IT UNIVERSITY OF COPENHAGEN

Biomedical event extraction as sequence labeling

A. Ramponi,[1,2] R. van der Goot,[3] R. Lombardo,[2] B. Plank[3]

EMNLP 2020

[1] *Department of Information Engineering and Computer Science, University of Trento*

[2] *Fondazione The Microsoft Research – University of Trento Centre for Computational and Systems Biology*

[3] *Department of Computer Science, IT University of Copenhagen*

Biomedical event extraction

Bio-event: biomedical "happening" involving bio-entities

entities:	PROTEIN					PROTEIN
text:	*STAT-4*	*activation*	*promotes*	*production*	*of*	*IL-10*

Biomedical event extraction

Bio-event: biomedical "happening" involving bio-entities

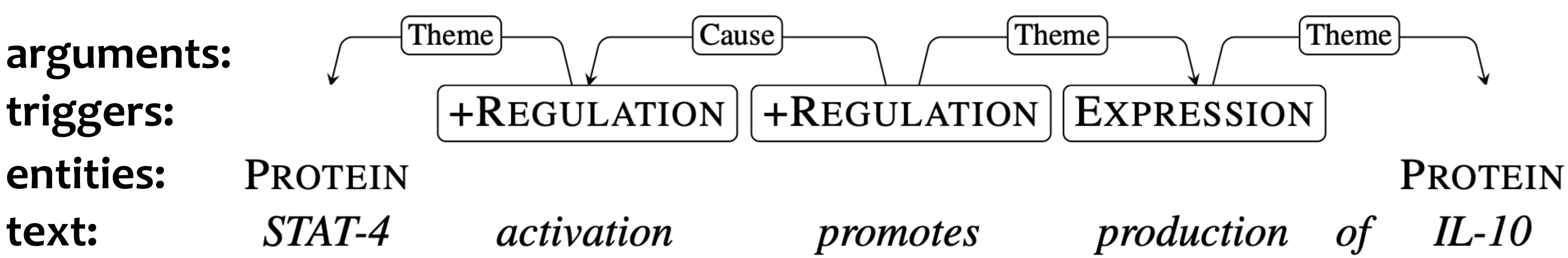
- **Triggers**: center of events, with a semantic type
  - e.g., "production" triggers an Expression event

triggers:		+REGULATION	+REGULATION	EXPRESSION		
entities:	PROTEIN					PROTEIN
text:	*STAT-4*	*activation*	*promotes*	*production*	*of*	*IL-10*

Biomedical event extraction

Bio-event: biomedical "happening" involving bio-entities

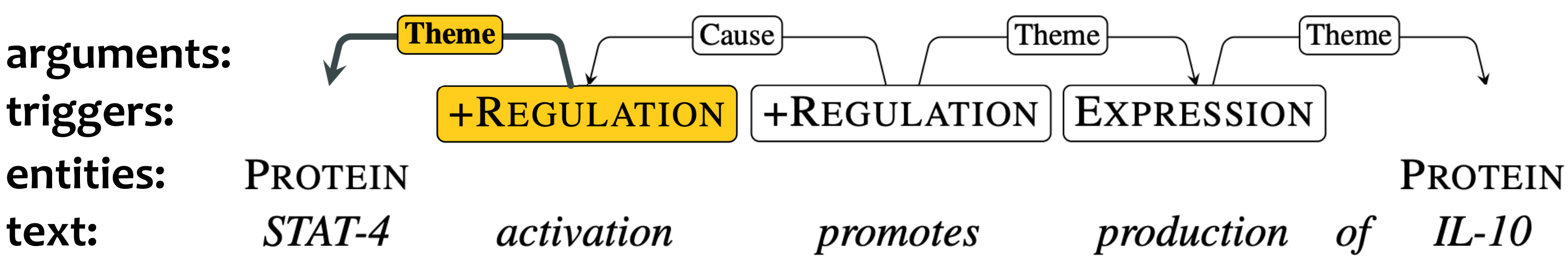
- **Triggers**: center of events, with a semantic type
  - e.g., "production" triggers an Expression event
- **Arguments:** participants of events, with a semantic type
  - e.g., "IL-10" is a Theme of the Expression event



Biomedical event extraction

Bio-event: biomedical "happening" involving bio-entities

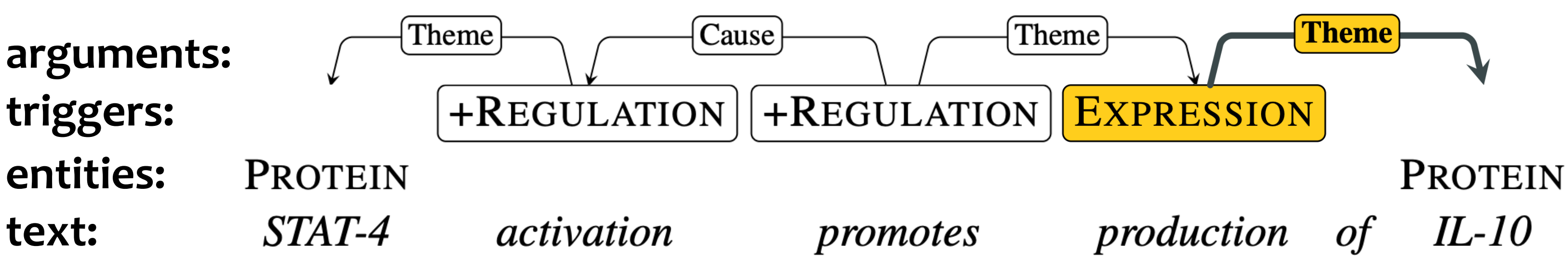
- **Triggers**: center of events, with a semantic type
  - e.g., "production" triggers an <u>Expression</u> event
- **Arguments:** participants of events, with a semantic type
  - e.g., "IL-10" is a <u>Theme</u> of the Expression event



Biomedical event extraction

Bio-event: biomedical "happening" involving bio-entities

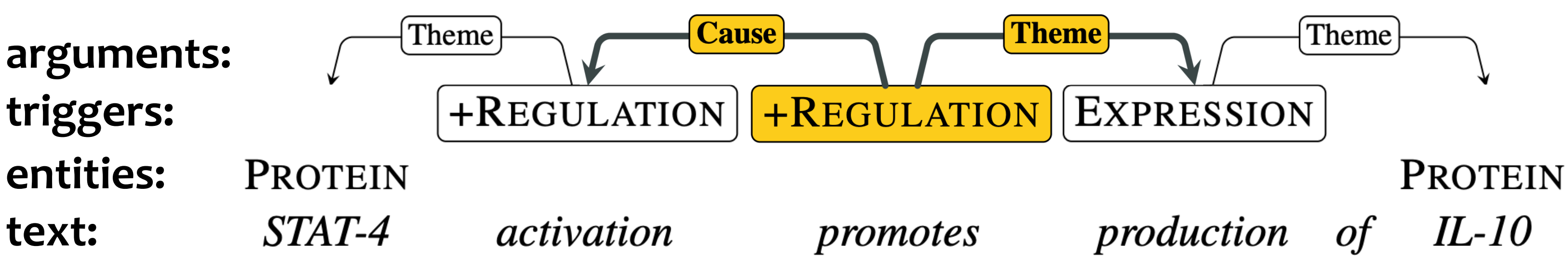
- **Triggers**: center of events, with a semantic type
  - e.g., "production" triggers an <u>Expression</u> event
- **Arguments:** participants of events, with a semantic type
  - e.g., "IL-10" is a <u>Theme</u> of the Expression event



Biomedical event extraction

Bio-event: biomedical “happening” involving bio-entities

- **Triggers**: center of events, with a semantic type
  - e.g., “production” triggers an Expression event
- **Arguments:** participants of events, with a semantic type
  - e.g., “IL-10” is a Theme of the Expression event



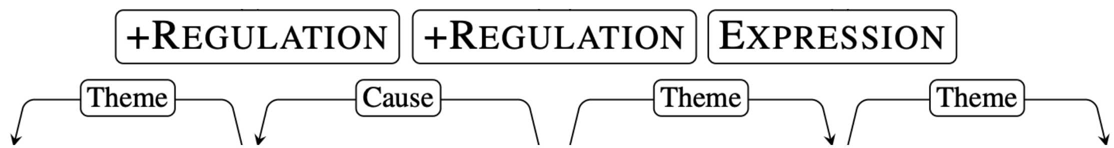
Motivation

SOTA systems work as locally-optimized **classifier pipelines**

- e.g., CNN (Bjorne and Salakoski, 2018), KB TreeLSTM (Li et al., 2019)

Classifier$_1$ (*triggers*)

Classifier$_2$ (*arguments*)



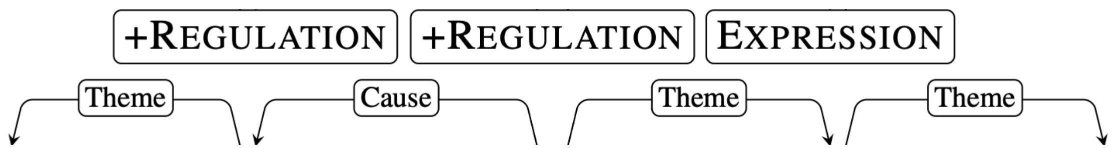
Motivation

SOTA systems work as locally-optimized **classifier pipelines**

- e.g., CNN (Bjorne and Salakoski, 2018), KB TreeLSTM (Li et al., 2019)

Classifier$_1$ (*triggers*)

Classifier$_2$ (*arguments*)



Biomedical **E**vent **E**xtraction as **S**equence **L**abeling ( )

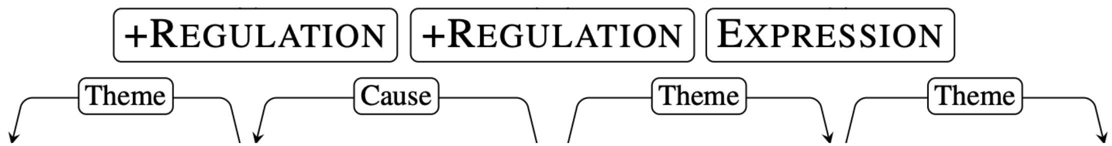
Motivation

SOTA systems work as locally-optimized **classifier pipelines**

- e.g., CNN (Bjorne and Salakoski, 2018), KB TreeLSTM (Li et al., 2019)

Classifier$_1$ (*triggers*)

Classifier$_2$ (*arguments*)



Biomedical **E**vent **E**xtraction as **S**equence **L**abeling ( )

- Linearization of event structures as word-level tagging

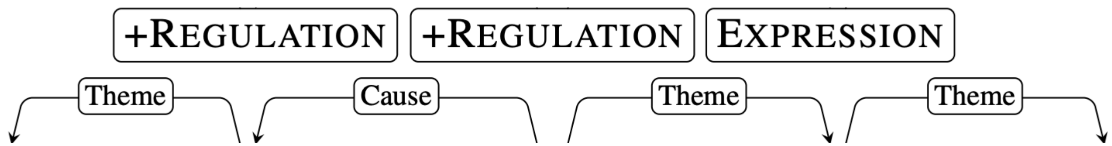
Motivation

SOTA systems work as locally-optimized **classifier pipelines**

- e.g., CNN (Bjorne and Salakoski, 2018), KB TreeLSTM (Li et al., 2019)

Classifier$_1$ (*triggers*)

Classifier$_2$ (*arguments*)



Biomedical **E**vent **E**xtraction as **S**equence **L**abeling ( )

- Linearization of event structures as word-level tagging
- Joint modeling of triggers and arguments via multi-task learning

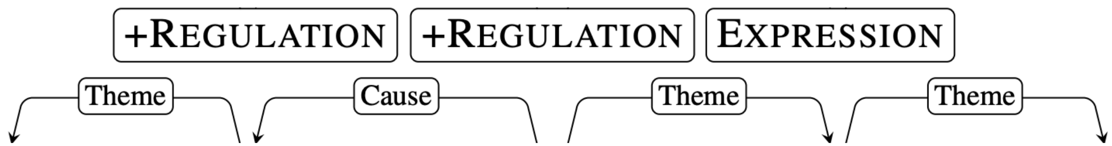
Motivation

SOTA systems work as locally-optimized **classifier pipelines**

- e.g., CNN (Bjorne and Salakoski, 2018), KB TreeLSTM (Li et al., 2019)

Classifier$_1$ (*triggers*)

Classifier$_2$ (*arguments*)



Biomedical **E**vent **E**xtraction as **S**equence **L**abeling ( )

- Linearization of event structures as word-level tagging
- Joint modeling of triggers and arguments via multi-task learning
- Handling of multiple labels per token via multi-label decoding

Linearization of event structures

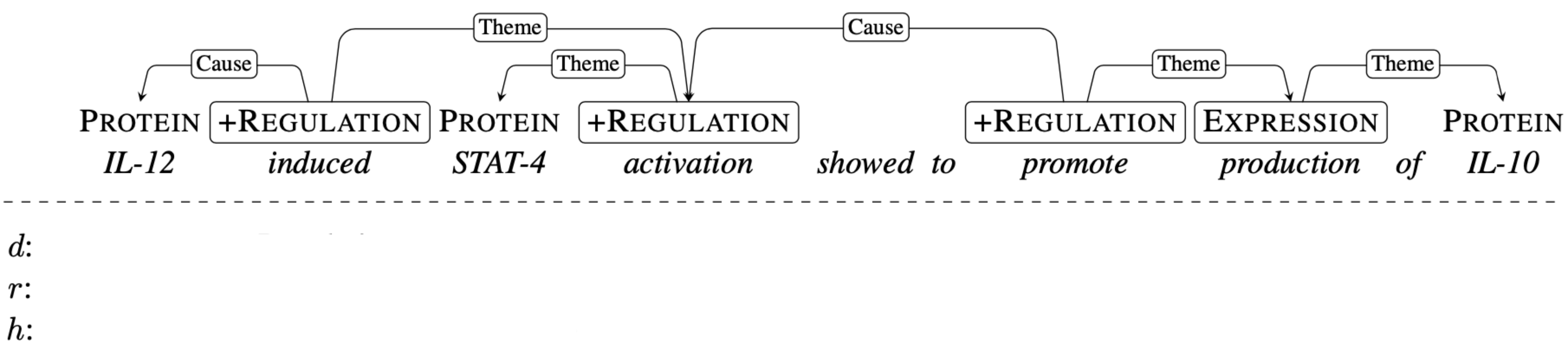
Each token x_i has a structured label $y_i = \langle d, r, h \rangle$, where:

- $\boldsymbol{d}$ (*dependent*): mention type of the token
- $\boldsymbol{r}$ (*relation*): argument role type of the token
- $\boldsymbol{h}$ (*head*): event type and position the token is argument

Linearization of event structures

Each token x_i has a structured label $y_i = \langle d, r, h \rangle$, where:

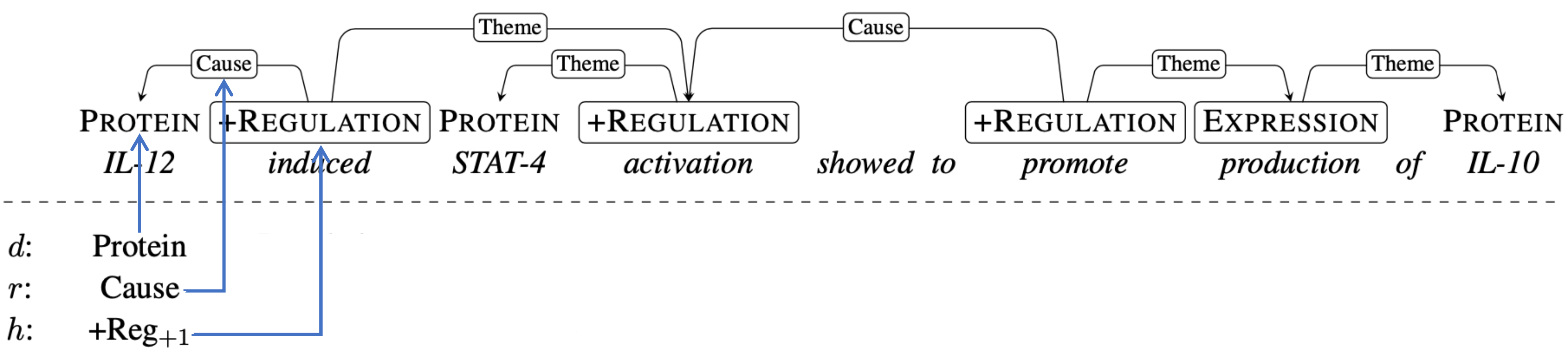
- $\boldsymbol{d}$ (*dependent*): mention type of the token
- $\boldsymbol{r}$ (*relation*): argument role type of the token
- $\boldsymbol{h}$ (*head*): event type and position the token is argument



Linearization of event structures

Each token x_i has a structured label $y_i = \langle d, r, h \rangle$, where:

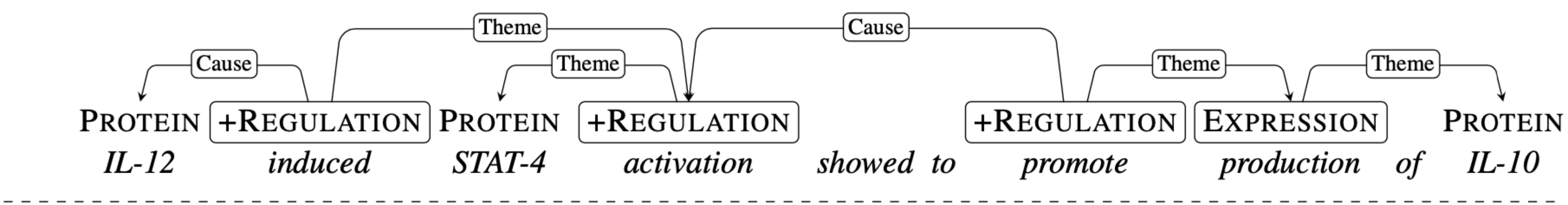
- $\boldsymbol{d}$ (*dependent*): mention type of the token
- $\boldsymbol{r}$ (*relation*): argument role type of the token
- $\boldsymbol{h}$ (*head*): event type and position the token is argument



Linearization of event structures

Each token x_i has a structured label $y_i = \langle d, r, h \rangle$, where:

- $\boldsymbol{d}$ (*dependent*): mention type of the token
- $\boldsymbol{r}$ (*relation*): argument role type of the token
- $\boldsymbol{h}$ (*head*): event type and position the token is argument

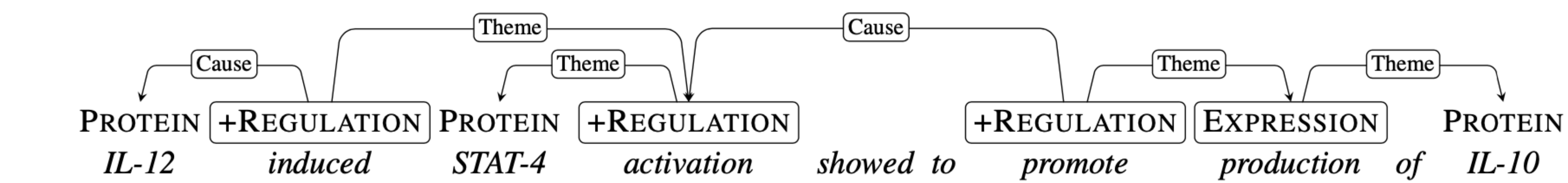


d:	Protein	+Regulation
r:	Cause	
h:	+Reg$_{+1}$	

Linearization of event structures

Each token x_i has a structured label $y_i = \langle d, r, h \rangle$, where:

- $\boldsymbol{d}$ (*dependent*): mention type of the token
- $\boldsymbol{r}$ (*relation*): argument role type of the token
- $\boldsymbol{h}$ (*head*): event type and position the token is argument

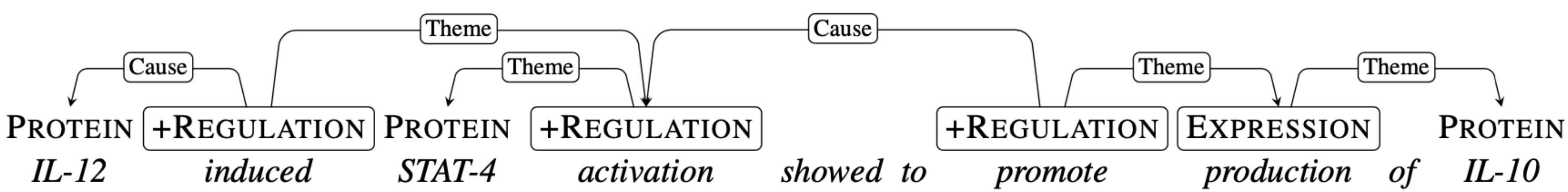


d:	Protein	+Regulation	Protein
r:	Cause		Theme
h:	+Reg$_{+1}$		+Reg$_{+1}$

Linearization of event structures

Each token x_i has a structured label $y_i = \langle d, r, h \rangle$, where:

- $\boldsymbol{d}$ (*dependent*): mention type of the token
- $\boldsymbol{r}$ (*relation*): argument role type of the token
- $\boldsymbol{h}$ (*head*): event type and position the token is argument



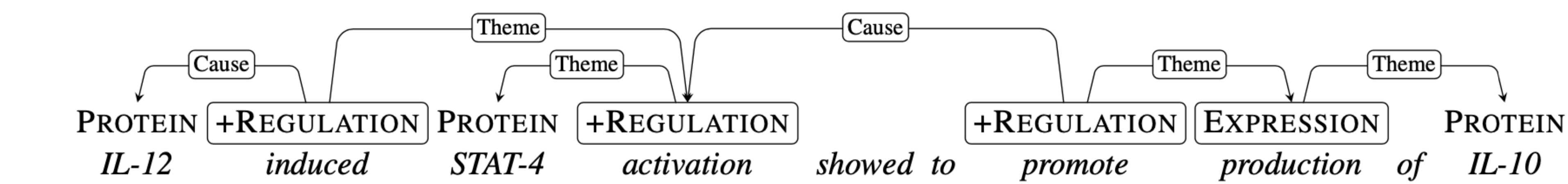
d:	Protein	+Regulation	Protein	+Regulation
r:	Cause		Theme	Theme, Cause
h:	$+\text{Reg}_{+1}$		$+\text{Reg}_{+1}$	$+\text{Reg}_{-1}$, $+\text{Reg}_{+1}$

multi-label!

Linearization of event structures

Each token x_i has a structured label $y_i = \langle d, r, h \rangle$, where:

- $\boldsymbol{d}$ (*dependent*): mention type of the token
- $\boldsymbol{r}$ (*relation*): argument role type of the token
- $\boldsymbol{h}$ (*head*): event type and position the token is argument



	IL-12	induced	STAT-4	activation	showed	to	promote	production	of	IL-10
d:	Protein	+Regulation	Protein	+Regulation			+Regulation	Expression		Protein
r:	Cause		Theme	Theme, Cause				Theme		Theme
h:	$+\text{Reg}_{+1}$		$+\text{Reg}_{+1}$	$+\text{Reg}_{-1}$, $+\text{Reg}_{+1}$				$+\text{Reg}_{-1}$		Exp_{-1}

multi-label!

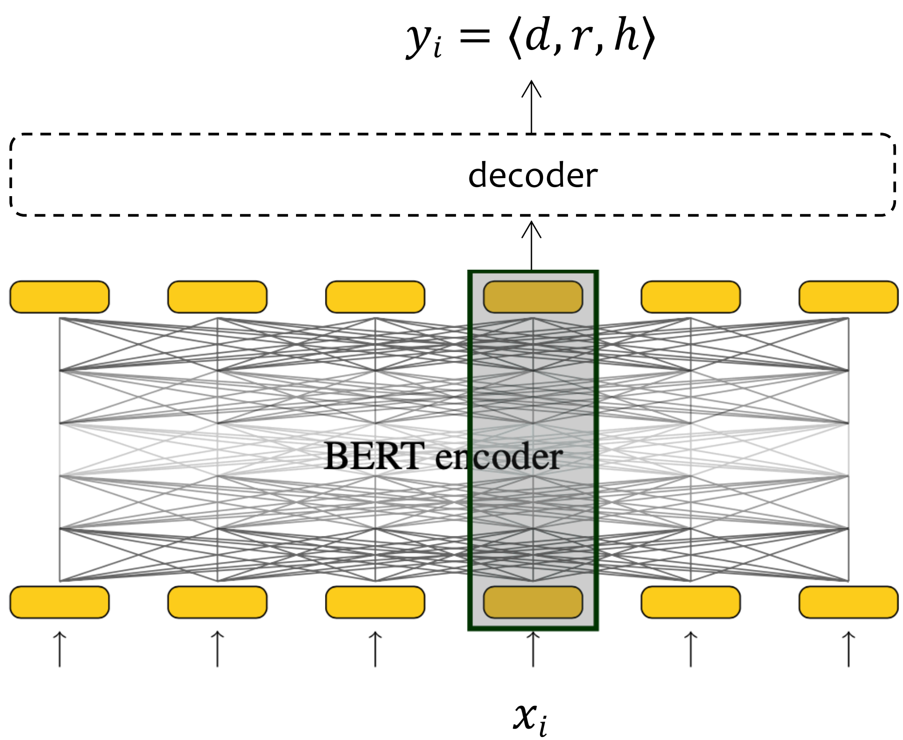
The model and learning strategies

Shared BERT encoder, private decoder(s)

Single task (ST)

- $y_i = \langle d, r, h \rangle$

dependent **r**elation **h**ead



The model and learning strategies

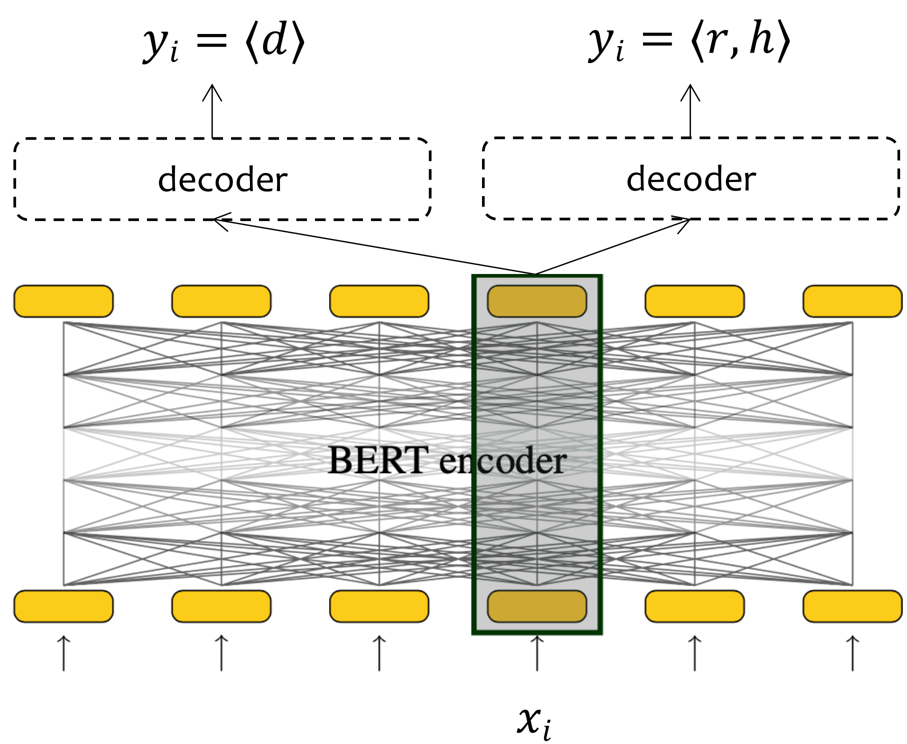
Shared BERT encoder, private decoder(s)

Single task (ST)

- $y_i = \langle d, r, h \rangle$

Multi-task (MT)

- $\boldsymbol{y_i = \langle d \rangle, \langle r, h \rangle}$



The model and learning strategies

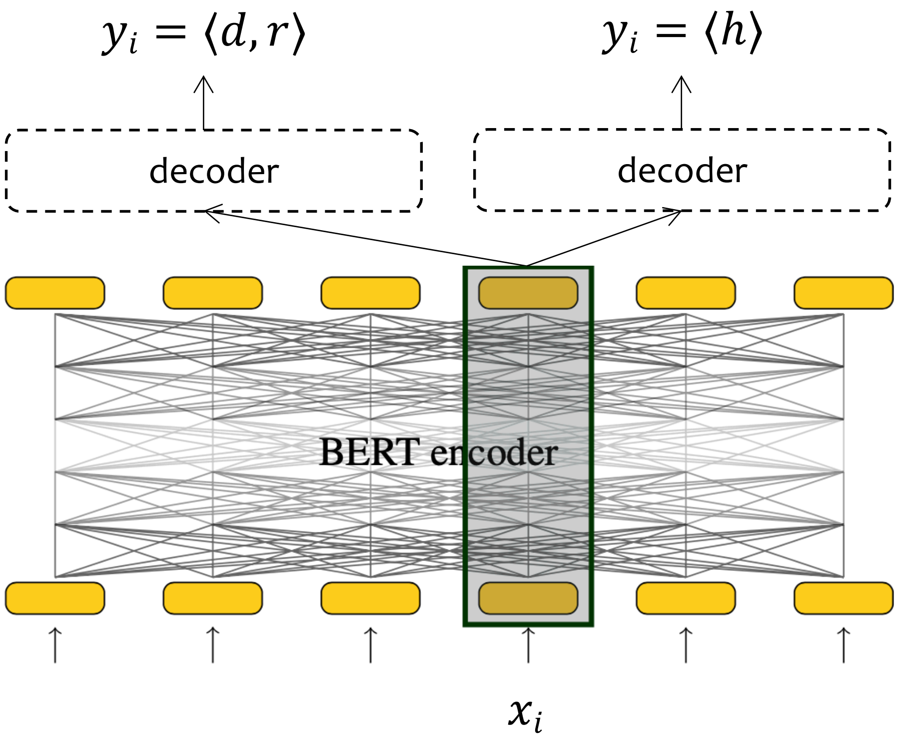
Shared BERT encoder, private decoder(s)

Single task (ST)

- $y_i = \langle d, r, h \rangle$

Multi-task (MT)

- $y_i = \langle d \rangle, \langle r, h \rangle$
- $\boldsymbol{y_i = \langle d, r \rangle, \langle h \rangle}$



The model and learning strategies

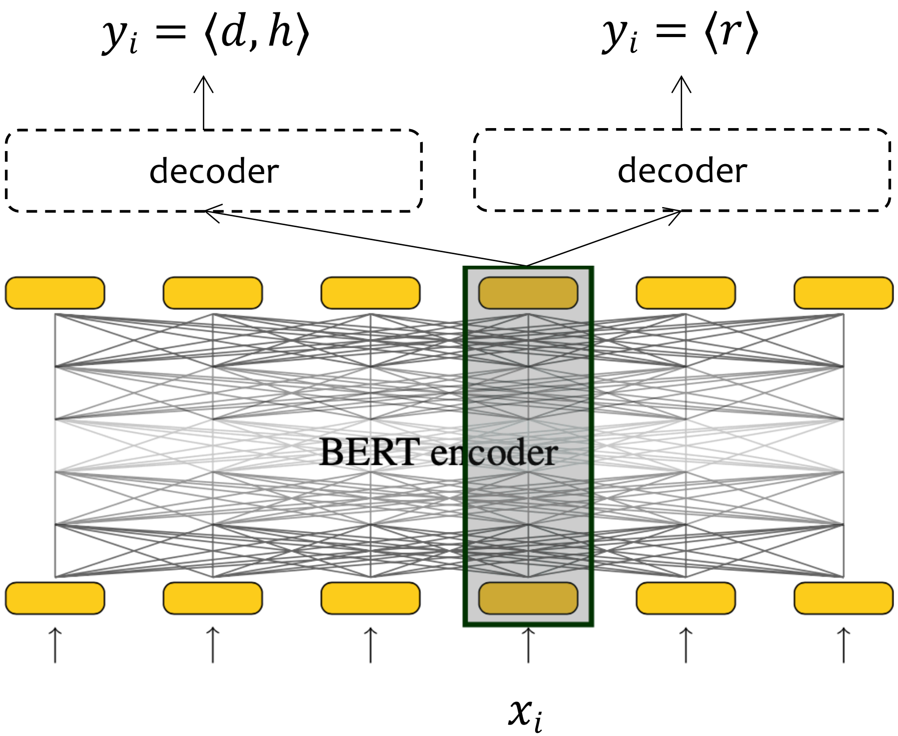
Shared BERT encoder, private decoder(s)

Single task (ST)

- $y_i = \langle d, r, h \rangle$

Multi-task (MT)

- $y_i = \langle d \rangle, \langle r, h \rangle$
- $y_i = \langle d, r \rangle, \langle h \rangle$
- $\boldsymbol{y_i = \langle d, h \rangle, \langle r \rangle}$



The model and learning strategies

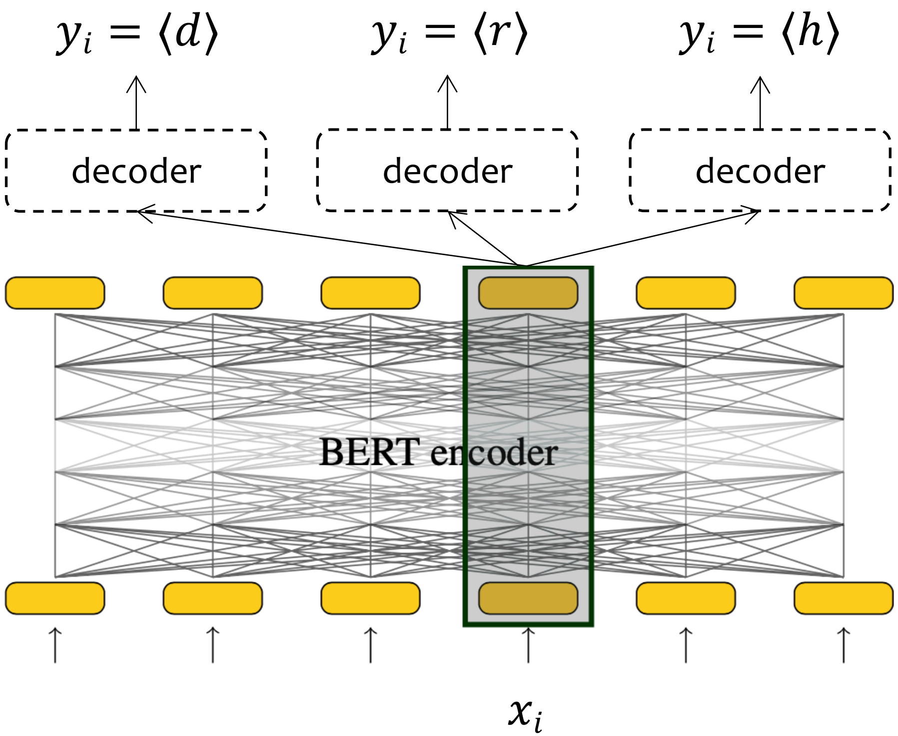
Shared BERT encoder, private decoder(s)

Single task (ST)

- $y_i = \langle d, r, h \rangle$

Multi-task (MT)

- $y_i = \langle d \rangle, \langle r, h \rangle$
- $y_i = \langle d, r \rangle, \langle h \rangle$
- $y_i = \langle d, h \rangle, \langle r \rangle$
- $\boldsymbol{y_i = \langle d \rangle, \langle r \rangle, \langle h \rangle}$

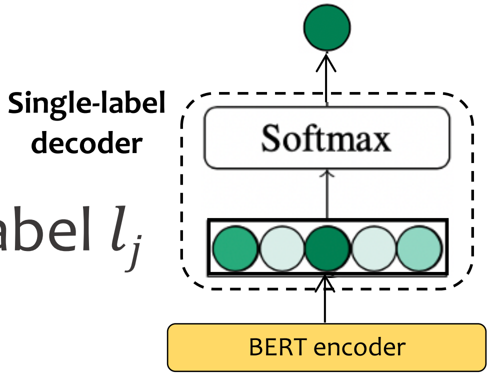


Single- and multi-label decoding

After encoding, each token x_i is given:

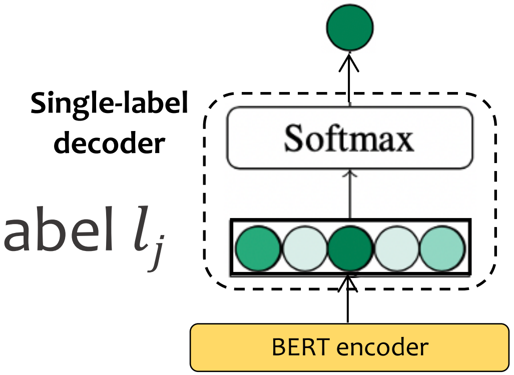
Single-label decoder: the highest scoring label l_j

- Suitable for predicting dependent d



Single- and multi-label decoding

After encoding, each token x_i is given:

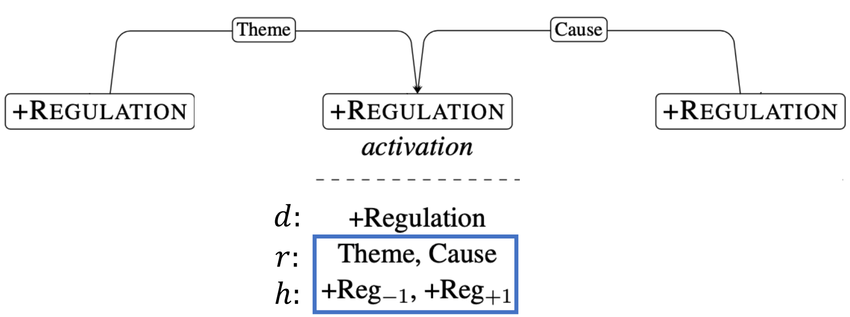


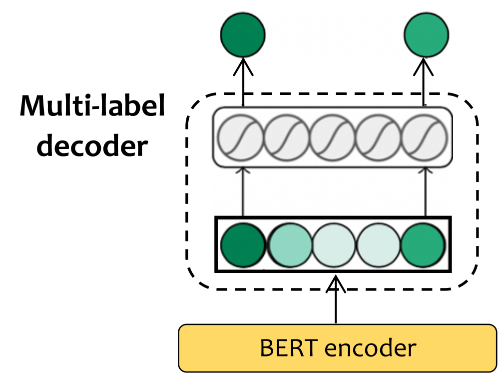
Single-label decoder: the highest scoring label l_j

- Suitable for predicting dependent d

Multi-label decoder: all labels l_j with probability $P(l_j) > \tau$

- Suitable for predicting relation r and head h





Experiments and evaluation

Genia11 benchmark

- Largest biomedical event extraction dataset
- Both abstract and full-text documents
- Simple and complex events (37.2% nested)

Experiments and evaluation

Genia11 benchmark

- Largest biomedical event extraction dataset
- Both abstract and full-text documents
- Simple and complex events (37.2% nested)

Model selection

- Single-task vs multi-task strategies
- Addition of multi-label decoding

Experiments and evaluation

Genia11 benchmark

- Largest biomedical event extraction dataset
- Both abstract and full-text documents
- Simple and complex events (37.2% nested)

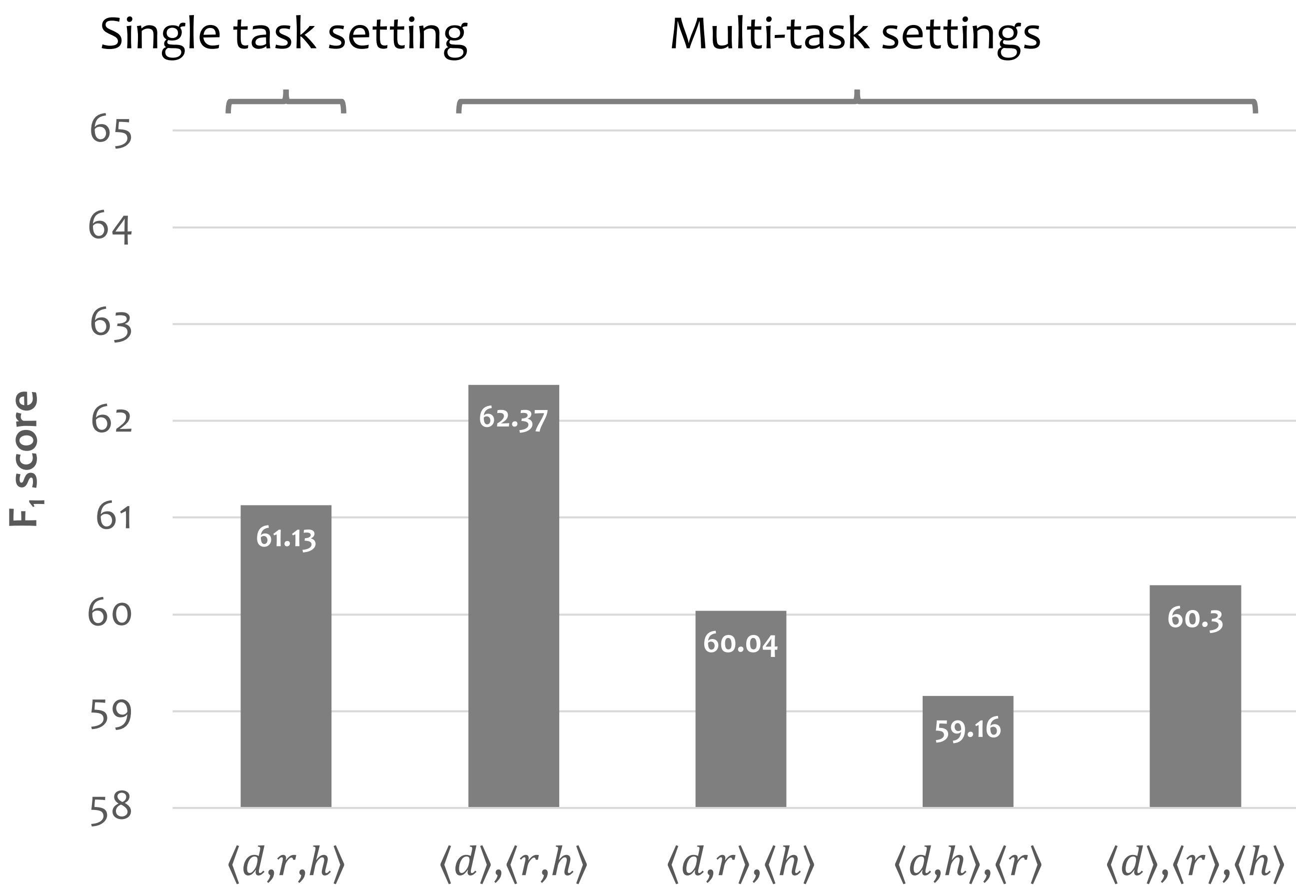
Model selection

- Single-task vs multi-task strategies
- Addition of multi-label decoding

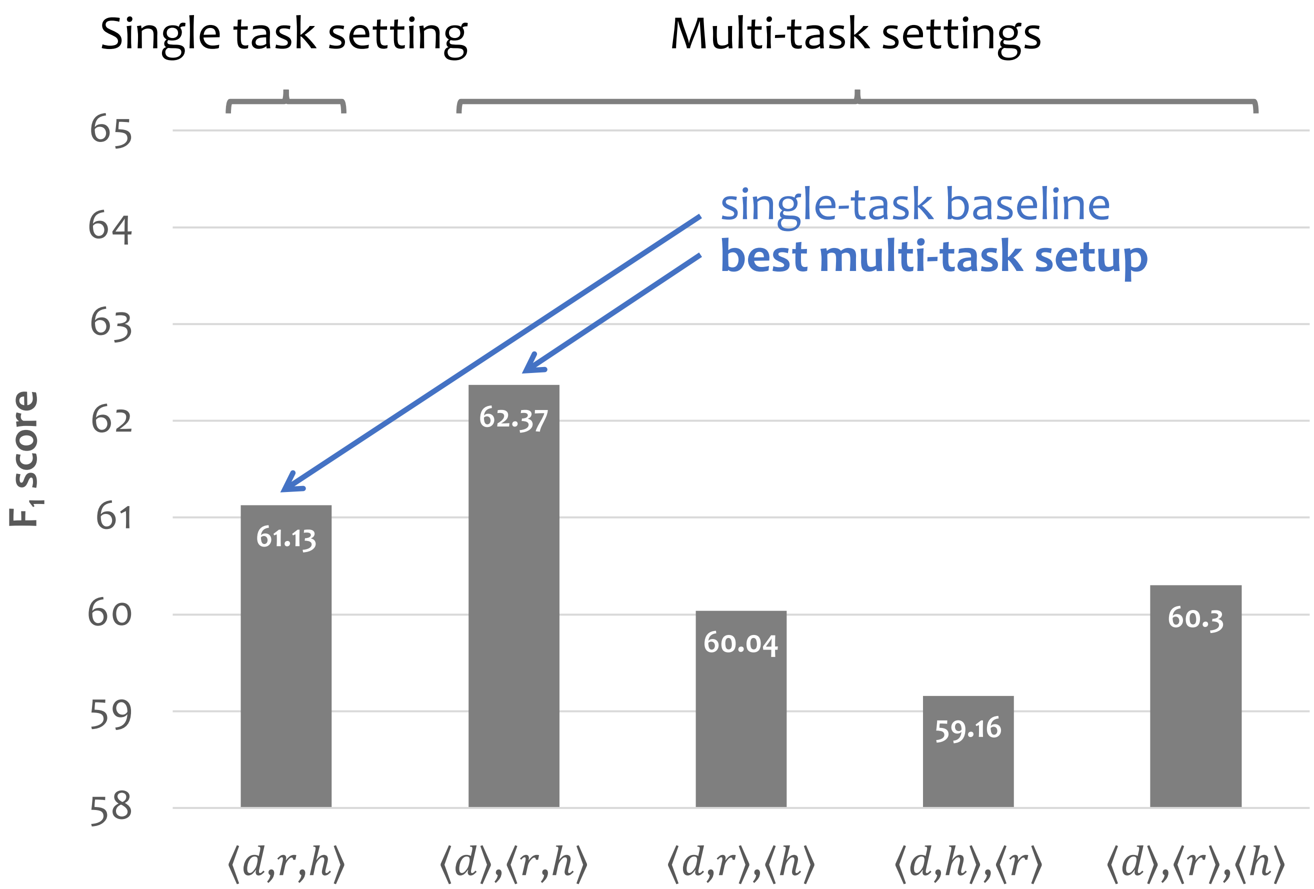
Evaluation

- Accuracy and speed comparison

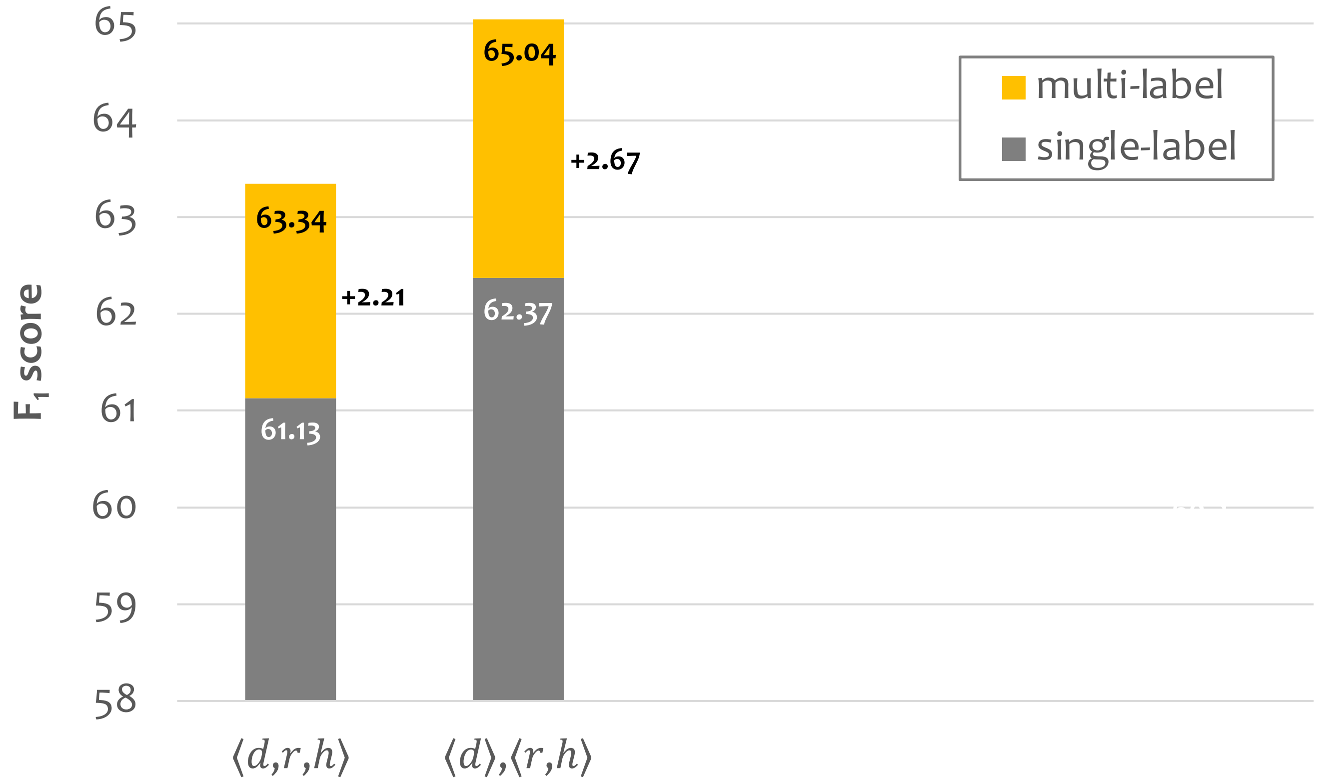
Model selection



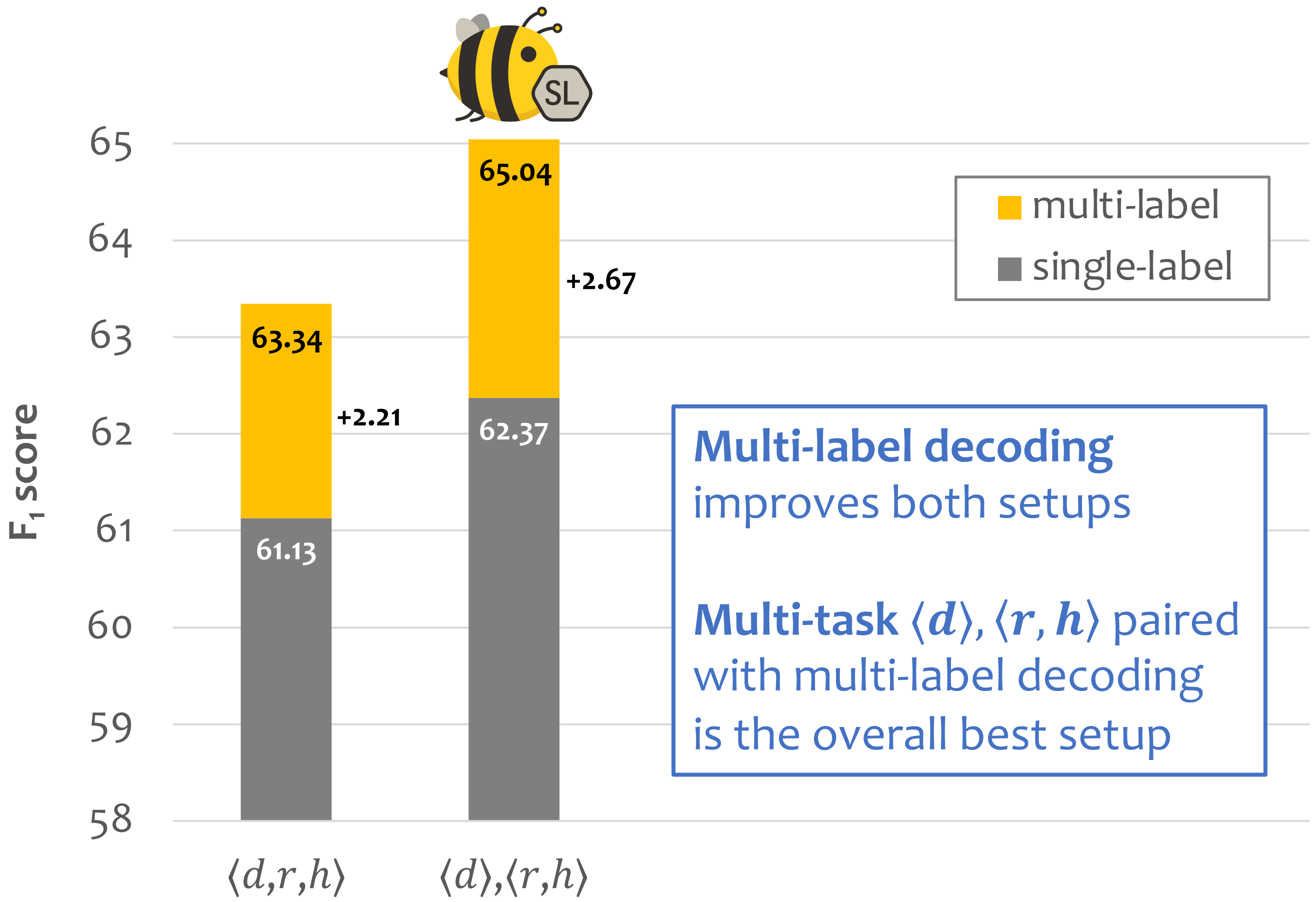
Model selection



Model selection



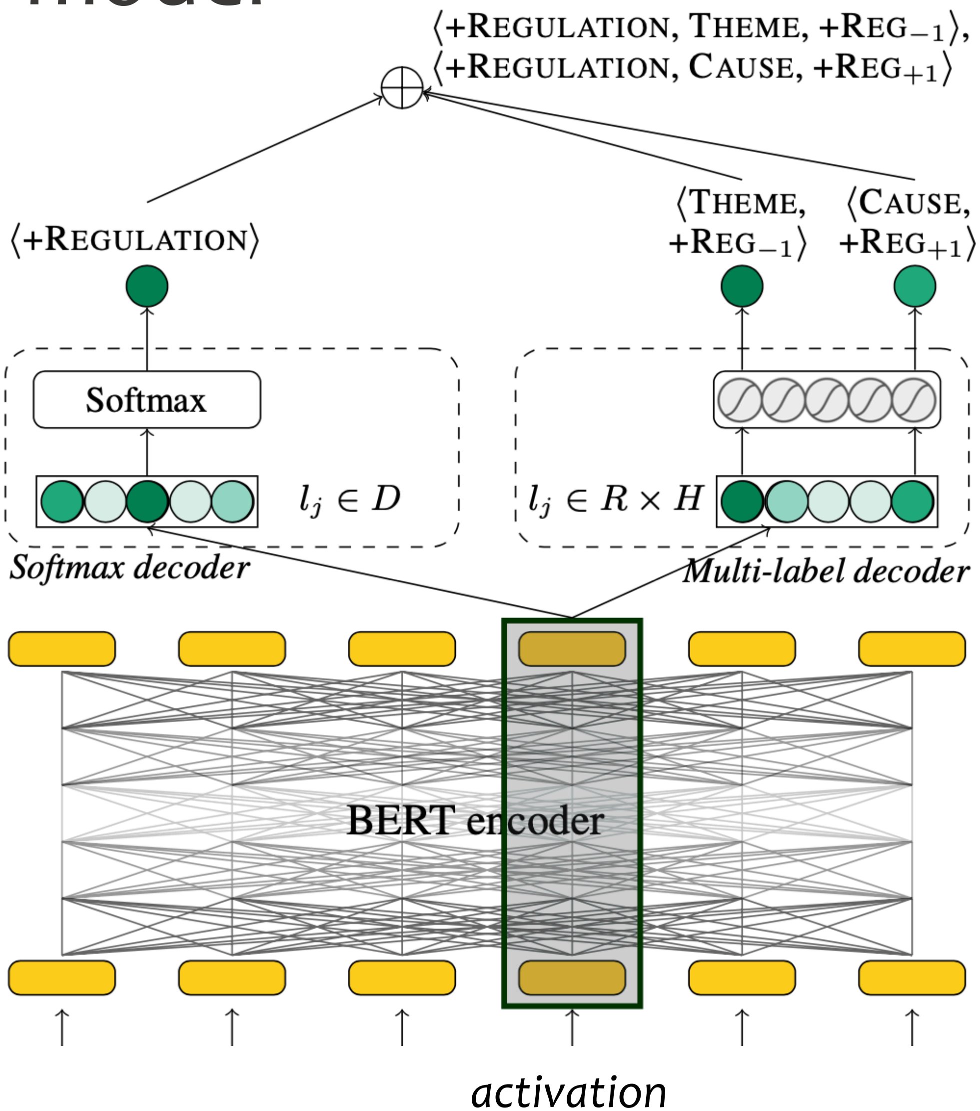
Model selection



Multi-label decoding improves both setups

Multi-task $\langle \boldsymbol{d} \rangle, \langle \boldsymbol{r}, \boldsymbol{h} \rangle$ paired with multi-label decoding is the overall best setup

The final model



Comparison to the state-of-the-art

Work	Method	P	R	F1
Riedel et al. (2011)	FAUST – Model combination (joint+parsing)	64.75	49.41	56.04
Miwa et al. (2012)	EventMine – SVM pipeline (+coref)	63.48	53.35	57.98
Venugopal et al. (2014)	BioMLN – SVM pipeline & MLN (joint)	63.61	53.42	58.07
Majumder et al. (2016)	Stacked generalization	66.46	48.96	56.38
Björne and Salakoski (2018)	TEES – CNN pipeline (single model)	64.86	50.53	56.80
Björne and Salakoski (2018)	TEES – CNN pipeline (5x ensemble)	68.76	49.97	57.87
Björne and Salakoski (2018)	TEES – CNN pipeline (mixed 5x ensemble)	69.45	49.94	58.10
Li et al. (2019)	BiLSTM pipeline	62.18	48.44	54.46
Li et al. (2019)	Tree-LSTM pipeline	64.56	50.28	56.53
Li et al. (2019)	KB-driven Tree-LSTM pipeline	67.01	52.14	58.65
BEESL	Multi-task neural sequence labeling	69.72	53.00	**60.22**

Comparison to the state-of-the-art

Work	Method	P	R	F1
Riedel et al. (2011)	FAUST – Model combination (joint+parsing)	64.75	49.41	56.04
Miwa et al. (2012)	EventMine – SVM pipeline (+coref)	63.48	53.35	57.98
Venugopal et al. (2014)	BioMLN – SVM pipeline & MLN (joint)	63.61	53.42	58.07
Majumder et al. (2016)	Stacked generalization	66.46	48.96	56.38
Björne and Salakoski (2018)	TEES – CNN pipeline (single model)	64.86	50.53	56.80
Björne and Salakoski (2018)	TEES – CNN pipeline (5x ensemble)	68.76	49.97	57.87
Björne and Salakoski (2018)	TEES – CNN pipeline (mixed 5x ensemble)	69.45	49.94	58.10
Li et al. (2019)	BiLSTM pipeline	62.18	48.44	54.46
Li et al. (2019)	Tree-LSTM pipeline	64.56	50.28	56.53
Li et al. (2019)	KB-driven Tree-LSTM pipeline	67.01	52.14	58.65
BEESL	Multi-task neural sequence labeling	69.72	53.00	**60.22**

	sents/min
TEES (*single*)	$255_{\pm 1}$
TEES (*ensemble*)	$101_{\pm 1}$
BEESL	$499_{\pm 3}$

*on a consumer grade CPU

Analysis and discussion

Joint modeling via multi-task learning is important

- +1.57% compared to two independent classifiers $\langle d \rangle \rightarrow \langle r, h \rangle$

Analysis and discussion

Joint modeling via multi-task learning is important

- +1.57% compared to two independent classifiers $\langle d \rangle \rightarrow \langle r, h \rangle$

Multi-label decoding is stable

- Values of the threshold τ in $[0.3, 0.7]$ only minimally alter results

Analysis and discussion

Joint modeling via multi-task learning is important

- +1.57% compared to two independent classifiers $\langle d \rangle \rightarrow \langle r, h \rangle$

Multi-label decoding is stable

- Values of the threshold τ in $[0.3, 0.7]$ only minimally alter results

Analysis of the errors

- Mainly due to ambiguous or generic words acting as event triggers

Analysis and discussion

Joint modeling via multi-task learning is important

- +1.57% compared to two independent classifiers $\langle d \rangle \rightarrow \langle r, h \rangle$

Multi-label decoding is stable

- Values of the threshold τ in $[0.3, 0.7]$ only minimally alter results

Analysis of the errors

- Mainly due to ambiguous or generic words acting as event triggers

Impact of non-gold entity mentions

- Empirical results show the robustness to noisy, predicted entities

Summary and conclusions

- From a highly structured problem to a tagging problem
  - Novel linearization approach of event structures
  - Joint learning using multi-task learning and multi-label decoding

Summary and conclusions

- From a highly structured problem to a tagging problem
  - Novel linearization approach of event structures
  - Joint learning using multi-task learning and multi-label decoding

- Accurate and efficient solution
  - SOTA results on standard GENIA 11 benchmark (**+1.57% F_1**)
  - High speed efficiency (**5x sents/min**)
  - Viable solution for large-scale real-world scenarios

Summary and conclusions

- From a highly structured problem to a tagging problem
  - Novel linearization approach of event structures
  - Joint learning using multi-task learning and multi-label decoding
- Accurate and efficient solution
  - SOTA results on standard GENIA 11 benchmark (**+1.57% F_1**)
  - High speed efficiency (**5x sents/min**)
  - Viable solution for large-scale real-world scenarios
- Linearization approach useful for other NLP tasks
  - e.g., enhanced dep. parsing, fine-grained NER, semantic parsing