Inflectional Language Modeling with Random Forests for ASR

Ilya Oparin
Dept. of Computer Science and Ingineering, University of West Bohemia,
Plzeň, Czech Republic,
Speech Technology Center,
St.Petersburg, Russia
ilya@speechpro.com

Abstract

In this paper we show that the Random Forest (RF) approach can be successfully implemented for language modeling of an inflectional language for Automatic Speech Recognition (ASR) tasks. While Decision Trees (DTs) perform worse than a conventional trigram language model (LM), RFs outperform the latter. WER (up to 3.4% relative) and perplexity (10%) reduction over the trigram model can be gained with morphological RFs. Further improvement is obtained after interpolation of DT and RF LMs with the trigram one (up to 15.6% perplexity and 4.8% WER relative reduction).

1 Introduction

In this paper we study the application of the decision tree (DT) and random forest (RF) approaches to language modeling of Czech as an inflectional language. The DT mechanism for estimating probabilities of words following each other has long been known (L.R. Bahl et al., 1989) as an alternative to the N-gram approach. The DTs suffer from training data fragmentation and absence of theoretically founded grow-stopping criteria (P. Xu and F. Jelinek, 2004). However, with the recent advances in language modeling that extended the use of decision trees to that of random forests, this direction of research was brought back to the spotlight (P. Xu and F. Jelinek, 2004).

A random forest is a collection of DTs that include randomization in the tree-growing algorithm. The underlying assumption is that while one DT does not generalize well to unseen data, a set of randomized DTs might perform better. Greedy algorithms are used at the stage of DT construction for choosing best questions to split data. As a result, trees are only locally but not globally optimal (with respect to training data). Randomized DTs are not locally optimal, but the collection of them may be closer to a global optimum and thus provide better results.

The process of DT construction is fully unsupervised and basically follow the framework introduced in (P. Xu and F. Jelinek, 2004).

2 Morphological Decision Trees

2.1 Morphological Features as Predictors

In word-based DTs questions like "Does the previous word belong to the set of words $\{ w_b, w_f, w_q, \ldots \}$?" are asked at each node. In morphological DTs we want to ask questions about morphological features of word predictors. We expect it to be particularly useful for morphologically-rich languages as Czech and Russian. In this study morphological feature types are wordform itself (W); word lemma, i.e. initial form of the word (L); word stem (S); part-of-speech - POS (P); full morphological tag (T) and inflection (I). Thus, the questions may be in the form "Is the full morphological tag of the predictor *animate singular noun in accusative case*?".

2.2 Results

2.2.1 Perplexity

The recognition of spoken lectures held in Czech is our target task. The transcriptions of three lectures on different subjects in the domain of information technology were chosen as the testing data (IRP, ISS, MUL). The setup is described in details in (I. Oparin et al., 2008).

First we evaluate the performance of different LMs with perplexity. Since our training data is very close in topics and style to the testing data, the results give insight in real performance of the models even though the size of the training data is small. Table 1 represents perplexities for stand-alone models on three different testing lectures.

Model	IRP	ISS	MUL
Trigram	317	212	258
Word DT	433	253	336
Morph DT	413	252	320
Word RF	360	221	280
Morph RF	298	190	237

Table 1: Perplexity for the stand-alone models.

Model	IRP	ISS	MUL
Trigram	317	212	258
Word DT	302	198	245
Morph DT	296	197	240
Word RF	292	191	234
Morph RF	272	179	220

Table 2: Perplexity for the interpolated models.

Model	IRP	ISS	MUL
1-best	63.1	70.2	58.3
Trigram	63.8	70.9	59.2
Word DT	63.8	70.7	59.1
Morph DT	64.1	69.7	59.1
Word RF	64.2	70.9	59.2
Morph RF	64.7	71.9	59.7

Table 3: Accuracy for the stand-alone models.

Model	IRP	ISS	MUL
Word DT	64.5	71.5	59.6
Morph DT	64.5	71.3	59.8
Word RF	64.5	71.6	59.8
Morph RF	64.8	72.3	60.1

Table 4: Accuracy for the interpolated models.

We can see that individual DTs perform worse than the standard interpolated Kneser-Ney trigram model (trained with SRILM toolkit). Word RF does not show steady perplexity improvement on its own but rather performs in the same way as the trigram model. Little improvement of 2.6% for ISS data can not be considered noteworthy for perplexity experiments. This can be explained by the fact that in our framework we do not make use of any smoothing and backoff techniques that are known successful for language modeling. However, with morphological RFs we achieve a noteworthy improvement of perplexity over 10%.

Perplexity results after the interpolation of the trigram model with different DT-based ones are presented in Table 2. All DT-based models show steady perplexity improvement in interpolation with the trigram model. The best result (15.5% relative perplexity improvement) is again obtained with the morphological RFs.

2.2.2 Word Accuracy Estimation

Word recognition accuracy for different stand-alone models is shown in Table 3. Large bigram LM is used to generate 500-best lists that are subsequently rescored with DT and RF models in the second pass. Row *1-best* corresponds to the 1-best accuracy for 500-best lists without any rescoring. Trigram LM is taken as the baseline. Following the results represented in Table 1, individual DTs do not directly improve the accuracy. However, both morphological and word RFs do. Table 4 shows results for the DT-based models after the interpolation with the trigram one. The difference with the perplexity results presented in Table 2 is mostly in the lower improvement of the results with the interpolation of RFs.

3 Conclusions

In this paper we studied language modeling of an inflectional language with decision trees and random forests for recognition of spoken lectures. Both approaches were tested with taking pure lexical and morphological information into account. Our experiments proved that decision trees do not outperform a classical trigram model. The perplexity and WER improvement is possible only with the interpolation of DT models with a trigram one. RFs, on the contrary, directly improve the baseline. We got even larger improvement with the interpolation of RFs with the conventional trigram model. The best results are always gained with the morphological RFs.

References

I. Oparin and O. Glembek and L. Burget and J. Černocký. 2008. Morphological Random Forests for Language Modeling of Inflectional Languages. *Submitted to SLT'08*, Goa, India.

L.R. Bahl, P.F. Brown, P.V. de Souza and R.L. Mercer. 1989. A Tree-Based Statistical Language Model for Natural Language Speech recognition. *Computer, Speech and Language*, vol. 37, pp. 1001-1008.

P. Xu and F. Jelinek. 2004. Random Forests in Language Modeling. *Proceedings of EMNLP'04*, Barcelona, Spain, pp. 325-332.