Why Misinformation is Created?
Detecting them by Integrating Intent Features

Authors: Bing Wang, Ximing Li, Changchun Li, Bo Fu, Songwen Pei, Shengsheng Wang

CIKM 2024
BOISE
OCTOBER 21-25

Background

Social media platforms are inevitably full of misinformation, causing damage

Over-the-counter cold and cough medications are being pulled from drugstore shelves in an effort to start the “next plandemic” or force people to get the COVID-19 vaccine.

Misinformation

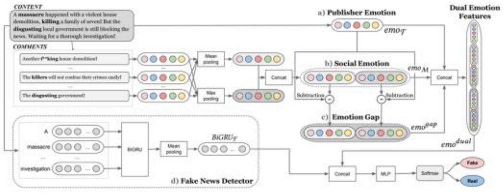
COVID-19 vaccines are safe for people who have existing health conditions, including conditions that have a higher risk of getting serious illness with COVID-19.

Misinformation Detection

Previous Works on MD

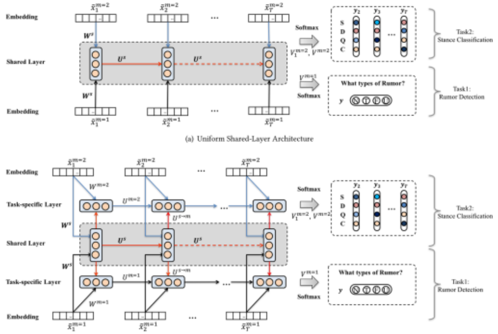
The cutting-edge MD methods extracting more discriminative features by incorporating influential aspects from psychology and sociology perspectives

Emotion Feature



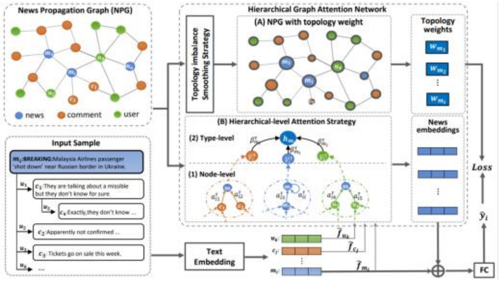
Zhang, Xueyao, et al. "Mining dual emotion for fake news detection." *Proceedings of the web conference 2021*. 2021.

Stance Feature



Ma, Jing, et al. "Detect rumor and stance jointly by neural multi-task learning." Companion proceedings of the the web conference 2018. 2018.

Propagation Structure



Gao, Li, et al. "Topology imbalance and relation inauthenticity aware hierarchical graph attention networks for fake news detection." ACL 2022.

Motivation

Misinformation is created by specific intents, which are often negative, and harmful

Real information is more objective with the **straightforward intent** of sharing

Over-the-counter cold and cough medications are being pulled from drugstore shelves in an effort to start the "next plandemic" or force people to get the COVID-19 vaccine.

Misinformation

Intent: conspiracy theories

COVID-19 vaccines are safe for people who have existing health conditions, including conditions that have a higher risk of getting serious illness with COVID-19.

Intent: popularize commonsense

Motivation

Misinformation is created by specific **intents**, which are often negative, and harmful

Real information is more objective with the **straightforward intent** of sharing

Over-the-counter cold and cough medications

Reason **the intent of articles and form the corresponding** **intent features**

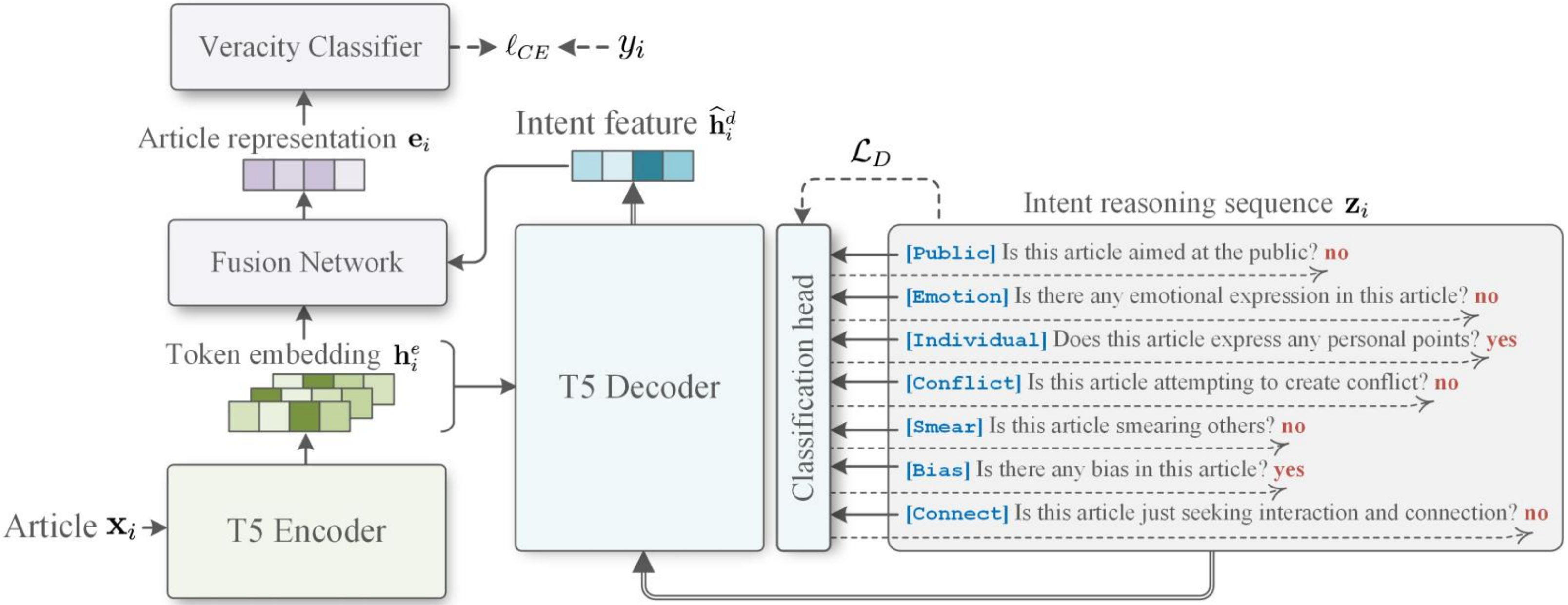
COVID-19 vaccines are safe for people who have existing health conditions, including conditions that have a higher risk of getting serious illness with COVID-19.

Intent: popularize commonsense

Method: Detecting Misinformation by Integrating Intent featuRes

✓ **Four components of DM-INTER:**

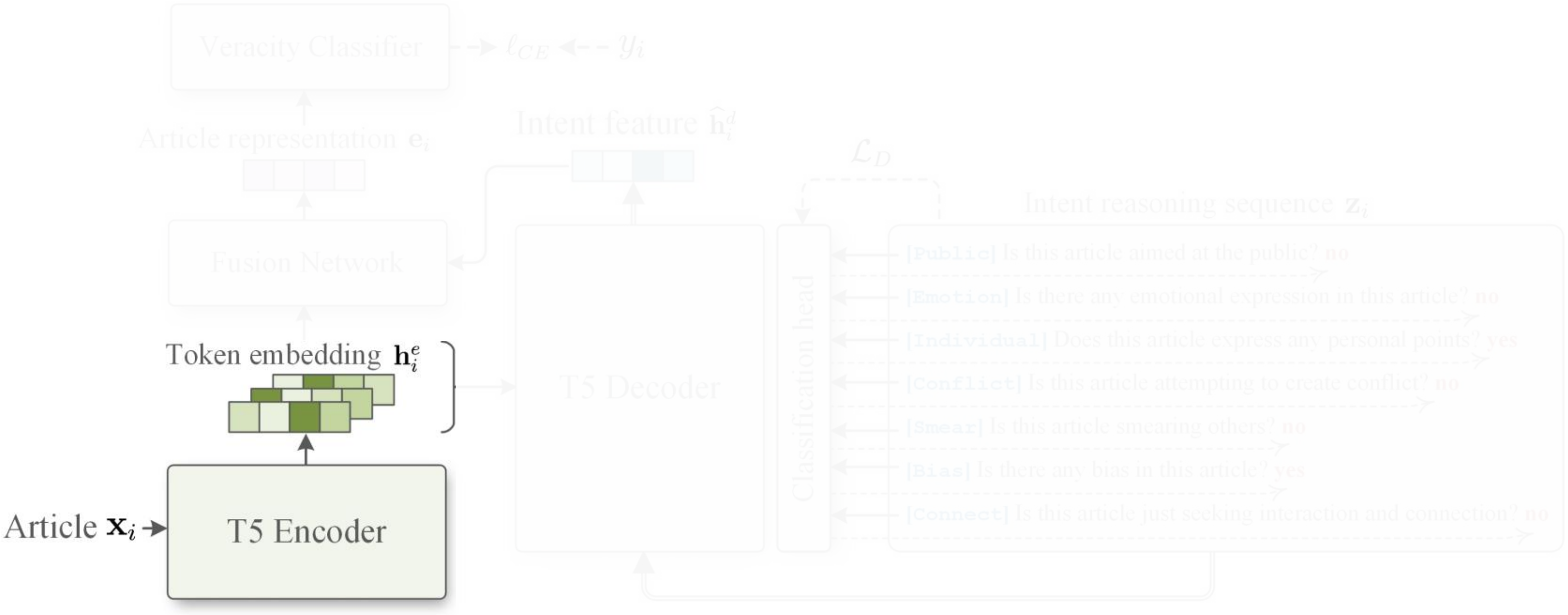
❶LM encoder, ❷LM decoder, ❸fusion network, ❹veracity classifier



Method: Detecting Misinformation by Integrating Intent featuRes

❶ LM encoder

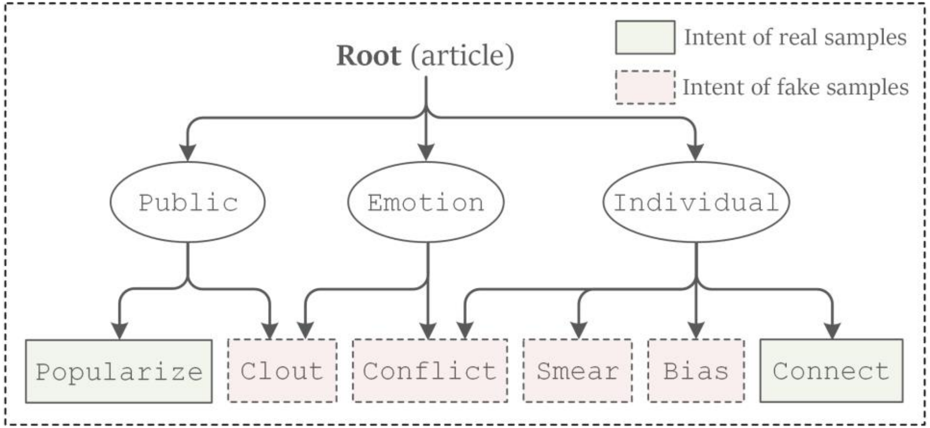
Specified by the **T5 encoder** extracts the hidden token embeddings of a given article.



Method: Detecting Misinformation by Integrating Intent featuRes

❷ LM decoder

To reason the potential intent, we refer to some psychological concepts and present an **intent hierarchy**



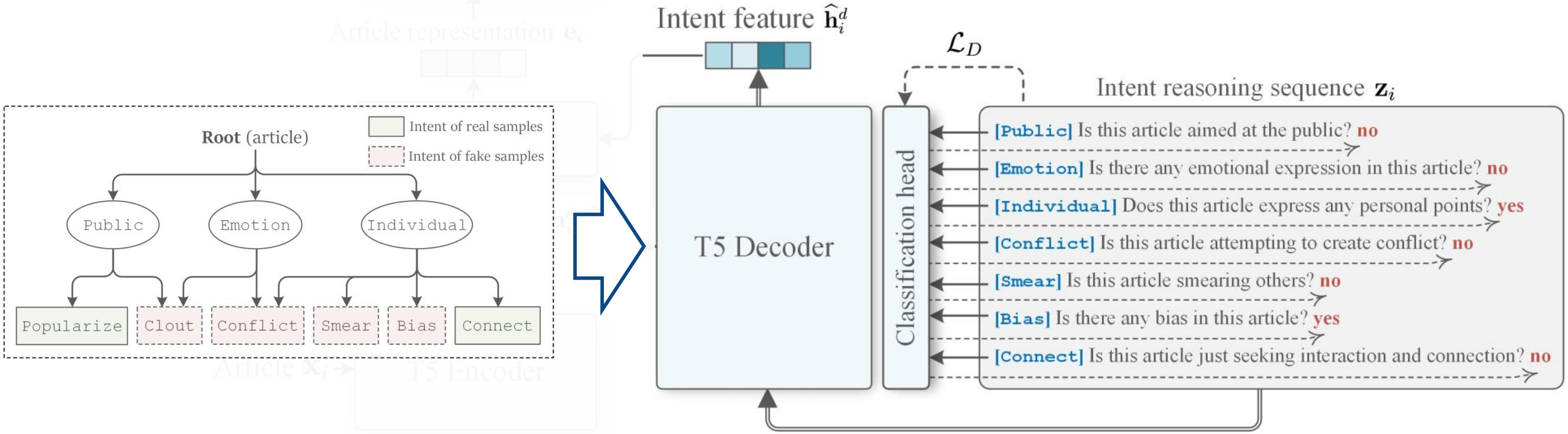
Intent Recognition

Hierarchical Classification

Method: Detecting Misinformation by Integrating Intent featuRes

❷ LM decoder

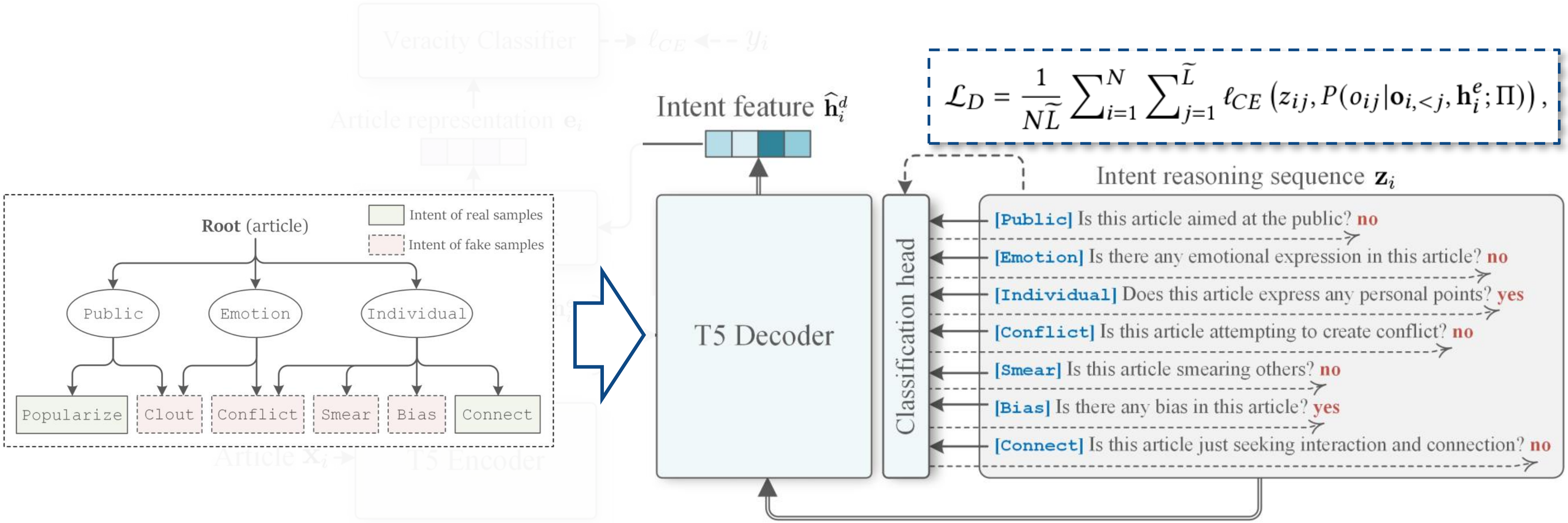
- ✓ We progressively prompt a **T5 decoder** to reason one or multiple paths on the hierarchy, and obtain a textual **intent reasoning sequence**
- ✓ the T5 decoder also outputs the token embeddings, and directly adopt the average pooled representation, which can be seen as the **intent feature**



Method: Detecting Misinformation by Integrating Intent featuRes

❷ LM decoder

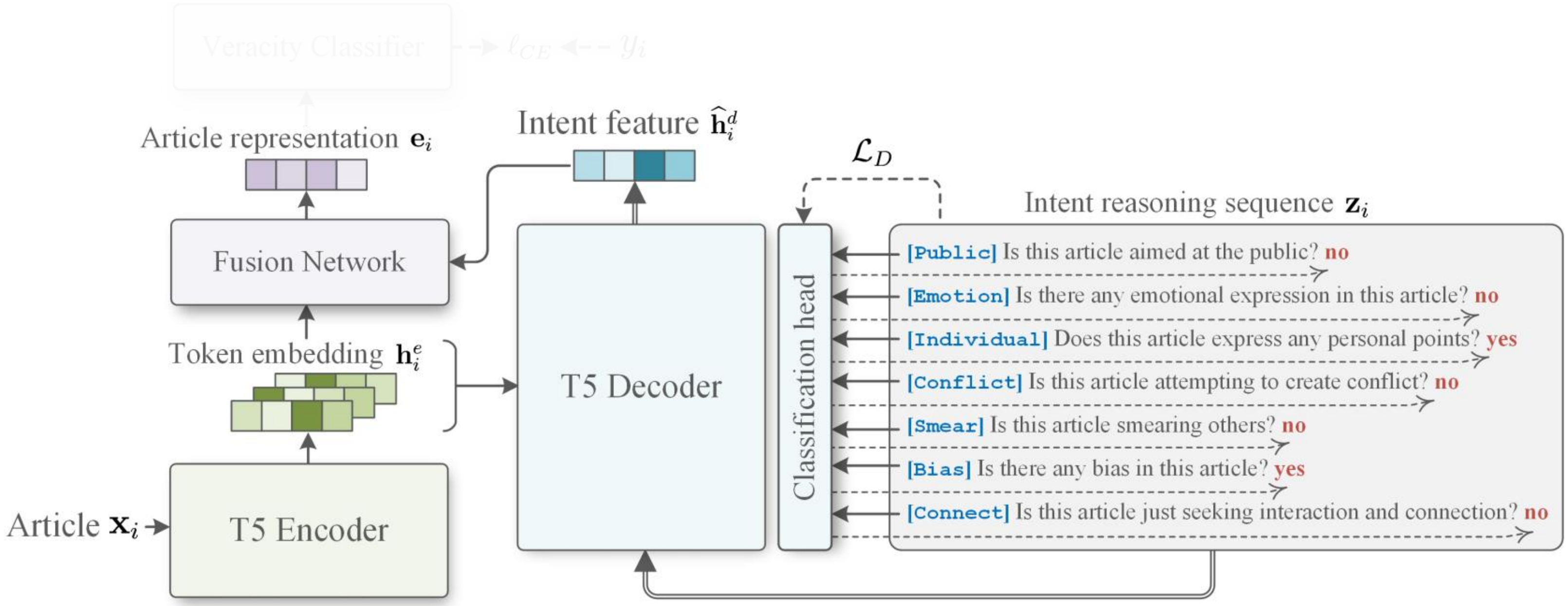
✓ We involve training the decoder with **the self-training method**, to enable the decoder to generate more accurate answers



Method: Detecting Misinformation by Integrating Intent featuRes

❸ Fusion network

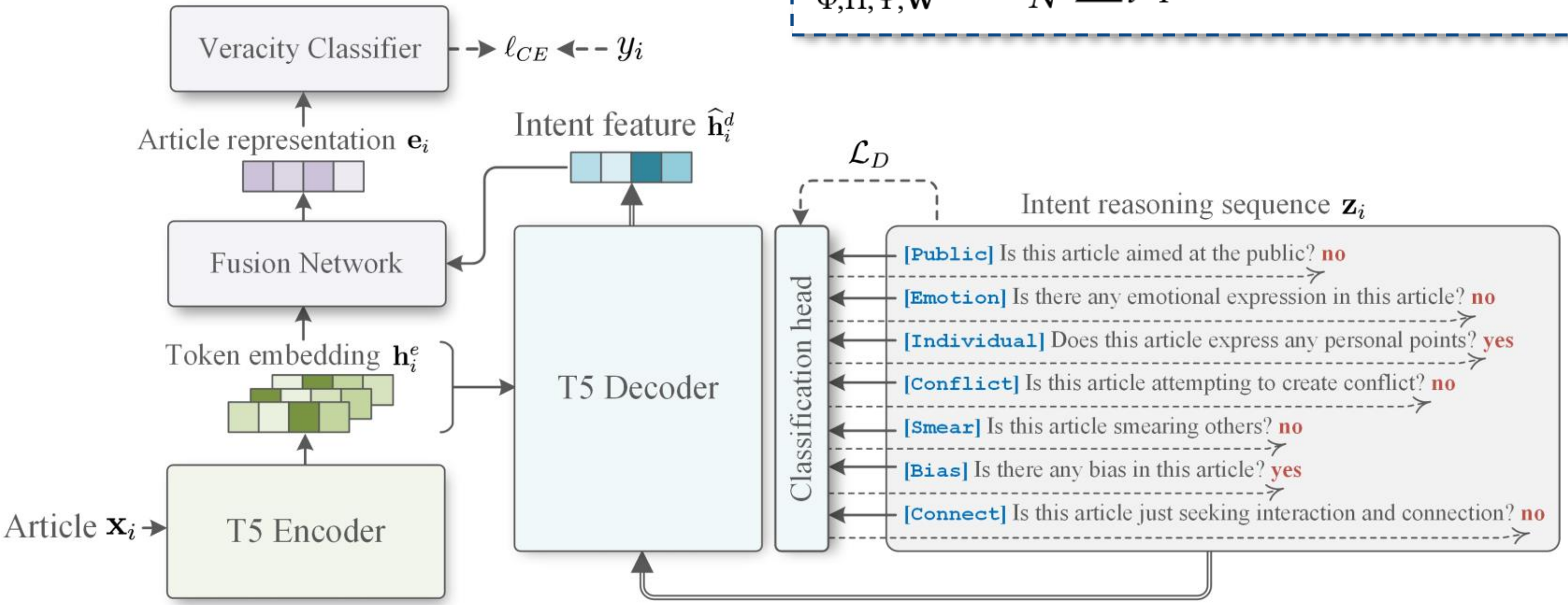
We adopt a **multi-head attention network** to fuse them to obtain the overall article representation



Method: Detecting Misinformation by Integrating Intent featuRes

❹ Veracity classifier Predict the final veracity predictions

$$\min_{\Phi,\Pi,\Psi,\mathbf{W}} \mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \alpha_i \left(\ell_{CE}\left(\mathbf{e}_i\mathbf{W}, y_i\right) + \beta\mathcal{L}_D\right),$$



Method: Detecting Misinformation by Integrating Intent featuRes

❹ Veracity classifier

Predict the final veracity predictions

$$\min_{\Phi,\Pi,\Psi,\mathbf{W}} \mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \alpha_i \left(\ell_{CE}(\mathbf{e}_i\mathbf{W}, y_i) + \beta\mathcal{L}_D\right),$$

Adaptive Weight Assigning

✓ **Error propagation**

If an intent in the intent hierarchy is reasoned incorrectly, then its child intents will also be incorrect.

$$\alpha_i^E = \frac{T}{\sum_{t=1}^{T} \|a_{it} - \widehat{a}_{it}\|_2^2}.$$

✓ **Veracity inconsistency**

Each intent corresponds to a veracity label, when the reasoned intent of an article fails to align with its veracity label, indicating an incorrect reasoning for this sample

$$\alpha_i^V = \begin{cases} 1, & \text{veracity consistency}, \\ 0, & \text{veracity inconsistency}. \end{cases}$$

Evaluation

✓ Experimental Settings

3 datasets, **5** baseline models, and **7** metrics

3 Datasets

❶ *GossipCop*, ❷ *PolitiFact*, and ❸ *Snopes*

Dataset	# Train		# Validation		# Test	
	Fake	Real	Fake	Real	Fake	Real
GossipCop	2,024	5,039	604	1,774	601	1,758
PolitiFact	1,224	1,344	170	186	307	337
Snopes	2,288	838	317	116	572	210

5 Baseline Models

❶ BERT (*bert-base-uncased*)

❷ T5 (*T5-base*)

❸ EANN learns event-invariant features

❹ BERT-EMO introduces emotional signals

❺ CED generate intra/inter category features

Evaluation: Performance Comparison

Method	Macro F1	Accuracy	Precision	Recall	$F1_{real}$	$F1_{fake}$	AUC
Dataset: *GossipCop*							
$BERT_{base}$ [4] (∼*110M*)	78.23±0.45	83.78±0.80	79.00±1.45	77.69±0.59	89.21±0.69	67.24±0.45	86.58±0.33
BERT + EANN [46]	78.59±0.84	84.47±0.66	80.37±1.46	77.42±1.36	89.80±0.55	67.39±1.59	86.89±0.45
BERT + BERT-EMO [57]	78.63±0.47	84.62±0.39	79.75±0.93	77.10±1.01	89.83±0.59	67.23±1.03	86.75±0.37
BERT + CED [52]	78.33±0.40	83.77±0.68	78.85±1.26	77.94±0.25	89.17±0.57	67.49±0.25	86.31±0.46
$T5_{base}$ [30] (∼*220M*)	78.44±0.33	84.56±0.27	80.61±0.50	76.92±0.33	89.93±0.20	66.96±0.51	87.56±0.36
+ DM-INTER (ours)	79.45±0.62*	85.33±0.49*	81.92±1.08*	77.82±0.88*	90.44±0.39	68.46±1.12*	87.46±0.12
T5 + EANN [46]	78.60±0.35	84.40±0.18	80.06±0.51	77.51±0.74	89.74±0.18	67.47±0.77	87.49±0.54
+ DM-INTER (ours)	79.73±0.66*	85.59±0.48*	82.38±0.91*	77.96±0.85	90.63±0.66*	68.83±1.00*	87.97±0.22
T5 + BERT-EMO [57]	78.46±0.39	84.48±0.26	80.45±0.89	77.11±1.08	89.84±0.31	67.08±1.01	87.53±0.36
+ DM-INTER (ours)	79.55±0.53*	85.19±0.26*	81.34±0.63*	78.26±1.01*	90.29±0.21	68.81±1.07*	88.10±0.15*
T5 + CED [52]	78.83±0.64	84.49±0.52	80.09±0.89	77.86±0.82	89.77±0.39	67.90±1.04	87.57±0.36
+ DM-INTER (ours)	79.78±0.60*	85.50±0.38*	82.01±0.86*	78.24±0.91	90.53±0.27*	69.03±1.08*	88.01±0.47
Dataset: *PolitiFact*							
$BERT_{base}$ [4] (∼*110M*)	59.46±0.98	60.02±0.73	60.26±0.91	59.82±0.88	62.38±1.58	59.55±1.02	64.28±1.22
BERT + CED [52]	60.11±0.59	60.33±0.85	60.55±0.93	60.35±0.70	61.08±1.41	59.16±1.84	64.71±0.88
$T5_{base}$ [30] (∼*220M*)	59.09±1.32	59.53±0.98	59.71±0.92	59.42±1.13	61.23±1.70	56.95±1.83	63.81±1.14
+ DM-INTER (ours)	60.31±0.89*	60.67±0.78*	60.72±0.86*	60.47±0.86*	63.22±1.40*	57.40±1.31	64.98±1.10*
T5 + CED [52]	59.19±0.97	59.43±0.89	59.39±0.97	59.28±0.89	61.70±1.37	56.69±1.27	63.63±0.85
+ DM-INTER (ours)	61.27±1.11*	61.42±0.93*	61.41±0.82*	61.33±0.74*	63.09±1.43*	59.45±0.71*	65.73±1.16*
Dataset: *Snopes*							
$BERT_{base}$ [4] (∼*110M*)	62.28±1.21	71.55±1.57	63.27±1.32	62.05±1.22	43.67±1.66	80.89±1.61	69.48±1.32
BERT + CED [52]	62.68±0.78	71.91±1.44	63.59±1.28	62.29±0.94	44.17±0.97	81.20±1.41	70.42±0.74
$T5_{base}$ [30] (∼*220M*)	62.51±0.91	72.19±1.26	63.73±1.11	62.03±1.10	43.51±1.44	81.52±1.26	70.70±0.27
+ DM-INTER (ours)	64.21±0.82*	73.85±1.54*	66.08±0.95*	63.56±1.17*	45.71±1.53*	82.72±1.04*	70.59±0.59
T5 + CED [52]	62.70±0.28	72.32±1.43	63.23±1.00	62.52±0.45	44.82±1.31	80.58±1.47	69.15±0.90
+ DM-INTER (ours)	64.51±1.05*	74.14±0.66*	66.21±1.72*	63.80±0.80*	46.35±1.04*	82.67±1.05*	70.44±1.08*

Observation 1

In general, this paper proposes a plug-and-play method that consistently enhances the performance of its baseline models across almost all settings.

Evaluation: Performance Comparison

Method	Macro F1	Accuracy	Precision	Recall	F1$_{real}$	F1$_{fake}$	AUC
Dataset: *GossipCop*							
BERT$_{base}$ [4] (~*110M*)	78.23±0.45	83.78±0.80	79.00±1.45	77.69±0.59	89.21±0.69	67.24±0.45	86.58±0.33
BERT + EANN [46]	78.59±0.84	84.47±0.66	80.37±1.46	77.42±1.36	89.80±0.55	67.39±1.59	86.89±0.45
BERT + BERT-EMO [57]	78.63±0.47	84.62±0.39	79.75±0.93	77.10±1.01	89.83±0.59	67.23±1.03	86.75±0.37
BERT + CED [52]	78.33±0.40	83.77±0.68	78.85±1.26	77.94±0.25	89.17±0.57	67.49±0.25	86.31±0.46
T5$_{base}$ [30] (~*220M*)	78.44±0.33	84.56±0.27	80.61±0.50	76.92±0.33	89.93±0.20	66.96±0.51	87.56±0.36
+ DM-INTER (ours)	79.45±0.62*	85.33±0.49*	81.92±1.08*	77.82±0.88*	90.44±0.39	68.46±1.12*	87.46±0.12
T5 + EANN [46]	78.60±0.35	84.40±0.18	80.06±0.51	77.51±0.74	89.74±0.18	67.47±0.77	87.49±0.54
+ DM-INTER (ours)	79.73±0.66*	85.59±0.48*	82.38±0.91*	77.96±0.85	90.63±0.66*	68.83±1.00*	87.97±0.22
T5 + BERT-EMO [57]	78.46±0.39	84.48±0.26	80.45±0.89	77.11±1.08	89.84±0.31	67.08±1.01	87.53±0.36
+ DM-INTER (ours)	79.55±0.53*	85.19±0.26*	81.34±0.63*	78.26±1.01*	90.29±0.21	68.81±1.07*	88.10±0.15*
T5 + CED [52]	78.83±0.64	84.49±0.52	80.09±0.89	77.86±0.82	89.77±0.39	67.90±1.04	87.57±0.36
+ DM-INTER (ours)	79.78±0.60*	85.50±0.38*	82.01±0.86*	78.24±0.91	90.53±0.27*	69.03±1.08*	88.01±0.47
Dataset: *PolitiFact*							
BERT$_{base}$ [4] (~*110M*)	59.46±0.98	60.02±0.73	60.26±0.91	59.82±0.88	62.38±1.58	59.55±1.02	64.28±1.22
BERT + CED [52]	60.11±0.59	60.33±0.85	60.55±0.93	60.35±0.70	61.08±1.41	59.16±1.84	64.71±0.88
T5$_{base}$ [30] (~*220M*)	59.09±1.32	59.53±0.98	59.71±0.92	59.42±1.13	61.23±1.70	56.95±1.83	63.81±1.14
+ DM-INTER (ours)	60.31±0.89*	60.67±0.78*	60.72±0.86*	60.47±0.86*	63.22±1.40*	57.40±1.31	64.98±1.10*
T5 + CED [52]	59.19±0.97	59.43±0.89	59.39±0.97	59.28±0.89	61.70±1.37	56.69±1.27	63.63±0.85
+ DM-INTER (ours)	61.27±1.11*	61.42±0.93*	61.41±0.82*	61.33±0.74*	63.09±1.43*	59.45±0.71*	65.73±1.16*
Dataset: *Snopes*							
BERT$_{base}$ [4] (~*110M*)	62.28±1.21	71.55±1.57	63.27±1.32	62.05±1.22	43.67±1.66	80.89±1.61	69.48±1.32
BERT + CED [52]	62.68±0.78	71.91±1.44	63.59±1.28	62.29±0.94	44.17±0.97	81.20±1.41	70.42±0.74
T5$_{base}$ [30] (~*220M*)	62.51±0.91	72.19±1.26	63.73±1.11	62.03±1.10	43.51±1.44	81.52±1.26	70.70±0.27
+ DM-INTER (ours)	64.21±0.82*	73.85±1.54*	66.08±0.95*	63.56±1.17*	45.71±1.53*	82.72±1.04*	70.59±0.59
T5 + CED [52]	62.70±0.28	72.32±1.43	63.23±1.00	62.52±0.45	44.82±1.31	80.58±1.47	69.15±0.90
+ DM-INTER (ours)	64.51±1.05*	74.14±0.66*	66.21±1.72*	63.80±0.80*	46.35±1.04*	82.67±1.05*	70.44±1.08*

Observation 2

The average improvement over the baseline models is

Snopes>PolitiFact>GossipCop

This phenomenon indicates that the positive impact of DM-INTER is more pronounced in scenarios with limited training data

Evaluation: Ablative Study

Method	F1	Acc.	P.	R.	$F1_{real}$	$F1_{fake}$
Dataset: *GossipCop*						
$T5_{base}$ [30]	78.44	84.56	80.61	76.92	89.93	66.96
+ DM-INTER	79.45	85.33	81.92	77.82	90.44	68.46
3 w/o $\mathcal{L}_D$	78.80	84.56	80.34	77.65	89.85	67.75
2 w/o hierarchy	78.93	84.68	80.65	77.70	89.91	67.92
4 w direct query	78.01	84.22	80.12	76.56	89.69	66.32
1 w/o weights	**79.21**	**84.82**	**80.68**	**78.17**	**90.00**	**68.43**
Dataset: *PolitiFact*						
$T5_{base}$ [30]	59.09	59.53	59.71	59.42	61.23	56.95
+ DM-INTER	60.31	60.67	60.72	60.47	63.22	57.40
3 w/o $\mathcal{L}_D$	59.63	60.43	60.53	60.05	**64.23**	55.03
2 w/o hierarchy	59.73	60.11	60.31	59.95	61.59	**57.84**
4 w direct query	59.16	59.59	59.86	59.51	60.97	57.34
1 w/o weights	**60.10**	**60.83**	**61.19**	**60.58**	63.40	56.81
Dataset: *Snopes*						
$T5_{base}$ [30]	62.51	72.19	63.73	62.03	43.51	81.52
+ DM-INTER	64.21	73.85	66.08	63.56	45.71	82.72
3 w/o $\mathcal{L}_D$	63.63	**74.11**	**66.00**	62.74	44.12	**83.13**
2 w/o hierarchy	63.86	72.97	64.79	63.30	**45.76**	81.98
4 w direct query	62.68	72.09	63.61	62.20	43.96	81.39
1 w/o weights	**63.96**	73.37	65.41	**63.38**	45.61	82.33

Observation 1

In general, the performance of all ablation experiments is consistently lower than that of our comprehensive model DM-INTER

Observation 2

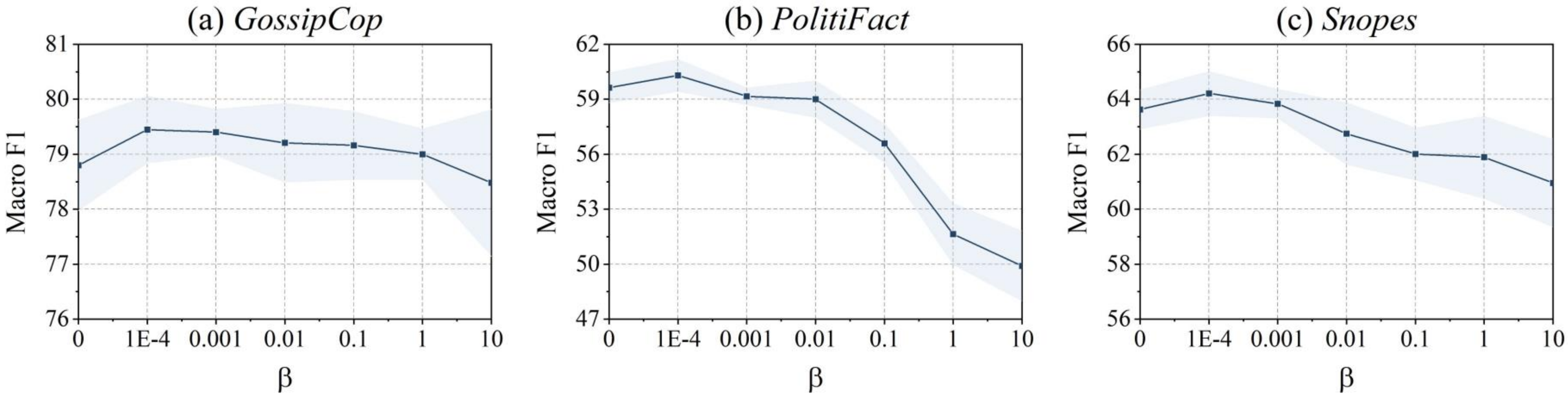
The performance of the four ablation versions can be ranked as:

w direct query < w/o $\mathcal{L}_D$ < w/o hierarchy < w/o weights

shows their order of importance

Evaluation: Sensitivity Analysis

$$\min_{\Phi,\Pi,\Psi,\mathbf{W}} \mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \alpha_i \left(\ell_{CE}\left(\mathbf{e}_i\mathbf{W}, y_i\right) + \beta\mathcal{L}_D\right),$$



Observation

The model is sensitive to β, and the model consistently shows the best performance on all datasets when β is approximately 0.0001

Evaluation: Case Study

Article: Warning : This article contains spoilers! So many spoilers! Highly detailed, movie-ruining spoilers! “Somewhere out there, there’ s an 8-year-old girl dreaming of becoming a criminal,” Debbie Ocean, played by Sandra Bullock, tells her mirrored reflection in one of the standout moments of “ Ocean’ s 8, ” the highly anticipated sequel to Steven Soderbergh’ s iconic heist films. “You’ re doing this for her.” The film gives budding bad girls everywhere role models to look up to, but just how...	**Article**: The Coachella Valley Music and Arts Festival has announced the dates for its 20th anniversary and how to get pre-sale tickets. The festival is also offering a new upgrade for car camping. So when is Coachella 2019? The festival happens April 12-14 and April 19-21 at the Empire Polo Club in Indio. As it has done in recent years, promoter Goldenvoice will put a limited number of passes on sale early. This year you can get Coachella 2019 passes for Weekend 1 and 2 starting at 11 a.m. Pacific Friday, June 1.	**Article**: Alex Rodriguez flatly denied a report that claimed he was threatening to cut child-support payments for his two daughters over a legal dispute with his ex-wife’ s brother. “I have always paid far more than the maximum in child support and that will never change,” the former New York Yankees star said in a statement to Page Six. “It’ s highly offensive to me that my former brother-in-law, who has been trying to pursue a frivolous case against me for four years and has gotten absolutely nowhere with it ...
Veracity Label: *fake* **Prediction**: *fake*	**Veracity Label**: *real* **Prediction**: *real*	**Veracity Label**: *fake* **Prediction**: *fake*
Reasoning sequence: [Public] Is this article aimed at the public? **yes** [Emotion] Is there any emotional expression in this article? **yes** [Individual] Does this article express any personal points? **no** [Popularize] Is this an article aimed at popularization? **no** [Clout] Is this an article aimed at pursuing attention? **yes** [Conflict] Is this article attempting to create conflict? **no**	**Reasoning sequence**: [Public] Is this article aimed at the public? **yes** [Emotion] Is there any emotional expression in this article? **no** [Individual] Does this article express any personal points? **no** [Popularize] Is this an article aimed at popularization? **yes** [Clout] Is this an article aimed at pursuing attention? **no**	**Reasoning sequence**: [Public] Is this article aimed at the public? **no** [Emotion] Is there any emotional expression in this article? **no** [Individual] Does this article express any personal points? **no** [Nointent] This article does not convey any intents.

Observation

In summary, these cases consistently demonstrate the effectiveness of DM-INTER in reasoning intents

Takeaways

Motivation

We present to investigate the intents expressed by articles and utilize them to identify misinformation.

Method: DM-INTER

We design an intent hierarchy based on several psychological studies and use it to progressively reason intents with a pre-trained auto-regressive decoder.

Experiments

Our experimental results can indicate that DM-INTER can improve the performance of the baseline models.

Thanks.

Why Misinformation is Created? Detecting them by Integrating Intent Features

Authors: Bing Wang (wangbing1416@gmail.com),

Ximing Li, Changchun Li, Bo Fu, Songwen Pei, Shengsheng Wang

CIKM 2024
BOISE
OCTOBER 21-25