

课程论文题目

刘越锋

摘要

假新闻的肆意传播已成为一个严重的社会问题，并已造成严重后果。同时，由于假新闻内容的是人故意捏造的，以及它的传播模式及其复杂，使得假新闻难以拦截，这突出了自动检测假新闻的重要性。提出了一种用于假新闻检测的子图推理范式，该范式通过揭示新闻传播网络的哪些子图对于新闻验证最重要，从而提供了可解释性，并通过去除与任务无关的信息，同时提高了基于图的检测模型的泛化和辨别能力。还提出了一种增强子图生成方法，并通过开发分层路径感知的核心图注意力网络对生成的子图进行细粒度建模。

关键词：子图推理；假新闻；可解释性；社交网络

1 引言

假新闻是“新闻机构故意发布的、可证实的虚假新闻”^[1]，它产生的原因是假新闻往往吸引更多的读者、匿名用户故意传播噪音和误导信息。假新闻的广泛传播会误导公众，并“为一些政党带来不公正的政治、经济或心理利益”^[2]。同时，由于假新闻内容的是人故意捏造的，以及它的传播模式及其复杂，使得假新闻难以拦截，这突出了自动检测假新闻的重要性。人工智能时代的到来为使用强大的数据驱动的强化学习方法来解决这个问题提供了新的可能性。

2 相关工作

为了通过子图预测新闻文章的标签 y ，提出了一个用于假新闻检测的子图推理框架 (SureFact)。从数学上讲，该框架表示为：

$$\mathbb{P}(y | G) \approx \sum_{\substack{g_{1 \sim M} \leftarrow G \\ g_m \subset G}} \underbrace{\mathbb{P}(y | g_{1 \sim M})}_{\text{Subgraph Modeling}} \underbrace{\mathbb{P}(g_{1 \sim M} | G)}_{\text{Subgraph Generation}} \quad (1)$$

也就是包含两个模块：增强子图生成模块 $\mathbb{P}(g_{1 \sim M} | G)$ 和细粒度子图建模模块 $\mathbb{P}(y | g_{1 \sim M})$

2.1 Reinforced Subgraph Generation

增强子图生成模块估计概率分布 $\mathbb{P}(g_{1 \sim M} | G)$ 通过最大化假新闻检测的预期精度来检测重要子图。生成子图的关键是学习 $\mathbb{P}(g_{1 \sim M} | G)$ 没有子图的基本真值标签。建议通过最大化假新闻检测的预期精度来生成子图，这可以用强化学习 (RL) 来实现。这里可以使用 RL 直接优化第一个模块（增强子图生成）的原因如下：

$$\sum_{y \in \{0,1\}} \mathbb{P}(y | G) \mathcal{R}^*(y) \approx \sum_{g_{1 \sim M} \subset G} \mathbb{P}(g_{1 \sim M} | G) \mathcal{R}(g_{1 \sim M}) \quad (2)$$

$$\mathcal{R}(g_{1 \sim M}) = \sum_{y \in \{0,1\}} \mathbb{P}(y | g_{1 \sim M}) \mathcal{R}^*(y) \quad (3)$$

2.2 Fine-Grained Subgraph Modeling

细粒度子图建模模块学习 $\mathbb{P}(y | g_{1 \sim M})$ 与 HP-KGAT 网络相结合，该网络通过有效整合异构路径信息并联合建模子图内和子图间关系来预测标签 y 。

3 本文方法

3.1 本文方法概述

对假新闻检测的工作主要分为两类：基于新闻内容的方法 (Content-based methods)、基于知识的方法 (Knowledge-based methods)

3.2 Content-based methods

基于内容的方法主要利用新闻文章中的文本或视觉内容进行新闻验证。假设假新闻可能有不同的语言风格，例如，夸张，更多的情感词汇，更多的语法错误等。这些方法可以早期发现假新闻，但准确性可能有限，因为仅仅考虑了新闻内容。示例模型:SVM, GRU-2, RFC, DTC。

3.3 Knowledge-based methods

基于知识的方法利用关于实体关系的外部知识或关于在线帖子的社会知识假设信息，假新闻也可以通过考虑其他方面，例如，用户评论和它是如何通过互联网传播。它们通常会导致更高的精度，但需要收集更多的数据，而且模型通常更复杂。示例模型:B-TransE, KCNN, GCAN, KAN。

4 复现细节

4.1 部分算法伪代码

Mutual-Reinforcement-Based Evidence Ranking

Input: 1. 待验证的新闻文章 S ,
2. 新闻的在线帖子 P 的集合以及帖子之间的评论/转发关系,
3. 发布帖子 P 的在线用户 U 。

Output: 新闻的预测标签：假的 ($y=1$) 或真的 ($y=0$)

Graph Construction: Mutual-Reinforcement-Based Evidence Ranking、Topic-Aware Claim-Evidence Association

Mutual-Reinforcement-Based Evidence Ranking:

Step1:Attribute saliency computation.
Step2:Mutual reinforcement evidence graph building.
Step3:Iterative saliency propagation.

Topic-Aware Claim-Evidence Association:

Step1:Topic modeling.
Step2:Key claim and evidence extraction.
Step3:Graph construction.

4.2 预处理模型

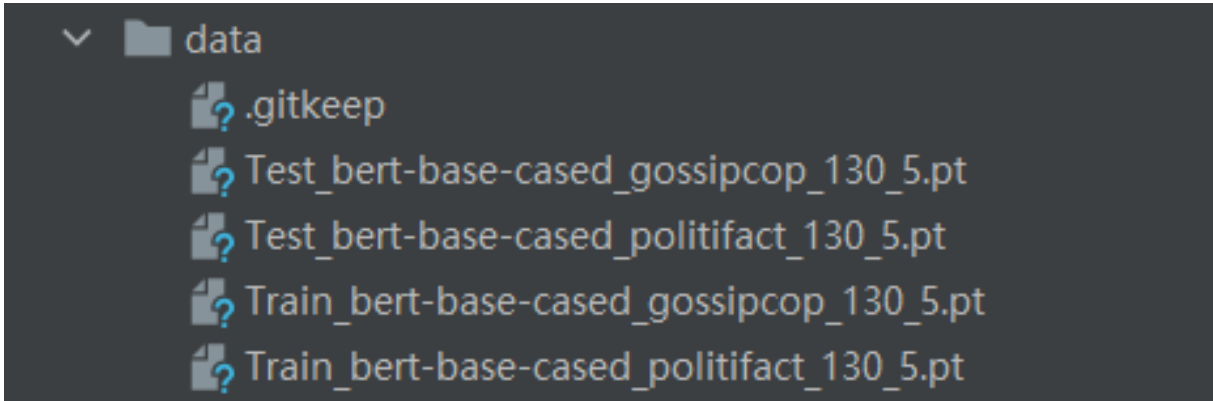


图 1: 预处理模型

5 实验结果分析

5.1 Baselines

与八个基线进行比较，这些基线可以分为两组：第一组（G1）是基于内容的方法，它利用新闻的文本或视觉内容来检测假新闻。G1 包含四个基线：SVM（Yang 等人 2012）、GRU-2（Ma 等人 2016）、RFC（Kwon 等人 2013 年）和 DTC（Castillo、Mendoza 和 Poblete, 2011 年）。第二组（G2）由知识感知方法组成，该方法通过利用诸如知识图和关于在线帖子的社会知识之类的辅助知识来检测假新闻。该组包括四种方法：B-TransE（Pan 等人 2018）、KCNN（Wang 等人 2018a）、GCAN（Lu 和 Li 2020）和 KAN（Dun 等人 2021）。

5.2 评估标准

评估标准包括精密度（Pre）、召回率（Rec）、F1 评分、准确度（Acc）和 ROC 曲线下面积（AUC）。进行了 5 倍的交叉验证，并报告了平均性能。

5.3 实验结果

实验结果的模型如图 2 所示：

名称	修改日期	类型	大小
G_12-06_23_FF_batch2_20_lr5e-05_K11_Sig0.1_pt	2022/12/7 7:42	PT 文件	427,074 KB
P_12-05_18_FF_batch2_20_lr5e-05_K11_Sig0.1_pt	2022/12/5 18:26	PT 文件	427,074 KB

图 2: 实验结果模型

实验结果如图 3，4 所示：

```
[Eval] Epoch 19 Pre: 0.9483, Rec: 0.8333      Acc: 0.8833, F1: 0.8871
, AUC: 0.9618
[15-11-2022 19:16:58] INFO: Dev acc: 0.8833333333333333

[Train] Epoch 19      Pre: 1.0000, Rec: 0.9962      Acc: 0.9979, F1: 0.9981
, AUC: 1.0000
```

图 3: 实验结果示意:PolitiFact

```
[07-12-2022 08:32:42] INFO: Start eval!

[Eval] Epoch 19 Pre: 0.9739, Rec: 0.9661      Acc: 0.9850, F1: 0.9700
, AUC: 0.9967
[07-12-2022 08:34:42] INFO: Dev acc: 0.985014985014985

[Train] Epoch 19      Pre: 1.0000, Rec: 1.0000      Acc: 1.0000, F1: 1.0000
, AUC: 1.0000
```

图 4: 实验结果示意:GossipCop

八条基准线的性能汇总如图 5 所示:

		PolitiFact					GossipCop				
		Pre	Rec	F1	Acc	AUC	Pre	Rec	F1	Acc	AUC
G1 (Content-based)	B-TransE	0.7739	0.7658	0.7641	0.7694	0.8340	0.7369	0.7330	0.7340	0.7394	0.7995
	BERT-C	0.8249	0.8465	0.8413	0.8327	0.8632	0.7863	0.7619	0.7628	0.8123	0.8371
	RoBERTa-C	0.8904	0.8784	0.8759	0.8843	0.9179	0.8134	0.8092	0.8276	0.8154	0.8623
G2 (Spcial-based)	DTC	0.7397	0.7269	0.7362	0.7364	0.7618	0.7129	0.6794	0.6929	0.7156	0.7007
	RFC	0.7536	0.7450	0.7422	0.7582	0.8082	0.6978	0.6833	0.6534	0.6845	0.7366
	GCAN	0.8169	0.8208	0.8364	0.8372	0.8089	0.7824	0.8027	0.7757	0.7506	0.8066
	dEFEND	0.9016	0.8953	0.8879	0.8846	0.8904	0.7217	0.8015	0.7538	0.8098	0.8401
	BERT-S	0.8243	0.9013	0.8453	0.8619	0.9131	0.8459	0.8516	0.8542	0.8475	0.8533
	RoBERTa-S	0.8529	0.9052	0.8729	0.8804	0.9218	0.8441	0.8594	0.8397	0.8289	0.8625
	BERTweet	0.8440	0.9028	0.8658	0.8781	0.9155	0.8508	0.8624	0.8626	0.8478	0.8619
	FinerFact	0.9185	0.9043	0.9148	0.9088	0.9303	0.8609	0.8739	0.8657	0.8373	0.8644
Ours	SureFact	0.9506*	0.9283*	0.9392*	0.9436*	0.9413*	0.8796*	0.8835*	0.8811*	0.8658*	0.8797*

图 5: Performance

5.4 结果分析

可以看出，FinerFact 在两个数据集中都始终优于基线。例如，FinerFact 在 F1 得分和准确度方面分别比最具竞争力的基准 KAN 好 6.3%、5.1% 和 9.7.5%。这证明了这个细粒度推理框架的有效性，它使模型能够通过识别和连接不同类型的细微线索来做出预测。同时，与 GCAN 相比，FinerFact 将两个数据集上的 F1 分别提高了 8.3% 和 9.8%。这意味着基于核心注意力的方法可以更好地模拟新闻文章和证据之间的互动。此外，结合外部知识的方法（G2）通常比基于内容的方法（G1）表现更好，这说明了外部知识在假新闻检测中的有用性。

6 总结与展望

6.1 总结

假新闻检测的子图推理框架通过增强假新闻检测模型的泛化和鉴别能力，实现更好的解释性并提高准确性。Hierarchical Path-Aware Kernel Graph Attention Network 网络通过对多个异构子图进行细粒度建模来准确检测假新闻。引入了一种 curriculum-based 的优化方法，通过逐渐增加学习难度和确保端到端训练，可确保得到收敛性更好的解决方案。

6.2 展望

修改训练代码，使用 batch 加快训练速度。考虑在更大规模的数据集上进行训练与验证。

参考文献

[1] PECKHAM O. AI Squares Off Against Fake News[J/OL]., 2019. <https://www.datanami.com/2019/09/02/ai-squares-off-against-fake-news/>.

- [2] RASHKIN H, CHOI E, JANG J Y, et al. Truth of varying shades: Analyzing language in fake news and political fact-checking[J]., 2017: 2931-2937.