InFillmore: Frame-Guided Language Generation with Bidirectional Context

Jiefu Ou, Nathaniel Weir, Anton Belyy, Felix Yu, Benjamin Van Durme
jouaa@connect.ust.hk, {nweir,abel,fyu17,vandurme}@jhu.edu

Johns Hopkins University

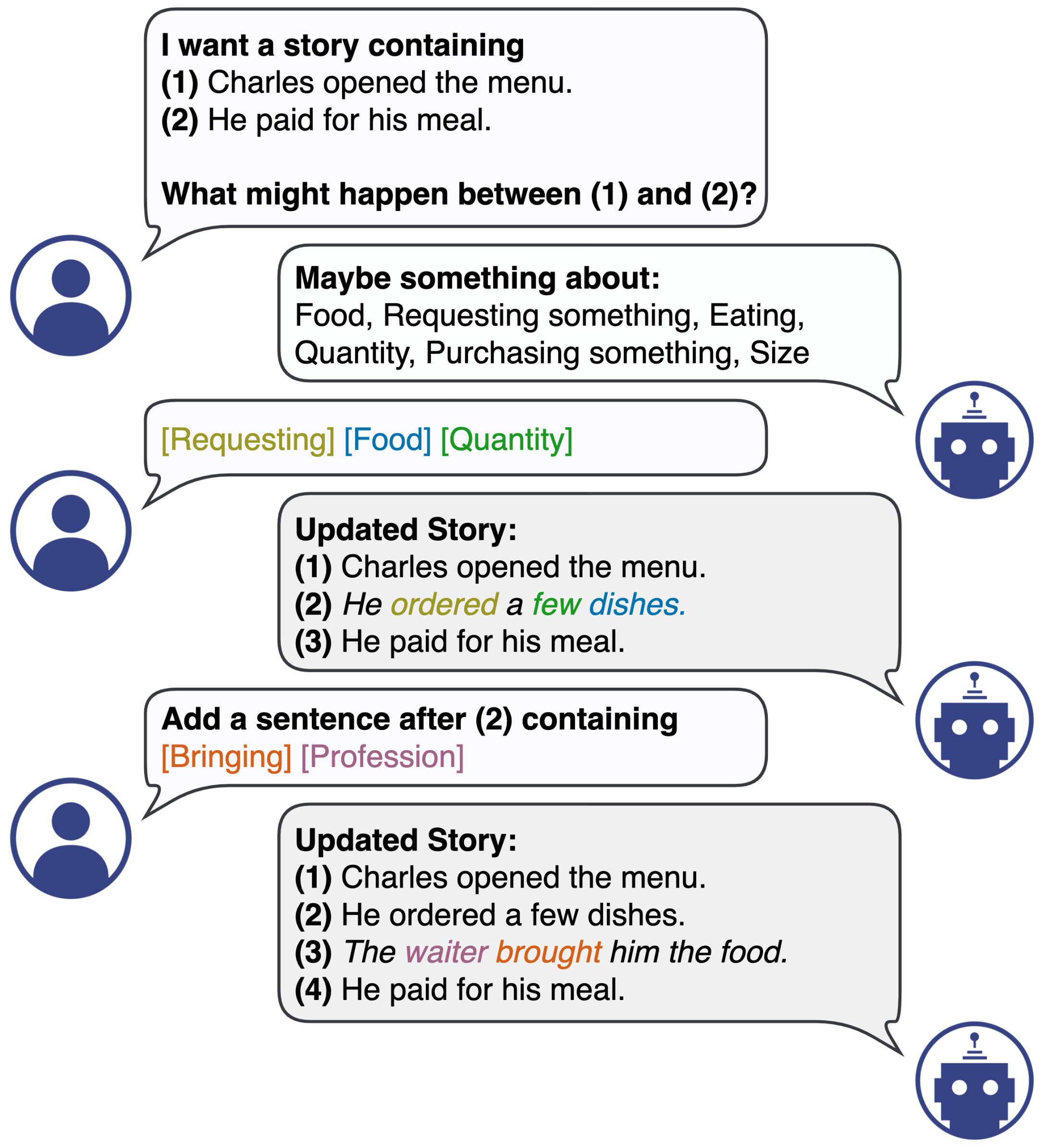
Motivation

- State-of-the-art text infilling models are very good at producing humanlike text and have been proposed as core components of **automatic** and **interactive story generation**.
- However, they have not yet been equipped with a mechanism to explicitly **control for underlying semantic content**.
- We propose a way for humans (or content planning models) to **specify discrete semantic content** while conditioning on **surrounding textual context**.

Data

- **FrameNet**: We use the 1221 frames defined in FrameNet as a signal that constraints or conditions the infilling model. We hypothesize that a sequence of FrameNet frames provides enough signal to control for semantic content in infilling models.
- **ROCStories**: We use this dataset of 5-sentence short stories to evaluate the performance of our proposed guidance methods.

Interactive Story Generation



I want a story containing
(1) Charles opened the menu.
(2) He paid for his meal.

What might happen between (1) and (2)?

Maybe something about:
Food, Requesting something, Eating, Quantity, Purchasing something, Size

[Requesting] [Food] [Quantity]

Updated Story:
(1) Charles opened the menu.
(2) *He ordered a few dishes.*
(3) He paid for his meal.

Add a sentence after (2) containing
[Bringing] [Profession]

Updated Story:
(1) Charles opened the menu.
(2) He ordered a few dishes.
(3) *The waiter brought him the food.*
(4) He paid for his meal.

Evaluation Methodology

- **Automatic Evaluation**: we examine the performance of both FFL and LCD by measuring the rate at which they produce sentences that trigger the desired frame(s) (**Frame Fidelity**) and by measuring the perplexity score of the framefilling-trained language model on test examples (**Perplexity**).
- **Human Evaluation**: In addition to the automatic evaluation, we conduct two human evaluations that ask annotators to tell apart model- and human-generated sentences (**Indistinguishability**) and rank model-generated sentences relative to one another (**Relative Plausibility**).

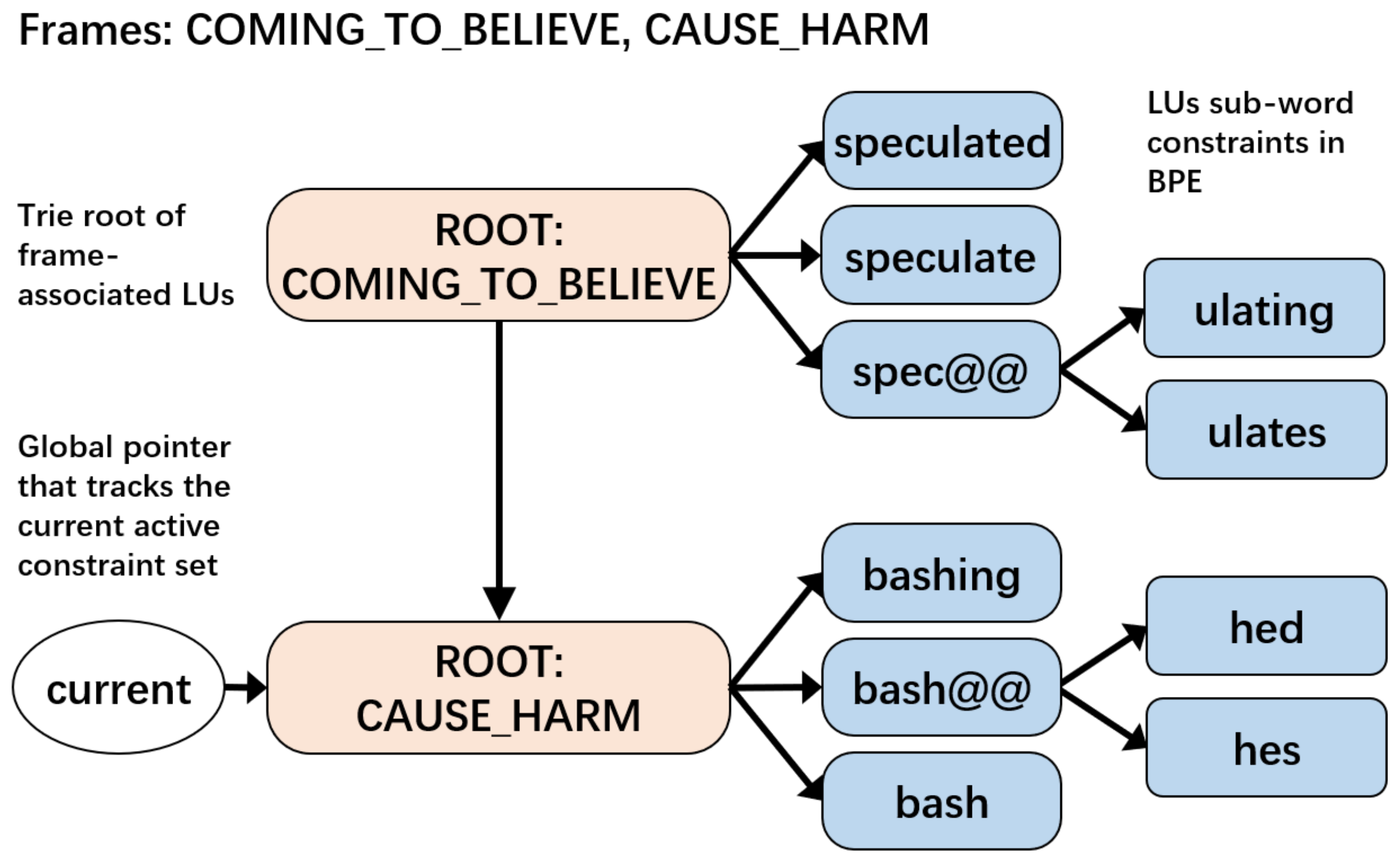
Two Approaches for Frame Guidance

- **Fine-Tuned "Framefilling" (FFL)**: we fine-tune a recent infilling model (ILM) with a frame guided denoising objective. A training instance is formed by randomly masking out spans of text with **[blank]**, which is appended with frame ID tokens $F_1, F_2, \ldots$ (e.g. **[Food]**) as guiding signals, followed by golden span infill. The frame ID & infill per span are seperated with **[sep]**, as shown in the example below, where **S-FFL** and **A-FFL** stands for infilling conditioned on a single frame/all frames respectively. Instances are then fed to a unidirectional language model (in our case, GPT-2).

Story	Charles went shopping. He bought fruit. Then he left.
ILM	Charles went shopping. [blank] Then he left. *[sep] He bought fruit.*
S-FFL	[sep] [Food] He bought fruit.
A-FFL	[sep] [Commerce_buy] [Food] He bought fruit.

- **Lexically Constrained Decoding (LCD)**: The frame guidance/constraints are provided through lexical units (LUs) of frames at decoding phase only. Given a sequence of frame ID tokens $F_1, F_2, \ldots, F_n$, we build a corresponding sequence of disjunctive lexical constraint sets $C_1, C_2, \ldots, C_n$, where C_i consists of all LUs of F_i with their morphological variants.

  As shown in the figure below, LCD represents a sequence of disjunctive constraint sets as a list of tries, one per frame, each covering a set of disjunctive lexical units (with morphological variants) based on the Byte Pair Encoding. During decoding, the progress through the trie is recorded and the output is forced to contain one and only one of the LUs per frame.


Frames: COMING_TO_BELIEVE, CAUSE_HARM
Trie root of frame-associated LUs
ROOT: COMING_TO_BELIEVE
speculated
speculate
spec@@
LUs sub-word constraints in BPE
ulating
ulates
Global pointer that tracks the current active constraint set
current
ROOT: CAUSE_HARM
bashing
bash@@
bash
hed
hes

Conclusion

- We propose the use of FrameNet to control for semantic content in an infilling model.
- We introduce two extensions of neural text generation that use FrameNet frames as guiding signal.
- Experiments on the sentence infilling task demonstrate that both our extensions enable explicit manipulation of semantics at the frame level with competitive generation quality.

Acknowledgements

This work was supported by DARPA KAIROS and NSF grant no. BCS-2020969.

The demo of our approach is available at

`https://nlp.jhu.edu/demos/infillmore`