Supplementary Materials to Modelling Sentence Generation from Sum of Word Embedding Vectors as a Mixed Integer Programming Problem

Lyndon White, Roberto Togneri, Wei Liu and **Mohammed Bennamoun**
The University of Western Australia
35 Stirling Highway, Crawley, Western Australia
lyndon.white@research.uwa.edu.au
{roberto.togneri, wei.liu, mohammed.bennamoun}@uwa.edu.au

These supplementary materials show additional examples of the performance of our method against the works of Iyyer, Boyd-Graber, and Daumé III [1] and Bowman, Vilnis, Vinyals, *et al.* [2], as of our well as on sentences with ambiguous order. Bare in mind, exact reproduction is not the goal of either prior work; nor truly is it a goal of out work. Our goal being the regeneration of sentences while preserving meaning – exact reproduction does of course meet that goal. The examples that follow should highlight the differences in the performance of the methods.

Tables 1 to 3 show quantitative examples; including comparison to the existing works. In these tables ✗ and ✓ are used to show correctness of the output in the selection (Sel.) and in the ordering (Ord.) steps.

The sentences shown in Table 1, are difficult. The table features long complex sentences containing many proper nouns. These examples are sourced from Iyyer, Boyd-Graber, and Daumé III [1]. The output from their DT-RAE method is also shown for contrast. Only 3C is completed perfectly by our method. Of the remainder the MIP word ordering problem has no solutions, except in 3D, where it is wrong, but does produce an ordered sentence. In the others the language model constraints does not return any feasible ($P(\tau) > 0$) ordering solutions. This failure may be attributed in a large part to the proper nouns. Proper nouns are very sparse in any training corpus for language modelling. The Kneser-Ney smoothed trigrams back-off only down to bigrams, so if the words of the bigrams from the training corpus never appear adjacently in the training corpus, ordering fails. This is largely the case for very rare words. The other significant factor is the sentence length.

The sentences in Table 2, are short and use common words – they are easy to resynthesis. These examples come from Bowman, Vilnis, Vinyals, *et al.* [2]. The output of their VAE based approach can be compared to that from our approach. Of the three there were two exact match's, and one failure.

Normally mistakes made in the word selection step result in an unorderable sentence. Failures in selection are likely to result in a BOW that cannot be grammatically combined e.g. missing conjunctions. This results in no feasible solutions to the word ordering problem.

The examples shown in Table 3 highlight sentences where the order is ambiguous – where there are multiple reasonable solutions to the word ordering problem. In both cases the word selection performs perfectly, but the ordering is varied. In 5A, the Ref. BOW+Ord. sentence and the overall Sel. BOW+Ord. sentence in word order but not in word content. This is because under the trigram language model both sentences have exactly identical probabilities, so it comes to which solution is found first, which varies on the state of the MIP solver. In 5B the word order is switched – "from Paris to London" vs "to London from Paris", which has the same meaning. But, it could also have switched the place names. In cases like this where two orderings are reasonable, the ordering method is certain to fail consistently for one of the orderings. Though it is possible to output the second (and third etc.) most probable ordering, which does ameliorate the failure somewhat. This is the key limitation which prevents this method from direct practical applications.

4A Reference	we looked out at the setting sun .	**Sel.**	**Ord.**
Ref. BOW+Ord.	we looked out at the setting sun .	–	✓
Sel. BOW+Ord.	we looked out at the setting sun .	✓	✓
VAE Mean	they were laughing at the same time .		
VAE Sample1	ill see you in the early morning .		
VAE Sample2	i looked up at the blue sky .		
VAE Sample3	it was down on the dance floor .		
4B Reference	i went to the kitchen .	**Sel.**	**Ord.**
Ref. BOW+Ord.	i went to the kitchen .	–	✓
Sel. BOW+Ord.	i went to the kitchen .	✓	✓
VAE Mean	i went to the kitchen .		
VAE Sample1	i went to my apartment .		
VAE Sample2	i looked around the room .		
VAE Sample3	i turned back to the table .		
4C Reference	how are you doing ?	**Sel.**	**Ord.**
Ref. BOW+Ord.	how are you doing ?	–	✓
Sel. BOW+Ord.	how 're do well ?	✗	✗
VAE Mean	what are you doing ?		
VAE Sample1	are you sure ?		
VAE Sample2	what are you doing, ?		
VAE Sample3	what are you doing ?		

Table 2

A comparison of the output of the Two Step process proposed in this paper, to the example sentences generated by the VAE method of Bowman, Vilnis, Vinyals, *et al.* [2].

References

[1] M. Iyyer, J. Boyd-Graber, and H. Daumé III, “Generating sentences from semantic vector space representations”, in *NIPS Workshop on Learning Semantics*, 2014.

[2] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, “Generating sentences from a continuous space”, *ArXiv preprint arXiv:1511.06349*, 2015.

[3] L. White, R. Togneri, W. Liu, and M. Bennamoun, “Generating bags of words from the sums of their word embeddings”, in *17th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*, 2016.

5A Reference	it was the worst of times , it was the best of times .	**Sel.**	**Ord.**
Ref. BOW+Ord.	it was the worst of times , it was the best of times .	–	✓
Sel. BOW+Ord.	it was the best of times , it was the worst of times .	✓	✗
5B Reference	please give me directions from Paris to London .	**Sel.**	**Ord.**
Ref. BOW+Ord.	please give me directions to London from Paris .	–	✗
Sel. BOW+Ord.	please give me directions to London from Paris .	✓	✗

Table 3

A pair of example sentences, where the correct order is particularly ambiguous.

3A	**Reference**	name this 1922 novel about leopold bloom written by james joyce .	**Sel.**	**Ord.**
	Ref. BOW+Ord.	written by name this . novel about 1922 bloom leopold james joyce	–	✖
	Sel. BOW+Ord.	written novel by name james about leopold this bloom 1922 joyce .	✓	✖
	DT-RAE Ref.	name this 1906 novel about gottlieb_fecknoe inspired by james_joyce		
	DT-RAE Para.	what is this william golding novel by its written writer		
3B	**Reference**	ralph waldo emerson dismissed this poet as the jingle man and james russell lowell called him three-fifths genius and two-fifths sheer fudge .	**Sel.**	**Ord.**
	Ref. BOW+Ord.	sheer this as james two-fifths emerson fudge lowell poet genius waldo called russell the and ralph and him . dismissed jingle three-fifths man	–	✖
	Sel. BOW+Ord.	him “ james great as emerson genius ralph the lowell and sheer waldo three-fifths man fudge dismissed jingle russell two-fifths and gwalchmai 2009 vice-versa ____ prominent called 21.25 explained	✗	✖
	DT-RAE Ref.	henry_david_thoreau rejected this author like the tsar boat and imbalance created known good writing and his own death		
	DT-RAE Para.	henry_david_thoreau rejected him through their stories to go money well inspired stories to write as her writing		
3C	**Reference**	this is the basis of a comedy of manners first performed in 1892 .	**Sel.**	**Ord.**
	Ref. BOW+Ord.	this is the basis of a comedy of manners first performed in 1892 .	–	✓
	Sel. BOW+Ord.	this is the basis of a comedy of manners first performed in 1892 .	✓	✓
	DT-RAE Ref.	another is the subject of this trilogy of romance most performed in 1874		
	DT-RAE Para.	subject of drama from him about romance		
3D	**Reference**	in a third novel a sailor abandons the patna and meets marlow who in another novel meets kurtz in the congo .	**Sel.**	**Ord.**
	Ref. BOW+Ord.	kurtz and another meets sailor meets the marlow who abandons a third novel in a novel in the congo in patna .	–	✗
	Sel. BOW+Ord.	kurtz and another meets sailor meets the marlow who abandons a third novel in a novel in the congo in patna .	✓	✗
	DT-RAE Ref.	during the short book the lady seduces the family and meets cousin he in a novel dies sister from the mr.		
	DT-RAE Para.	during book of its author young lady seduces the family to marry old suicide while i marries himself in marriage		
3E	**Reference**	thus she leaves her husband and child for aleksei vronsky but all ends sadly when she leaps in front of a train .	**Sel.**	**Ord.**
	Ref. BOW+Ord.	train front of child vronsky but and for leaps thus sadly all her she she in when aleksei husband ends a . leaves	–	✖
	Sel. BOW+Ord.	she her all when child for leaves front but and train ends husband aleksei leaps of vronsky in a sadly micro-history thus , she the	✗	✖
	DT-RAE Ref.	however she leaves her sister and daughter from former fiancé and she ends unfortunately when narrator drives into life of a house		
	DT-RAE Para.	leaves the sister of man in this novel		

Table 1

A comparison our method, to the example sentences generated by the DT-RAE method of Iyyer, Boyd-Graber, and Daumé III [1]. Ref. BOW+Ord. shows the word ordering step on the reference BOW. the Sel. and Ord. columns indicate if the output had the correct words selected, and ordered respectively. With ✓ indicating correct and ✗ indicating incorrect. ✖ indicates not only that ordering was not correct, but that the MIP problem had no feasible solutions at all. DT-RAE Ref. shows the result of the method of Iyyer, Boyd-Graber, and Daumé III [1], when the dependency tree of the output is provided to the generating process, whereas in DT-RAE Para. an arbitrary dependency tree is provided to the generating process. Note that the reference used as input to Sel. BOW+Ord. and Ref. BOW+Ord. sentence was varied slightly from that used in Iyyer, Boyd-Graber, and Daumé III [1] and White, Togneri, Liu, *et al.* [3], in that terminating punctuation was not removed, and nor were multiword entity references grouped into single tokens.