Baltic J. Modern Computing, Vol. 0 (2016), No. 0, pp. 00–00

Neural Machine Translation using Bitmap Fonts

David Aldón Mínguez, Marta R. Costa-jussà, José A. R. Fonollosa

Universitat Politècnica de Catalunya, Barcelona

david.aldon@est.fib.upc.edu, {marta.ruiz,jose.fonollosa}@upc.edu

Abstract. Recently, translation systems based on neural networks are starting to compete with systems based on phrases. The systems which are based on neural networks use vectorial representations of words. However, one of the biggest challenges that machine translation still faces, is dealing with large vocabularies and morphologically rich languages. This work aims to adapt a neural machine translation system to translate from Chinese to Spanish, using as input different types of granularity: words, characters, bitmap fonts of Chinese characters or words. The fact of performing the interpretation of every character or word as a bitmap font allows for obtaining more informed vectorial representations. Best results are obtained when using the information of the word bitmap font.

1 Introduction

Deep learning (or neural networks) allows to solve problems that require the processing of big amounts of data[1]. Neural networks try to simulate a common feature of human beings, the accumulated experience. In brief, neural networks are not more than a simplified and artificial model that try to emulate the features of a human brain, which mostly consist of:

1. Processing units that exchange data or information.
2. Use them to recognize patterns, including bitmap fonts, manuscript and time sequences.
3. Have the ability of learning and improving its operation mode.

[1] http://blog.cit.upc.edu/?p=986

Machine translation (MT) can be defined a set of algorithms that aims at transforming a source language to a target language. Since the decade of the 90s, statistical translation systems, among which phrase-based systems, have prevailed over others. These statistical translation systems maximize the probability of target phrases given source phrases.

Recently, systems based on neural networks, which use vector representations of words, have began to have a great relevance [Kalchbrenner and Blunsom2013, Cho et al.2014, Sutskever et al.2014]. The big obstacle that these systems arise is the inability to deal with large vocabularies. This problem originates because of the architecture of these systems, not to mention their computational cost. In this work, we are introducing the bitmap font information of the input unit (either word or character) in order to provide the neural MT system with an informed initialization instead of random [Bahdanau et al.2015].

The rest of the paper is structured as follows. Section 2 reviews the related work, both in the specific task of Chinese-Spanish and in neural MT. Section 3 explains the baseline neural MT system, and our contribution of integrating translation unit (words or characters) bitmap fonts. Section 4 describes the experimental framework and the results obtained. Finally, Section 5 concludes.

2 Related Work

This section does a brief overview of works that approach Chinese-Spanish translation and previous works in neural MT.

2.1 Chinese to Spanish

Surprisingly, there are not many publications in the field of automatic translation between the pair Chinese-Spanish despite being two of the most spoken languages in the world, occupying the first position and the third respectively [Costa-jussà2015].

A work that was done was the creation of a pseudo corpus which is intended to translate English to Chinese or to Spanish and create an artificial corpus for the association of Spanish-Chinese [Banchs et al.2006]. The problem was tried to resolve Chinese-Spanish with Spanish-English and English-Chinese corpora. As reference system, they use a system based on n-grams that differs from the phrases mainly in the translation model.

In 2008, a IWSLT[2] (International Workshop on Spoken Language Translation) evaluation between these two languages was performed. There were two tasks for Spanish-Chinese. The first task was based on a direct translation for Spanish-Chinese. The second task was motivated by the fact that there is little corpus between Spanish-Chinese, but many among the Chinese-English and English-Spanish, so the task proposed consisted in translating from Chinese-Spanish by pivoting through English. As a final result, the second task, the pivot technique, performed better than direct translation because of the larger corpus provided. [Costa-jussà et al.2012] show a comparison between two types of standard pivots (pseudo corpus and cascade) using English and the direct system. These results show that the pivot and direct techniques do not differ much in their results, but that the technical pivot cascade is slightly better than the pseudo corpus.

Differently, [Costa-jussà and Centelles2016] presents the first rule-based MT system for Chinese to Spanish. Authors describe a hybrid method for constructing this system taking advantage of available resources such as parallel corpora that are used to extract dictionaries and lexical and structural transfer rules.

Additionally, to all this research, there is the Chispa Android application and web service[3], that can be useful to tourists or traveling between countries, which use these languages [Centelles et al.2014].

2.2 Neural Machine Translation

Text translation via deep learning relies on an autoencoder structure [Bahdanau et al.2015] to translate from a source to a target language. The autoencoder is trained using translated texts. Source words are mapped to a small space. The new representation of words is encoded in a summary vector (a representation of source sentence) using a recurrent neural network. Then, the summary vector is decoded into the target language. From 2013, there were different groups proposing competitive architectures that have progressed towards this new approach of neural MT [Kalchbrenner and Blunsom2013, Sutskever et al.2014]. And in 2015, [Bahdanau et al.2015] proposed to use gated recursive units (i.e. attention-based mechanism) that allows a better performance on long sentences. This same attention-based mechanism is also used to describe the content of bitmap fonts [Xu et al.2015]. [Lamb and Xie2015] propose a general Convo-

[2] http://iwslt2010.fbk.eu

[3] http://www.chispa.me

lutional Neural Network (CNN) encoder model for MT that fits within in the framework of encoder-decoder models.

3 Theoretical description of the system

In this section, we introduce the neural MT baseline system together with the main technique that we propose to enhance it.

3.1 Baseline System

As a baseline system, see Figure 1, we use the neural network system proposed by [Bahdanau et al.2015] which is mainly based on a encoder-decoder with the attention-based mechanism. For simplification, the Figure does not show the attention-based mechanism.

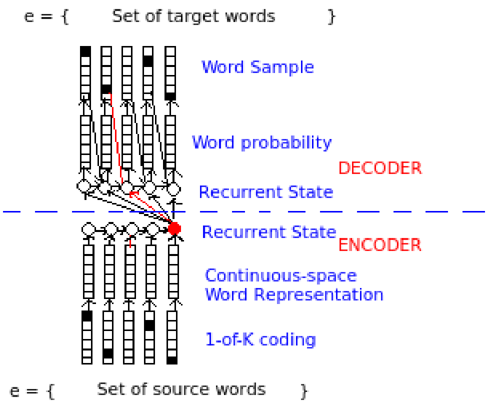
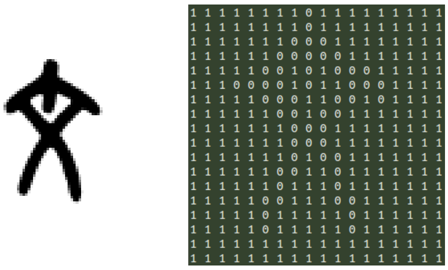


Fig. 1. Encoder/decoder neural MT scheme.

The encoder takes a source sentence, and encodes each word in 1-of-K coding vectors, then trains word embeddings which will be codified into a summary vector through the recurrent network with attention. Then, the decoder applies the reverse process obtaining the destination sentence.

3.2 Adding Word Bitmap fonts

In this study, the previous system will be enhanced to be able to use word bitmap fonts to add further information to the neural system. Given that the baseline representations of vectors are random bits 0/1 level (with the 1-of-K coding), we propose to create a more informative representation. We represent Chinese words by means of 2-dimensional bitmap which reflects the shape of the written word characters. See Figure 2 for illustration.



```
1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1
1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1
1 1 1 1 1 0 0 1 0 1 0 0 0 1 1 1 1 1
1 1 1 0 0 0 0 1 0 1 1 0 0 0 1 1 1 1
1 1 1 1 1 0 0 0 1 1 0 0 1 0 1 1 1 1
1 1 1 1 1 1 0 0 1 0 0 1 1 1 1 1 1 1
1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 0 1 0 0 1 1 1 1 1 1 1
1 1 1 1 1 1 0 0 1 1 0 1 1 1 1 1 1 1
1 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1
1 1 1 1 1 0 0 1 1 1 0 0 1 1 1 1 1 1
1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 1
1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

Fig. 2. Chinese Word Representation

Like this, we are converting Chinese words to bitmap fonts. Then we can get the vector of bits representing the bitmap fonts obtained from each word. In this way it is not only providing more information to the system, but it is contributing with much smarter information than a set of random values, due to the characteristic that offer the neural networks learning patterns. This new bitmap font vector becomes the initialization of word embeddings used in the encoder. See the integration of this new encoding in the system in Figure 3.

4 Experiments and Results

In this work, we use the Chinese-Spanish parallel corpus, United Nations Corpus (UN) [Rafalovitch and Dale2009]. Corpus statistics are shown in Table 1. Statistics for Chinese are shown both with word and character segmentations. In the case of word segmentation, the size of the vocabulary is similar to the target vocabulary, while in the case of using Chinese characters, we have a much lower vocabulary.

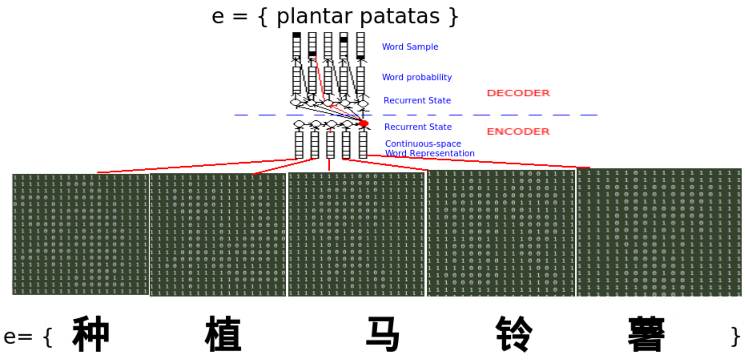


Fig. 3. New Encoder-Decoder

L	Set	S	W	V
ES	Train	58.6K	2.3M	22.5K
	Dev	990	43.4K	5.4k
	Test	1K	44.2K	5.5K
ZH Words	Train	58.6K	1.6M	17.8K
	Dev	990	33K	3.7K
	Test	1K	33.7K	3.8K
ZH Characters	Train	58.6K	2.8M	3.8K
	Dev	990	53.9K	1.7K
	Test	1K	55.1K	1.7K

Table 1. Corpus details. Number of sentences (S),words (W), vocabulary (V). M stands for millions and K stands for thousands.

The neural-based system was built using the software available in github [4]. We generally used settings from previous work: networks have an embedding of 529 and a dimension of 1024. We used a vocabulary size of 20000 in Spanish, 3500 for Chinese when using characters and 15000 for Chinese when using words.

Given that we are experimenting with either Chinese words or characters, we also tried with both word/character initialization. In the case of words, the bitmap fonts had less resolution than characters because the size is the same. Table 2 shows the results in terms of BLEU.

[4] http://github.com/nyu-dl/dl4mt-tutorial/

System	BLEU
Characters	5.52
Characters +Bitmap	5.72
Words	5.55
Words +Bitmap	**8.49**

Table 2. BLEU results. In bold, best results.

Results show that using bitmap fonts as initialization is much better than using a random initialization, since much more information is provided to the neural system. When using character bitmap fonts the improvement is of 0.2 BLEU points, while using word bitmap fonts the improvement is of almost of 3 BLEU points. In any case, it is observed that it is better to use words than characters as translation units.

	Type	Sentence
1	Src	55/14． 大会议事规则第
	Words	Convenio sobre la emprendidas tratado
	+Bitmap	Reforma en **la Asamblea General**
	Ref	reglamento de **la Asamblea General**
2	Src	（e）小口径弹药；
	Words	e) el Medio comunicación Mundial ;
	+Bitmap	e) La información **pequeño calibre ;**
	Ref	e) Municiones de **pequeño calibre ;**
3	Src	58/147． 妇女的家庭暴力行为
	Words	ampliamente de los derechos humanos
	+Bitmap	discriminación **contra la mujer**
	Ref	violencia **contra la mujer** en el hogar
4	Src	（e）法律和司法体制；
	Words	e) acceso a la tecnología y la gestión de la información;
	+Bitmap	e) la **jurídicas** o **judiciales ;**
	Ref	e) las instituciones **jurídicas** y **judiciales ;**

Table 3. Example Sentences. Source (Src), Baseline (Words), Bitmap fonts (+Bitmap), Reference (Ref)

Table 3 shows some examples of the kind of improvements that the neural MT system with the new initialization is capable of. Examples show how it improves the adequacy and fluency of the translations in general.

5 Conclusions

This work has presented an alternative to the representation of 1-of-K coding using bitmap fonts instead. We have tried to take advantage of the representation of patterns.

Neural performance in this task is far from state-of-the-art results [Costa-jussà et al.2012]. The fact of not achieving comparable performance to the standard phrase-based system may be due to the fact that we are using a small dataset. However, this study shows a significant improvement when using a smarter initialization of the neural word vectors from standard neural MT system. The bitmap font initialization definitively provides more information to the neural systems than the random 1-of-K vectors. When comparing Chinese words or characters, the performance is similar, but it makes a big difference introducing bitmap fonts of words instead of bitmap fonts of characters. However, experiments on larger corpus would be required and are left for further work.

Software for experiments reported in this paper is freely available in github[5].

Acknowledgements

This work is supported by the 7th Framework Program of the European Commission through the International Outgoing Fellowship Marie Curie Action (IMTraP-2011-29951) and also by the Spanish Ministerio de Economía y Competitividad and the European Regional Development Fund, contract TEC2015-69266-P (MINECO/FEDER, UE).

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *CoRR*.

Rafael Banchs, Josep Maria Crego, Patrik Lambert, and José B. Mariño. 2006. A feasibility study for chinese-spanish statistical machine translation. In *in Procedings of the 5th International Symposium on Chinese Spoken Language Processing (ISCSLP)*.

Jordi Centelles, Marta R. Costa-jussà, and Rafael E. Banchs. 2014. Chispa on the go: A mobile chinese-spanish translation service for travellers in trouble. In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 33–36.

[5] https://github.com/aldomin/NMTbitMaps

Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder–decoder approaches. In *CoRR*.

Marta R. Costa-jussà and Jordi Centelles. 2016. Description of the chinese-to-spanish rule-based machine translation system developed using a hybrid combination of human annotation and statistical techniques. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 15.

Marta R. Costa-jussà, Carlos A. Henríquez Q., and Rafael E. Banchs. 2012. Evaluating indirect strategies for chinese-spanish statistical machine translation. *Journal Of Artificial Intelligence Research*, 45:762–780.

Marta R. Costa-jussà. 2015. Traducción automática estadística entre chino y castellano. *Komputer Sapiens*, 1:16–36.

Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent convolutional neural networks for discourse compositionality. In *CoRR*.

Andrew Lamb and Michael Xie. 2015. Convolutional encoders for neural machine translation. In *CoRR*.

Alexandre Rafalovitch and Robert Dale. 2009. United Nations General Assembly Resolutions: A Six-Language Parallel Corpus. In *Proc. of the MT Summit XII*, pages 292–299, Ottawa.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *CoRR*.

Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *CoRR*.