# A Context-Dependent Gated Module for Incorporating Symbolic Semantics into Event Coreference Resolution


Tuan Lai [1], Heng Ji [1], Trung Bui [2], Quan Hung Tran[2], Franck Dernoncourt [2], Walter Chang [2]

[1] UIUC, [2] Adobe Research


## Introduction

- Event coreference resolution is the task of clustering event mentions in a text that refer to the same real-world events.
- *Is using BERT-based models alone enough for event coreference resolution?*
- We argue that it is still highly beneficial to utilize symbolic features (e.g., event types, attributes, arguments) for the task.

... we are seeing these soldiers $\{$head out$\}_{\text{ev1}}$ ...
... these soldiers **were set to** $\{$leave$\}_{\text{ev2}}$ in January ...
ev1 (`Movement:Transport`): Modality = ASSERTED
ev2 (`Movement:Transport`): Modality = OTHER

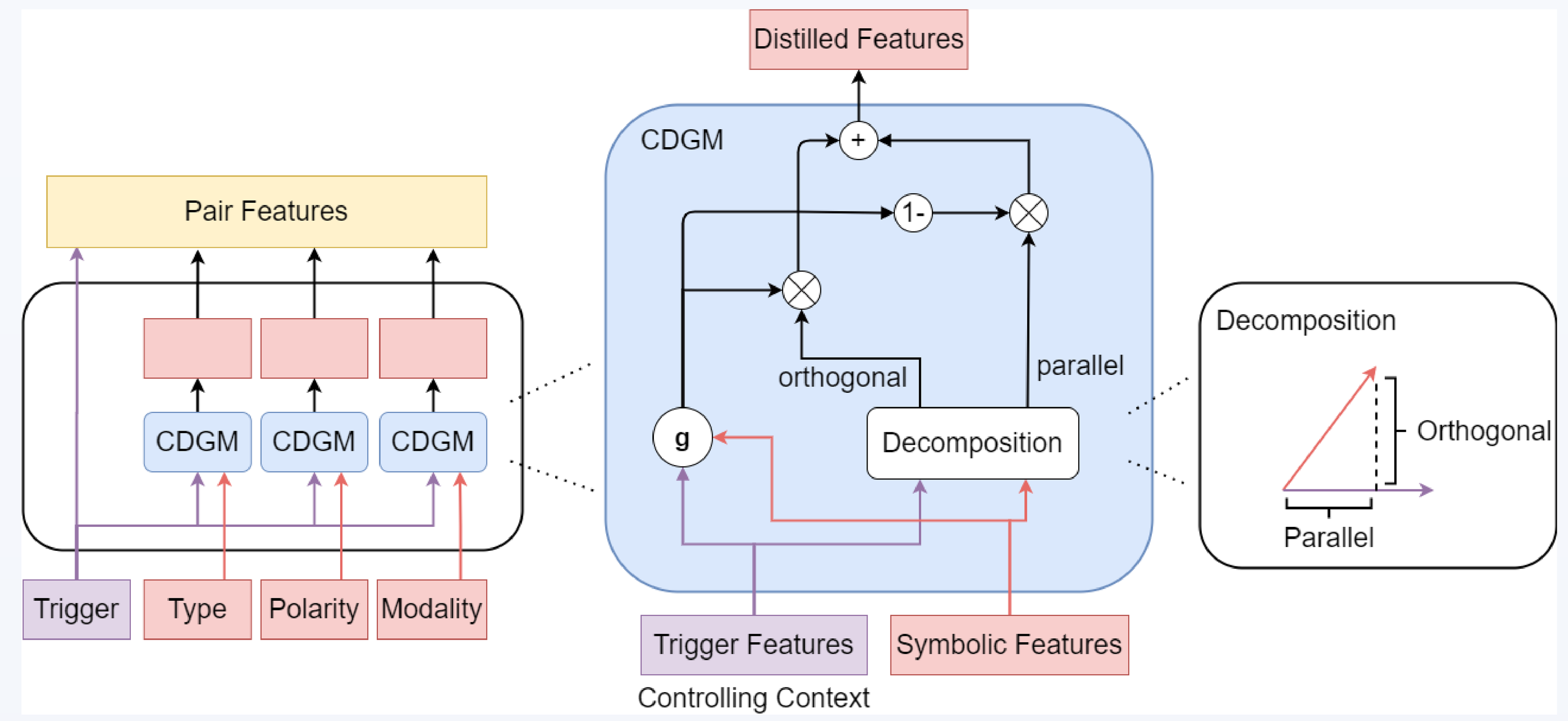
An example of using the modality attribute to improve coreference resolution

## Contributions

- However, the automatically extracted symbolic features can be **noisy** and **contain errors.**
- We propose
  1. **A novel gated module** to adaptively control the information flows from the input symbolic features.
  2. **A simple regularization method** to intentionally add more noise to the symbolic features during training.

## Context-Dependent Gated Module (CDGM)

- Trigger features are used as the main controlling context.
- CDGM takes trigger features and symbolic features as inputs and returns the distilled version of the symbolic features.
- The distilled representation is a combination of:
  - A parallel component ~ "old" information
  - An orthogonal component ~ "new" information




## Introduce Random Noise in Training

- The quality of event attribute prediction on training data is much higher than that on test data.

→ Intentionally add noise to symbolic features during training.

- For epsilon% of the time, change the predicted attribute label to some different label.
  - Epsilon is larger when the performance of classifier is lower

| Dataset | Features | Categorical Values | Acc. (Train) | Acc. (Dev) | Acc. (Test) |
|---|---|---|---|---|---|
| ACE 2005 | Event Type | 33 subtypes | 0.999 | 0.945 | 0.953 |
| | Polarity | Positive, Negative | 0.999 | 0.994 | 0.988 |
| | Modality | Asserted, Negative | 0.999 | 0.856 | 0.884 |
| | Genericity | Generic, Specific | 0.999 | 0.865 | 0.872 |
| | Tense | Past, Present, Future, Unspecified | 0.984 | 0.802 | 0.763 |
| KBP 2016 | Type | 38 subtypes | 0.960 | 0.874 | 0.818 |
| | Realis | Actual, Generic, Other | 0.979 | 0.845 | 0.840 |

List of symbolic features we consider in this work

**Algorithm 1:** Noise Addition for Symbolic Features

**Input:** Document $D$
**Hyperparameters**: $\{\epsilon_1, \epsilon_2, \cdots, \epsilon_K\}$
**for** $i = 1 \ \dots \ k$ **do**
  **for** $u = 1 \ \dots \ K$ **do**
    With prob. $\epsilon_u$, replace $c_i^{(u)}$ by $\hat{c}_i^{(u)} \sim Uniform(N_u)$
  **end**
**end**

| Notation | Description |
|---|---|
| $D$ | The input document |
| $k$ | Number of event mentions in the input document |
| $K$ | Number of categorical symbolic features |
| $c_i^{(u)}$ | The original (predicted) value of the $u$-th symbolic feature of the $i$-th event mention. |
| $\hat{c}_i^{(u)}$ | The newly sampled value for $c_i^{(u)}$ |
| $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_K\}$ | Hyperparameters controlling the noise addition frequencies. |

## End-to-End Results

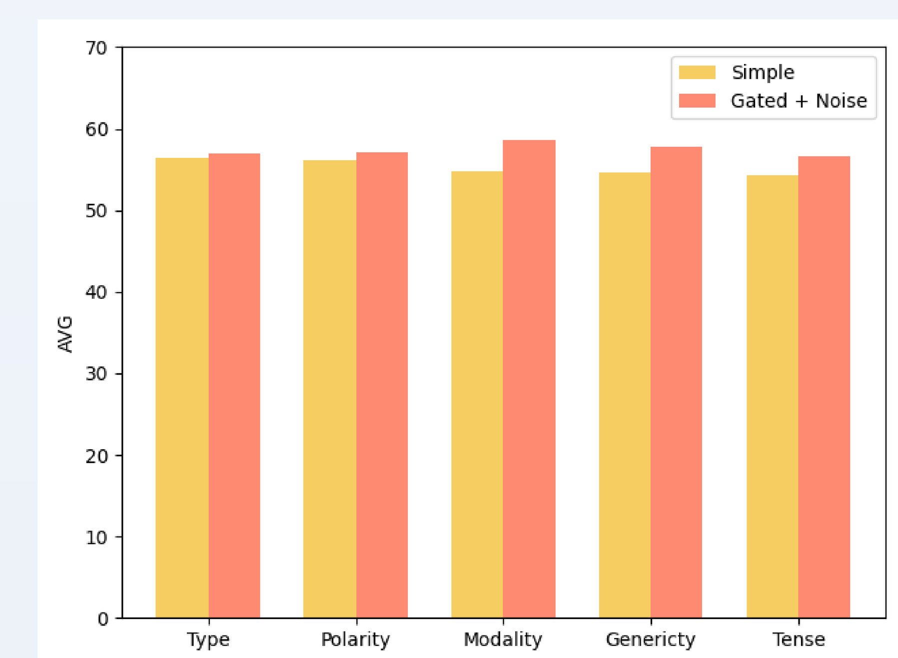
| | CoNLL | AVG |
|---|---|---|
| Joint Learning (Lu et al. 2017) | 35.77 | 33.08 |
| E3C (Lu et al. 2020) | 41.97 | 38.66 |
| SpanBERT | 40.57 | 37.59 |
| SpanBERT + All Features (Simple Concatenation) | 41.40 | 38.58 |
| SpanBERT + All Features (CDGM + Noisy Training) | **43.55** | **40.61** |

Overall F-score (%) on KBP 2016 Dataset

| | CoNLL | AVG |
|---|---|---|
| SSED + MSEP (Peng et al. 2016) | 53.80 | 51.38 |
| SpanBERT | 58.93 | 55.78 |
| SpanBERT + All Features (Simple Concatenation) | 57.55 | 54.79 |
| SpanBERT + All Features (CDGM) | 58.99 | 56.32 |
| SpanBERT + All Features (Noisy Training) | 60.43 | 57.85 |
| SpanBERT + All Features (CDGM + Noisy Training) | **62.07** | **59.76** |

Overall F-score (%) on ACE 2005 Dataset

## Contributions of Different Symbolic Features




Our methods consistently perform better than the simple concatenation strategy across all feature types

Please refer to our original paper for more analysis and results.