Discourse-Aware Neural Rewards for Coherent Text Generation

PAUL G. ALLEN SCHOOL W UWNLP

Antoine Bosselut, Yejin Choi
Paul G. Allen School of CSE,
University of Washington

Asli Celikyilmaz, Jianfeng Gao, Po-Sen Huang
Deep Learning Technology Center
Microsoft Research AI

Xiaodong He
JD AI Research

JD.COM Microsoft Research AI

Motivation

- Fine-tuning generation policies with RL improves performance on the metric used as reward
- Underlying assumption that reward metrics highly correlated with quality of generated text

	CIDEr	B-4	R-L	M
MLE	94.0	29.6	52.6	25.2
CIDEr	**106.3**	31.9	54.3	25.5
BLEU-4	94.4	**33.2**	53.9	24.6
ROUGE-L	97.7	31.6	**55.4**	24.5
METEOR	80.4	25.3	51.3	**25.9**

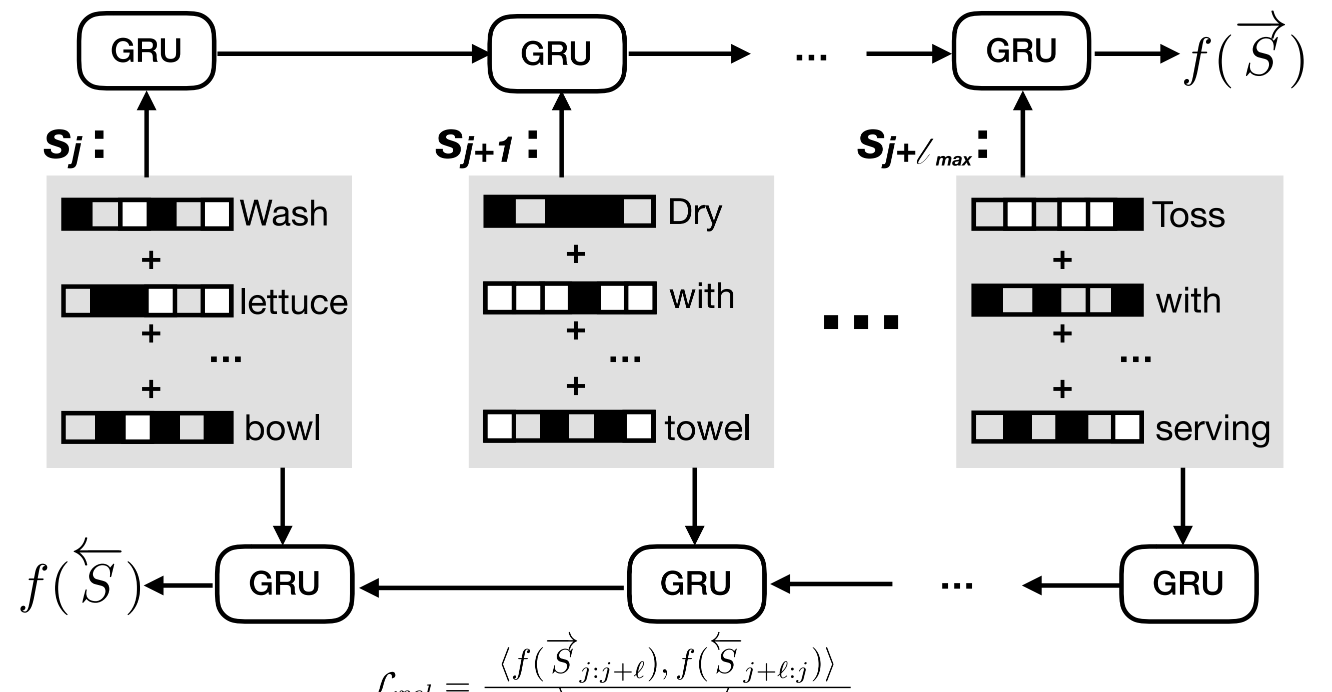
Image Captioning results from Rennie et al., 2017

- Most reward metrics only match localized *n*-gram patterns between generated and gold text
- Discourse structure is **NOT** evaluated by these metrics, but is crucial for coherent long text generation
- **Solution:** Train a neural network to score production of desired discourse structure during generation, and use score as a reward for the model in self-critical learning

Neural Teacher

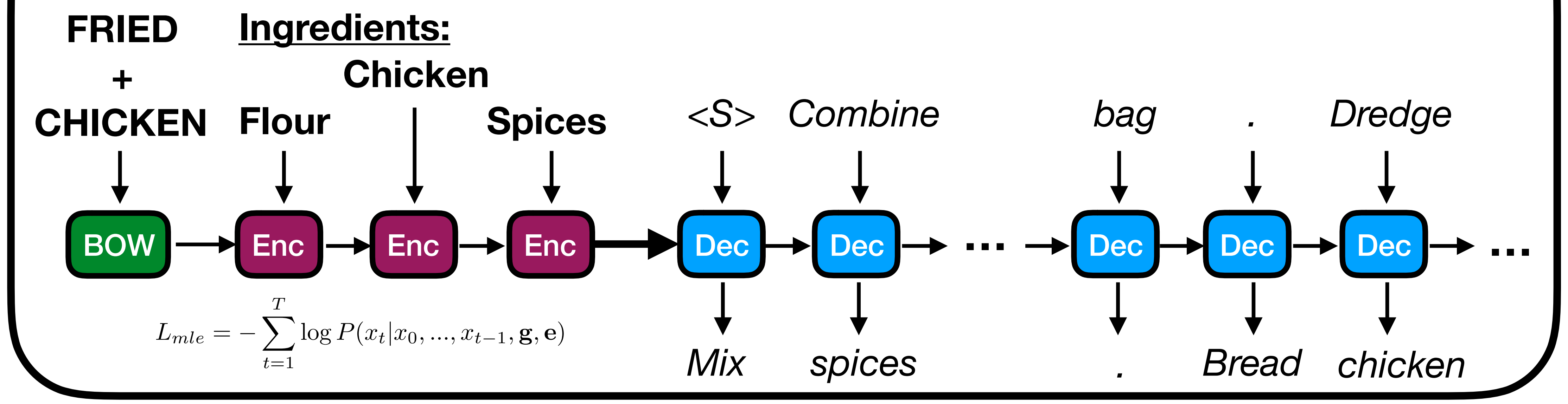
- Ordering structure can be used as approximation for discourse structure (Barzilay and Lapata, 2005)
- Teacher learns to score sentence order as analogue for discourse

Training:
1) Sample subsequence
2) Encode ordered BOWs
3) Reverse sentences
4) Encode reverse
5) Maximize difference



$$\mathcal{L}_{rel} = \frac{\langle f(\overrightarrow{S}_{j:j+\ell}), f(\overleftarrow{S}_{j+\ell:j}) \rangle}{\|f(\overrightarrow{S}_{j:j+\ell})\| \|f(\overleftarrow{S}_{j+\ell:j})\|}$$

Maximum Likelihood Training

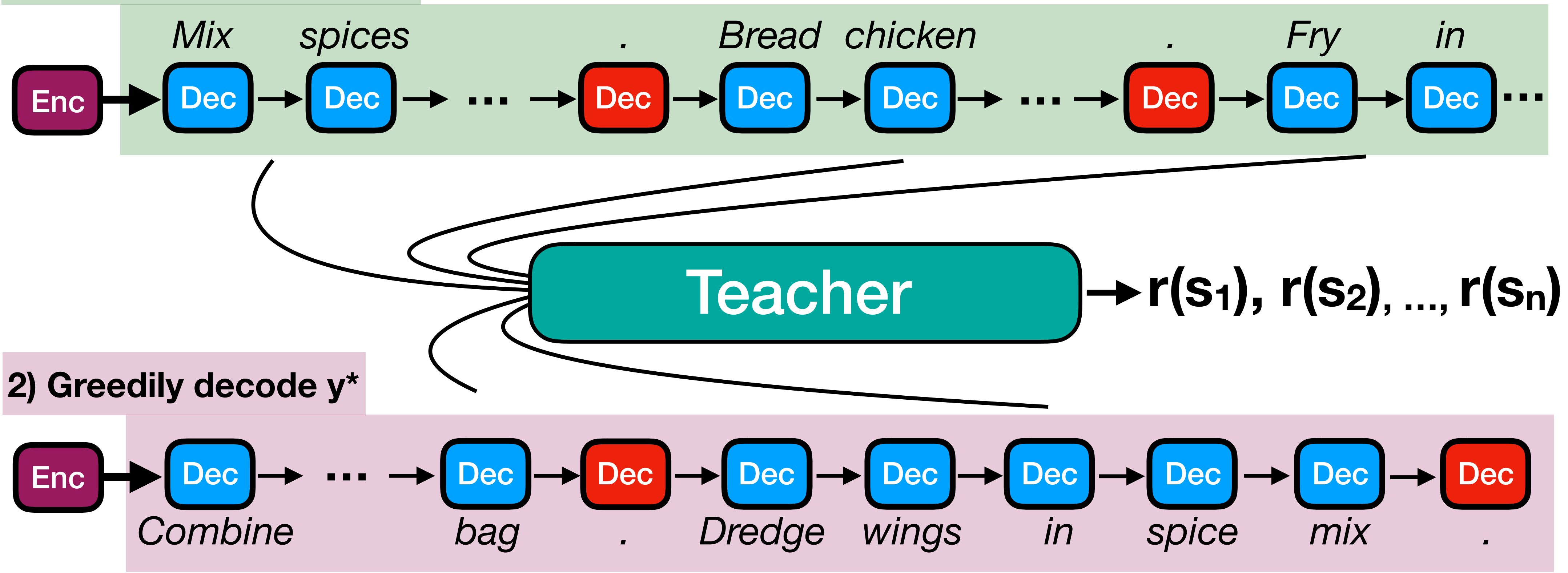


$$L_{mle} = -\sum_{t=1}^{T} \log P(x_t | x_0, \ldots, x_{t-1}, \mathbf{g}, \mathbf{e})$$

Self-critical Training[2]

1) Sample a sequence, $\hat{y}$

2) Greedily decode y*



3) Compute rewards from teacher

$$r(s_j) = \sum_{\ell=\ell_{min}}^{\ell_{max}} \left(\frac{\langle f(\overrightarrow{S}_{j-\ell:j}), f(\overleftarrow{S}_{j-\ell:j}) \rangle}{\|f(\overrightarrow{S}_{j-\ell:j})\| \|f(\overleftarrow{S}_{j-\ell:j})\|} - \frac{\langle f(S_{j-\ell:j}), f(\overleftarrow{S}_{j-\ell:j}) \rangle}{\|f(S_{j-\ell:j})\| \|f(\overleftarrow{S}_{j-\ell:j})\|} \right)$$

4) Make reward from decoded sequence the baseline for model reward

$$r_t = \sum_{j=1}^{|S|} \mathbb{1}(y_t \in \hat{s}_j)(r(\hat{s}_j) - r(s_j^*))$$

sampled

decoded

5) Apply REINFORCE with baselined reward

$$\mathcal{L}_{rl} = -\sum_{t=1}^{T} r_t \log P(\hat{y}_t | \hat{y}_0, \ldots, y_{t-1}, \mathbf{g}, \mathbf{e})$$

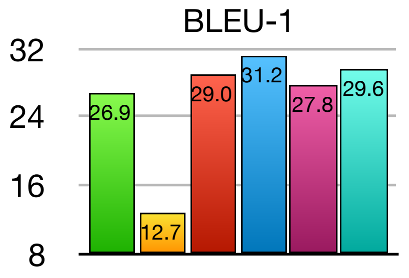
Mixed Loss[3]

Title: Jellied Horseradish
Ings: horseradish, sugar, vinegar, fruit pectin
Generated Recipe:
Add sugar and sugar. **Add sugar and cook. Add sugar and cook. Add sugar and cook.** Remove from heat and **add sugar**. Fold in whipped cream. Chill. **Add sugar** and lemon juice. Add sugar.

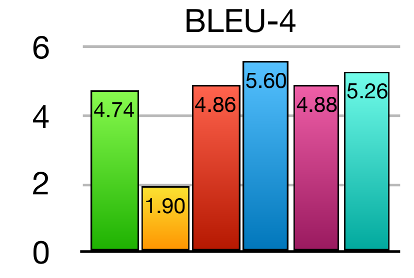
$$\mathcal{L}_{mix} = \gamma \mathcal{L}_{rl} + (1 - \gamma)\mathcal{L}_{mle}$$

RL Loss MLE Loss

Generation



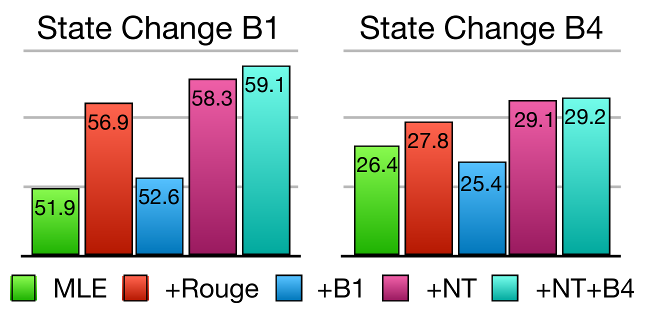
Takeaways:
- Fine-tuning with word metrics improves performance on word metrics
- Neural Teacher better than pure MLE training
- Mixed loss needed for stability

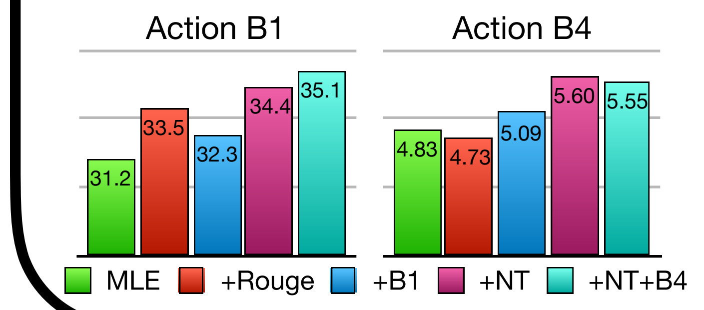


MLE +CIDEr (1.0) +Rouge (0.9984) +B1 (0.97) +NT (0.97) +NT+B4 (0.97)

Discourse

Process:
- Use action and state change lexicons from Bosselut et al., 2018.
- Retain words that map to a verb in lexicon
- Compute scores on ordered action sequence
- Map ordered action words to state changes
- Compute scores on state change sequence

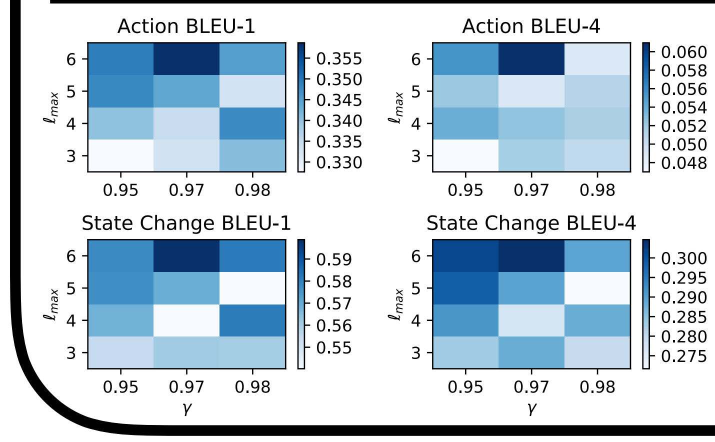


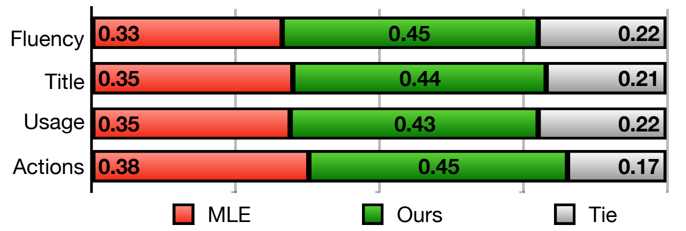


Takeaways:
- Fine-tuning word metrics does not improve more abstract discourse scores
- Neural Teacher better than MLE training or fine-tuning with word-level scores
- Fine-tuning with **BOTH** Teachers and word scores does best

Insights

Gold	MLE	Mixed MLE + RL (Ours)
Stir all together. **Boil** until it drops from spoon. Add vanilla. **Beat until it thickens.** Pour into buttered pan.	Mix all ingredients. Cook on low for 10 minutes. Remove from fire and add flavoring. **Beat until creamy. Beat until creamy. Beat until creamy. Beat until creamy. Add vanilla. Beat until creamy.** Add vanilla. **Beat until creamy. Add vanilla.** Shape into balls. Wrap in waxed paper and chill.	**Combine sugar, cream and butter. Boill** until firm. Remove from fire and add flavoring. Cool. Beat until creamy. **Add vanilla and beat until stiff.** Form into balls and roll in sugar. Wrap in wax paper and chill.
Core apples; Peel top third of each. Place apples in a shallow baking dish. Combine walnuts and raisins; **Stuff cavities of apples with nut mixture.** Combine water, honey, and cinnamon in a small saucepan. **...**	Core apples and pare. Cut apples in half. Place in a baking dish. Combine remaining ingredients and pour over apples. Bake at 350 degrees for 1 hour.	Core apples and pare. Place apples in a baking dish. Combine remaining ingredients and mix well. **Spoon into cavities of apples.** Bake in 350 f oven for 1 hour. Serve hot.
Bring all but cheese to a boil; Simmer until veggies are tender. Sprinkle each serving with cheese.	In large saucepan, brown hamburger, onion, and celery in margarine. Add remaining ingredients, except cheese. Heat to boiling. Reduce heat and simmer, covered, for 30 minutes. Stir in cheese **...**	Combine all ingredients except cheese, celery, and onion in a large dutch oven; Bring to a boil. Reduce heat; Cover and simmer for 30 minutes. **Stir in macaroni, cheese, and cheese.** Heat through. Serve over hot noodles.





References:
[1] Barzilay and Lapata, *Modeling Local Coherence: An Entity-based Approach.* In *ACL,* 2005.
[2] Rennie et al., *Self-critical Sequence Training for Image Captioning*. In *CVPR*, 2017.
[3] Paulus et al., *A Deep Reinforced model for Abstractive Summarization.* In *ICLR,* 2018.