Tübingen-Oslo system at SIGMORPHON shared task on morphological inflection. A multi-tasking multilingual sequence to sequence model.

Taraka Rama
Department of Informatics
University of Oslo
Oslo, Norway
`tarakark@ifi.uio.no`

Çağrı Çöltekin
Department of Linguistics
University of Tübingen
Tübingen, Germany
`ccoltekin@sfs.uni-tuebingen.de`

Abstract

In this paper, we describe our three submissions to the inflection track of SIGMORPHON shared task. We experimented with three models: namely, sequence to sequence model (popularly known as seq2seq), seq2seq model with data augmentation, and a multilingual multi-tasking seq2seq model that is multilingual in nature. Our results with the multilingual model are below the baseline in the case of both high and medium datasets.

1 Introduction

Morphological inflection is the task of predicting the target inflected form from a lemma and a bundle of inflectional features. For instance, given the Norwegian lemma *hus* "house" and the morphological features N, DEF, PL the task is to predict *husene* "houses".

The SIGMORPHON shared task for 2018 (Cotterell et al., 2018) provided three data scenarios consisting of high (10000), medium (1000), and low (100) examples. This paper described the three systems that we submitted to the inflection track in the SIGMORPHON shared task. All our models are based on encoder-decoder model introduced by Faruqui et al. (2016) for the morphological inflection task. We trained our models on all the data sizes and tested on the test datasets provided by the organizers.

2 Background

The morphological (re)inflection task has been studied mainly in last two SIGMORPHON shared tasks (Cotterell et al., 2016, 2017). Most of the morphological inflection models are variants of sequence to sequence models applied by Faruqui et al. (2016) to morphological reinflection.

The input to the model is the source word prepended with relevant morphological tags, the output of the model is the target word for the inflection task. For re-inflection task, the input includes the target tags as well. The success of the system seems to depend highly on 'training data enhancement'. For different tracks (with different restrictions on data used) of the 2016 shared task, Kann and Schütze (2016) developed new techniques to increase the number of training instances. The methods used mostly work well for re-inflection task, since the re-inflection task is symmetric, and one can invert the source and target forms. In the subsequent year's shared task for 2017 (Cotterell et al., 2017), multiple authors explored new data enhancement techniques (Kann and Schütze, 2017; Bergmanis et al., 2017; Silfverberg et al., 2017) to improve the performance of the seq2seq models in medium and low resource scenarios. The work presented in this paper is based on the work of the simple encoder-decoder system of Faruqui et al. (2016).

3 Models

In this section, we describe the three different models and the feature representations used in our experiments.

Morphological features In this paper, we enumerated all the possible features in Unimorph (Kirov et al., 2018) and encoded the feature bundle as multi-hot feature vector. We experimented with both one-hot feature vectors and multi-hot feature vectors. In our development experiments, we found that multi-hot feature vectors have lower dimension than one-hot feature vectors and yielded similar results.

Seq2seq-baseline This model consists of two parts: bidirectional encoder and decoder. In this model, each character is represented as a one-hot vector whereas the morphological features are rep-

resented as multi-hot feature bundle. The encoder consists of LSTM cells that transform a sequence into a continuous vector. The final time step's hidden state and the cell state are used to initialize the decoder LSTM network. The decoder LSTM network predicts a character at each time step by passing the output of the decoder LSTM through a softmax layer. The output of the softmax layer is a predicted character that is input along with the multi-hot morphological feature vector to the next timestep. We intended this model to be the baseline model in our experiments.

Augment-Seq2seq   This model is a variation of the baseline encoder-decoder model where the training data is augmented with random strings generated with weights proportional to the character probabilities. This model is similar to the data augmentation model of Silfverberg et al. (2017) who generate new training instances by randomly sampling characters from unigram distributions. In our model, we generate a training instance of the same length as the original training instance. We also experimented with

Seq2seq-MTL-global   In this model, we train a single encoder-decoder model which is trained to perform both language identification and language modeling as auxiliary tasks apart from generating the target inflection. The encoder LSTM is trained to predict the next character in the source word at each time step. The final hidden state of the encoder is trained to predict the language of the example. This model differs from the other seq2seq models in that the model is multilingual (or global) and attempts to predict target inflections for all the languages in the test dataset. The seq2seq-mtl-global model is similar to the model of Kann et al. (2018) and Bergmanis et al. (2017) who train their attention enhanced encoder-decoder model using an auxiliary autoencoder objective. In contrast, our model uses both prediction of subsequent character and language prediction as auxiliary tasks.

3.1   Experimental settings

We trained our models at all the three resource settings: high, medium, and low. In all our experiments, the maximum length of both source and target strings are fixed to 30 and padded with zeroes at the end. Both the encoder and decoder LSTM units consisted of 256 hidden units. All the models were trained with Adam (Kingma and Ba, 2014)

with minibatches of size 32 or 128 depending on the size of the data; and, used a early-stop with a patience of 5 to prevent overfitting.

4   Results

Participating in the competition with less than three weeks at hand, we did not have much time to explore the hyperparameter settings required to tune our models. In our development experiments, we found that the baseline seq2seq model performed the best among the tested models. We observed similar results with the test dataset also. We present the average accuracies of all the models at high and medium datasets in table 1. Our results are lower than the baseline system. We also present the top-5 and the bottom-5 languages' accuracies of the three models on high and medium data sizes in table 2. We did not present the results for low sized datasets since all the models had accuracies lower than 5%. Both the seq2seq and augmented-seq2seq systems performed the worst on languages such as Zulu, Swahili, and Basque. On the other hand, the MTL system seemed to perform worse on the languages that have close orthography and substantial amount of borrowing such as Hindi, Urdu, and Persian.

Data size	Seq2seq	Augment-seq2seq	Seq2seq-MTL-global
High	63.048	56.598	49.521
Medium	30.979	29.722	20.973

Table 1: Average accuracies for high and medium datasets with three different models.

5   Conclusion

In conclusion, our global multi-tasking model requires more effort to improve the results for languages with low accuracies. As part of future work, we plan to work on incorporating embeddings and attention which are part of the winning systems from the shared tasks of 2016 and 2017. We observed that the multi-tasking model's auxiliary objective was easier to achieve than the main objective. Therefore, we need to explore ways to regularize the network, for instance, by weighing the individual loss components. Finally, the output softmax layer of the decoder has to be made sensitive to the language of the example in the training data to prevent softmax from yielding low values due to the high dimension of the target of the softmax.

	Seq2seq		Augment-Seq2seq		Seq2seq-MTL-global	
High	Adyghe	97.4	Crimean-Tatar	92	Crimean-Tatar	95
	Bashkir	97.1	Bashkir	90.9	Kabardian	94
	Crimean-Tatar	97	Friulian	89	Classical-Syriac	94
	Tatar	97	Adyghe	88.2	Tatar	93
	Yiddish	94	Azeri	88	Bashkir	91.6
	Navajo	12.4	Albanian	10.2	Swahili	4
	Khaling	8.5	Quechua	9.7	Persian	3.7
	Basque	6.2	Basque	9.6	Hindi	2.5
	Swahili	5	Swahili	5	Zulu	2.4
	Zulu	2.7	Zulu	2.5	Arabic	0.1
Medium	Kabardian	81	Turkmen	82	Classical-Syriac	90
	Haida	78	Kabardian	79	Kabardian	75
	Neapolitan	78	Classical-Syriac	71	Turkmen	66
	Kashubian	74	Friulian	66	Khakas	66
	Greenlandic	70	Neapolitan	65	Uzbek	66
	Basque	1.4	Quechua	1.9	Swahili	1
	Zulu	1.2	Russian	1.8	Hindi	0.8
	Swahili	1	Italian	1.5	Navajo	0.8
	Finnish	0.2	Zulu	1.5	Zulu	0.8
	Kazakh	0	Finnish	1.2	Arabic	0

Table 2: Top-5 and bottom-5 languages at which the three models perform the best and worse for high and medium datasets.

References

Toms Bergmanis, Katharina Kann, Hinrich Schütze, and Sharon Goldwater. 2017. Training data augmentation for low-resource morphological inflection. In *Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*, pages 31–39, Vancouver. Association for Computational Linguistics.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Arya D. McCarthy, Katharina Kann, Sebastian Mielke, Garrett Nicolai, Miikka Silfverberg, David Yarowsky, Jason Eisner, and Mans Hulden. 2018. The CoNLL–SIGMORPHON 2018 shared task: Universal morphological reinflection. In *Proceedings of the CoNLL–SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection*, Brussels, Belgium. Association for Computational Linguistics.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sandra Kübler, David Yarowsky, Jason Eisner, and Mans Hulden. 2017. Conll-sigmorphon 2017 shared task: Universal morphological reinflection in 52 languages. In *Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*, pages 1–30, Vancouver. Association for Computational Linguistics.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, David Yarowsky, Jason Eisner, and Mans Hulden. 2016. The SIGMORPHON 2016 shared task—morphological reinflection. In *Proceedings of the 14th Annual SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 20–22. Association for Computational Linguistics.

Manaal Faruqui, Yulia Tsvetkov, Graham Neubig, and Chris Dyer. 2016. Morphological inflection generation using character sequence to sequence learning. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*,

pages 634–643, San Diego, California. Association for Computational Linguistics.

Katharina Kann, Johannes Bjerva, Isabelle Augenstein, Barbara Plank, and Anders Søgaard. 2018. Character-level supervision for low-resource pos tagging. In *Proceedings of the Workshop on Deep Learning Approaches for Low-Resource NLP*, pages 1–11. Association for Computational Linguistics.

Katharina Kann and Hinrich Schütze. 2017. The lmu system for the conll-sigmorphon 2017 shared task on universal morphological reinflection. In *Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*, pages 40–48, Vancouver. Association for Computational Linguistics.

Katharina Kann and Hinrich Schütze. 2016. MED: The LMU system for the SIGMORPHON 2016 shared task on morphological reinflection. In *Proceedings of the 14th Annual SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 62–70. Association for Computational Linguistics.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Christo Kirov, Ryan Cotterell, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sebastian Mielke, Arya D. McCarthy, Sandra Kübler, David Yarowsky, Jason Eisner, and Mans Hulden. 2018. UniMorph 2.0: Universal Morphology. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Miikka Silfverberg, Adam Wiemerslage, Ling Liu, and Lingshuang Jack Mao. 2017. Data augmentation for morphological reinflection. In *Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*, pages 90–99, Vancouver. Association for Computational Linguistics.