

Towards Fine-Grained Reasoning for Fake News Detection

Yiqiao Jin, Xiting Wang, Ruichao Yang, Yizhou Sun, Wei Wang, Hao Liao, Xing Xie

摘要

The detection of fake news often requires sophisticated reasoning skills, such as logically combining information by considering word-level subtle clues. In this paper, we move towards fine-grained reasoning for fake news detection by better reflecting the logical processes of human thinking and enabling the modeling of subtle clues. In particular, we propose a fine-grained reasoning framework by following the human's information-processing model, introduce a mutual reinforcement-based method for incorporating human knowledge about which evidence is more important, and design a prior-aware bi-channel kernel graph network to model subtle differences between pieces of evidence. Extensive experiments show that our model outperforms the state-of-the-art methods and demonstrate the explainability of our approach.

关键词: fake news detection; fine-grained reasoning;

1 引言

社交媒体虽然极大的促进了信息的有效分享，但同时也可能为虚假新闻传播提供了温床。快速的，不受限制的虚假新闻传播正逐渐成为国际社会问题并造成不良影响。因此，人们对虚假新闻的检测越来越重视。此文引入了 **FinerFact**，一种用于虚假新闻检测的可解释的细粒度推理框架。论文设计了一种基于相互强化的有效证据排序方法，以及一种用于对多组证据进行细粒度推理的先验感知双通道内核图网络。大量实验表明，**FinerFact** 优于目前最先进的方法，并证明了可解释性。本次课程的论文复现工作拟通过对 **FinerFact** 算法进行复现并在 **PolitiFact** 和 **GossipCop** 数据集上检测其性能。

2 相关工作

假新闻检测方法可分为两大类：基于内容的和基于知识感知的。

2.1 基于内容的检测

基于内容的方法主要利用新闻文章和相关帖子中的文本或视觉内容进行新闻验证^[1-10]。这些方法可以在早期阶段检测到假新闻^[11-12]。然而，他们的效果并不十分出色，因为它们没有使用新闻验证的辅助知识。

2.2 基于知识感知的检测

有些学者知识感知方法利用辅助知识进行新闻验证^[13-19]。这些方法通常利用有关实体关系的外部知识^[20-24]或关于在线推文的社交信息^[25-28]用于假新闻检测。虽然现有方法已经证明了异质社会关系和外部信息的有效性^[1]，但它们要么没有对新闻内容与不同类型的知识数据之间的相互作用进行建模，要么没有在粗粒度（例如，句子或帖子）级别对其进行建模，这限制了它们的性能。本文通过提出一种先验感知的双通道核图网络来解决这个问题，该网络可以实现细粒度推理并提高检测精度。

3 本文方法

3.1 本文方法概述

受人类信息处理启发，此文设计了一个用于假新闻检测（FinerFact）的细粒度推理框架。这使我们能够通过更好地反映人类思维的逻辑过程来检测假新闻，从而增强可解释性并为整合人类知识提供基础。FinerFact 的整体架构如图 1所示：

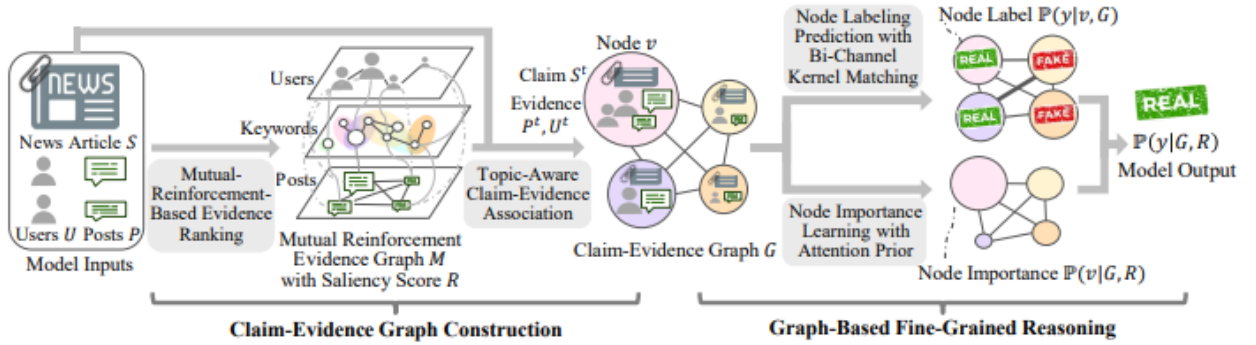


图 1: FinerFact framework for fake news detection

我们的架构主要分为两大模块，第一个模块是声明-证据图构建模块，此模块目的是要对新闻中的解关键声明及其相关证据（例如，支持的帖子和用户）做检索，然后将不同结构的数据做整合，构建出以主题为核心的全连接图。第二个模块是基于图的细粒度推理，对应于人类信息处理模型的检索子过程，其中人们根据其决策关联重新激活特定信息片段。在假新闻检测中，该模块通过考虑微妙的线索来实现证据关系的细粒度建模。通过基于人类信息处理模型设计框架，我们方法的整体工作流程类似于人类使用的逻辑过程，大多数中间结果是用户可以理解的，这为用户信任提供了良好的基础。

3.2 声明-证据图构建模块

我们的图构建方法包含两部分：1）基于相互增强的证据排序，通过结合人类知识（过滤噪声）来区分重要证据和噪声；2）主题感知的声明-证据关联，受记者对高质量新闻的知识的启发（for Excellence in Journalism 2005），提取关键声明并将其与相应的证据（组织事实）相关联。

3.3 基于图的细粒度推理

在构建完声明-证据图 G 之后，我们对微妙的线索进行建模，并通过基于细粒度图的推理有效地利用它们进行预测。我们的方法基于核图注意力网络（KGAT）（Liu 等人，2020）。我们选择这种方法，因为它可以有效地对语句之间的细微差异进行建模，并将学习到的信息传播到图上。然而，KGAT 不能直接应用于我们的声明-证据图，因为它只能处理文本输入，不能整合学习的显著性 R ，而显著性包含了人类关于哪些证据是重要的知识。为了解决这些问题，我们提出了一种先验感知双通道 KAT，它将 KGAT 扩展到 1）同时对来自文本（帖子）和社交（用户）输入的微妙线索进行建模，具有两个连接的通道；2）整合现有知识，将有关重要证据的现有知识与注意力先验相结合。

表 1: 数据集

	#True	#Fake	#Total
PolitiFact	443	372	815
GoosipCop	4219	3393	7612

4 复现细节

4.1 数据集与实验参数

本文所使用的数据集来源于 Twitter 爬取，包含着与新闻有关的推文及相关信息（如转发、评论等）。其中 PolitiFact 包含 815 条新闻，其中 443 条为 True, 372 条为 Fake; GoosipCop 包含 7612 条新闻，其中 4219 条为 True, 3393 为 Fake。此数据集涉及用户敏感信息，根据相关政策无法公开，但仍可以使用公开数据集如 FakeNewsNet¹ 来进行实验。

4.2 与已有开源代码对比

此文作者在 github 上发布有开源代码²，但只有第二大模块的部分而没有第一大模块的。因此，我的主要工作则是对模块一进行构建，也就是声明-证据图的构建。相关代码已整合发至课程网站。

4.3 实验环境搭建

此实验对硬件有一定要求，需要在 1050 显卡及以上，内存 16G 及以上。至于如何一步步搭建环境，可以参考文件中的 README.md。此实验需用到 python 和 pytorch，其中环境配置随版本更替、硬件不同可能会遇到麻烦。建议使用 pycharm 配合虚拟环境分别对第一、第二部分建立专门的环境。

MRG

This is the implementation of the Mutual Reinforcement Graph in **FinerFact**

It is recommended that you use PyCharm to run the code so that you can easily set your working directory and customize the environments you run the code.

About installation

- First run

```
conda install pytorch==1.7.1 torchvision==0.8.2 torchaudio==0.7.2 cudatoolkit=10.2 -c pytorch
```

 conda is preferred thanks to its stability
- Then run `pip install -r requirements.txt` to install other dependencies
- Note that you may need to install **pytorchgeometric** manually by consulting its official doc. This is because the dependencies of **pytorchgeometric** is constantly changing.

图 2: 操作指引示意

更加详细的内容请在 readme.md 文件中查阅。

4.4 创新点

本实验是一次结合传统机器学习（例如，特征工程和相互强化图）和深度学习的好处的初步尝试。传统的机器学习速度快、可预测、需要的标签少得多，并且可以结合根据多年经验和对数据集的仔细分析总结的人类知识。同时，深度学习更强大、更容易实现，可以识别隐藏在大规模数据集中的复杂模式。为了结合它们，我们将传统的机器学习放在第一个模块中，通过有效地识别候选证据来减轻深

¹<https://github.com/KaiDMML/FakeNewsNet>

²<https://github.com/Ahren09/FinerFac>

度学习的负担。然后我们在第二个模块中使用深度学习来进行精确的、数据驱动的、细粒度的推理，这是传统机器学习很难做到的。我们希望在未来探索更多这样的组合

5 实验结果分析

由于部分参数的缺失，本实验关于显著性系数 R 的计算并不完全按照原文，实验结果如图 3 所示。

	PolitiFact					GossipCop				
	Pre	Rec	F1	Acc	AUC	Pre	Rec	F1	Acc	AUC
原文结果	0.9196	0.9037	0.9172	0.9092	0.9384	0.8615	0.8779	0.8685	0.8320	0.8637
实验结果	0.9024	0.8863	0.8911	0.8762	0.9104	0.8435	0.8602	0.8479	0.8058	0.8514

图 3: 实验结果

我们采用的评价标准包括 Precision(Pre), Recall(Rec), F1 分数, Accuracy(Acc) 以及 Area Under the ROC curve(AUC)。

6 总结与展望

我们提出了 FinerFact，一个用于可解释的假新闻检测的细粒度推理框架。我们设计了一种基于相互增强的高效证据排序方法和一种先验感知的双通道核图网络，用于对多组证据进行细粒度推理。实验结果表明了该方法的有效性。尽管使用传统机器学习方法来提取特征被证明是 very 有效的，但是我们是否对原数据进行了过多的处理而使其失去了一些关键的证据信息（如传播网络的拓扑结构），未来可以尝试对数据做减法，提取关键子图进行推理。此外，随着人们对短视频的日益依赖，多模态检测成为必然，人们对早期检测、可解释性和用户隐私保护的要求也越来越高。总之，虚假新闻检测正在高速更新迭代，还有很多方向可以取研究。

参考文献

[1] YUAN C, MA Q, ZHOU W, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection[C] // 2019 IEEE international conference on data mining (ICDM). 2019: 796-805.

[2] KALIYAR R K, GOSWAMI A, NARANG P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach[J]. Multimedia tools and applications, 2021, 80(8): 11765-11788.

[3] AFROZ S, BRENNAN M, GREENSTADT R. Detecting hoaxes, frauds, and deception in writing style online[C] // 2012 IEEE Symposium on Security and Privacy. 2012: 461-475.

[4] KWON S, CHA M, JUNG K, et al. Prominent features of rumor propagation in online social media[C] // 2013 IEEE 13th international conference on data mining. 2013: 1103-1108.

[5] PRZYBYLA P. Capturing the style of fake news[C] // Proceedings of the AAAI conference on artificial intelligence: vol. 34: 01. 2020: 490-497.

- [6] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[J]., 2016.
- [7] ZELLERS R, HOLTZMAN A, RASHKIN H, et al. Defending against neural fake news[J]. Advances in neural information processing systems, 2019, 32.
- [8] QI P, CAO J, YANG T, et al. Exploiting multi-domain visual information for fake news detection[C]// 2019 IEEE international conference on data mining (ICDM). 2019: 518-527.
- [9] GUPTA A, LAMBA H, KUMARAGURU P, et al. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy[C]//Proceedings of the 22nd international conference on World Wide Web. 2013: 729-736.
- [10] JIN Z, CAO J, ZHANG Y, et al. Novel visual and statistical image features for microblogs news verification[J]. IEEE transactions on multimedia, 2016, 19(3): 598-608.
- [11] WEI L, HU D, ZHOU W, et al. Towards propagation uncertainty: Edge-enhanced bayesian graph convolutional networks for rumor detection[J]. arXiv preprint arXiv:2107.11934, 2021.
- [12] PELRINE K, DANOVITCH J, RABBANY R. The surprising performance of simple baselines for misinformation detection[C]//Proceedings of the Web Conference 2021. 2021: 3432-3441.
- [13] RUCHANSKY N, SEO S, LIU Y. Csi: A hybrid deep model for fake news detection[C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017: 797-806.
- [14] SHU K, CUI L, WANG S, et al. defend: Explainable fake news detection[C]//Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019: 395-405.
- [15] JIN Z, CAO J, ZHANG Y, et al. News verification by exploiting conflicting social viewpoints in microblogs[C]//Proceedings of the AAAI conference on artificial intelligence: vol. 30: 1. 2016.
- [16] MA J, GAO W, WONG K F. Rumor detection on twitter with tree-structured recursive neural networks [C]//. 2018.
- [17] CHO K, VAN MERRIËNBOER B, BAHDANAU D, et al. On the properties of neural machine translation: Encoder-decoder approaches[J]. arXiv preprint arXiv:1409.1259, 2014.
- [18] WANG H, ZHANG F, XIE X, et al. DKN: Deep knowledge-aware network for news recommendation [C]//Proceedings of the 2018 world wide web conference. 2018: 1835-1844.
- [19] WANG Y, MA F, JIN Z, et al. Eann: Event adversarial neural networks for multi-modal fake news detection[C]//Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining. 2018: 849-857.
- [20] XU W, WU J, LIU Q, et al. Mining Fine-grained Semantics via Graph Neural Networks for Evidence-based Fake News Detection[J]. arXiv preprint arXiv:2201.06885, 2022.

- [21] DUN Y, TU K, CHEN C, et al. Kan: Knowledge-aware attention network for fake news detection[C] // Proceedings of the AAAI conference on artificial intelligence: vol. 35: 1. 2021: 81-89.
- [22] PAN J Z, PAVLOVA S, LI C, et al. Content based fake news detection using knowledge graphs[C] // The Semantic Web–ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part I 17. 2018: 669-683.
- [23] SILVA A, LUO L, KARUNASEKERA S, et al. Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data[C] // Proceedings of the AAAI conference on artificial intelligence: vol. 35: 1. 2021: 557-565.
- [24] HU L, YANG T, ZHANG L, et al. Compare to the knowledge: Graph neural fake news detection with external knowledge[C] // Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021: 754-763.
- [25] LU Y J, LI C T. GCAN: Graph-aware co-attention networks for explainable fake news detection on social media[J]. arXiv preprint arXiv:2004.11648, 2020.
- [26] NGUYEN V H, SUGIYAMA K, NAKOV P, et al. Fang: Leveraging social context for fake news detection using graph representation[C] // Proceedings of the 29th ACM international conference on information & knowledge management. 2020: 1165-1174.
- [27] KHOO L M S, CHIEU H L, QIAN Z, et al. Interpretable rumor detection in microblogs by attending to user interactions[C] // Proceedings of the AAAI conference on artificial intelligence: vol. 34: 05. 2020: 8783-8790.
- [28] BIAN T, XIAO X, XU T, et al. Rumor detection on social media with bi-directional graph convolutional networks[C] // Proceedings of the AAAI conference on artificial intelligence: vol. 34: 01. 2020: 549-556.