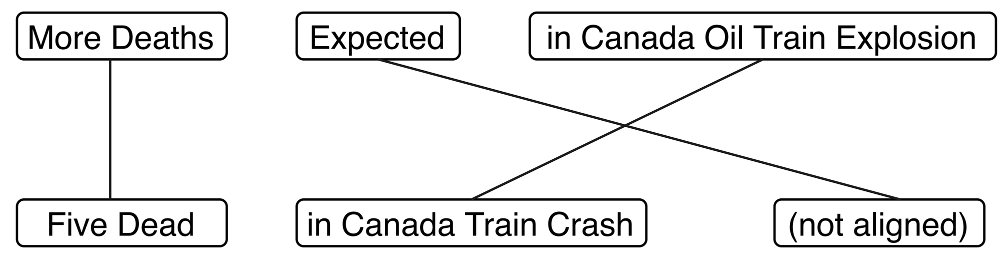
Logic Constrained Pointer Networks for Interpretable Textual Similarity

Subhadeep Maji[1*†], Rohan Kumar[1*], Manish Bansal[1], Kalyani Roy[2], Pawan Goyal[2]

[1]Flipkart; [2]Dept. of CSE, IIT Kharagpur, India

1 Problem Statement

- Measuring semantic textual similarity (STS) is pivotal for text understanding.
- Alignment of segments across sentence pairs provides interpretability to STS.



- Many tasks such Machine Translation, Paraphrase Recognition can utilize alignment of parts of sentence pairs.

2 Contributions

- Gated Pointer Networks for alignment.
- Bidirectionality in Pointer Network alignments.
- Side-supervision to network with First Order Logic statements.

3 Bidirectional Pointer Network : Model Architecture

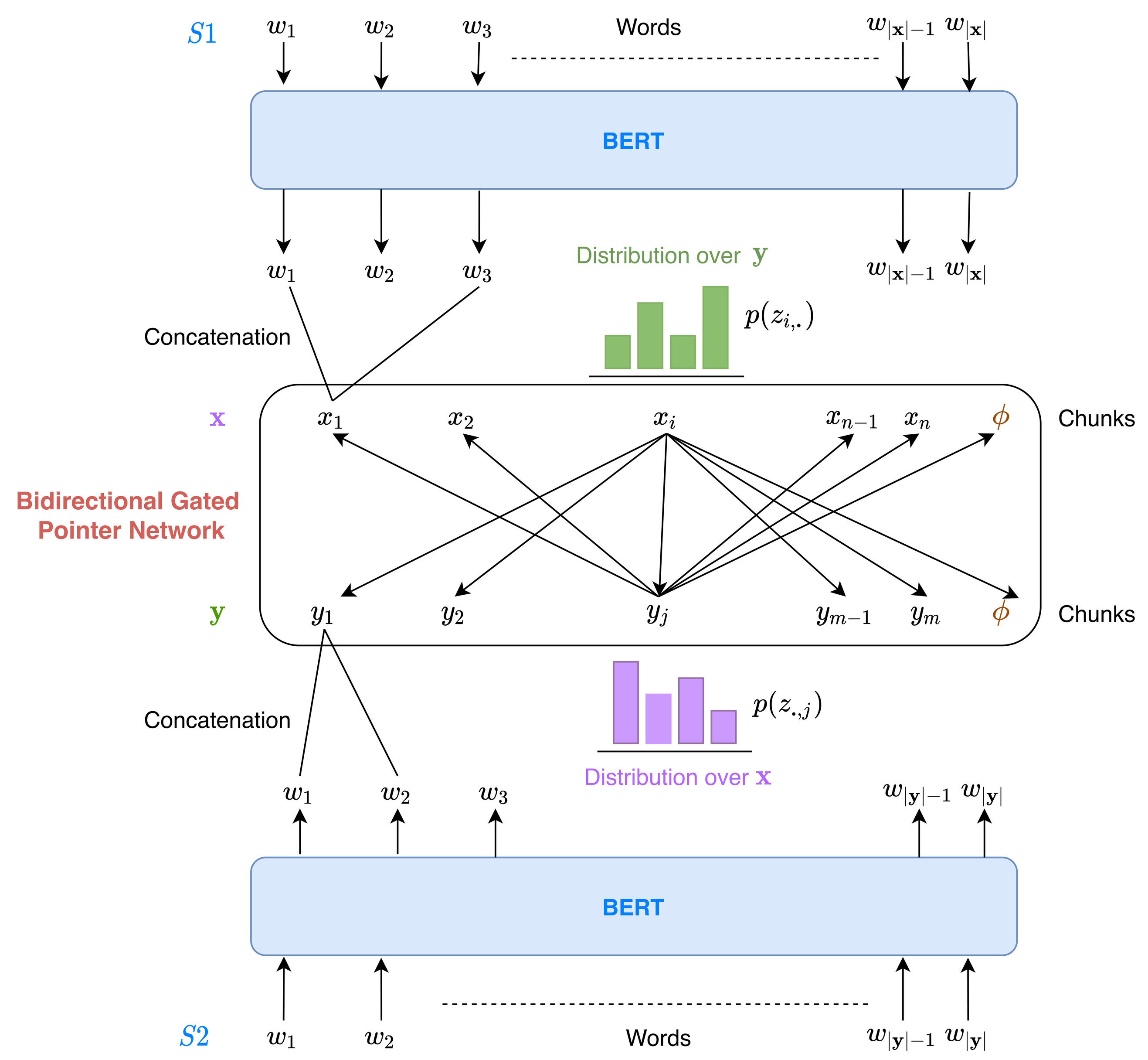


Figure 1:Block level illustration of our Bidirectional Pointer Network with BERT-based chunk embeddings. For ease of illustration, we show the non-aligned chunk ϕ on both sentences and omit logic constraints.

- **Pointer Networks** adapt attention mechanism to 'point' from a chunk x_i in x to a chunk y_j in y. The alignment $\theta_{i,j}$ is modelled as:

$$\theta_{i,j} = v^T f(W_1 x_i + W_2 y_j + W_3 x_i \otimes y_j)$$

where W_1, W_2, W_3 and v are model parameters, f is a nonlinearity.

- Since not all x_i's may align to some constituent in y (and vice versa), we introduce **gating functions** $g_i^{\mathbf{x}}$ and $g_j^{\mathbf{y}}$. This Gated Pointer Network (PN) models the alignment as:

$$p(z_{i,j} = 1) \propto g_i^{\mathbf{x}} g_j^{\mathbf{y}} \theta_{i,j}$$
$$p(z_{i,\phi} = 1) \propto (1 - g_i^{\mathbf{x}})$$
$$p(z_{\phi,j} = 1) \propto (1 - g_j^{\mathbf{y}})$$

where $z_{i,j} \in \{0, 1\}$ indicates whether x_i is aligned to y_j.

- Gated PN models alignments for x_i as a distribution, $p(z_{i,.})$ but alignments for y_j are unconstrained, i.e. y_j could be aligned to many x_i's. This violates one-to-one alignment:

$$p(z_{i,j} = 1) = \text{softmax}([g_i^{\mathbf{x}} g_j^{\mathbf{y}} \theta_{i,j}; (1 - g_i^{\mathbf{x}})])$$

- Bidirectional Gated PN (Bi-PN): We address the shortcomings of gated PN by enforcing **bidirectionality** i.e.,both $z_{i,.}$ and $z_{.,j}$ are distributions for all i, j. We Pose alignment $p(z_{i,j})$ as optimal transport problem and use approximate solution. Two-way alignment distribution:

$$2\left(-\sum_{i=1}^{n}\sum_{j=1}^{m} a_{i,j}\log(p_{i,j})\right) + \left(-\sum_{i=1}^{n} a_{i,\phi}\log(p_{i,\phi})\right) + \left(-\sum_{j=1}^{m} a_{\phi,j}\log(p_{\phi,j})\right)$$

- To further improve the model, we introduce side-supervision with **First Order Logic** (FOL). Rules are derived from ConceptNet knowledge base and Syntactic heuristics.

Code: `https://github.com/manishb89/interpretable_sentence_similarity`

*Equal contribution †Now at Amazon

- Logic rules to constrain pointer network decisions.

$$\theta'_{i,j} = \theta_{i,j} + \rho m_{i,j}$$

where $m_{i,j}$ is a positive constant if rules are true for an alignment. The importance of FOL rules is controlled using the ρ hyperparameter.

4 Experiments

- **SemEval-2016** Task-2 datasets for interpretable Semantic Textual Similarity (iSTS). Two domains - News Headlines and Flickr Images

Comparison with baseline methods and model ablations:

Model	Dataset	
	Headlines	Images
Inspire [Kazmi and Schüller, SemEval, 2016]	81.94	86.70
UWB [Konopiket al., SemEval, 2016]	89.87	89.37
ILP [Li and Srikumar, EMNLP, 2016]	92.57	87.38
M1 [Glove + PN]	89.70	88.34
M2 [Glove + Bi-PN]	91.48	90.88
M3 [BERT + Bi-PN]	96.63	93.81
M4 [BERT + Bi-PN + FOL]	**97.73**	**96.32**

Table 1:Average F1 score of alignments on the test set.

Cross-domain Experiments:

- M4 achieves F1 scores of **96.16** and **94.80** on headlines and images datasets, respectively.

Qualitative Examples:

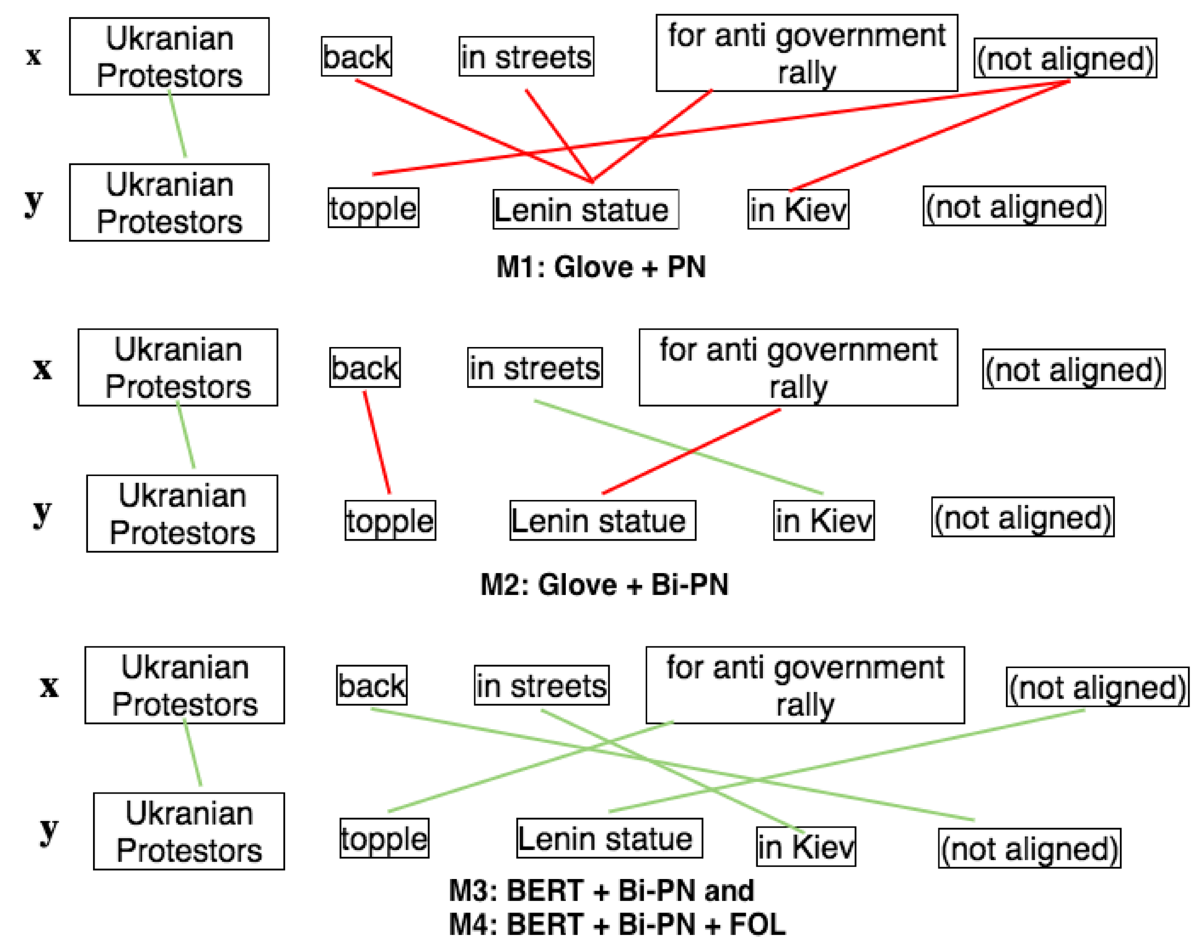


Figure 2:Importance of modelling bidirectionality.

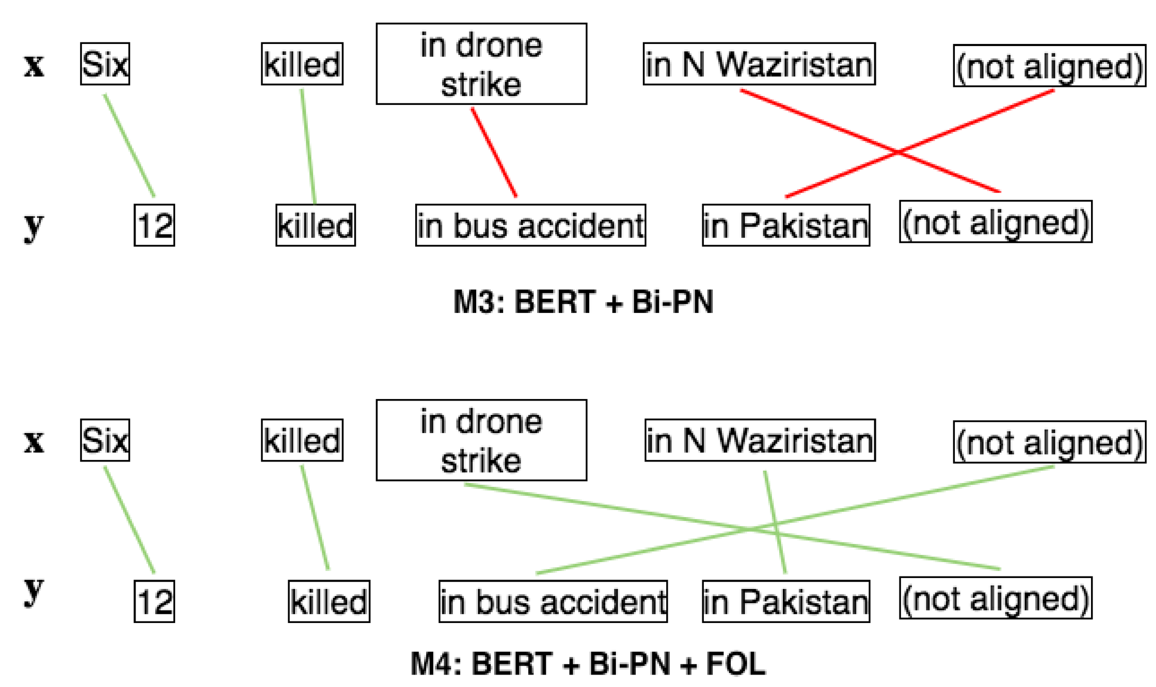


Figure 3:Effect of including external knowledge in the form of FOL statements in the model.

(Note : The correct alignments are shown in green and the incorrect ones are shown in red)

5 Conclusion

- We introduce a novel logic statement constrained gated pointer network model to align constituents of the two sentences.
- Our model achieves a large performance improvement over existing methods and also performs well in cross-domain evaluation.

References

[1] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT. pp. 4171–4186 (2019)

[2] Li, T., Srikumar, V.: Augmenting neural networks with first-order logic. In: ACL. pp. 292–302. Association for Computational Linguistics, Florence, Italy (Jul 2019)

[3] Vinyals, O., Fortunato, M., Jaitly, N.: Pointer networks. In: NeurIPS. pp. 2692–2700 (2015)