Algorithm	AskUbuntu			English		
	@1	@5	@10	@1	@5	@10
TFIDF	8.3	17.5	22.5	10.0	18.1	21.6
BM25	7.3	17.1	21.8	10.0	18.9	23.2
IB	8.1	18.1	22.6	10.1	18.4	22.7
DFR	7.7	17.8	22.4	10.5	19.0	23.0
LMD	5.6	14.1	19.0	10.9	20.1	24.2
LMJ	8.3	17.5	22.5	10.3	18.5	22.1
CNN	**11.5**	**24.8**	**31.4**	**11.6**	**23.0**	**26.9**
BOW-CNN	10.9	22.6	28.7	11.3	21.4	26.0

Table 3: Question *title* retrieval performance (Accuracy@k) for different algorithms.

Algorithm	AskUbuntu			English		
	@1	@5	@10	@1	@5	@10
TFIDF	16.9	31.3	38.3	25.9	42.0	48.1
BM25	18.2	33.1	39.8	29.4	45.7	52.5
IB	14.9	28.2	34.8	25.4	42.3	48.0
DFR	18.0	32.6	39.2	28.6	45.4	52.5
LMD	13.7	26.8	34.4	23.0	40.2	46.0
LMJ	18.3	33.4	40.7	28.5	45.7	52.3
CNN	20.0	33.8	40.1	17.2	29.6	33.8
BOW-CNN	**22.3**	**39.7**	**46.4**	**30.8**	**47.7**	**54.9**

Table 4: Question *title* + *body* (*all*) retrieval performance for different algorithms.

best IR baseline (LMJ) in terms of Accuracy@1, which represents an improvement of 21.9%. Since the BOW representation we use is closely related to TFIDF, an important comparison is the performance of BOW-CNN vs. TFIDF. In Tables 3 and 4, we can see that BOW-CNN consistently outperforms the TFIDF model in the two datasets for both cases *title* and *all*. These findings suggest that BOW-CNN is indeed combining the strong semantic representation power conveyed by the convolutional-based representation to, jointly with the BOW representation, construct a more effective model.

Another interesting finding is that CNN outperforms BOW-CNN for short texts (Table 3) and, conversely, BOW-CNN outperforms CNN for long texts (Table 4). This demonstrates that, when dealing with large input texts, BOW-CNN is an effective approach to combine the strengths of convolutional-based representation and BOW.

Impact of Initialization of BOW Weights. In the BOW-CNN experiments whose results are presented in tables 3 and 4 we initialize the elements of the BOW weight vector t with the IDF of each word in V computed over the question set Q. In this section we show some experimental results that indicate the contribution of this initialization.

In Table 5, we present the performance of BOW-CNN for the English dataset when different configurations of the BOW weight vector t are used. The first column of Table 5 indicates the type of initialization, where *ones* means that t is initialized with the value 1 (one) in all positions. The second column informs whether t is allowed to be updated (*Yes*) by the network or not (*No*). The numbers suggest that letting BOW weights free to be updated by the network produces better results than fixing them to IDF values. In addition, using IDF to initialize the BOW weight vector is better than using the same weight (ones) to initialize it. This is expected, since we are injecting a prior knowledge known to be helpful in IR tasks.

t initial	t updated	Title		All	
		@1	@10	@1	@10
IDF	Yes	**11.3**	**26.0**	**30.8**	**54.9**
IDF	No	10.6	25.3	29.7	**54.9**
Ones	Yes	10.7	24.2	26.3	51.2

Table 5: BOW-CNN performance using different methods to initialize the BOW weight vector t.

5 Conclusions

In this paper, we propose a hybrid neural network architecture, BOW-CNN, that combines bag-of-words with distributed vector representations created by a CNN, to retrieve semantically equivalent questions. Our experimental evaluation showed that: our approach outperforms traditional bow approaches; for short texts, a pure CNN obtains the best results, whereas for long texts, BOW-CNN is more effective; and initializing the BOW weight vector with IDF values is beneficial.

Acknowledgments

Dasha Bogdanova's contributions were made during an internship at IBM Research. Her work was partially supported by Science Foundation Ireland through the CNGL Programme (Grant 12/CE/I2267) in the ADAPT Centre (www.adaptcentre.ie) at DCU.

References

Gianni Amati and Cornelis Joost Van Rijsbergen. 2002. Probabilistic models of information retrieval based on measuring the divergence from randomness. *ACM Transactions on Information Systems (TOIS)*, 20(4):357–389.

James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Des-