St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences

Development of Factored Language Models for Automatic Russian Speech Recognition

Irina Kipyatkova,
Alexey Karpov

30 May, 2015, Moscow, Russia

Introduction

- Russian is a morphologically rich inflective language. Words in Russian can inflect for a number of syntactic features:
  - case
  - number
  - gender
  - etc.
- This leads to a large number of possible word forms and consequent problems connected with sparseness of data.
- Rich morphology of Russian results in extremely large vocabulary.

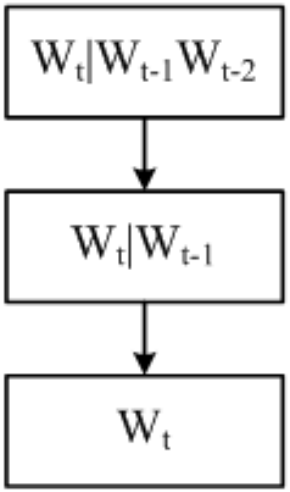
Factored Language Modeling

- Factored language models (FLM) for the first time was introduced in order to deal with morphologically rich Arabic language in [J. A. Bilmes. K. Kirchhoff. “Factored language models and generalized parallel backoff”. in Proc. Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 2, Stroudsburg, PA, USA, pp. 4–6, 2003.].
- This model incorporates various morphological features (factors). So, a word is viewed as a vector of k factors:

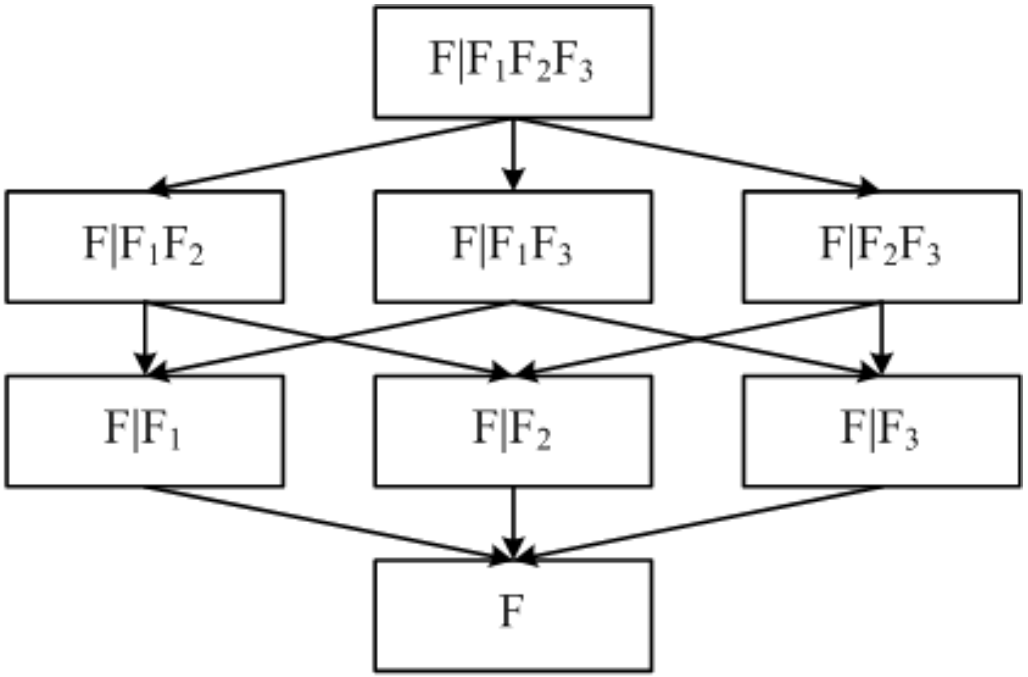
$$w_i = (f_i^1, f_i^2, .., f_i^k)$$

- Factors of a given word can be word, morphological classes, stems, roots, and other grammatical features.
- Probabilistic language model is constructed over sets of the factors.

n-gram and FLM Backoff Trees



Backoff path for a 3-gram language model over words



Backoff graph for with three parent variables F_1, F_2, F_3

In **[Vergyri, D., Kirchhoff, K., Duh, K., Stolcke, A.: Morphology-Based Language Modeling for Arabic Speech Recognition. In: Proceedings of ICSLP 2004, pp. 2245-2248 (2004)]**, it was shown that automatic method, which uses Genetic Algorithms (GAs) for optimization of the factor set, backoff path, and smoothing techniques, performs better than manual search in terms of perplexity.

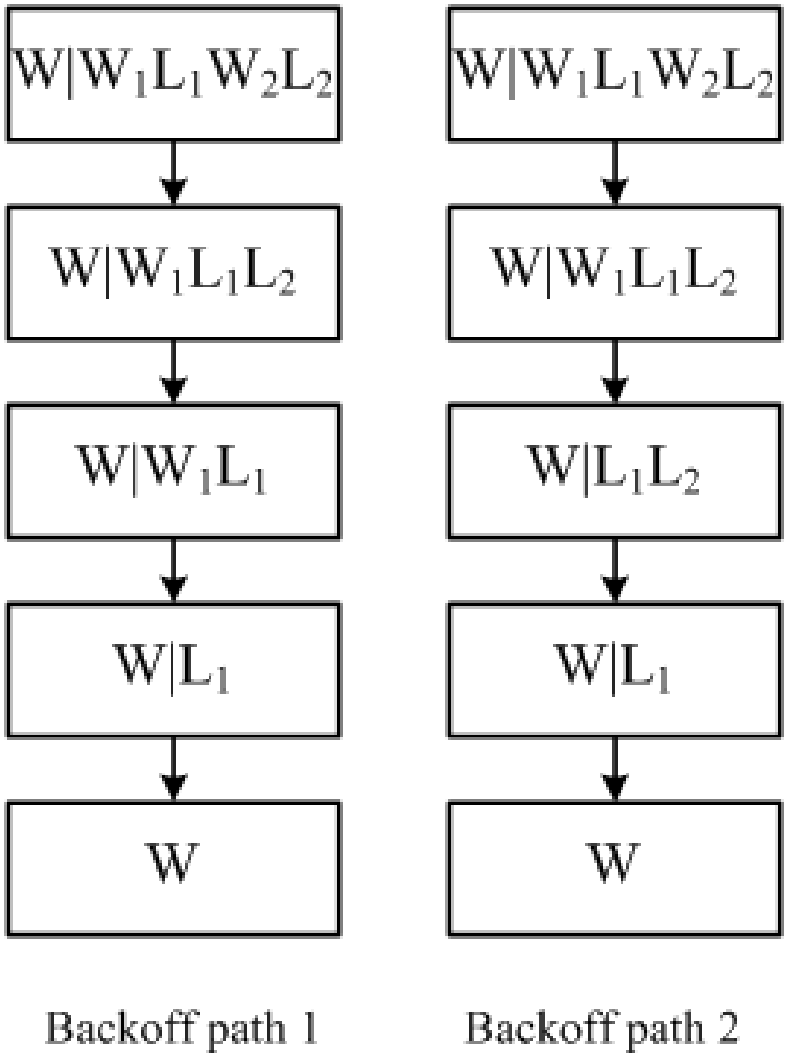
Morphological Analysis and Factor Selection

- “VisualSynan” software from the AOT project [A. Sokirko. “Morphological modules on the website www.aot.ru”. In Proc. “Dialogue-2004”, Protvino, Russia, pp. 559–564, 2004] was used for obtaining morphological word features.
- Factors that were used: the word (W), its lemma (L), stem (S), part-of-speech (P), and morphological tag (M).
- Every word in the text corpus was replaced with the vector, for example:
  - ‘схеме’ (“scheme”) -> {W-схеме: L-схема: S-схем: P-сущ: M-bc}

Fixed Backoff Paths

- 4 models with the word plus one of the other factors were created:
  - W+L
  - W+S
  - W+M
  - W+P

- 2 fixed backoff paths:
  1. The first drop was of the most distant word and factor, then – of the less distant ones.
  2. The first drop was of the words in time-distance order, then drop of the factors in the same order.



Backoff paths for a model W+L

Perplexity of Different FLMs with Different Backoff Paths

- Models were created with the help of the SRI Language Modeling Toolkit (SRILM).
- Vocabulary size is 150K words.
- Perplexity of the obtained models was calculated on text data consisting of phrases (33M words in total) from online newspaper "Фонтанка.ru" (www.fontanka.ru).
- Perplexity of the baseline 3-gram model is 553.

Factors	Good-Turing		Unmodified Kneser-Ney		Modified Kneser-Ney		Witten-Bell		Natural	
	BO path 1	BO path 2	BO path 1	BO path 2	BO path 1	BO path 2	BO path 1	BO path 2	BO path 1	BO path 2
W+M	573	696	593	724	566	691	749	898	761	916
W+L	557	597	550	603	**529**	577	826	1007	747	779
W+P	572	636	649	755	623	729	725	727	734	762
W+S	617	685	617	701	595	672	879	1098	824	895

Continuous Russian Speech Corpus

- **Characteristics of speech recordings:**
  - 22 hours of continuous speech and dialogs
  - The corpus recorded in an acoustic studio with 44.1 KHz, 16 bits. SNR≈35dB
  - A stereo pair of Oktava MK-012 cardioid microphones
  - Presonus Firepod sound board
- **Training corpus:**
  - 327 phonetically-balanced and meaningful phrases
  - 50 native speakers from St. Petersburg (male and female voices are fifty-fifty)
- **Test corpus:**
  - 100 continuously pronounced phrases from the on-line newspaper «Фонтанка.ru» (www.fontanka.ru).
  - 5 speakers

WER Obtained After Rescoring of 20-best List with FLMs with Different Backoff Paths

- The decoder Julius ver. 4.2 for speech recognition
- WER obtained with the **baseline 3-gram** LM was **26.54%**.
- The OOV rate for the test set was 1.1%.

Factors	Good-Turing		Unmodified Kneser-Ney		Modified Kneser-Ney		Witten-Bell		Natural	
	BO path 1	BO path 2	BO path 1	BO path 2	BO path 1	BO path 2	BO path 1	BO path 2	BO path 1	BO path 2
W+M	27.87	28.00	27.79	28.09	28.15	28.16	**27.30**	27.55	27.58	27.40
W+L	28.45	28.78	28.37	28.82	28.28	28.99	27.83	28.39	27.88	28.67
W+P	28.61	28.61	28.58	28.71	28.63	28.88	27.72	28.48	28.33	28.91
W+S	29.93	30.24	29.78	30.19	30.02	30.28	29.01	29.46	28.90	29.91

WER Obtained After Rescoring of 20-best List with FLMs Interpolated with 3-gram Model

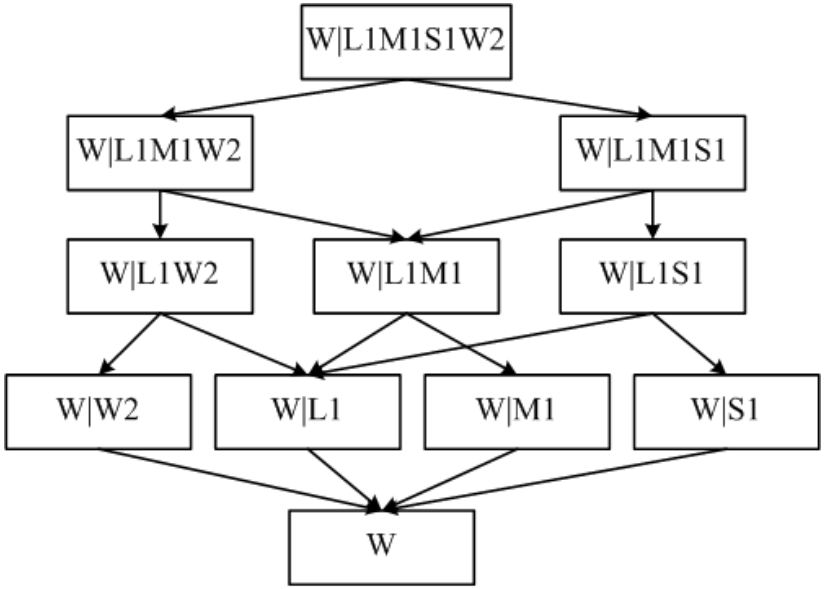
- WER obtained with the **baseline 3-gram** LM was **26.54%**.

Factors	Good-Turing		Unmodified Kneser-Ney		Modified Kneser-Ney		Witten-Bell		Natural	
	BO path 1	BO path 2	BO path 1	BO path 2	BO path 1	BO path 2	BO path 1	BO path 2	BO path 1	BO path 2
3gram & W+M	25.00	24.89	24.57	24.93	**24.44**	24.78	24.94	25.22	25.41	25.36
3gram & W+L	25.51	25.54	25.54	25.67	25.58	25.43	25.51	25.47	24.98	25.47
3gram & W+P	25.21	25.28	25.30	25.32	25.07	25.24	25.47	25.64	25.60	25.43
3gram & W+S	25.97	25.92	26.03	25.86	25.88	25.90	26.05	25.86	25.49	25.90

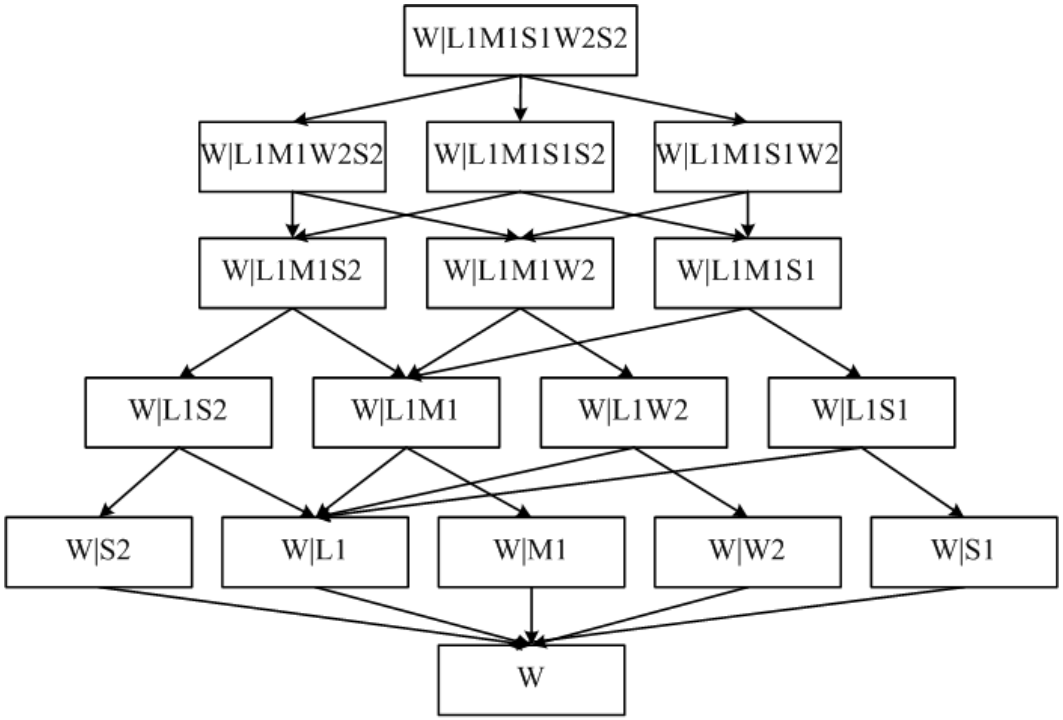
Creation FLM with parallel generalized backoff

- Since the creation of FLM with parallel generalized backoff requires a large amount of memory, we used a part of the text corpus, which contains 100M words.
- We apply the GA to find the best backoff graph using 5 factors: word, lemma, stem, part-of-speech, and morphological tag; time context of 2; and the following discounting methods: Good-Turing, Unmodified Kneser-Ney, Modified Kneser-Ney, Witten-Bell, and Natural.
- GA was implemented using population size of 10 and the maximum number of generation equal to 20.

Backoff paths for two best models



Backoff graph for FLM 1



Backoff graph for FLM 2

- Both best models used four factors: lemma, morphological tag, stem, word,
- and three discounting methods on different stages of backing-off: Unmodified Kneser-Ney, Modified Kneser-Ney, and Witten-Bell.

WER Obtained After Rescoring of 20-best List

- Oracle WER for this N-best list was 16.63%

Language model	WER, %
Baseline 3-gram	26.54
FLM 1	27.94
FLM 2	28.56
FLM 1 & 3-gram	**24.53**
FLM 2 & 3-gram	24.74

Conclusion

- Rich morphology of Russian complicates the creation of language models. FLMs can help to include additional information in language model and thereby to improve Russian speech recognition system.
- We have investigated an application of FLM for N-best lists rescoring for Russian speech recognition. We made a comparison of influence of different factor sets on speech recognition results. We obtained relative WER reduction of 8% comparing to the baseline system.
- In further research we plan to investigate FLMs with other factors, for example, word ending, part of sentence, and other linguistic factors.

Thank you for your attention!