław	23.93	16.6
<b><i>t-REx_bert (with 8 from nbest)</i></b>	<b><i>University of Wrocław</i></b>	<b>23.4</b>	<b>17.0</b>
t-REx_fbsx_50	University of Wrocław	25.09	23.3
t-REx_fbsx_150	University of Wrocław	23.67	19.1

# Conclusion

Fine tuned Flair LM + beam search gave UWr 1st place in PoEval 2020 Task 1

Lattice extensions and BERT rescoring give additional WER improvements

We aren't done yet, many idea left to try

Code and models available (shortly) at:

<https://github.com/adamjankaczmarek/poleval2020>