



IM²C 2020 Greater China Problem B (Autumn) (English 简体 繁體)
Credibility Assessment in Online News

“A lie can travel halfway around the world while the truth is putting on its shoes.” Such situation as the adage puts has never been truer than in the era of online media. Misinformation can quickly reach thousands or millions of readers, thanks to inattentive, irresponsible or malicious sharing and algorithms that optimize the engagement. The concern over the credibility in online news is growing worldwide. Assessment of credibility in online news has been a pressing challenge for the health of information ecosystem where every citizen in the digital age is influenced.

Suppose your team is invited to participate in a research led by a group of international researchers aiming to develop a set of indicators for representing credibility in online news. Such indicators will allow publishers, media platforms and the public to both understand and communicate what aspects of news article contribute to its credibility and why.

Based on the work by scholars on distinguishing misinformation and characterizing credible news, indicators for credibility at article level can be determined by analyzing the title and text of the article, and by examining external sources of the article as well. For example, the clickbait title and logical fallacies in the text typically signal misinformation in the article; whereas the fact-checking result and placement of ads and social calls related to external factors usually indicate the degree of credibility of the article. In consideration of what indicators to use and how to measure them, your team is expected to create the mathematical model for the assessment of the credibility in online news.

Your team can use the given source data containing 41 most shared articles on social media on the topics of climate and health. The dataset includes the evaluations by domain experts as well. (Source data: IMMC2020 Autumn Problem B- Data.xlsx)

Your team is required to address the following aspects of the task:

1. Develop a framework of indicators for assessment of online news credibility and define how to measure them.
2. With your indicators, develop a mathematical model assessing the credibility by examining both content (title and content) and context information (external sources) of the online articles. Applying your model on the articles in the given source, you may need to collect data for your indicators by asking people you can reach to read the sample articles in response to your indicators.
3. Given the above dataset, compare the assessment by your model with the domain experts' evaluation on credibility of the news article, and discuss the results. Does your model imply your indicators any advantages or disadvantages in predicting credibility?
4. Based on your findings, prepare a press release up to 2 pages to a news publisher or social media on how to improve the public awareness and facilitate for improving credibility in online news.

Submission

Your solution paper should include a 1-page Summary Sheet. The body cannot exceed 20 pages for a maximum of 23 pages with the Summary Sheet and Two-page Press Release inclusive. The appendices and references should appear at the end of the paper and do not count towards the 23 pages limit.

Glossary

Misinformation: information that is false or written with the intention to deceive or mislead.

News credibility: the truthfulness of news and fact-based reporting.

Clickbait title: a title with a certain kind of web content designed to entice its readers into clicking.

Logical fallacies: arguments that is poor but tempting, misleading people to find them more convincing than is rational.

Fact-checking: the act of checking factual claims in the article in order to determine the veracity of the fact-based statement in the article, which could be conducted by individuals such as journal editor or approved organization.

Placement of ads and social calls: mounting ads on media page of the article, and/or adding exhortations to share the content on social media, email the article or join a mailing list or social media group, by appearing in pop-up windows, covering up article content, or distracting through additional animation, video and audio.

Acknowledgement goes to Dr. Amy X. ZHANG of MIT CSAIL, Dr. Aditya Ranganathan of UC Berkeley Institute for Data Science et al. for sources by their research on credibility in online news articles.



IM²C 2020 中华赛 B 题（秋季赛）(English 简体 繁體)

网络新闻可信度测评

西谚云：“真相还在穿鞋的时候，谎言可走遍半个世界”，这情形再没有比网络媒体时代来得更真实了。谬误信息可以迅速触及成千上万乃至数百万的读者，这要拜随心所欲、不负责任或恶意的转发以及优化社会参与的算法之赐。对网络新闻可信度的关注在全球范围内日益增长。网络新闻可信度测评已成为关乎信息生态系统健康的紧迫挑战，每一位身居其中的数字时代公民都受到影响。

您的团队受邀参与一项由一组国际研究者领导的研究，目的是开发一套可表示网络新闻可信度的指标。针对新闻文章的哪些方面会影响其可信度，这样的指标将有助于出版商、社交媒体平台和公众对此问题的理解和交流，并探讨这些影响的原因。

基于学者在识别谬误信息和表征新闻可信性方面的工作，在篇章层面表达新闻可信度的指标可通过分析文章本身的标题和文本以及检查文章的外部信息来源来确定。例如，点击诱饵标题和文本中的逻辑谬误通常传达文章中的谬误信息；而与文章的外部因素相关的事实检查的结果以及设置广告和社交信息邀请，通常也在一定程度上可透露文章的可信度。您的团队需要考虑使用哪些指标以及如何对这些指标进行测量，以创建你们的数学模型对网络新闻的可信度进行测评。

您的团队可以使用给定的数据资源，其中包含了 41 篇有关气候与健康主题的文章，这些文章在社交媒体上都受到广泛的转发。数据文件还包含了有关领域专家对新闻可信度的评估。

（数据资源：见文件 IMMC2020 Autumn Problem B- Data.xlsx）

您的团队有以下任务：

1. 建立测评网络新闻可信度的指标框架，并定义如何测量这些指标。
2. 利用您的指标体系，通过检查网络新闻的内容（标题和文本内容）和文章语境信息（外部信息源），建立测评网络新闻可信度的数学模型。当您把模型应用于给定的数据文件中的文章时，您可能需要为指标收集数据，您可请您能接触到的人士阅读这些样本文章，并对您的指标作出回答。
3. 使用给定的数据资源，将您的模型所给出的评估与领域专家对新闻报道的可信度评价进行比较，并讨论比较的结果。您的模型是否表明您的指标在网络新闻可信度测评方面有任何优势或不足？
4. 根据您的研究发现，请就如何提高公众意识以及如何改善网络新闻可信度，写一篇不超过两页的新闻稿给你所知的新闻出版商或社交媒体。

提交

您的方案论文应包含 1 页的摘要。正文不可超过 20 页，包括摘要和新闻稿则最多不超过 23 页。

附录和参考资料应置于正文末尾，且不计入 23 页的限制。

词汇表

谬误信息：虚假信息，或意图欺骗或误导的信息。

新闻可信度：新闻的真实性及基于事实的报道。

点击诱饵标题：具有某种网络内容的标题，其设计旨在吸引读者点击。

逻辑谬误：立论贫乏却具诱惑力，误导人们觉得其比理性更令人信服。

事实检查：为了确定文章中基于事实的陈述的准确性而检查文章中的事实性主张的行为，可以由诸如期刊编辑等个人或经批准的组织来执行。

设置广告和社交信息邀请：在文章网页安放广告；及/或添加劝导邀请，在社交媒体或通过电邮转发该文章，或加入电邮或社交媒体群组。方式通常是弹出窗口，掩盖文章内容，或通过动画、视频和音频来转移注意力。

感谢 本命题及数据集来源于麻省理工学院计算机科学及人工智能实验室（CSAIL）Amy X. ZHANG 博士、加州大学伯克利分校数据科学研究所 Aditya Ranganathan 博士等学者的在线新闻文章的可信度研究。



IM2C 2020 中華賽 B 題 (秋季賽) (English 簡體 繁體)
網絡新聞可信度測評

西諺云：“真相還在穿鞋的時候，謊言可走遍半個世界”，這情形再沒有比網絡媒體時代來得更真實了。謬誤信息可以迅速觸及成千上萬乃至數百萬的讀者，這要拜隨心所欲、不負責任或惡意的轉發以及優化社會參與的算法之賜。對網絡新聞可信度的關注在全球範圍內日益增長。網絡新聞可信度測評已成為關乎信息生態系統健康的緊迫挑戰，每一位身居其中的數字時代公民都受到影響。

您的團隊受邀參與一項由一組國際研究者領導的研究，目的是開發一套可表示網絡新聞可信度的指標。對於新聞文章的哪些方面會影響其可信度，這樣的指標將有助於出版商、社交媒體平臺和公眾對此問題的理解和交流，並探討這些影響的原因。

基於學者在識別謬誤信息和表征新聞可信性方面的工作，在篇章層面表達新聞可信度的指標可通過分析文章本身的標題和文本以及檢查文章的外部信息源來確定。例如，點擊誘餌標題和文本中的邏輯謬誤通常傳達文章中的謬誤信息；而與文章的外部因素相關的事實檢查的結果以及設置廣告和社交信息邀請，通常也在一定程度上可透露文章的可信度。您的團隊需要考慮使用哪些指標以及如何對這些指標進行測量，以創建你們的數學模型對網絡新聞的可信度進行測評。

您的團隊可以使用給定的數據資源，其中包含了 41 篇有關氣候與健康主題的文章，這些文章在社交媒體上都受到廣泛的轉發。數據文件還包含了有關領域專家對新聞可信度的評估。（數據資源：見文件 IMMC2020 Autumn Problem B- Data.xlsx）

您的團隊有以下任務：

1. 建立測評網絡新聞可信度的指標框架，並定義如何測量這些指標。
2. 利用您的指標體系，通過檢查網絡新聞的內容（標題和文本內容）和文章語境信息（外部信息源），建立測評網絡新聞可信度的數學模型。當您將模型應用於給定的數據文件中的文章時，您可能需要為指標收集數據，您可請您能接觸到的人士閱讀這些樣本文章，並對您的指標作出回答。
3. 使用給定的數據資源，將您模型所給出的評估與領域專家對新聞報道的可信度評價進行比較，並討論比較的結果。您的模型是否表明您的指標在新聞可信度測評方面有任何優勢或不足？
4. 根據您的研究發現，請就如何提高公眾意識以及如何改善網絡新聞可信度，寫一篇不超過兩頁的新聞稿給新聞出版商或社交媒體。

提交

您的方案論文應包含 1 頁的摘要。正文不可超過 20 頁，包括摘要和新聞稿則最多不超過 23 頁。附錄和參考資料應置於正文末尾，且不計入 23 頁的限制。

詞匯表

謬誤信息：虛假信息，或意圖欺騙或誤導的信息。

新聞可信度：新聞的真實性及基於事實的報道。

點擊誘餌標題：具有某種網絡內容的標題，其設計旨在吸引讀者點擊。

邏輯謬誤：立論貧乏卻具誘惑力，誤導人們覺得其比理性更令人信服。

事實檢查：為了確定文章中基於事實的陳述的準確性而檢查文章中的事實性主張的行為，可以由諸如期刊編輯等個人或經批准的組織來執行。

設置廣告和社交信息邀請：在文章網頁安放廣告；及/或添加勸導邀請，在社交媒體或通過電郵轉發該文章，或加入電郵或社交媒體群組。方式通常是彈出窗口，掩蓋文章內容，或通過動畫、視頻和音頻來轉移注意力。

感謝 本命題及數據集來源於麻省理工學院計算機科學及人工智能實驗室（CSAIL）Amy X. ZHANG 博士、加州大學柏克萊分校數據科學研究所 Aditya Ranganathan 博士等學者的在線新聞文章的可信度研究。