**虚假新闻多模态识别比赛第一名解决方案**

**Qingbo& bird团队**

**摘要**：**随着互联网快速发展，多媒体新闻数量迅速增长，我们生活在这样一个信息横流的时代，不免会受到网络上虚假信息的干扰，那么如何通过算法来自动识别多媒体新闻的真假变得越来越重要。同时虚假新闻试图利用多媒体内容来吸引和误导读者，并迅速传播，这使得视觉内容成为假新闻的重要组成部分。针对这些问题，本文提出了一个基于Gdbts-DenseNet-Bert（GDB）联合抽取特征的识别网络模型，实现了准确全面的对虚假新闻进行多模态识别，并达到了很好的识别效果。**

1. **赛题分析.**

本赛题为中国科学院计算技术研究所，以及北京智源人工智能研究院联合举办的虚假新闻识别挑战赛赛道三：多模态。随着多媒体技术的发展，新闻当中通常都包含着文本与图片等多模态信息，不同模态之间既存在着增强关系，同时也包含着互补信息。如何充分利用多模态信息进行虚假新闻检测，仍是一项具有挑战的工作。

因此本赛题的任务为：给定一条新闻的多模态内容，包括文本、配图、用户特征等，要求参赛者判断该新闻属于虚假新闻还是真实新闻。

1. **虚假新闻多模态识别**

虚假新闻是指故意且可核实为虚假的新闻文章。随着多媒体技术的发展，新闻由单纯的文本类型向含有文本，图片以及视频等多种方式结合拓展，来吸引更多的读者。然而虚假新闻也是利用这些特性来扭曲事实，误导读者，比如通过篡改图片或者修改人物名称，事件时间地点等来误导读者，传播虚假信息。其中，我们将虚假新闻图片大致分为两类:篡改图片和误导图片。篡改图片是指经过修改后的虚假新闻图片。误导性图片是指没有经过任何修改的虚假新闻图片，但其内容具有误导性，这些也促使了虚假新闻多模态识别的快速发展。

早期检测 [4] [5] 是否为虚假新闻的方法是通过提取特征来检测新闻中附属的图片是否存在被修改的痕迹，这种只适合第一类篡改图片，而不适合第二类误导性图片的识别。因此，包括传统学习[6,7,8]和基于深度学习的模型[9,10]在内的多种假新闻检测方法被用来识别假新闻。通过对不同事件的充分验证，现有的深度学习模型由于其在提取方面的优势，已经达到了超越传统模型的效果。

[11] [12] 利用VGG19等经过预处理的卷积网络获取视觉信息，由于没有关联性信息，很难找到虚假新闻的语义特征。Qi P 等人[13] 构建了一个多分支的CNN-RNN网络来提取不同语义层次的特征，以充分捕捉假新闻图像在像素域的特征。该模型是一个多域视觉神经网络(MVNN)框架(如下图所示)，该框架通过结合频率域和像素域的信息来学习有效的视觉表征，用于假新闻的检测。该模型由三个主要部分组成:一个频域子网络和一个像素域子网络，分别用于在物理和语义层面捕获假新闻图像的特征，并利用融合子网络对这些特征进行动态融合。

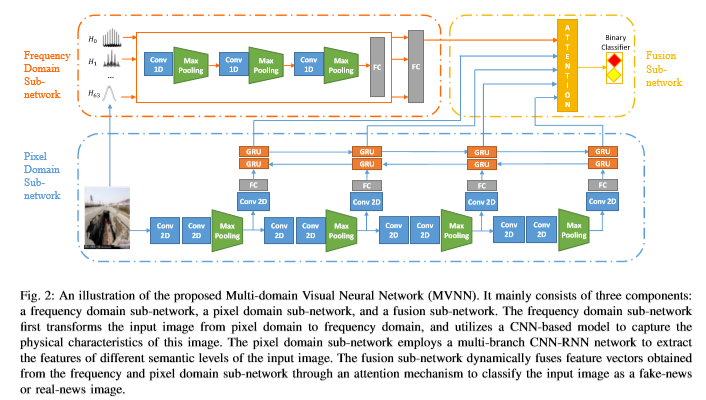


图2-1. MVNN框架图

大多数的网络倾向于学习新闻事件特定的特性，而这些特性不能被转移到不可见的事件中。为了解决这一问题，Wang Y 等人 [14] 提出了一种端到端的框架结构—事件对抗神经网络(EANN)，它具有事件不变性的特点。如下图所示，在该模型中，利用TextCNN抽取文本模态特征，利用VGG-19抽取视觉模态语义特征，并将两种模态特征拼接作为虚假新闻的内容特征表达。该特征一方面用来判断该新闻是否为虚假新闻，另一方面用来判断该新闻来自于哪一事件模型。

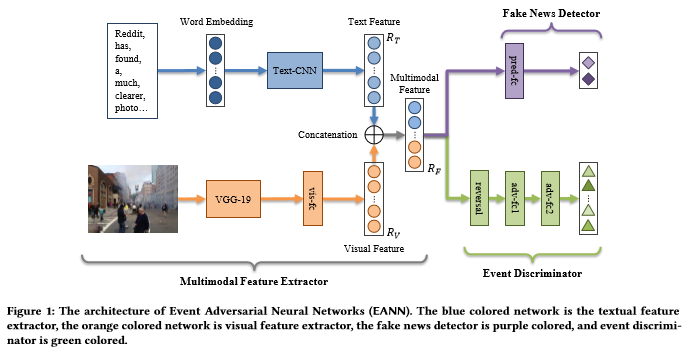


图2-2．EANN框架图

该模型有两个目标，一方面要使虚假新闻的检测结果要尽可能准，另一方面要使得事件判别器要尽可能的不准(对抗部分)，以保证所学到的特征是事件无关的。该论文用到了对抗学习的思想，本质上是多任务联合学习，但是由于比赛的训练样本标签只给出了是否为虚假新闻的字段，并没有标注出改新闻表示的事件。所以针对这里面的比赛数据无法使用论文中的模型框架。我们参试使用了改模型的特征提取层的神经网络模型，后面直接连接singmoid层，训练了一个分类模型，实际测试效果，出现了过拟合现象，线上效果并不是很好。

Dhruv 等人[15]提出一种端到端的网络，多模态变分自编码器(MVAE)，在该模型中，文本模态与视觉模态的拼接特征被编码为一个中间表达，利用重构损失保证编码后的中间表达可以解码回原状态，并用该中间表达的向量做虚假新闻检测。该编码器与一个二进制分类器相结合，用于虚假新闻的检测。主要由三个部分组成：编码器、解码器和虚假新闻检测模块。

如下图所示：

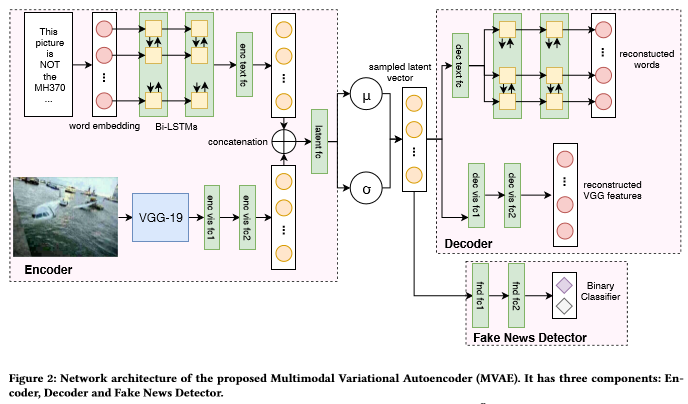


图2-3. MVAE框架图

该论文灵感来自于自动编码器的想法，试图学习共享。[]提出了变分自动编码器(VAE)的概念，从而为复杂的分布式系统提供了一种全新的生成模式。为了克服当前模型的局限性，提出了一种能够学习共享(视觉+文本)表示的多模态变分自动编码器，经过训练来发现tweets中不同模式之间的相关性。然后，VAE与一个分类器耦合以检测假新闻。VAEs通过对观测数据的边缘似然值进行优化，使变分自编码器具有学习概率潜在变量模型的能力。虚假新闻检测器利用从双模变分自动编码器获得的多模态特征来区分新闻是否虚假。

**3. 比赛思路.与模型的设计**

**3.1数据探索与分析**

训练样本一共38471条，正负样本比例接近1:1，给出的信息字段包括以下字段。

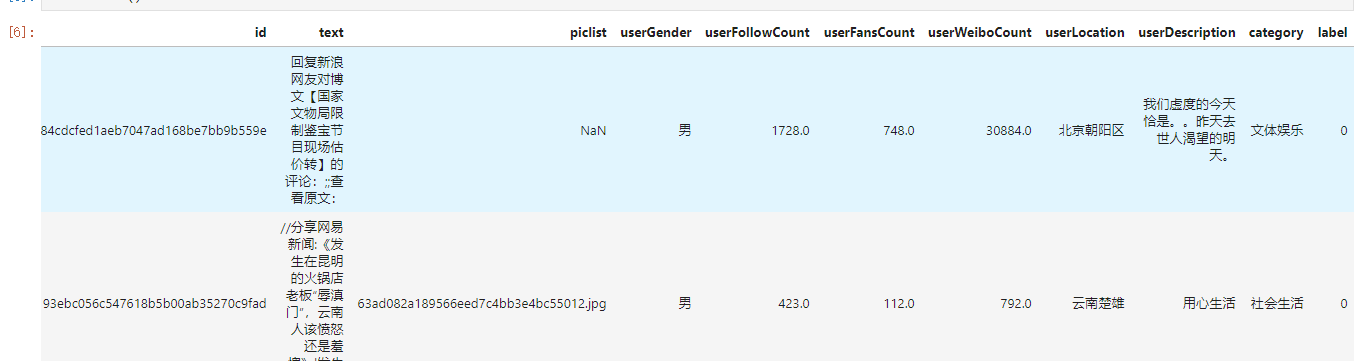


图3-1.部分信息字段

其中piclist为新闻中嵌入的图片路径，通过cv库可以读取相应的图像文件。我们通过python对原始特征数据以及构造的特征进行数据分析。

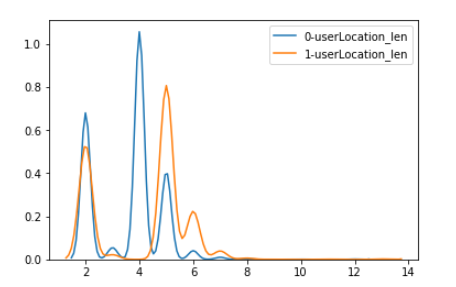


图1-2正负样本在userLocation字段长度上的分布差异

图3-2.字段长度地域分布差异

图3-2展示的是正负样本不同地域字段长度上面的分布差异，横坐标表示的是不同的地域，纵坐标表示的是字段长度的大小。



图3-3.虚假新闻在部分userLocation字段的分布差异

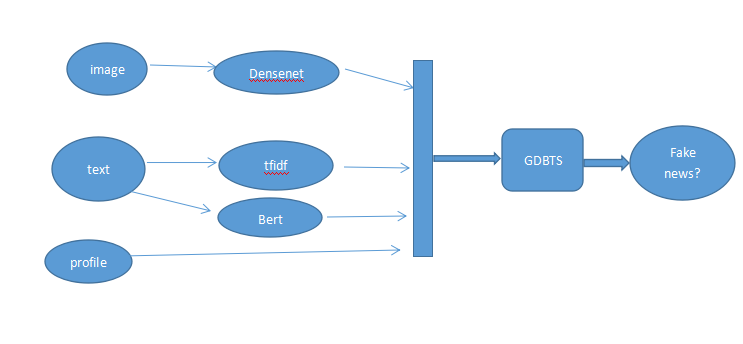
从图中可以看出样本中不同区域发的新闻有偏向，部分地区的多媒体新闻几乎全为假新闻，还有部分地区的多媒体新闻几乎全部为真新闻。

上面展示了我们部分对比赛数据分析的过程，同理我们还对其余的字段进行了数据分布上面的探索和分析。这些结构化的原始特征以及构造的相关统计特征，对我们后面使用GDBT-based模型训练帮助很大。

**3.2数据预处理**

本道赛题的数据预处理主要是针对非结构化数据的text文本数据字段，使用了jieba分词库对text进行了分词，为我们后面通过词袋模型提取ngram特征做准备。

**3.3模型部分**

多媒体新闻主要包含三类特征，第一类是图像特征，训练数据中含有图片的样本占了80%以上；第二类是文本特征；第三类是多媒体新闻的发布或者转发者的用户信息特征，比如粉丝数目，关注数，用户简介等用户画像特征。根据新闻多种特征信息，我们提出一种多模态虚假新闻识别模型，结构如下：图3-4.GDB模型框架

在这里我们使用了GDBT-based的模型[1]，其中针对图像特征，我们使用densenet121[2]预训练模型的最后一个全连接层的输出作为图像的语义特征。针对text文本字段我们使用了tfidf提取ngram特征。最后我们把图像的语义特征，ngram和bert[3]提取的文本特征以及用户画像特征拼接到一起输入GDBT-based模型，进行训练，最终得到一个虚假新闻判断模型。

**4总结**

在本次比赛中我们研究学习了对多媒体新闻从多模态的角度进行真假识别。提出了GDBT-based-DenseNet-Bert的GDB虚假新闻多模态识别模型，实验证明基于改模型从多模态角度提取。比单领域的特征提取效果更好。同时实验也证明了GDB框架识别虚假新闻的优势，最终我们线上的成绩f1达到了99%,领先第二名两个百分点。

**文献：**

1. Si Si, Huan Zhang, Sathiya Keerthi ,Dhruv Mahajan ,Inderjit S. Dhillon ,Cho-Jui Hsieh , Gradient Boosted Decision Trees for High Dimensional Sparse Output
2. Gao Huang,Zhuang Liu,Laurens van der Maaten,Densely Connected Convolutional Networks

[3][Jacob Devlin](https://arxiv.org/search/cs?searchtype=author&query=Devlin,+J), [Ming-Wei Chang](https://arxiv.org/search/cs?searchtype=author&query=Chang,+M), [Kenton Lee](https://arxiv.org/search/cs?searchtype=author&query=Lee,+K), [Kristina Toutanova](https://arxiv.org/search/cs?searchtype=author&query=Toutanova,+K) ,BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

[4] C. Boididou, S. Papadopoulos, D.-T. Dang-Nguyen, G. Boato, M. Riegler, S. E. Middleton, A. Petlund, Y. Kompatsiaris et al., “Verifying multimedia use at mediaeval 2016.” in MediaEval, 2016.

[5] C. Boididou, S. Papadopoulos, D.-T. Dang-Nguyen, G. Boato, and Y. Kompatsiaris, “The certh-unitn participation@ verifying multimedia use 2015.” in MediaEval, 2015.

[6] NiallJConroy,VictoriaLRubin,andYiminChen.2015. Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology 52,1(2015),1–4.

[7] ZhiweiJin,JuanCao,YongdongZhang,JiansheZhou,andQiTian.2017. Novel visual and statistical image features for microblogs news verification. IEEE transactions on multimedia19,3(2017),598–608.

[8] EugenioTacchini,GabrieleBallarin,MarcoLDellaVedova,StefanoMoret,and LucadeAlfaro.2017. Somelikeithoax:Automated fake news detectioninsocial networks. arXiv preprint arXiv:1704.07506 (2017).

[9] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting Rumors from Microblogs with RecurrentNeuralNetworks..InIJCAI.3818–3824.

[10] NataliRuchansky,SungyongSeo,andYanLiu.2017.CSI:A Hybrid Deep Model for Fake News Detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.ACM,797–806.

[11] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, “Multimodal fusion with recurrent neural networks for rumor detection on microblogs,” in Proceedings of the 2017 ACM on Multimedia Conference. ACM, 2017, pp. 795–816.

[12] K. Dhruv, G. Jaipal Singh, G. Manish, and V. Vasudeva, “Mvae: Multimodal variational autoencoder for fake news detection,” in Proceedings of the 2019 World Wide Web Conference. ACM, 2019.

[13] Qi P , Cao J , Yang T , et al. Exploiting Multi-domain Visual Information for Fake News Detection[J]. 2019.

[14]Wang Y , Ma F , Jin Z , et al. EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection[C]// the 24th ACM SIGKDD International Conference. ACM, 2018.

[15] Khattar Dhruv, Goud Jaipal Singh, Gupta Manish, and Varma Vasudeva. Mvae: Multimodal variational autoencoder for fake news detection. In Proceedings of the 2019 World Wide Web Conference. ACM, 2019.