大数据分析与应用-不良信息发现综述

陈婉月，杨敏，徐元超，杨洋

摘要：为净化网络环境，保护我国未成年人成长环境，也为了维护社会国家的稳定，我们需要对网络信息进行审查。针对网络信息中所包含的不良信息，尤其是不良文本与不良图像识别成为了一个迫切需要解决的问题。对不良信息发现的技术全面的做了分析与总结。实验结果表明，文中所提出的结合机器学习、深度学习等技术对不良文本、不良图像的识别能够有效的过滤互联网上的不良信息。

关键词：不良文本；不良文本变体；字符串相似度计算；不良图像；机器学习；深度学习；神经网络

**Big Data Analysis and Application - Summary of Bad Information Detection**

Chen Wanyue, Yang Min, Xu Yuanchao, Yang Yang

**Abstract**：In order to purify the network environment, protect the environment for the growth of minors in China, and to maintain the stability of the country, we need to review the network information. For the bad information contained in the network information, especially the bad text and bad image recognition has become an urgent problem to be solved. The technology for the discovery of bad information was comprehensively analyzed and summarized. The experimental results show that the combination of machine learning and deep learning techniques in the paper can effectively filter bad information on the Internet.

**Key words**：bad text；bad text varian ；string similarity calculation ；bad image；machine learning；deep learning；neural networks；

# 引言

## 研究背景和意义

随着全球信息技术革命的发展，作为引领信息技术革命的互联网渗透到社会生活的各个方面，并深刻改变着人们的生产生活方式。同时，全球各个国家纷纷推出自己的互联发展策略，进一步加速了互联网的发展[1]。近几年，移动互联网更是迎来了爆发式的发展。手机，平板电脑等便携式移动设备的发展，使得人们可以随时随地接入互联网。与此同时，这些发展也增加了人们对于实时性互动的需求，微博、微信和直播等平台成为移动互联网的主流应用。移动互联网时代的“终端即服务”成为终端产业的新模式。

根据中国互联网信息中心（CNNIC）发布的第40次《中国互联网络发展状况统计报告》显示[2]，截至2017年6月，中国网民的规模达7.51亿，半年增长率为2.7%。互联网普及率为54.3%，较2016 年底提升1.1 个百分点。其中，手机网民占比高达 96.3%，达7.24亿。在互联网终端中，移动互联网处于主导地位而且还处于不断的增长之中。

互联网特别是移动互联网的高速发展，也为人们的信息交流和发布提供了方便的途径。信息交流与发布的主要方式是社交类应用，信息交流和发布的形式也从长文本为主的书面文本形式转变为以短文本为主的口语化方式。CNNIC经过调查和研究，将社交应用分为以微信朋友圈、微博为代表的综合类社交应用和以豆瓣等为代表的垂直类社交应用[2]。其中，微信朋友圈、微博等以短文本为主的综合类社交应用的使用率排名占据前三。其次，弹幕文化，商品、新闻评论以及影评等短文本信息也在信息交流和发布中占有很主要的席位。以弹幕为例，弹幕作为观影评论，以实时的方式让人们在观看网络视频以及直播时进行评论交流，用户以青少年为主。CNNIC报告[2]显示，截至2017年6月，中国网络视频用户规模达5.65亿，网络视频用户使用率为75.2%。其中，手机视频用户规模为5.25亿，手机网络视频使用率为72.6%。弹幕作为网络视频的一部分，其并发数量也是不容小觑。

面对如此大量的网络资源，缺乏统一的监管，致使网络信息中会出现危害国家安全、色情等有害信息以及非法广告等阻碍人们获取有效信息。同时，在线下较为隐蔽的色情、非法广告等，也因互联网匿名、免费等因素在互联网上传播。在传统网站，贴吧，论坛和聊天室等以传统方式散布的色情、非法广告等不良信息，也逐渐“适应”移动互联网的发展，把不良信息通过微博，微博评论以及弹幕等方式散布，对人们获取信息造成障碍，对青少年造成不良影响。

据CNNIC发布的《2015年中国青少年上网行为研究报告》[3]显示，截至2015年12月，中国青少年网民的规模达到2.87亿，占全国青少年总数的 83.5%。其中，6-11岁占比11.5%，12-18岁占比40.4%，19-24岁占比48.1%。青少年上网设备中，手机的占比高达90%。智能手机的碎片化使用场景更受青少年青睐。另外，CNNIC数据显示，交流沟通类应用（除邮件外）和网络娱乐类应用在青少年中的使用率均高于网民总体水平。从数据中可知，青少年对于互联网尤其是移动互联网的使用比例高于总体水平，其中与青少年比较容易接触新鲜事物息息相关。但是，与此同时，青少年对于不良信息的接触也会增加。2013年8月至12月，浙江团省委对14至35周岁青少年做了一项有关手机网络环境的调查[4]。调查显示，在有过浏览色情、暴力信息记录的青少年中，有78.8%属于“被动浏览者”，比例远超“主动浏览者”。随着近几年智能手机的发展，微博、直播等平台的兴起，以及色情信息的多样化，被动浏览情色、暴力信息的情况可能更为严重。由于情色信息的巨大的利益趋势，这些信息的发布者使用各种各样的手段逃避监督，对这些信息的过滤造成了一定的困难。

互联网不良信息种类主要为以下几种：

1. 国家安全：藏独、疆独、台独、港独等
2. 涉贿犯罪：涉赌、涉毒、暴力恐怖、传销、诈骗、邪教等
3. 个人伦理：低俗、谩骂、人身攻击、色情、反伦理、垃圾广告等

基于互联不良信息载体过多，我们主要对不良文本和不良图片进行介绍。

## 文字过滤研究现状

对不良短文本识别的问题，其本质就是短文本的二分类问题，将文本分为正常短文本和不良文本短两类。然而，对网络上的短文本分类除了克服文本自身的难点外，还要解决网络短文本所具有的特性。对于普通的非网络短文本而言，目前的研究主要集中在解决短文本长度较短，需要大量的背景知识，特征稀疏等问题上。对于网络不良短文本的新特性，主要问题是如何在短文本中发现不良文本的各种特征。

对于普通短文本而言，当前研究中重点在于如何增加短文本信息量以及如何扩展文本特征。在短文本信息量增加的问题上，此前一部分研究者通过增加文本主题信息来提高分类准确度。文献[5]提出一种基于隐含文本主题模型（LDA）和K近邻的方法进行短文本分类。作者认为，概率主题可以增加短文文本语义信息和减少特征空间的稀疏性。同时根据“主题-词”矩阵和短文本之间有区分度的项之间的关系来计算主题间的相似度。文献[6]通过已经混合方法，将分布密集的项与TF-IDF值相结合，以期减少含有信息量较少的项对短文本分类的影响。文献[7]提出一种基于强特征词库的方法来对短文本进行分类。强特征词库中的特征是基于LDA主题模型和信息增益模型获得的。该强特征词库增加了短文本之间的语义信息，使短文本分类效果得到提升。文献[8]为短文本社交情感分类提出了主题级别最大熵模型（TME）。TME通过隐含主题建模、多情感标签和多个读者打分等产生主题级别的特征。最大熵原理中的过拟合问题也通过将特征映射到高维空间来解决。文献[9]通过大量背景文本产生短文本的LDA模型，并通过LDA将短文本的主题特征作为扩展特征加入到原短文本特征，从而提高短文本的分类准确率。文献[10-14]也通过一定的方法对短文本的分类进行增强。

现在，随着深度学习模型在文本领域的应用，也有一部分研究者通过词的分布式表示来构造短文本的特征向量，来解决短文本的信息量不足和稀疏性等问题。文献[15]通过词向量和卷积神经网络(CNN)形成一个统一框架来扩展短文本。首先利用快速聚类生成语义派系，然后，利用可变窗口的上下文对词嵌入的叠加组合，计算短文本中多尺度语义单元。在嵌入空间中，选择语义单元的受限最近词向量构成扩展矩阵。最后，将工程矩阵和扩展矩阵联合后的矩阵作为CNN的输入。文献[16]将循环神经网络(RNN)和卷积神经网络(CNN)相结合，来解决顺序短文本分类的问题。文献[17]假设短文本文档是“高斯-贝叶斯”框架分布的一个特例，然后利用聚类算法在潜入空间中扩展和丰富短文本特征。文献[18]同时结合了LDA和词向量两种方式对短文本进行扩充。通过大量与短文本主题一致的外部数据来建立LDA模型。将所有的外部数据，短文本以及主题词组合成一个大的数据集，并在这个大的数据集上训练关于主题词和所有词的词向量。然后所有的短文本都使用这些词向量表示进行训练和分类。文献[19-21]通过卷积神经网络对短文本进行分类，也取得了一定效果。

有些研究者也利用外部资源对短文本信息进行归一和扩充。文献[22]受词袋模型的启发，生成一个“内容袋模型”。利用大量分类知识库，对于短文本的每个分类学习一个内容模型，然后将每一个短文本转化为内容模型中相关内容的集合。最后使用基于内容相似度的机制来将短文本划分到最近似的类别。文献[23]利用公共搜索引擎对短文本进行丰富短文本信息。首先，将短文本作为公共搜索引擎的输入，得到搜索引擎的返回结果。然后，将返回结果作为原短文本的背景知识并从中提取特征向量。文献[24]提出一种基于维基百科的短文本表示方法，得到短文本中含有的维基百科的内容信息，并将这些信息和原来的短文本信息一同作为特征向量表示。文献[25]利用知网和领域关键词对文本进行分类。其中，利用概率主题模型获取分类主题对应的关键词，利用《知网》计算短文本和领域关键词之间的文本相似度实现分类。以上的研究工作都是通过增加短文本的语义信息来提高短文本的分类准确度。文献[26]通过对短文本的语义信息增强短文本分类。文献[27]通过同义词词林扩展来增强短文本之间的联系，提高短文本的分类精度。

对于不良文本中的不良信息的发现，也有一些研究者进行了研究。文献[28]针对单核SVM无法兼顾不良文本以及不良文本组合的问题，提出了一种新的组合核函数提高了对不良词汇以及词汇的组合的识别的准确率。文献[29]指出了中文和英文关于基于内容的垃圾邮件分类的不同。该文章通过实验验证明了同样的特征提取方法和分类方法再中文和英文的数据集上的得到的结果不同，有些甚至相差甚远。另外，中文文本和英文文本还有一个很大的不同，那就是中文文本需要分词。中文分词的准确性很大程度上影响到不良信息的识别，同时，中文变体词会给中文分词造成很大的困扰。文针对英文的变体词，文献[30]通过计算变异敏感词汇相似度的方法，对变体词进行识别，达到了不错的效果。对于中文的变体词，文献[31]利用变体词所在文本的语义和时间信息，在了解词性的基础上，做出了字词联合的变体词的规范化的研究。首先根据时间和语义结合的方法获取候选目标词，然后使用基于字词联合的词向量方法对目标候选词排序。该方法的不需要进行对数据进行标注，并取得了不错的效果。文献[32]详细地介绍了中文变体词的研究发展。作者在文章中指出，变体词的识别和规范化虽然取得了一定的进步，但是对于识别效果还是不太理想，对网络文本变体的演化规律的研究也还有一定的缺乏，对于变体词的研究是一个有重要意义的课题。

以上研究者对于短文本以及不良短文本的识别，从外部知识拓展、不良信息识别等角度提高了短文本分类的精度。但是，由于网络短文本信息来源复杂，主题杂乱，主题个数难以确定，使用主题方法的效果不是很好；外部知识扩展过分依赖外部知识库和在线搜索引擎；变体词识别对于出现的变体词进行了还原，但是由于不断有新的变体词出现，变体词识别精度难以把握。为了充分利用不良短文本特征，我们从词、文本整体以及语义等多个方面提取不良文本特征，提出了一种新的不良短文本特征提取方法。

## 图片识别研究现状

随着互联网应用的普及，图片成为了全球范围内重要的信息载体和传播信息的最主要渠道之一。但随之而来也产生了一些日益突出的问题，尤其是反动、色情、暴力等不良内容在互联网上的传播已经成为日益突出、有待解决的问题，网络内容的识别与过滤逐渐被各个国家重视。1998年美国通过了“Child Online Protection Act”（COPA）法案，开始加强对网络内容的监控与过滤，2003年美国最高法院裁定国家公共电子图书馆必须安装色情过滤器，日本等国家有类似规定。我国各部委也正积极合作打造网络视频管理新规范，大力支持这方面的研究工作。可以预见，不久的将来随着中国互联网的规范化，类似的规定将会出台。

目前互联网色情图片、视频过滤技术也在不断进步，从基于关键字、黑名单的方式逐渐向基于内容的过滤技术过渡。如基于颜色、纹理的低级语义过滤，以及基于人体形状的高级语义过滤技术。

基于内容的图片识别技术是正确识别色情图片的基础。较早将基于内容的图片识别技术与色情图片过滤进行研究的是Stanford计算机系的Wang，他构造的WIPE系统综合了图标过滤、图标检测直方图过滤、文理过滤和一个小波基性状匹配算法，放在前面的是一些快速过滤器，通过这些快速过滤器可以很快地将类似于图标、图表以及一些含肤素很少的图片作为正常图片滤出，只有那些肤色和皮肤文理都满足给定的图片才能进入到最后的形状匹配模块。在形状匹配之前，先计算出该图一系列力矩等特征向量，最后用计算得到的特征向量与存储起来的训练图特征向量进行匹配，只有匹配符合度达到一定程度的才被判断为色情图片。该系统处理速度较快，但需要构造一个数量足够打的训练图片特征库。由于色情图片和非色情图片种类繁多，如何才能有效组织色情图片和非色情图片，同时选取什么样特征才能表述这些种类繁多的色情与非色情图片，是该系统的难点之处。

加州大学伯克利分校的和设计实现了一个裸体人自动识别系统，用来判断图片中是否含有裸露人体。该系统由两部分组成肤色过滤模块和人体肢体组合模块。肤色过滤模块联合颜色和纹理属性来标记图片中的肤色区域人体肢体组合模块利用几何分析法检测肤色区域中的细长区域，并把它们组合成可能的人体肢体，再连接这些人体肢体以确定是否含有人体，从而判断图片中是否有裸露人体。该系统从另一个角度分析了检测人体的可能性，但试验结果却不让人满意，在幅测试库上的正确检测率只有，并且处理时间较长。后来提出了基于统计颜色模型的肤色检测方法，并利用计算得到的肤色掩码特征向量分类色情图片。该方法的提出显示出颜色信息可以作为肤色检测乃至色情图片过滤的重要因素，为低层次特征向高级语义特征的转化提供了途径。该方法需要得到肤色在颜色中的先验统计信息，因此只有当统计量足够大时，其统计信息才有意义。在色情图片为5241幅、非色情图片为13970幅的图库上测试正检率为85.8%，误检率为7.4%。随后研究出现了一个基于统计肤色检测的色情图片过滤系统，该系统提出了一阶树模型（TFOM）的肤色检测方法，得到肤色掩码图片后，再在肤色掩码上提取特征向量，最后通过训练一个多层感知器（MLP）来对提取的特征向量进行分类。2003年美国VIMA公司（其前身是Morpho软件公司）退出的ImageBeagI是一款用于检测硬盘和网络缓存文件中色情图片的软件。ImageBeagle利用PBIR（Perception-Based Image Recognition）技术，从图片中提取出150多个特征变量，然后在一个高维空间中对图片进行分类识别。该公司的webFilter软件也是一款用于网络过滤的软件，webFilter的识别引擎也是基于PBIR技术的。WebFilter的一个突出特点是它可以区分出色情图片和艺术作品中的裸露人体的区别。

国内方面，2001年中科院计算所的段丽娟等人提出一种基于计算机视觉和模式识别的色情图片过滤方法---多层次特定类型图片过滤法。该方法通过研究人类视觉对图片的分析机理，提出综合肤色模型检验、支持向量机分裂（SVM）和最近邻方法校验的多层次图片处理框架。实验结果表明，该方法能够达到85%以上的准确率，但检测模型有待于进一步完善。台湾国立成功大学的胡冠宇提出了一种检测人体特定部位的色情图片过滤器，该过滤器可以检测到裸露人体的胸前部位和私处部位，再结合脸部特写分类器，在1538幅测试图片库上可以达到89.79%的检测率。这一检测思路有别于上述各个系统，为色情图片过滤提供了新思路。上面的这些研究者们都提出了很多令人鼓舞的研究结果，但离真正的自动色情图片过滤应用还有很大的距离。

通过对国内外相关技术的调研发现，色情图片识别相关技术比较少，从其在应用中的重要地位来看，迫切需要我国集中技术、资金和人力，尽快开发出高水平的基于内容的色情图片识别平台，促进网络环境的健康，使广大网民远离色情图片内容的危害。

# 不良文本变体相似度计算

## 不良文本变体整理

语言学和认知科学方面的研究表明，语音和字型是人脑识别英语词汇的两个重要方面[15-18]。通过对词汇变异体变换情况分化我们发现词汇变异主要从语音和词汇两个方面来进行变异的。一方面，变异词汇是根据语音相似进行变异的，将词汇中的某个字母（串）变换成发音相似的另外的一个字母（串），这样人脑在读到这个变异词汇时，会根据变异词汇的发音情况将这个变异词汇还原；另一方面，有些变异词汇是基于字形来进行变换的，将某些字母替换成形状相似的另外的字母或者特殊字符，这样人在阅读到这个变异词汇时候会基于字形将运个变异词汇进行还原。故此，在进行变异体词汇识别时，需要从语音和字形两个因素考虑，进行变异词汇相似度计算，才能得到更加精确的结果。

我们对收集到的各种词汇变形体进行分析，整理出下图这样的变异情况，该图基本涵盖了不良词汇常见的所有变体情况。

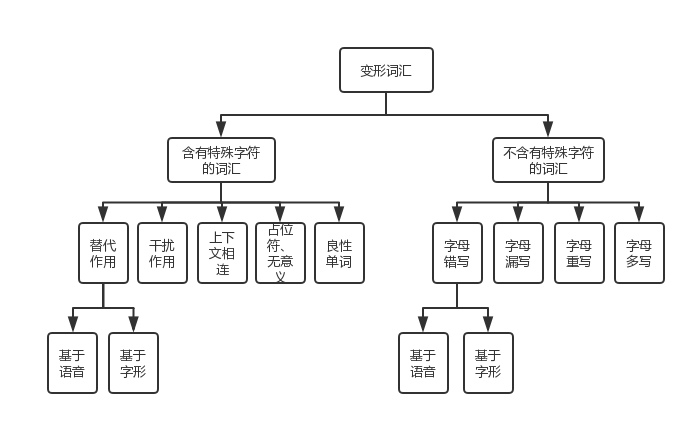


图 2‑1 不良词汇变形整理

上图对不良词汇的各种变形体进行了整理，现在对上图分类的各种情况，做解释说明。变形词汇分为含有特殊字符的情况和不含特殊字符的情况，含有特殊字符的情况是指不良词汇变形体中含有非字母的符号，比如“@”；不含特殊字符的变体，是指变形体中只含有字母，没有“#”“￥%”等字符。为了更清楚的对各种情况进行说明，在下表中，本文对各种情况进行了整理解释。

表格 1 不良词汇变形说明

|  |  |  |  |
| --- | --- | --- | --- |
| 变形词汇 | 变形情况 | 解释说明 | 举例说明 |
| 含有特殊字符情况 | 替代作用 | 特殊字符替代某个字母 | 例如：shit写成sh!t |
| 干扰作用 | 将字母换成特殊字符，以使用系统检测不到变体 | 例如：fuck写成f#$% |
| 上下文相连 | 将不良词汇写成两部分，中间用特殊字符相连接 | 例如：fuck写成fu-ck |
| 占位符、无意义 | 某个特殊字符表示此处位置有个英文字母 | 例如：fuck写成f\*\*\* |
| 良性词汇 | 某个特殊字符的存在是正常使用的情况 | 例如：微博中的@某人 |
| 不含有特殊字符 | 字母错写 | 将不良词汇的某个字母写成其他字母 | 例如：fuck写成fack |
| 字母漏写 | 将不良词汇串中的某个字母漏写 | 例如：fuck写成fck |
| 字母重写 | 将不良词汇串中的某些字母重写 | 例如：shit写成shiiiit |
| 字母多写 | 在不良词汇串中加入其它的字母 | 例如：shit写成shikt |

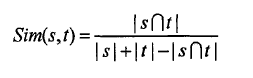
## 常用的字符串相似度计算方法

现有的对不良文本进行过滤的方法对于规矩的词汇（即用户没有故意误写的词汇）效果是不错的，但是现实生活中，恶意用户发表的不良敏感信息将词汇做变形处理，则现有的过滤不良信息的方法处理效果很差。论文[14]指出在线聊天室中，频繁出现的拼写错误使通过分类方法来检测不良信息的效果很差，因为这些敏感词经过拼写错误后，就很难再识别出来。但是，用户在阅读信息的时候，通过人脑对送些错误拼写的不良词汇自动纠正，会识别出不良信息来。

不良词汇的变形体在经过变形处理后，一般的精确匹配算法就很难将运些不良词汇变形体识别出来，为此在识别这些变形体的时候，就需要计算送些字符串的相似度。无论把不良文本信息过滤作为分类问题研究，还是直接对变异词汇进行识别，对变异词汇进行相似度的计算都是非常重要的。而相似度度量方法选择的优劣则直接影响着变异词汇识别的准确率。论文［19-20］指出Jaccard，N-gram和编辑距离是当前化种主要词汇相似度度量方法。

### 基于Jaccard的相似度计算

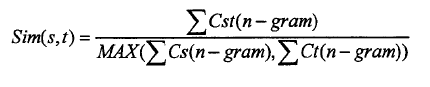
Jaccard方法经常用于计算两个集合之间的相似度，论文[21]使用Jaccard用于检测重复网页，论文[22]使用Jaccard用于相似字符串的检测。Jaccard的定义如下：



其中，表示源字符串向量s和目标字符串t中包含的相同非零项的个数，表示向量s中的非零项的个数，表示向量t中非零项的个数。Jaccard计算字符串相似度的时候，不考虑字符之间的顺序。

### 基于N-gram的相似度计算

目前，N-gram法在匹配技术中的运用较为广泛，其基本思想[23]是：先输入两个字符串，即源字符串s与目标字符串t，列出源字符串和目标字符串所有的N-gram，然后计算同现的N-gram的频率为源字符串s中n-gram的数量，为目标字符串t中n-gram的数量，则可以计算相似度：



### 基于编辑距离的相似度计算

编辑距离（Edit distance）ED[24]就是将一个字符串转换成另一个字符串所需要进行插入、删除、替换等相关编辑操作的最小次数。编辑距离被经常用于计算两个字符串之间的差异程度，例如：DNA分析、拼写检查、语音辨识及抄袭侦测等等。

例如：“kitten”和“sitting”的编辑距离李是3，计算过程如下：

1. kitten → sitten（将“s”替换为“k”）
2. sitten → sittin（将“i”替换为“e”）
3. sittin → sitting（在最后插入字母“g”）

字符串a和b的编辑距离为，具体计算定义如下：

最小编辑距离的算法描述如下：

Fuction Min-Edit\_Distance(target， source)

n = length(target);

m = length(source);

create distance matrix d[n，m];

d[0，0] = 0;

d[0，1] = 1， … d[0，m] = m;

d[1，0] = 1， … d[n，0] = n;

for each i from 1 to n do

for each j from 1 to m do

cost = targeti == sourcej?0:1;

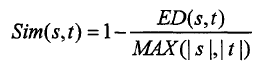
d[i，j] = min(d[i-1，j]+1，

d[I，j-1]+1，

d[i-1，j-1]+cost)

return d[n，m]

利用编辑距离计算相似度、并将其映射到[0，1]区间可得到下面的公式：



其中|s|表示字符串s的长度，|t|表示字符串ｔ的长度。

编辑距离计算字符串相似度对于变形体词汇效果不好，比如汁算“@$$”和“ass”的相似度是０，而人在阅读网络信息的时候遇到“@$$”时，会将“@$$”当作”ass”来理解

### 基于Yoon’s method的相似度计算

这个计算字符串相似度的方法是论文[]8提出来的，该方法一定程度上解决了含有特殊字符的词汇串相似度计算的问题，为了描述问题的方便，这里就将该方法称为Yoon’s method。该方法使用global alignment来计算两个字符串的相似度，对待特殊字符的时候将其转换成相似的字母。具体如下：

表格 2 字母与特殊字符的相似度分数

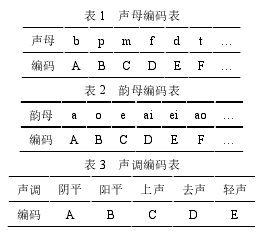
|  |  |  |
| --- | --- | --- |
| 字母 | 特殊字符 | 分数 |
| a | @ | 0.8 |
| s | $ | 0.8 |
| an | & | 1.2 |
| i | ! | 0.4 |

上表显示的是将殊字符和相似字母的分数，若两个字符相同则是1。那么字符串”fall”和字符串”f@ll”的分数就是3.8，相似度是3.8/4.0\*100＝95%。这种计算字符串相似度的方法一定程度上解决了“脏话”词汇的变形体，但是对于很多特殊字符则不能够处理。比如对于类似的变形体”f\*\*\*”（fuck的变异体）则不能够处理。

### 音码编码法

采用三位音码的的形式，利用字母对其中每一位音码进行编码。三位音码中，设第一位为声母位，第二位为韵母位，第三位为声调位。根据 1958 年第一届全国人民代表大会第五次会议批准公布推行的《汉语拼音方案》，其中包含声母 23个，韵母 34 个，为了方便编码，声调分为阴平、阳平、上声、去声和轻声。

表格 3音码编码



通过以上编码，可将汉字转换成一系列的字符序列，以便进行下一步的计算和比较。以“海洛因”为例，基于以上编码表得到的音码为“KDCHXDPTA”。对得到的音码编码再使用编辑距离进行相似度计算，就可以得到易混淆字符串的相似度。

以“海洛因”和“海诺因”为例，设 为 90%，其拼音分别 为 “ hailuoyin ” 和 “ hainuoyin ” , 转 换 成 音 码 之 后 为“KDCHXDPTA”和“KDCGXDPTA”，相似度为 88.89%，

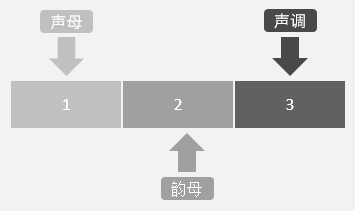


图 2‑2 音码编码形式

### 音形码编码法

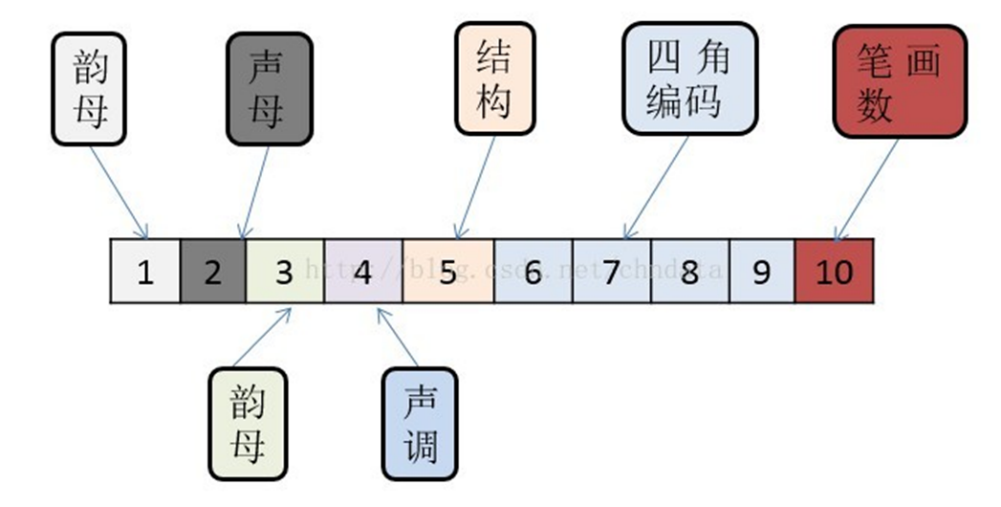


图 2‑3 音形码编码形式

整个音形码共分两部分，第一部分是音码部分，主要覆盖了韵母，声母，补码以及声调的内容。

第一位，是韵母位，通过简单的替代规则，将汉字的韵母部分映射到一个字符位。汉字的拼音中一共有24种韵母，其中部分为了后期计算目的，采用相同的字符来替代，以下是一张完整的匹配表：



图 2‑4 韵母编码

我们对于an和ang，所使用的是同一种转化，目的便是为了再后期计算相似度的时候，将这种差异弱化。对于没有这种需求的应用来说，完全可以自行生成映射表。

第二位是声母位，同样的，也是利用一张替换表来将声母转换成字符：

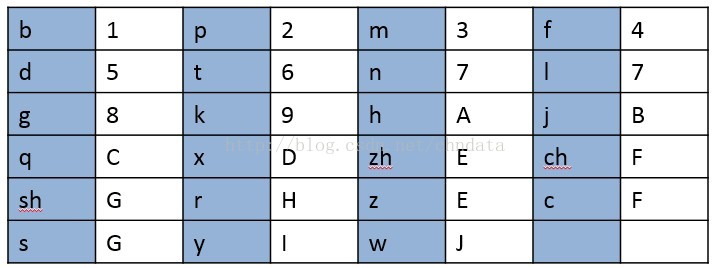


图 2‑5 声母编码

可以发现，z和zh等相似的平舌音卷舌音用的也是相同的转化。

第三位则是补码，通常用于当声母和韵母之间还有一个辅音的时候，采用的是韵母表相同的替代规则。

第四位是声调位，分别用1，2，3，4来替代汉字中的四声。

第二部分是字形码。

第一位被称为结构位，根据汉字的不同结构，用一个字符来表示该汉字的结构。



图 2‑6 字结构编码

接下来的四位，则依然是借用了四角编码，来描述该汉字的形态。由于四角编码表过长，在这里就不一一列举了。

最后一位，是汉字的笔画数位， 从一到九，分别代表该汉字的笔画为一到九，接下来是A代表10位，B代表11位，并依次类推。 Z代表35位，以及任何超过35位的都用z。

举例说明：汉字 “琅”，它的音形码编码是：

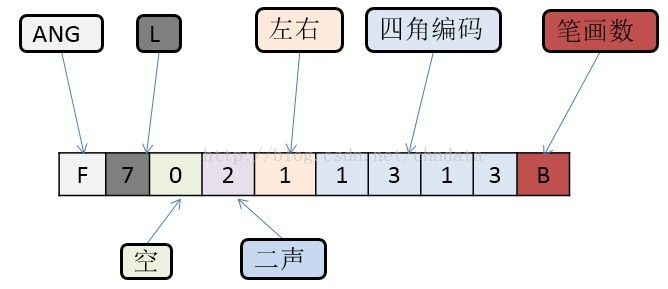


图 2‑7 “琅”字的编码

通过这样的方式，将汉字首先转换成了一系列的字符序列，这样我们就可以采用一定的办法，来计算他们的相似度。

字串相似度的计算可以通过直接将字符串中的每个汉字转化为音形码，再将所有音形码合并起来进行EditDistance算法比较，即可获得。

# 文本不良信息发现

## 文本识别技术的研究现状

信息内容预处理的结果好坏是影响过滤精度的决定性因素之一。对文本内容进行数据预处理,主要包括将信息源中的无关文本信息的清除,如删除描述信息的标签、脚本及非文本对象等,保留有用文本信息并将其描述特征分离出来进行量化,然后将能反映或有助判别内容性质的显性和隐性信息提取出来,使待过滤的内容能通过特征项有效表达。对于中文文本过滤来说,涉及到中文的分词、语法语义分析等自然语言处理过程和停用词处理、描述特征项维数约简等文档向量降维过程。与正常的web不同的是,含有不良信息内容的网页具有很多非内容的暗示性和隐性的特征。特别是近年,不良文本内容为躲避监控日渐呈现出表现多样性、隐蔽性和多变性等特点。给Web文本内容,尤其是中文文本的数据预处理和特征提算法进行不良文本信息过滤带来了巨大的挑战。传统文本挖掘的方法很难有效提取出能表达文本性质的特征信息来对文本进行性质判别[6]。

对于过滤系统无法对一些发生形式变化的词条进行识别的问题,相关研究提出可利用关联分析的方法挖掘不良文档中潜在的特征项,通过实验验证了该方法对特征词条的识别具有较高的查全率和准确率。相关研究对色情网页的页面布局、因特网内容选择平台等级评定应用、暗示性条文和文档内容四个方面的特征进行了详细的分析探讨,并提出了相应的提取特征项的方法。相关研究对互联网中表现多样性和多变性的非法文本具有的内容结构和用词形式两方面的特征进行了研究分析,并提出对此类文本的预处理方法。该方法充分利用此类信息中词出现的特征,在属性预选取中,将标点符号作为潜在特征词的触发信息进行文本预选属性集的提取,最后实验验证了该方法对增加合法文本与非法文本的区分度有较明显的效果。对多语种不良网页结构布局和暗示条文等进行了分析,提出了相关的挖掘和处理隐性特征的方法及通过字符编码方式的差异处理多语言混合不良内容。

## 语义识别技术

网络文本语义识别一般包括文本预处理、文本表示模型、信息抽取、文本分类、聚类及关联分析等步骤。

（1）文本预处理：文本信息挖掘的第一步是文本化处理。这里的文本预处理包括文本内容提取及文本预处理。文本内容提取是将半结构化的数据进行文本内容提取,完成网页到文本的转化。而文本预处理则根据语种而定,英文文本预处理包括去词根处理、词干提取及文本的分段、分句等;而中文文本预处理则包括中文词语切分、停用词去除等。

（2）文本表示模型:文本表示模型是将文本内容表示为一定的模型,方便计算,经典文本表示模型的代表是向量空间模型(vector space model),它是将文本进行词语统计,并根据词语的权重来将文本构造成一定的向量空间来表示的文本表示模型。TF-IDF是计算词权重的一个非常重要的计算方法,它是基于词向量的一种传统学习算法。TF-IDF假设词在文本中出现的频率是评价其权重的唯一标准。这种假设带来的弊端是明显的。最近几年,研究者又提出了基于概念的文本表示方法,用概念向量代替词向量,一个概念可以由多个词术语组成,这种概念文本表示方法考虑了一定的语义和词之间的依赖关系,可以在一定程度上提高文本处理的精度。

（3）信息抽取：信息抽取的目的是在非结构化的自然语言文本中抽取结构化的数据单元,从而使自由文本数据变成结构化的数据。传统信息抽取方法一般包括基于规则的半自动抽取方法及有指导的机器学习方法。网络文本信息挖掘从非结构化的自由语言文本中发现隐含的知识和信息,因此,信息抽取一般作为网络文本信息挖掘的前期处理,信息抽取效果的好坏将直接影响信息挖掘结果的优劣。基于信息抽取的网络文本信息挖掘逐渐发展成为这一领域的研究趋势。

（4）文本分类、聚类：文本分类和文本聚类是一般的浅层文本处理技术。对于大量文本数据的组织、开发,文本分类/聚类技术在文本信息处理中起着至关重要的作用,高效的文本分类/聚类算法是传统网络文本信息挖掘的基础。一般来说,文本分类是在有指导的文本类别区分方法,分为训练和测试两部分,训练是将预先设定好的类别交给分类器训练,然后利用训练好的分类器对测试语料进行分类测试。而文本聚类则是一种无指导的学习方法,没有预先类别的训练部分,仅依靠一定的学习算法对大量的文本进行类别区分。

（5）关联分析：关联分析是一种利用一定的关联规则对文本数据进行趋势预测分析的方法。其中,关联规则是数据中一种简单实用的规则,它属于描述型模式;发现关联规则的算法则属于无监督的学习方法l6]。最初的关联分析从超市购买模式上发展而来,是对规整数据立方体进行综合分析的一种方法。随着研究深入,关联分析的研究趋势从单一概念层次上的关联规则发现发展到多概念层次上的关联规则发现,研究对象也从单一规模的数据库发展到对多数据库、多维数据立方体上。

## 各个文本分类算法特点

分类算法的速度是决定不良信息发现速度的主要因素之一。目前常用的基于文本内容理解的基本信息分类算法主要有向量空间模型(VSM)、K-近邻(KNN)、支持向量机模型、贝叶斯决策等。

（1）向量空间模型:向量空间模型是基于统计的分类系统中广泛采用的文本计算模型,它把文档简化为以特征项的权重为分量的一个高维向量表示,把文本信息过滤过程简化为空间向量的运算,使得问题的复杂性大大降低。在进行文本过滤的过程中,首先对请求的页面数据进行加工,将其看成是一个由m个特征项及其权值表示的向量D,然后比较向量D和不良文本模板向量P的相似程度.其最突出的特点是可以方便地计算出数据库中不良文本向量和输入文本向量的相似度,然后通过两向量的相似性确定输入文本是否过滤目标。该方法可以简单地将文档与模板的表示统一起来,表示简洁、计算简单有效、可操作性好,使过滤系统的实现比较容易,因而在文本过滤中获得了广泛应用并取得了良好效果。缺点是使用的语义和语法要素较少,特征项之间互相独立,不能区分特征项出现在不同位置对表达文档主题性质能力的差异,不能保留文本结构信息,不能充分反映出文本总体面貌。而且,特征项权重难确定,文档的相关度计算量大,影响过滤速度。

（2）K-近邻分类算法（K-Nearest Neighbor algorithm），简称KNN算法，即是给定一个训练数据集，对新的输入实例，在训练数据集中找到与该实例最邻近的K个实例（也就是上面说的K个邻居），这K个实例的多数属于某个类，就把该输入实例分类到这个类中。KNN算法中，所选择的邻居都是已经正确分类的对象。该方法在定类决策上只依据最邻近的一个或者几个样本的类别来决定待分样本所属的类别。KNN算法本身简单有效，它是一种 lazy-learning 算法，分类器不需要使用训练集进行训练，训练时间复杂度为0。KNN 分类的计算复杂度和训练集中的文档数目成正比，也就是说，如果训练集中文档总数为 n，那么 KNN 的分类时间复杂度为O(n)。[KNN](https://baike.baidu.com/item/KNN/3479559)方法虽然从原理上也依赖于极限定理，但在类别决策时，只与极少量的相邻样本有关。由于KNN方法主要靠周围有限的邻近的样本，而不是靠判别类域的方法来确定所属类别的，因此对于类域的交叉或重叠较多的待分样本集来说，KNN方法较其他方法更为适合。

（3）SVM算法:SVM算法是基于有序风险最小化归纳法的统计学习方法。其核心思想是使用简单的线性分类器划分样本空间,通过在特征空间构建具有最大间隔的最佳超平面得到两类主题之间的划分准则,使期望风险的上界达到最小。对于在当前特征空间中线性不可分的模式,则使用一个核函数把样本映射到一个高维空间中,并在新的空间完成点积运算,使得样本能够线性可分。支持向量机在文本分类领域得到了比较成功的应用,具体过程如下:

根据训练集生成向量空间;

b)生成训练集的特征向量;

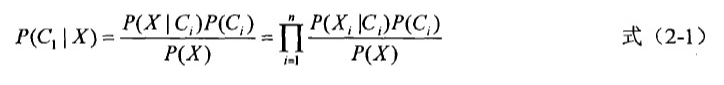
c)利用训练集的特征向量进行训练,产生决策函数;

d)对来自网络的文本内容向量化,生成该文本内容的特征向量;

e)利用决策函数对该特征向量进行分类,根据分类结果进行不良文本过滤决策。

基于支持向量机的过滤模型的优点是训练样本数量增减对分类器影响不大,可以通过有限的训练样本学习而获得较好的分类效果,可有效克服不良文本过滤中样本分布不平衡的问题;其主要缺点是计算量大、速度慢[4]。

（4）贝叶斯决策模型:贝叶斯决策模型通常是利用简单而有效的朴素贝叶斯理论建立模型。基于朴素贝叶斯决策模型的过滤算法的基本思想是根据以往的判断经验估计某一文档属于相关文档或无关文档的概率。它将训练实例S分解成特征向量x(x1,x2,x3,......)和决策类别变量c,并假定特征向量的各分量相对于决策变量是相对独立的。基于朴素贝叶斯决策模型在计算文档所属类别时,文档属于类别的概率等于文档中每个特征项属于类别的概率的综合表达,最后根据各类别的概率的计算结果选择概率最大的类别作为该文档的类别。算法中将信息源的文档分别分成正常文本和不良文本两类(c1,c2).),在检测新文档时,根据特征词分词,按下面的朴素贝叶斯公式计算该文档尤属于正常文本(c1)和不良文本(c2)的概率:



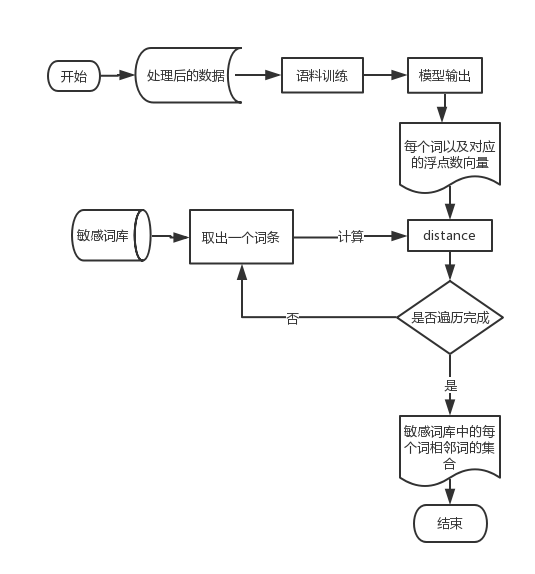
使用贝叶斯决策模型来判定文本,能够解决预处理算法对多语种兼容性问题,而且算法逻辑简单、易于实现,过滤过程中时间空间开销小、算法稳定、健壮性比较好。因此,贝叶斯决策模型在各种文本分类和过滤领域中得到了广泛的应用,特别是在垃圾邮件过滤领域。但是,这种模型并没有考虑特征项在文档中出现的频率,因为所有的权重都是二元的,而且特征项是建立在独立性假设基础上的。但独立性假设在许多实际问题中并不成立,所以在实际的内容过滤任务,特别是在复杂内容的过滤时,常会引起判别的误差。

## 不良关键词挖掘技术

网络上传播的关键词还存在时效性和流行性,当前流行的不良关键词可能不在系统的关键词库中,也可能系统的关键词库中的关键词有些己经失效,所以,系统具备不良关键词挖掘技术就十分重要,可以实时地更新系统的关键词,提高系统的中标率和准确率,不放过任何不良信息。网络不良关键词挖掘旨在提取文本集中的命名实体,抽取实体之间的关系,利用这些信息去结构化地表示文本语义,然后采用关联分析技术在文本集中进行知识模式的挖掘,系统已经采集了大量的中标网站,有大量的人工审核违规的文本可以作为样本参考,并且可以从每天中标的文字中发现新的违规文字,实时挖掘不良关键词。

（1）基于word2vec拓展关键词：word2vec工具于2013年末由Google公司发布,它是深度学习在自然语言领域的一项了不起的应用。word2vec在自然语言处理中广泛应用于文本聚类、同义词扩展、词性分析等等诸多方面口。word2vec训练大量的语料后可把语料文本转换为数学形式的词向量计算。

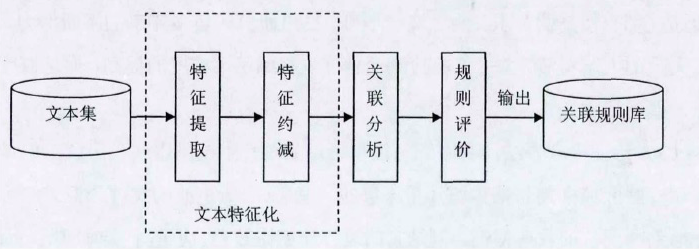
用一个普通向量来表示一个词，这样的普通向量表示形式, 通过词之间的距离(比如余弦相似度、欧式距离等确定它们之间的语义相似度。通过词向量将自然语言转化为可供机器认识的数学符号,即用词向量将一个词用一个向量来表示。通过向量空间的余弦相似度得出词与词之间的相似度,即可得出需求的敏感基础词库中各个词的相邻词的集合。这主要利用敏感基础词库和word2vec完成对敏感基础词库相关特征词的扩充工作。具体扩充过程如图3-1所示。



**3-1 敏感词库扩充过程**

（2）命名实体提取：命名实体提取和传统的命名实体识别的区别在于:命名实体识别是在未经过中文分词的文本中直接提取命名实体;命名实体提取重在研究在中文分词基础上如何利用命名实体识别相关技术完整地提取命名实体。命名实体识别的方法主要分为:基于规则的方法和基于统计的方法。较早的基于规则的命名实体识别方法多采用手工构造有限状态机的方法,使用模式和字符串相匹配。典型的系统有用于英语命名实体识别的爱丁堡大学的LTG系统,这些系统主要是基于规则的。

（3）文本关联分析：关联分析算法是数据挖掘中的一项重要技术,它属于无监督学习的方法,其基本思想是通过频度计算发现数据集中大于用户给定最小支持度和最小置信度的关联规则,应用范围非常广泛。文本关联分析是关联分析算法在文本数据集中的应用,包括文本集特征化、关联规则挖掘以及规则评价和冗余处理等几个部分,如图3-2所示。



**3文本关联分析原理框图**

文本特征化:对挖掘对象(一般为文本集)建立特征表示,即将文本转换成一种类似于数据库中记录的具有较严格结构且能反映文本完整内容特征的表示方式。目前采用的文本表示方法中,文本特征向量具有非常高的维数,为了便于后面的处理还需要进行维数约减。

关联分析:运用关联分析的各种方法来提取面向特定应用的知识模式

规则评价:对获取的关联规则进行质量评价。若满足事先定的要求,则存储或输出该关联规则,否则将这个关联规则删除或者与其它规则合并。

# 不良图片检测传统方法

## **基于数字签名**

### Photo DNA起源

2009年，微软与达特茅斯学院共同开发了PhotoDNA技术，该技术旨在发现和消除来自互联网的儿童色情图片。

### 工作原理

工作原理引入，如图4.1所示：

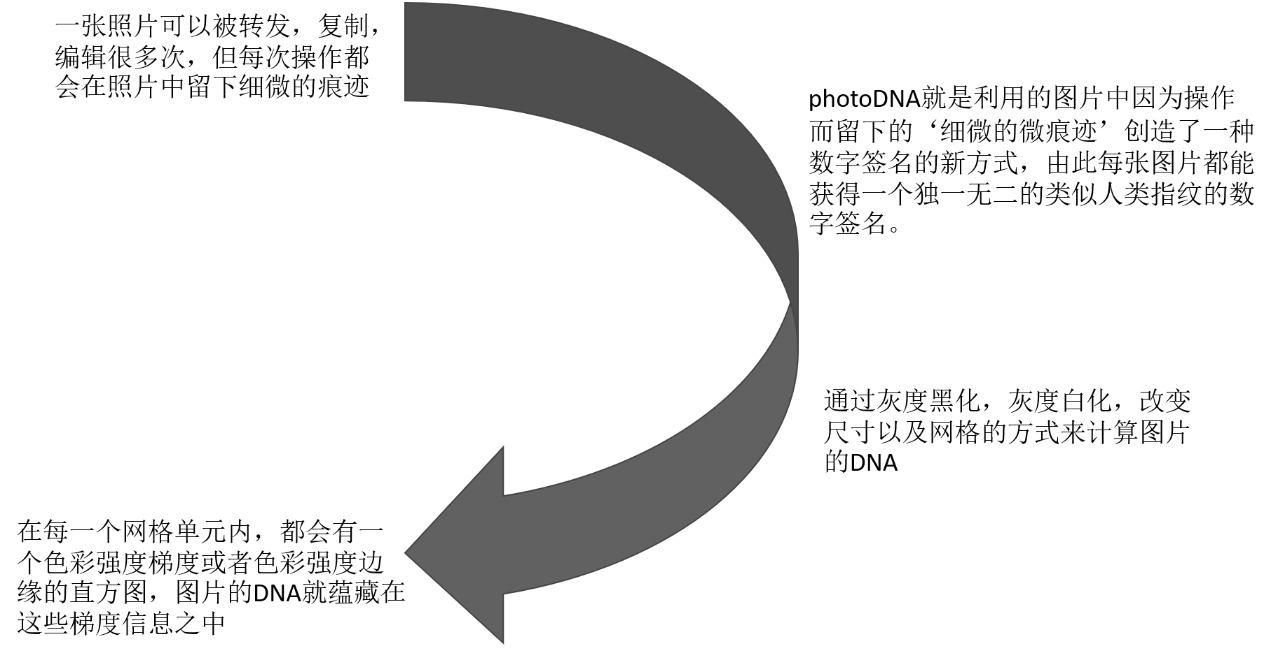


图4.1. photoDNA 工作原理

PhotoDNA技术能够通过分析图片中所蕴含的数字指纹（digital fingerprints），来寻找网络上的匹配图像，即各种类型的相似图像。该技术的基本原理类似于杀毒软件中常采用的病毒特征库机制。

### 效果

识别率达 99.7%，20 亿次检测中仅一次误报。

### 应用

微软将PhotoDNA技术首先应用在自家的SkyDrive、Hotmail和Bing搜索服务上，又将该技术捐给了NCMEC（美国失踪与受虐儿童援助中心）。在此之后，NCMEC就以PhotoDNA技术为基础建立了一整套完善的在线服务系统，用以打击儿童色情犯罪。

图片识别技术目前被全球超过70个公司采用，其中包括Facebook。微软和Facebook都有着极其庞大的用户基础和巨大的发展潜力，此类在线服务的跟进对优化互联网整体环境大有益处。全球两家顶尖的互联网公司同时采用顺应市场“游戏规则”的领先技术，如果 PhotoDNA 技术的应用领域扩展到罪犯嫌疑人识别、知识产权内容的保护，可以说潜力无限，更重要的是，无人工干涉的过程将避免更多的隐私官司。

## **基于内容检测**

图片识别在不良图片检测方面，等价于二分类的问题，把图片输入，输出结果为不良图片或者非不良图片。

### 特征提取

特征提取是基于内容的图片检测必须的步骤，在此介绍一下颜色特征提取，矩阵特征提取和sift特征提取的三个方法。

#### 颜色特征

基于不良信息中的色情图片检测，介绍一下颜色特征，是一种全局特征，描述了图像或图像区域所对应的景物的表面性质。

一般颜色特征是基于像素点的特征，此时所有属于图像或图像区域的像素都有各自的贡献。由于颜色对图像或图像区域的方向、大小等变化不7感，所以颜色特征不能很好地捕捉图像中对象的局部特征。另外，仅使用颜色特征查询时，如果数据库很大，常会将许多不需要的图像也检索出来。

颜色直方图是最常用的表达颜色特征的方法，颜色直方图能简单描述一幅图像中颜色的全局分布，即不同色彩在整幅图像中所占的比例，特别适用于描述那些难以自动分割的图像和不需要考虑物体空间位置的图像。缺点在于：它无法描述图像中颜色的局部分布及每种色彩所处的空间位置，即无法描述图像中的某一具体的对象或物体。

颜色直方图法是一种全局颜色特征提取与匹配方法，无法区分局部颜色信息。颜色集是对颜色直方图的一种近似首先将图像从 RGB颜色空间转化成视觉均衡的颜色空间（如 HSV 空间），并将颜色空间量化成若干个柄。然后用色彩自动分割技术将图像分为若干区域，每个区域用量化颜色空间的某个颜色分量来索引，从而将图像表达为一个二进制的颜色索引集。在图像匹配中，比较不同图像颜色集之间的距离和色彩区域的空间关系。

但是仅根据颜色检测容易出现误判，包括以下几种情况：

a). 漏出肤色较多的大头像

b). 肤色的卡通图片

c). 裸露的婴儿图片

为了防止以上误判，可以引入目标精确检测（比如人脸检测，肢体检测等）。

#### Haar特征

基于不良信息中的危险设备检测，在此介绍一下Haar特征提取方法。

Haar特征是一种反映图像的灰度变化的，像素分模块求差值的一种特征。简单分为三类：边缘特征、线性特征、中心特征和对角线特征，如图4.2所示。

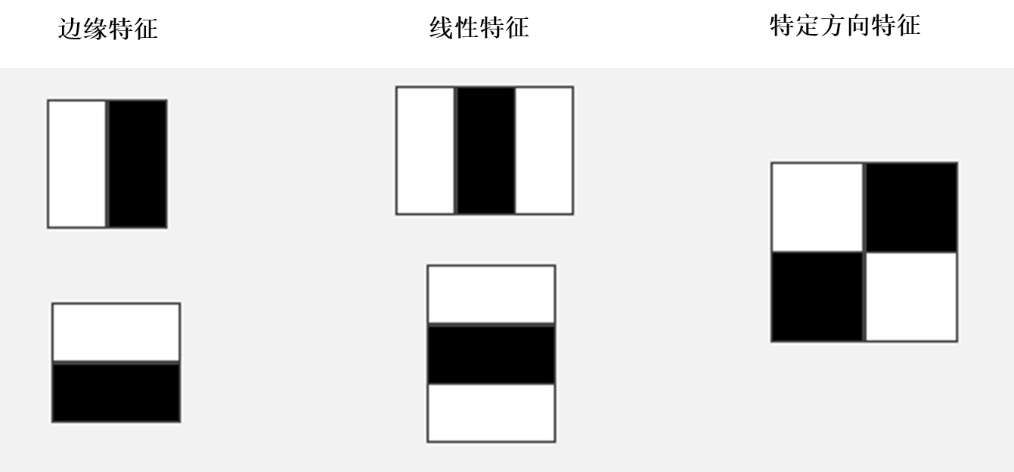


图4.2 矩形特征

用黑白两种矩形框组合成特征模板，在特征模板内用 黑色矩形像素和减去白色矩形像素和来表示这个模版的特征值。例如：脸部的一些特征能由矩形模块差值特征简单的描述，如：眼睛要比脸颊颜色要深，鼻梁两侧比鼻梁颜色要深，嘴巴比周围颜色要深等。但矩形特征只对一些简单的图形结构，如边缘、线段较敏感，所以只能描述在特定方向（水平、垂直、对角）上有明显像素模块梯度变化的图像结构。

#### SIFT特征提取

SIFT（Scale-invariant feature transform）是一种检测局部特征算法，该算法通过求一幅图中的特征点（interest points,or corner points）及其有关scale 和 orientation 的描述子得到特征并进行图像特征点匹配。

SIFT特征不只具有尺度不变性，即使改变旋转角度，图像亮度或拍摄视角，仍然能够得到好的检测效果。整个算法分为以下几个部分：

(1).构建尺度空间

这是一个初始化操作，尺度空间理论目的是模拟图像数据的多尺度特征。高斯卷积核是实现尺度变换的唯一线性核，于是一幅二维图像的尺度空间定义为：

4.1

(2).LoG近似DoG找到关键点<检测DOG尺度空间极值点>

为了寻找尺度空间的极值点，每一个采样点要和它所有的相邻点比较，看其是否比它的图像域和尺度域的相邻点大或者小。如图所示，中间的检测点和它同尺度的8个相邻点和上下相邻尺度对应的9×2个点共26个点比较，以确保在尺度空间和二维图像空间都检测到极值点。一个点如果在DOG尺度空间本层以及上下两层的26个领域中是最大或最小值时，就认为该点是图像在该尺度下的一个特征点,如图4.3所示。

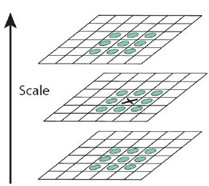


图4.3 尺度空间内的邻域

(3).除去不好的特征点

这一步本质上要去掉DoG局部曲率非常不对称的像素。通过拟和三维二次函数以精确确定关键点的位置和尺度，同时去除低对比度的关键点和不稳定的边缘响应点(因为DoG算子会产生较强的边缘响应)，以增强匹配稳定性、提高抗噪声能力。

(4.) 给特征点赋值一个128维方向参数

上一步中确定了每幅图中的特征点，为每个特征点计算一个方向，依照这个方向做进一步的计算，利用关键点邻域像素的梯度方向分布特性为每个关键点指定方向参数，使算子具备旋转不变性。

4.2

4.3

公式4.2，4.3分别为为(x,y)处梯度的模值和方向公式。其中L所用的尺度为每个关键点各自所在的尺度。至此，图像的关键点已经检测完毕，每个关键点有三个信息：位置，所处尺度、方向，由此可以确定一个SIFT特征区域。

(5). 关键点描述子生成

首先将坐标轴旋转为关键点的方向，以确保旋转不变性。以关键点为中心取8×8的窗口。

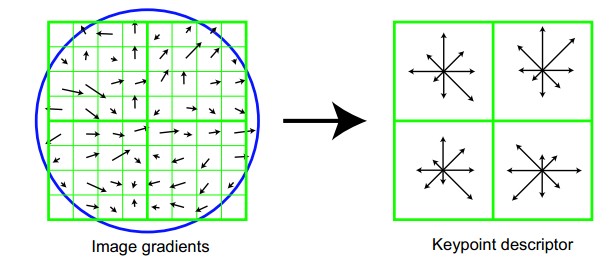


图4.4 特征点梯度方向(左)及加权到8个主方向的效果图(右)

图左部分的中央为当前关键点的位置，每个小格代表关键点邻域所在尺度空间的一个像素，利用公式求得每个像素的梯度幅值与梯度方向，箭头方向代表该像素的梯度方向，箭头长度代表梯度模值，然后用高斯窗口对其进行加权运算。

图中圆圈代表高斯加权的范围（越靠近关键点的像素梯度方向信息贡献越大）。然后在每4×4的小块上计算8个方向的梯度方向直方图，绘制每个梯度方向的累加值，即可形成一个种子点，如图右部分示。此图中一个关键点由2×2共4个种子点组成，每个种子点有8个方向向量信息。这种邻域方向性信息联合的思想增强了算法抗噪声的能力，同时对于含有定位误差的特征匹配也提供了较好的容错性。

(6). 根据SIFT进行Match

生成了A、B两幅图的描述子，（分别是k1\*128维和k2\*128维），就将两图中各个scale（所有scale）的描述子进行匹配，匹配上128维即可表示两个特征点match上了。

### 4.2.1 BOW-SVM模型

#### Bow特征提取

Bow（Bag of words）直译就是词袋，从名字可看出BOW其实是文本领域的处理方法。现在我们迁移到图像领域，文本可以把句子拆为单词，类似的，图像也将展示成视觉词汇，以假钞为例，转换过程如图4.5所示。

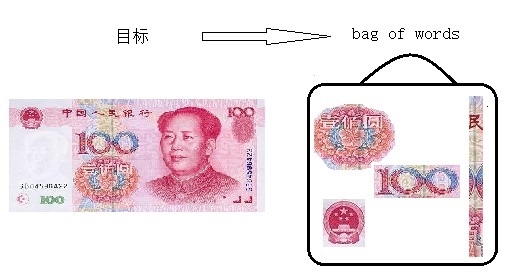


图4.5 视觉词汇提取

由于图像中的词汇不像文本文档中的那样是现成的，我们需要首先从图像中提取出相互独立的视觉词汇，这通常需要经过三个步骤：

第一步：利用SIFT算法，从每类图像中提取视觉词汇，将所有的视觉词汇集合在一起，如图4.6所示。



图4.6 视觉词汇集合

第二步：利用K-Means算法构造单词表。K-Means算法是一种基于样本间相似性度量的间接聚类方法，此算法以K为参数，把N个对象分为K个簇，以使簇内具有较高的相似度，而簇间相似度较低。SIFT提取的视觉词汇向量之间根据距离的远近，可以利用K-Means算法将词义相近的词汇合并，作为单词表中的基础词汇，假定我们将K设为4，单词表的构造过程如图4.7所示。

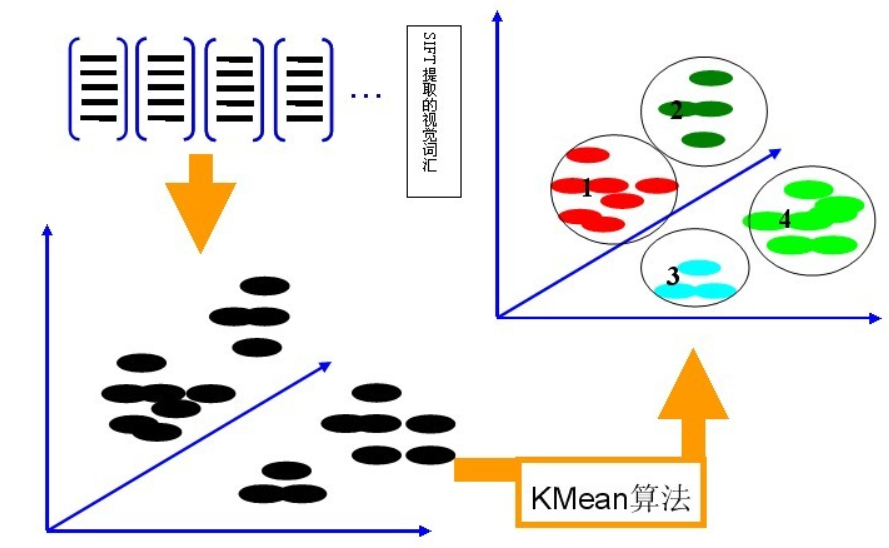


图4.7 单词表构造

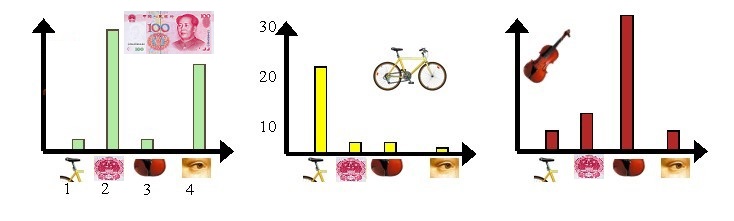
第三步：利用单词表的中词汇表示图像。利用SIFT算法，可以从每幅图像中提取很多个特征点，这些特征点都可以用单词表中的单词近似代替，通过统计单词表中每个单词在图像中出现的次数，可以将图像表示成为一个K=4维数值向量。如图4.8所示****

图4.8 图像向量表示

#### 训练SVM分类器

SVM是很常见的机器学习分类器模型，在此用一张流程图介绍该算法，如图4.9所示。

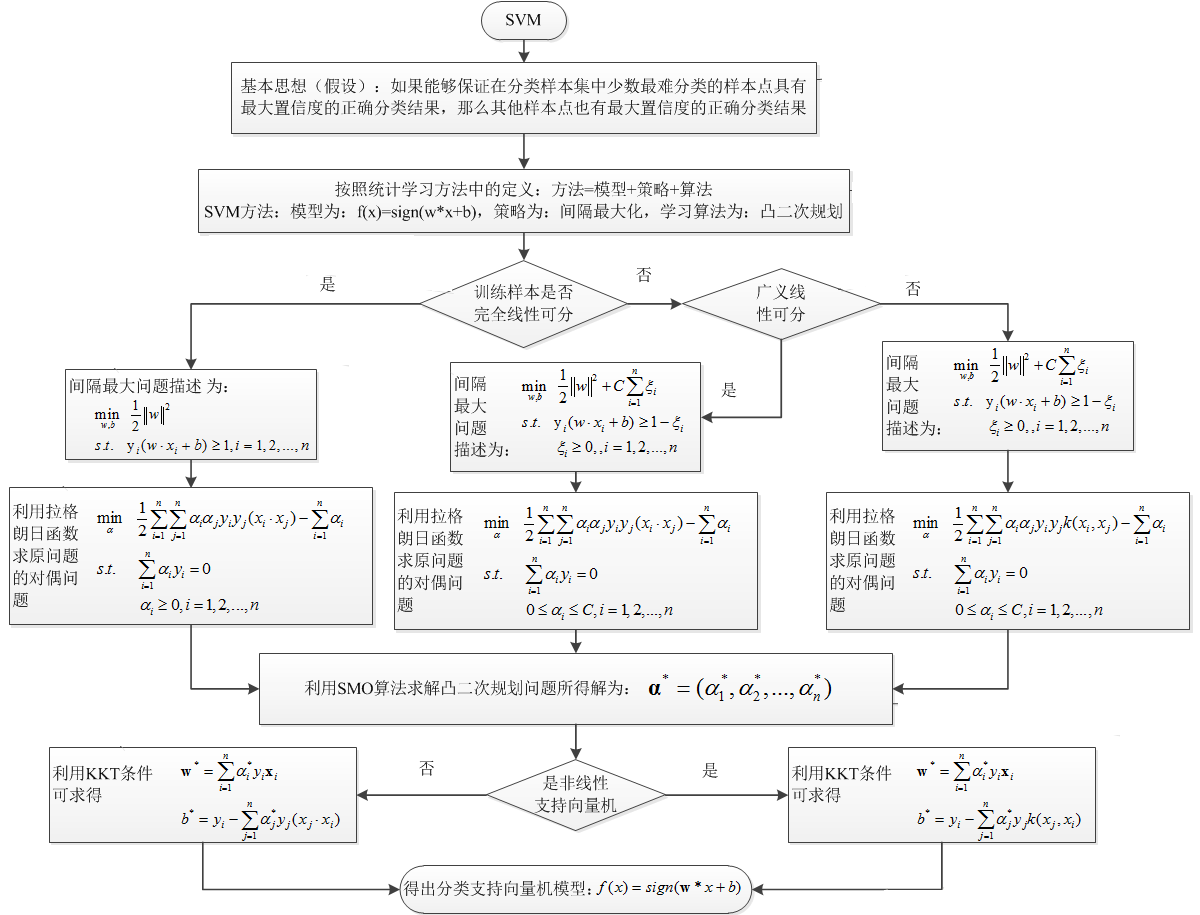


图4.9 SVM模型训练流程图

综合以上两部分，不良图片检测分类器就做好了，总体流程如图4.10所示

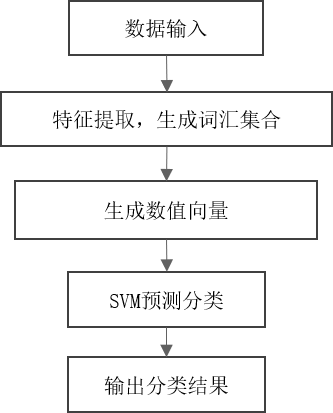


图4.10 BOW-SVM训练流程图

### Adaboost分类器

Adaboost（adaptive boosting）即自适应提升算法，是一种集成式分类器，包括很多弱分类器。Boosting 是一类算法的总称，这类算法的特点是通过训练若干弱分类器，然后将弱分类器组合成强分类器进行分类，总体结构如图4.11所示。

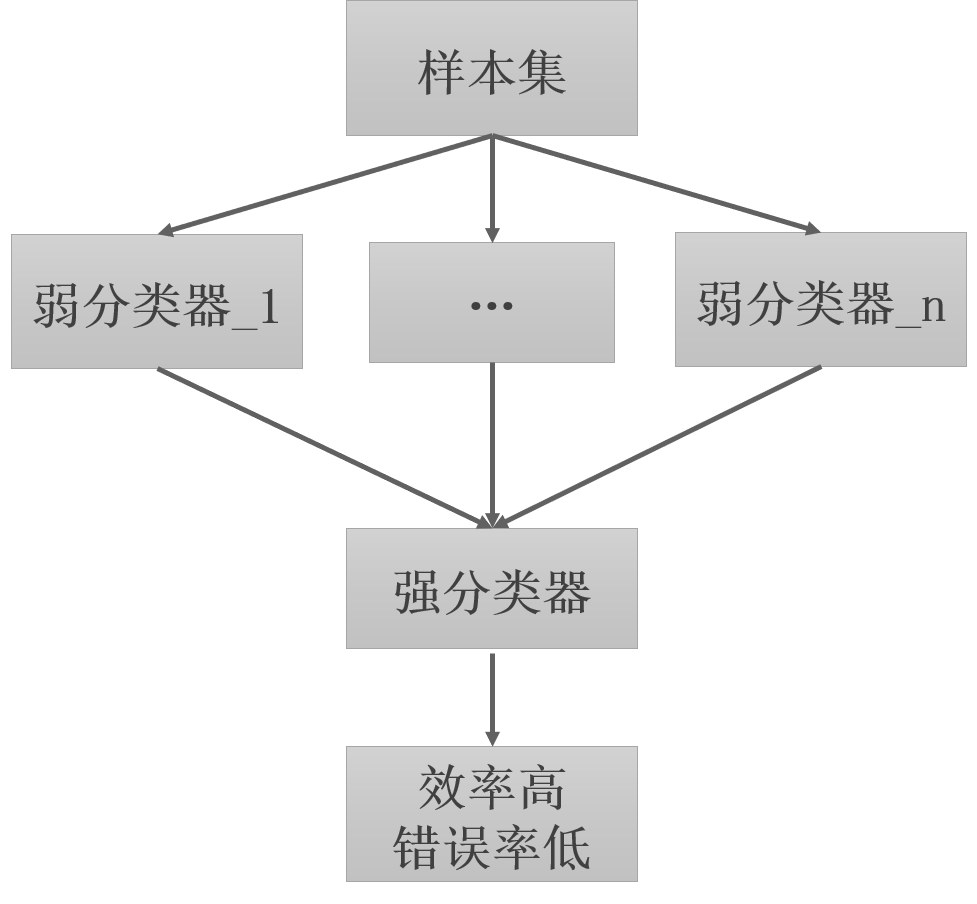


图4.11 总体结构

之所以加弱分类器，因为弱分类器训练起来很容易，将弱分类器集成起来，往往可以得到很好的效果。俗话说，"三个臭皮匠，顶个诸葛亮"，就是这个道理。这类 boosting 算法的特点是各个弱分类器之间是串行训练的，当前弱分类器的训练依赖于上一轮弱分类器的训练结果。各个弱分类器的权重是不同的，效果好的弱分类器的权重大，效果差的弱分类器的权重小。

AdaBoost 算法有其独特的优点，那就是可以将不同的分类算法组合起来，形成强分类器。这就可以充分利用不同分类算法的优势进行建模。也可以将同一算法的不同设置进行组合，这样训练的模型比单一设置模型的训练精度高。

#### 分类器训练

AdaBoost 的核心就是不断迭代训练弱分类器，并计算弱分类器的权重。需要注意的是，弱分类器的训练依赖于样本权重。每一轮迭代的样本权重都不相同，依赖于弱分类器的权重值和上一轮迭代的样本权重。具体过程如下：

(1).训练当前迭代最优弱分类器,最优弱分类器是错误率最小的那个弱分类器；

(2).计算最优弱分类器的权重，最优弱分类器的权重只与该弱分类器的错误率有关；

(3).根据错误率更新样本权重，样本权重的更新与当前样本权重和弱分类器的权重有关；

(4).迭代终止条件。

不断重复(1),(2),(3)步骤，直到达到终止条件为止。终止条件是强分类器的错误率低于最低错误率阈值或达到最大迭代次数。详细的流程图如图4.12所示（左边为弱分类器，右边为强分类器）：

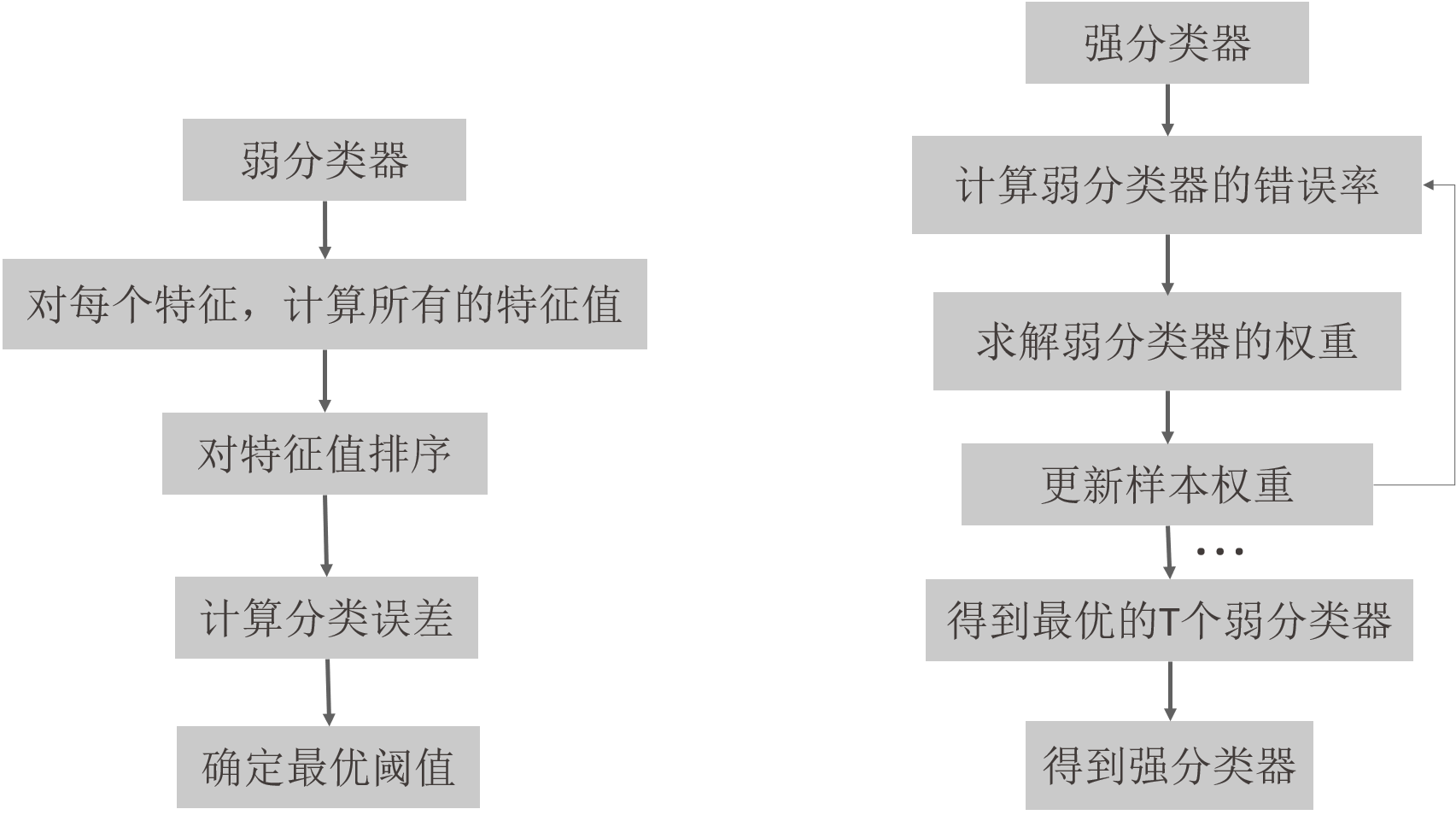


图4.12 adaboost算法详细流程图

#### 优缺点

优点：

     （1）Adaboost提供一种框架，在框架内可以使用各种方法构建子分类器。可以使用简单的弱分类器，不用对特征进行筛选。

     （2）Adaboost算法不需要弱分类器的先验知识，最后得到的强分类器的分类精度依赖于所有弱分类器。无论是应用于人造数据还是真实数据，Adaboost都能显著的提高学习精度。

     （3）Adaboost算法不需要预先知道弱分类器的错误率上限，且最后得到的强分类器的分类精度依赖于所有弱分类器的分类精度，可以深挖分类器的能力。Adaboost可以根据弱分类器的反馈，自适应地调整假定的错误率，执行的效率高。

     （4）Adaboost对同一个训练样本集训练不同的弱分类器，按照一定的方法把这些弱分类器集合起来，构造一个分类能力很强的强分类器，即“三个臭皮匠赛过一个诸葛亮”。

缺点：

     在Adaboost训练过程中，Adaboost会使得难于分类样本的权值呈指数增长，训练将会过于偏向这类困难的样本，导致Adaboost算法易受噪声干扰。此外，Adaboost依赖于弱分类器，而弱分类器的训练时间往往很长。

# 基于深度神经网络的不良图片检测

基于深度神经网络的不良图片检测的基本原理为图片分类，本章在经过大量调研的基础上主要介绍卷积神经网络原理以及结构，并着重介绍其工作流程。然后讲述AlexNet、VGGNet、ResNet、DenseNet和InceptionNet这五种经典的卷积网络结构，并重点介绍了InceptionNet在不良图片检测中的应用。



## 神经网络的结构



### 人工神经元模型

神经网络的设计是类比人类大脑进行设计的，神经生物学家将人工神经网络视为一种解释神经生物现象的研究方法。人类大脑通过对数以亿计的神经元进行合理组合，能够产生比计算机快好几个数量级的运算速度。

神经元作为构成人工神经网络结构中的最小单位，也是后续章节中设计人工神经网络的基础，图5.1给出了人工神经元的模型。

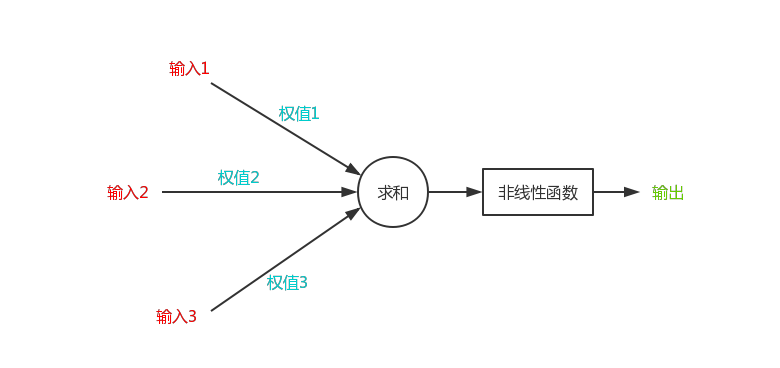


图5.1 人工神经元示意图

图5.1中为我们展示了一个标准神经元的三种基本组成单元：

1、连接：每一个连接都连接到求和节点，每一个连接都有一个权重。由图5.1所示，第ϳ个连接上的信号乘以连接权重作为该连接上的输入值。人工神经元的权值取值固定在一个范围内，作为待学习的参数。

2、加法器：加法器上的主要是将所有连接传递过来的输入值做累加和操作。结果作为激活函数的输入值。

3、激活函数：激活函数接收加法器的输出值，并将输出值做非线性变换并将输出结果压缩到小范围。通常单个神经元的输出范围为单位闭区间。

图5.1中的神经元还应该加上外部偏置量（bias），记为。偏置同样作为一个待学习的参数，适当提高或减小激活函数的输入值。

一般用下面一组方程来描述神经元

(5-1)

(5-2)

其中，，，…，是输入单元，，，…，是神经元个连接的权值，是加法器的输入，为偏置，激活函数为φ( )，是输出。

### 激活函数类型

激活函数（activation function）[33]，也称激励函数，主要作用是用来对输入值做变换，变换到一个固定的区间。激活函数用于函数之后，用来加入一些非线性的因素。下面我们给出三种常见的激活函数：

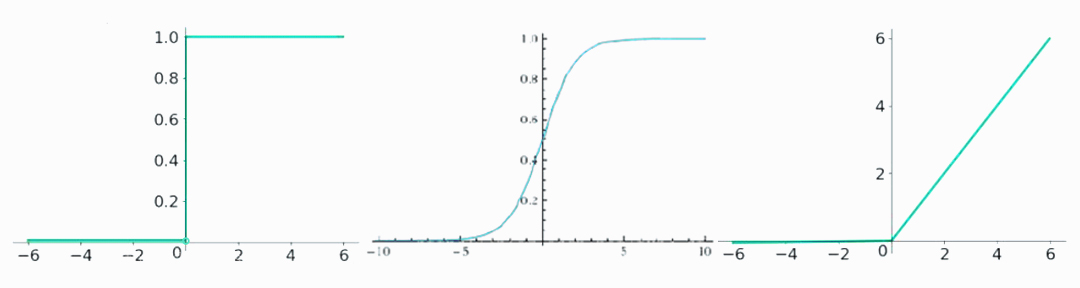
1、阈值函数。函数图像如图5.2（a）所示。将输入值映射成为两个固定的值。可惜的是阈值函数不连续，在实际中并不常用。

2、Sigmoid函数[34]。函数图像如图5.2（b）所示，该函数的图像平滑且连续，经常用于构建人工神经网络。函数的定义如下式5-3所示，其中a是Sigmoid函数的倾斜参数。

(5-3)

3、ReLU函数[35]。函数图像如图5.2（c）所示，Re LU函数（rectified linear units）是目前大部分卷积神经网络CNN（convolutional neural networks）中常用的激活函数，该函数的形式是y=max(x,0)，在该函数的原点左侧部分斜率是0，在原点右侧则是一条斜率为1的直线。从函数图像看出，这显然是非线性的函数，x小于0时输出为0，x大于0时输出就是输入值。

三种激活函数的函数图像如图5.2所示：



1. 阈值函数 (b) Sigmoid函数 (c) ReLU函数

图5.2 三种激活函数的图形

### 前馈网络结构

深度前馈网络（deep feedforward network）[]，也称为多层感知机（multilayer perceptron, MLP）[36]或者前馈神经网络（feedforward neural network），是目前典型的深度学习模型。这种模型之所以被称为前馈的，是因为网络中信息从原始输入开始，流动方向经过输入层神经元，中间经过隐藏层，到达输出层神经元。在模型本身和模型的输出之间没有反馈（feedback）连接。

构建网络的目的就是为了拟合出某个函数，使得其无限接近于实际输出。拿分类器举例子，函数接收输入x，输出预测的类别y ，于是网络就可以定义一个映射，通过拟合训练样本进而学习参数的值，得到最优的拟合度。前馈网络对于学习机器学习和深度学习是极其重要的，它在很多重要的领域中扮演重要角色。例如：对图像中的物体进行识别的卷积神经网络就是一种专用的神经网络，前馈网络也是循环神经网络的基石，循环神经网络在语音和自然语言处理应用中占统治地位。

前馈网络这种模型的拓扑结构是一个有向无环图，并且这个有向无环图描述了函数是如何复合在一起的。例如：我们有三个函数、和连接在一个链上以形成。这种链式结构是神经网络中最常用的结构。在这种结构中，称为的第一层（first layer），称为第二层（second layer）。拓扑结构的深度称为模型的深度（depth）。网络拓扑结构的最后一层被称为输出层（output layer）。在训练的过程中，每一轮迭代都让去匹配的值。训练数据中每个样本都有一个标签。

这些网络被称为神经网络是因为它们早起的发明是受神经生物科学的启发。网络中包含多个隐藏层，这些隐藏层的神经元个数决定了模型的宽度。向量的每个元素都类似于一个神经元的作用。我们可以把层看成由许多并行操作的单元组成，每个单元在某种意义上类似一个神经元，结构如图5.1所示。

用于计算这些表示的函数的选择也受到神经科学的启示。事实上，现代的神经网络研究理论更多来自数学和实践的指引，并且人工神经网络并不能完美地给大脑建模。不要将神经网络看成是对人脑的模拟，看作是为了实现近似模型的算法。它从大脑中运行机制获取灵感，但并不能模拟大脑功能。

沿用单隐藏层前馈网络的分析，当隐藏层个数超过2层时，成为多层前馈神经网络，其网络的拓扑结构特点是：多隐层、全连接、有向且无环。多层网络结构如图5.3所示：

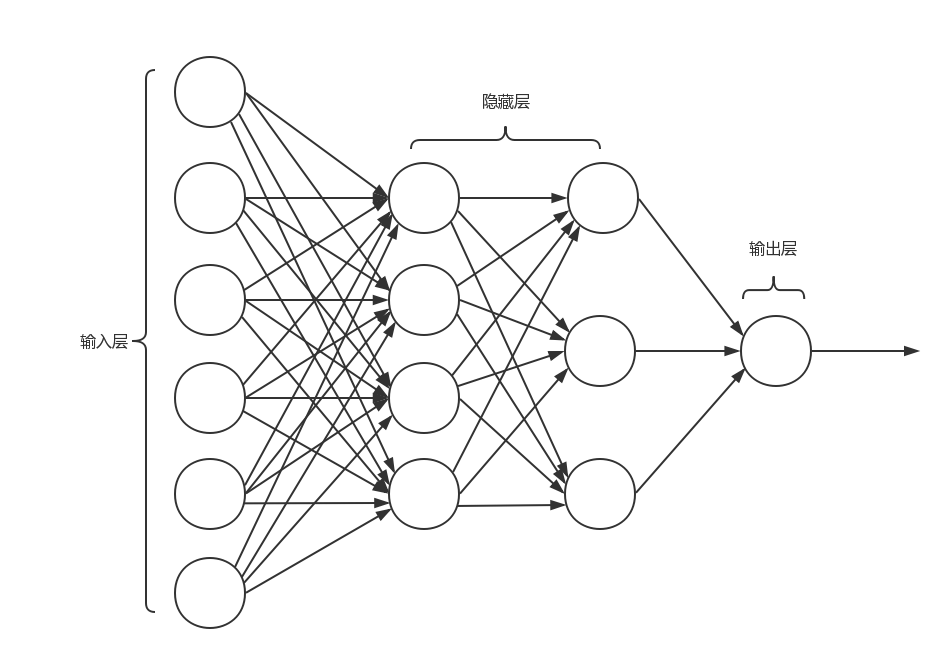


图5.3 四层神经网络结构

## 卷积神经网络的结构

上一节主要介绍了全连接网络。全连接网络的优点是每个输入维度上的信息都会传递到后面的每一个节点中，能最大程度上保证每个节点都不会遗失该维度上的信息。但是它的缺点也很明显，由于整个网络都是全连接的，所以参数ω和b特别多，这就使得网络的训练过程非常缓慢。对于图像这种输入维度动辄上百万的数据（以像素为维度），全连接网络的性能大打折扣，因此出现了卷积神经网络。

卷积神经网络（convolutional neural network, CNN）[37]结构专门用来处理网格结构的数据。例如图像数据，因为图像可以看作是二维的像素网格。“卷积神经网络”的由来是因为该网络使用了卷积（convolution）数学运算，卷积是一种特殊的线性运算。卷积网络是指那些在网络中使用卷积运算来替代一般的矩阵乘法运算的神经网络。



### 卷积运算与卷积核

在通常形式中，卷积是对两个实变函数的一种数学运算。卷积的数学定义为：

(5-4)

在卷积网络的术语中，卷积的第一个参数（在这个例子中，函数x）通常叫做输入（input），第二个参数（函数ω）叫做核函数（kernel function）。输出有时被称作特征映射（feature map）。图5.4演示了一个在二维张量上的卷积运算。

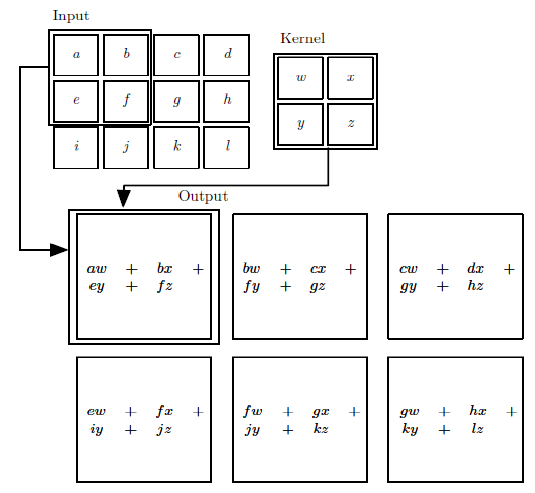


图5.4 一个二维卷积的例子

### 池化

卷积网络中一个典型复合运算含有三级。在第一级中，这一层并行地计算多个卷积产生一组线性激活响应。在第二级中，每一个线性激活响应将会通过一个非线性的激活函数，例如整流线性激活函数。这一级有时也被称为探测级（detector stage）。在第三级中，我们使用池化函数（pooling function）[38]来再一次调整该层的输出。

池化函数将像素点相邻输出的总体特征来代替在该位置的输出。举个例子，最大池化（max pooling）函数将输出相邻区域内的最大像素点的值，还有一些常用的池化函数包括：平均池化、L2范数和基于中心像素距离的加权平均函数。当在网络中采用各种各样的池化函数后，即使原始输入图像做出微小平移，池化能够帮助输入近似不变。这里所说的平移不变性是指当原始输入图像进行微小平移时，再经过池化函数后大多数输出并不会发生改变。图5.5用了一个例子来说明这是如何实现的。利用池化技术实现的平移不变性是一个相当有用的性质，图像检测只关心某个特征是否出现，但并而不关心它出现的具体位置。例如，检测图像中是否包含汽车时，并不一定知道车轮的精确位置，只要前后相对有两个平行的车轮即可。

使用池化使得这一层学得的特征必须具有对微小平移的不变性。实验证明，池化技术能够最大限度地提高网络的特征提取。算法对空间区域进行池化操作产生了平移不变性。在对分离参数做卷积后得到的输出进行池化时，网络能够学习到应该对于哪种变换具有不变性（如图5.5所示）。

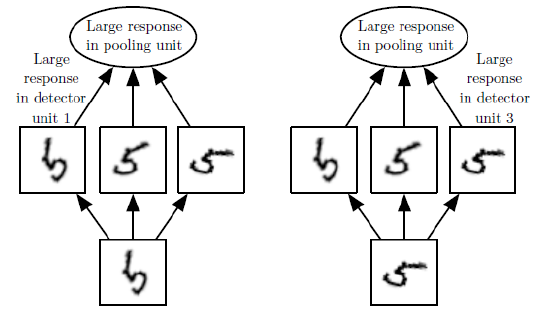


图5.5 学习不变性的示例

实际应用中，池化操作在处理不同像素的输入时具有至关重要的作用。通常在对不同像素的图像做分类时，Softmax分类层的输入一般是固定的大小，为了保证输入大小固定，最主要的做法就是调整池化层的偏置大小，最终分类层能接收到相同数量的特征而不必理会原始输入的大小。举个例子，网络结构中最后一个池化层假设输出四个统计特征，每个特征对应着图像的一个象限，这就与图像的原始输入大小没有关联。

### 卷积神经网络结构

CNN经典的结构主要包含三个重要组成：局部感受野、共享权重和池化。

1、局部感受野（local receptive fields）[39]：上一小节讲到的全连接网络中，网络结构被构造成纵向排列的一层层的神经元。卷积网络与全连接网络不同的是，将输入数据看作是一个个矩形排列的神经元，如图5.6所示。图像每一个方格中的值对应于输入的像素光强度：

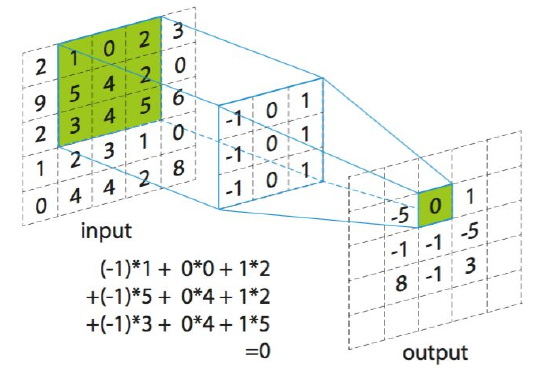


图5.6 局部感受野示例

上图中输入图像的阴影区域称为隐藏层的局部感受野，具体点说就是输入像素上的一个小窗口。小窗口上有很多连接，每个连接都要学习一个权重，而且每个隐藏层还要学习一个总的偏置。可以把这个特定的隐藏神经元看作是在学习分析它的局部感受野。

2、共享权重和偏置（shared weights）：每一个隐藏层上的全部神经元应该在输入图像的不同位置提取相同的特征。接下来解释为什么要这样做。把权重和偏置设想成隐藏神经元可以挑选的东西，特别是在一个特定的局部感受野的边缘。

这种能力在图像的其它位置同样是有用的。因此，在图像中应用相同的特征检测器是非常有用的。卷积网络能保持图像的平移不变性，稍微移动图像中一辆车的位置，它仍然是一辆车的图像。一般情况下，一组共享的权重和偏置被称为一个卷积核（也叫滤波器）。卷积核最大的优点就是极大地减少了参与卷积运算的参数量。

如下图5.7所示，为了找到鸟嘴，一个激活函数A需要检测图像左侧有没有鸟嘴，另外一个激活函数B需要检测另外一张图像中间有没有类似的鸟嘴。其实，鸟嘴都可能具有同样的特征，只需要一个激活函数C就可以了，这个时候，就可以共享同样的权值参数（也就是卷积核）。

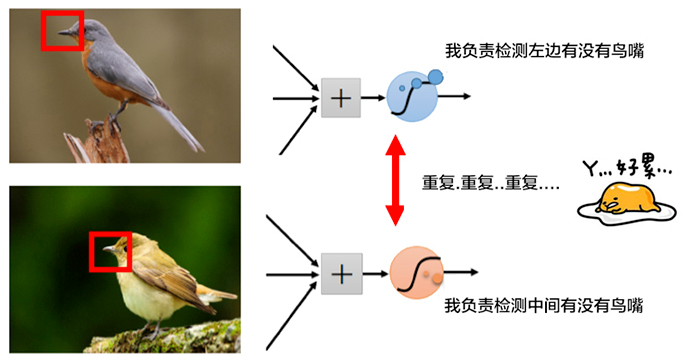


图5.7 非权值共享检测鸟嘴

如果使用了权值共享（共同使用一个卷积核），那么将可以大大减少卷积核的数量，加快运算速度。如图5.8所示。

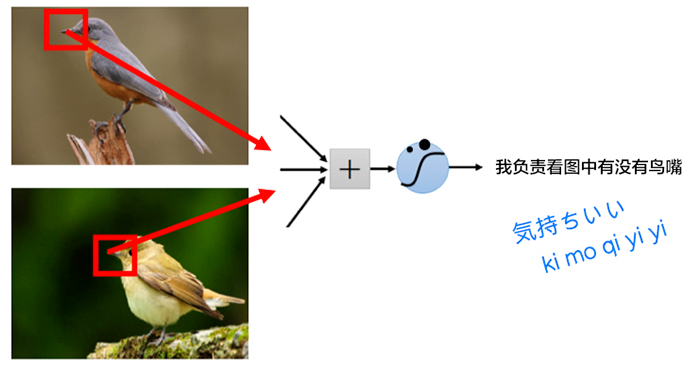


图5.8 权值共享检测鸟嘴

3、池化（pooling）：也就是上一节讲的池化层。池化层一般出现在卷积层之后，它的主要作用是简化输出信息。

综合在一起，就构成了卷积神经网络的特定结构。接下来举一个卷积网络的实例，图5.9是一个典型的卷积神经网络LeNet-5[40]结构示意图：

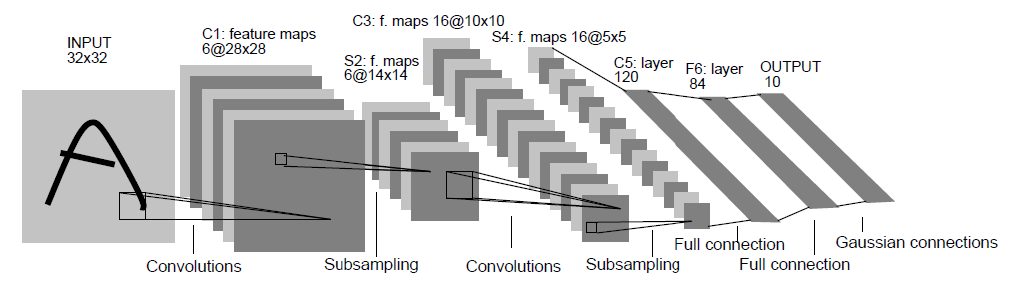


图5.9 LeNet-5的整体模型结构

LeNet-5网络结构配置如表5.1所示：

表5.1 LeNet-5网络模型的具体参数配置

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Layer | Type | Size | Params | Output |
| 1 | Conv | 6-5×5 | 156 | 28×28 |
| 2 | Pool | — | — | 14×14 |
| 3 | Conv | 16-5×5 | 1516 | 10×10 |
| 4 | Pool | — | — | 5×5 |
| 5 | Conv | 120-5×5 | 48120 | 1×1 |
| 6 | Full | 84 | 10164 | 1×1 |
| Output | — | | 10 | 10×1 |

这个经典网络被成功应用于手写字体识别数据集MNIST[41]上，达到了惊人的效果，错误率仅为0.95%。网络的的原始输入图像为32×32像素，Conv层后接一个Pooling层来从而降低提取的特征图（Feature Map）的大小，最后接一个Softmax全连接层输出结果。

## 经典卷积神经网络的结构

2012年到2015年是深度学习快速发展的四年，先后出现了以AlexNet[42]、VGGNet[43]、Google Inception Net[44]、ResNet[45]和DenseNet[46]为代表的经典卷积神经网络。它们分别在相应年份获得了ILSVRC[47]大赛中分类项目的冠军。以上五种经典的网络结构以其巧妙的结构设计和超高的准确率，对卷积网络的发展产生巨大的推动作用。



### AlexNet

2012年，一种全新的深度神经网络模型Alex Net横空出世。Alex Net是在LeNet结构基础上增加了更深层的模型。Alex Net在当年的大赛中表现相当亮眼，它将top-5的错误率突破性地下降至16.4%，远远超越了第二名的26.2%，以绝对的优势赢下了比赛。Alex Net第一次在实际应用中成功实验了Dropout 等新技术。Alex Net获胜的意义非常重大，象征着神经网络从低迷到高速增长的转折点。随后逐渐加固了卷积神经网络在计算机视觉领域的统治地位。

Alex Net进一步发扬LeNet关于深度卷积的思想，尝试了把CNN的重要思想应用到更深的结构层次。Alex Net破天荒地使用了以下新技术：

1、首次使用ReLU函数作为神经元的激活函数。ReLU函数有比Sigmoid函数更好的性质，成功解决了训练深度网络时的梯度消失问题。

2、首次使用Dropout[48]机制。训练网络时随机舍弃了一部分神经元。Dropout最早只是在论文中进行了论证，Alex Net首次将其应用到深度网络中。实验充分地验证了它的效果。

3、首次使用Max Pooling。在此之前的卷积网络模型中大多使用平均池化，为了避免平均池化常见的模糊效果，Alex Net首次使用了最大池化。并且实际实现网络的时候，Alex Net中使用的滑动窗口步长比核的还小，以增加池化层输出的重复，使得提取的特征更丰富。

4、增加LRN层。抑制响应较小的神经元，提高网络的泛化能力。

5、数据增强。随机从原始图像中截取224×224像素区域，在进行水平翻转，相当于扩充了上千倍的数据量。还对原始图像做一个高斯扰动，给原始数据增加一些噪声。这些操作都能降低错误率。

Alex Net网络的结构如图5.10所示，结构的前面5层为卷积层，后面3层为全连接层。可以看出中间有8层需要训练参数，每一层都用ReLU作为激活函数。网络的输出层是由Softmax层作为分类器，共有1000类输出。前两个卷积层后跟一个LRN层，Max pooling池化层出现在两个LRN层和最后一个卷积层后。

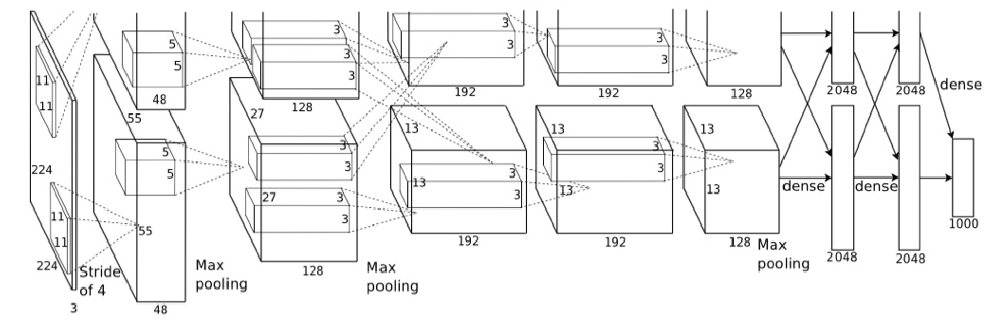


图5.10 Alex Net的网络结构

### VGGNet

VGGNet是牛津大学和Google公司共同发明的卷积神经网络。VGGNet经过循环累加卷积和池化块（一个块包含3×3的卷积核和2×2的池化核），成功地构建起19层深的卷积神经网络。VGGNet迈出了探索卷积网络的深度和性能之间的关系的坚实一步。VGGNet相比Alex Net错误率又下降很多，并取得2014年比赛的第二名。VGGNet的成功证明了卷积网络的结构可以非常简洁。网络一致使用了相同尺寸的卷积核和最大池化核。VGGNet能被简单地移植到其他图像数据上，具有很强的泛化能力。VGGNet至今依然是最常见的网络，很多研究都用VGGNet来提取图像特征。表5.2是VGGNet各级别的网络结构图。

在设计卷积段的时候，用到了一个很巧妙的设计，就是每个段里连续出现多个尺寸一样的3×3的卷积层堆叠。就具体设计而言，每个卷积段内的卷积核数量相等，并随着段的增加卷积核数量也翻倍，这有助于提取深度特征。如图5.11所示，由两个3×3的卷积层串联起来使用效果相当于一个5×5的卷积层，进一步地，将3个3×3的卷积层串联起来效果相当于一个7×7的卷积层。这当然是一个有用的设计，尽管最终效果是一样的，但是串联的3×3卷积层有比一个7×7的卷积层更少的参数量，计算一下发现只有后者的3×3×3/7×7=55%。这种设计最重要的一点是，串联的3个卷积层比只有1个卷积层的模型做了更多的非线性变换，串联模块使用了多次ReLU激活函数，也就使得卷积网络对特征的学习能力更强。

表5.2 VGG-16网络结构图

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Conv 3×3 | Conv 3×3 | Conv 3×3 | Pooling | Output size |
| Input | 3×224×224 | | | | |
| Block1 | slide: 1  padding: 1  kernel size: 3  kernel: 64 | slide: 1  padding: 1  kernel size: 3  kernel: 64 | — | type: Max  slide: 2  kernel: 2 | 112×112×64 |
| Block2 | slide: 1  padding: 1  kernel size: 3  kernel: 128 | slide: 1  padding: 1  kernel size: 3  kernel: 128 | — | type: Max  slide: 2  kernel: 2 | 56×56×128 |
| Block3 | slide: 1  padding: 1  kernel size: 3  kernel: 256 | slide: 1  padding: 1  kernel size: 3  kernel: 256 | slide: 1  padding: 1  kernel size: 3  kernel: 256 | type: Max  slide: 2  kernel: 2 | 28×28×256 |
| Block4 | slide: 1  padding: 1  kernel size: 3  kernel: 512 | slide: 1  padding: 1  kernel size: 3  kernel: 512 | slide: 1  padding: 1  kernel size: 3  kernel: 512 | type: Max  slide: 2  kernel: 2 | 14×14×512 |
| Block5 | slide: 1  padding: 1  kernel size: 3  kernel: 512 | slide: 1  padding: 1  kernel size: 3  kernel: 512 | slide: 1  padding: 1  kernel size: 3  kernel: 512 | type: Max  slide: 2  kernel: 2 | 7×7×512 |
| Full connect layer-1 | | 4096 output | | | 1×1×4096 |
| Full connect layer-2 | | 4096 output | | | 1×1×4096 |
| Full connect layer-3 | | 1000 output | | | 1×1×1000 |
| Softmax | | | | | |

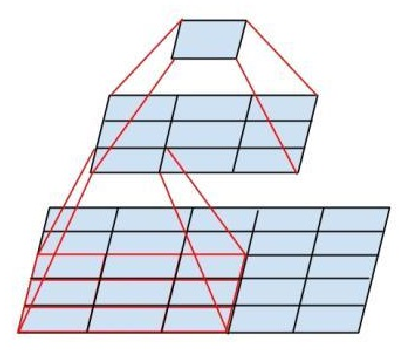


图5.11 两个串联的3×3卷积层

表5.3展示了网络每一级别的参数量。直观地从数据走势来看，从前往后每一级网络的层次逐渐变深，然而参数量却没有显著增加，主要原因是卷积使用了共享参数，主要的参数量消耗在了最后3个全连接层。也间接说明了卷积共享参数的重要性。

表5.3 VGGNet各级别网络参数量

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 网络类型 | A, A-LRN | B | C | D | E |
| 参数量 | 133 | 133 | 134 | 138 | 144 |

VGGNet网络在训练时先训练前面类型的简单网络，再重复利用网络中的权重来初始化后面几个复杂的模型，这样做的好处是让训练过程收敛的更快。网络训练时还使用了其他方法，比如VGGNet还使用数据增强将原始数据缩放到不同尺寸，这样就能成倍增加数据量，增加数据量是防止模型过拟合的重要方法。VGGNet分别训练了16层和19层的网络结构，这两种结构均能达到7.5%的错误率。

### ResNet

ResNet（残差神经网络）是由微软研究院提出的，以不可思议的错误率（仅3.57%）获得了2015年识别大赛的冠军。模型通过重复利用“残差块”成功地训练了152层深的神经网络。

ResNet以其独特的“跳跃式连接”的结构可以快速地构造出非常深的神经网络模型，并且这种跳跃式连接能加速模型的训练，进而提高模型的准确率。理论和实验结果表明，ResNet在达到最出众的准确率的同时参数量却比同类模型低，表现十分亮眼。ResNet同时还是个移植性很好的结构，甚至可以不用修改直接用在Inception Module中。

ResNet借鉴了“skip-connect”的方式，这种思想实现不相邻的层之间的跨越连接。以往认识的前馈网络只是对原始输入数据仅做一次非线性变换，而跳跃连接则会保留一定比例的原始输入。前一层的数据会有一定比例不进行特征向量的矩阵乘法和激活函数的非线性变换，直接传输到下一层。

ResNet同样也是允许原始输入信息直接传递到后面的层中。某一层神经网络的输入是x，模型最终期望输出为H(x)，如果直接将原始输入数据x传到输出作为结果的一部分，那么模型需要学习的目标就是变成了F(x)=H(x)-x。ResNet使用“跳跃连接”等于改变了学习目标，模型学习的不再是一个完整地输出H(x)，而是输出与输入之间的差别H(x)-x，这种差别也叫残差（Residual）。图5.12所示是一个ResNet的残差学习单元。

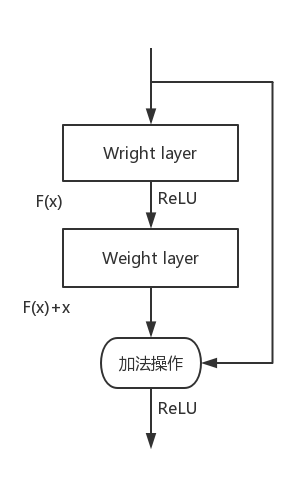


图5.12 ResNet的残差学习模块

全连接网络或者前面介绍的前馈卷积网络层与层之间传递数据时，大多会存在信息丢失等问题。ResNet网络通过“跳跃连接”直接将信息绕过中间层传递到不相邻的层中，这在一定程度上保护了信息的完整性。

训练过程中整个网络只需要学习输入和输出差别的那一部分，从而简化了训练的难度。不同于一般直连的前馈网络，ResNet模型中前面层会多出很多分支将输入直接传递给后面的层作为输入。后面的网络层不再学习完整的模型，只需要学习残差。

同Inception模块类似，ResNet模型首先定义Residual Block（残差单元），然后不断累加残差单元得到最终的模型。一般除了有两层的Residual Block，还有三层的Residual Block。图5.13中左边的两层Residual Block包含两个3×3卷积，它们有相同的输出通道数。而3层的Residual Block则借鉴了InceptionNet中的1×1卷积，并且在一个3×3卷积的前后使用，两种学习模块如图5.13所示：

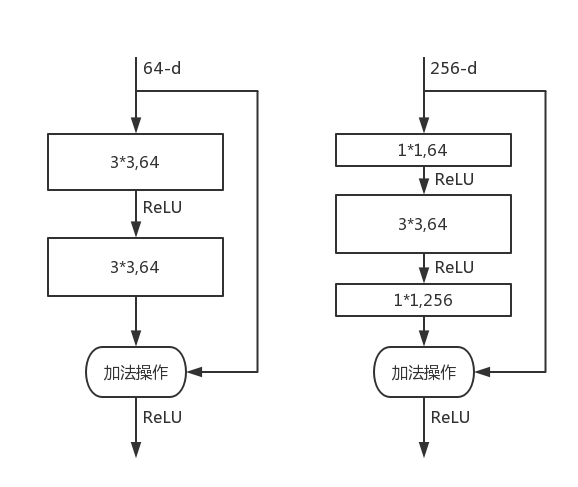


图5.13 两种ResNet残差学习模块

表5.4是ResNet在不同深度时的网络配置。从表中可以清晰地看出各种网络结构的基础结构很类似，主要是由不同尺寸的Residual Block累加而成的。Residual Block之间会加入一个卷积层和一个池化层。

Residual Block内部的一系列卷积层提取的特征图尺寸都是一样的。观察训练集上训练ResNet网络的过程发现，不断加深的层数并没有导致训练集上的误差增大，并且在测试集上的表现良好。

ResNet一经推出，就被其他网络模型借鉴，由Inception Net借鉴得到的模型在ImageNet数据集上取得了更低的错误率。

表5.4 ResNet不同层数时的网络配置

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Kernel param | Repeat | Output | Param(float) |
| Input | 1×128×128 | | | |
| F1 | 7×7, 64  Padding: 3  Stride: 2 |  | 64×64×64 | 3,200 |
| F2 | 3×3, 64  MaxPooling  Stride: 2 |  | 64×32×32 |  |
| L1 | 3×3, 64 | 3 | 64×32×32 | 221,568 |
| 3×3, 64 |
| L2 | 3×3, 128 | 4 | 128×16×16 | 1,106,944 |
| 3×3, 128 |
| L3 | 3×3, 256 | 6 | 256×8×8 | 6,786,048 |
| 3×3, 256 |
| L4 | 3×3, 512 | 3 | 512×4×4 | 12,979,200 |
| 3×3, 512 |
| T1 | C/32×C/32, Average Pooling | | | |
| T2 | S classes | | | |
| Total params: 21,093,760 | | | | |

### DenseNet

稠密网络DenseNet在2016年被提出，并且获得了2017年CVPR的最佳论文，主要是吸取了ResNet和NIN网络结构的思想，通过加深网络结构但又要防止梯度消失，于是有了稠密连接的方式，缩短了前层和后层的连接，能够最大化的去提升信息的流动，强化了特征的传播和特征的复用，这样就能够大幅度减少网络参数的数量，最终又能提升分类效果。DenseNet网络分成了多个稠密块，保证各层的特征图的尺寸统一，层与层之间进行了稠密的连接又保证信息的传输。

如图5.14所示，白色圈代表输入，黑色圈表示的是网络的特征层，稠密块中每一层的输入都是来自前面所有层的输出，构建了一种跨层连接的方式来连通网络中的前后层，由于网络存在着大量的密集连接，所以称为稠密块(dense block)。相比之前的深度卷积神经网络，假设神经网络有M层时，那么就会有M个连接，但是稠密连接的卷积神经网络将会有M×(M+1)/2个连接。这种方法建立了卷积层与卷积层直接的联系，从而充分利用了各层的特征。当进行反向传播的时候，网络深层的梯度信号都会向前进行传播，这种机制就不会因为网络变深，使得接近输入层的梯度消失的情况出现。

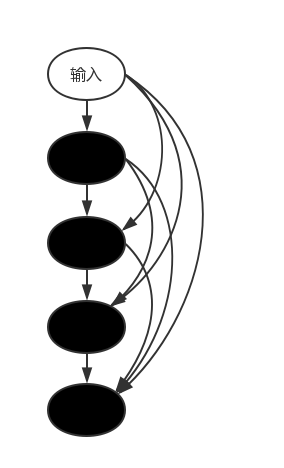


图5.14 稠密块结构图

针对物体识别的任务，DenseNet系列的网络模型包含了121层，169层，201层以及264层的网络结构，网络能够达到如此深的原因就是加强了前后层之间的联系，减轻了深层网络在训练过程中梯度消失难以优化的问题，避免了梯度消失。不同层的网络结构都有稠密块的机制，在稠密块机制中每一层输出的特征层数量相比其他网络少了很多，一定程度上减少了参数的数量，能够让特征传递的更有效。

### InceptionNet

Google公司开发的InceptionNet网络结构出现在2014年的比赛中，跟VGGNet同一年。在当时的比赛中，InceptionNet以微弱的优势战胜了VGGNet，并取得第一名。当年比赛时用的InceptionNet版本通常被称为Inception V1，该版本提出了创造性的观点就是既有效限制了计算量和参数量，同时还能获得优秀的分类性能。Inception V1版本具有22层深的网络，比当时其他网络都要深。模型增加了深度，但是又保证了计算量只有数十亿次运算，和只有百万级的参数量，一举成为当时最优秀的深度模型。

Inception V1模型的主要目的之一就是降低参数量。研究发现，首先模型随着参数的增长呈现爆发式的增长，那么需要提供给模型学习的数据量就越大，然而高质量的数据非常稀缺。其次随着参数的增加，耗费的计算资源也呈现出指数级的增长。Inception V1模型之所以准确率高且参数少，主要是因为：模型层数更深也就使得模型表达能力更强，模型还首创性地用平均池化层代替最后的全连接层。由前馈网络的结构可知，全连接层几乎占据了神经网络中绝大部分的参数量。而且全连接会引起模型过拟合，删除了全连接层后模型收敛得更快，并且减轻了过拟合。模型首创性地使用全局平均池化层，进而取代全连接层的做法借鉴了“网络中还有网络”的思想。Inception V1同样精心设计了Inception Module模块可以很容易构造出深层网络，也提高了参数的利用率。深入到网络模型内部，可以看到模块本身如同大网络中包含了许多个小网络，Inception Module模块结构经过反复堆叠在一起形成一个大网络。

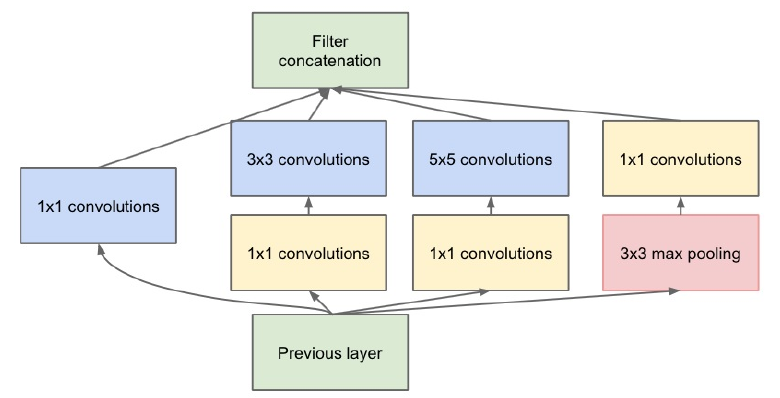


图5.15 Inception Module结构图

大网络中的每一个分支网络对输入进行1×1的卷积运算。事实证明，1×1卷积是一个相当实用的模型，1×1卷积最主要的优势就是它可以跨通道组织信息。由图5.15所示，模型的的所有分支都用到了1×1卷积，除了第一个 Inception Module，其他分支还使用了3×3和5×5的卷积。这样经过一系列复杂的组合就能提高网络的表达能力。模型的所有分支在最后一步通过一个聚合操作进行合并，合并后的最终结果传递给最后的分类层。

通常认为大型神经网络的有效连接方式应该是稀疏的，稀疏连接对减轻过拟合并降低计算量的影响至关重要。InceptionNet设计的主要目的之一就是找到最优的稀疏结构单元。想要得到一个好的稀疏结构，网络模型就要把相关性高的一些神经元节点连接在一起，不相关或者相关性低的节点尽量保持不连接。对于人工提取特征的数据集来说可能需要用机器学习算法对神经元节点进行聚类操作。但是我们知道在图像数据中，邻近像素点或者相邻区域的数据具有天然的高度相关性，这也是为什么相邻的像素点会被卷积操作连接在一起。当网络深度不断加大，模型中会出现多个卷积核，于是特征图上计算同一区域，但不同通道的卷积核的输出结果相关性极高。这时1×1的卷积就派上用场了，它能很轻松地将一组在同一区域但在不同通道上，相关性又极高的特征连接起来。回过头来看，Inception Module中频繁地使用1×1的卷积操作能够增加网络稀疏性的原因。

在整个网络中最理想的是后面的Inception模块可以提取更抽象的特征。一般网络会设计多个堆叠的Inception模块，而且越往后的Inception模块中的卷积的密集程度应该逐渐降低，从而实现学习更大尺寸的特征。为了使卷积密集程度逐渐下降，在设计Inception Module时，1×1卷积的数量最多，3×3和5×5卷积次之。因此越往后的Inception Module中，3×3 卷积和5×5卷积的比例越大，也就能学习更大尺寸的特征。Inception V2[49]学习了VGGNet的想法。使用了两个3×3的卷积代替5×5卷积的大卷积。伴随着V2版本的问世，正则化方法Batch ormalization（BN）的提出和实践表明，大型卷积神经网络分批次进行训练速度比一次性训练整个数据集要快上数倍，同时模型收敛得更快且分类准确率也能得到大幅提高。通过这样一种非常有效的正则化方法，最终V2网络取得的top-5错误率为4.8%，已经优于人眼水平。

Inception V3[50]在V2版本上引入了两方面的改造：一是引入“小卷积”的思想，具体做法是将较大的二维卷积拆分为两个较小的一维卷积，比如将6×6的卷积拆分成1×6卷积和6×1卷积。这样做有很多好处，一方面大大减少了参数量，减轻了过拟合；另一方面，Inception V3继续在 Inception模块中寻找优化方法，除了在整个模型中使用小网络，还在小网络中继续增加小网络。

Inception V4[53]相比V3主要是采用了ResNet的思想。

用于训练识别不良图像检测器的框架如图5.16所示。将在网络上获取的图像通过数据增强处理后分为训练集和测试集。首先使用训练集训练设计好的卷积神经网络，等待网络收敛且误差达到要求后，将测试集中的图片输入训练好的神经网络中判别其是否为不良图像，最后输出测试集的准确率。

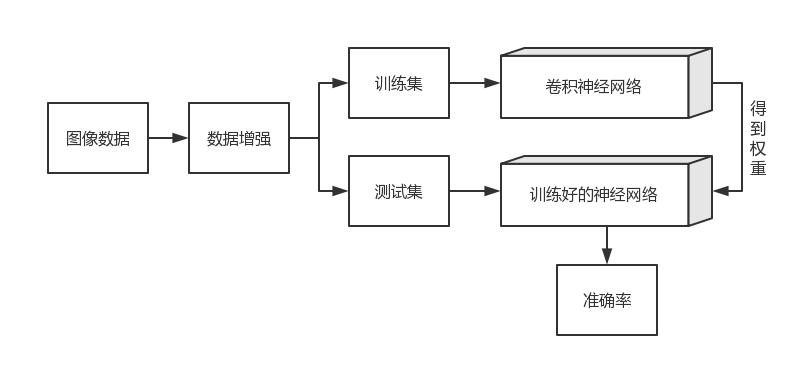


图5.16 不良图像检测框架

为了提高神经网络的性能会增加网络的大小，包括增加网络的深度(网络的层数)以及宽度(每层神经元的数量)。当拥有大量数据时这是最简单易实现的方法，但这种方法有两种缺陷。首先，网络过大意味着大量的参数，这使得网络更容易发生过拟合，降低准确率。特别是在获得训练集的成本很高时，简单扩大网络的大小是不经济的。另一个缺陷就是网络尺寸过大会侵占过多的计算资源。因为计算资源总是有限的，过大的尺寸会造成计算资源几何级数的浪费。

如图5.17所示为基于Inception模块的卷积神经网络模型。

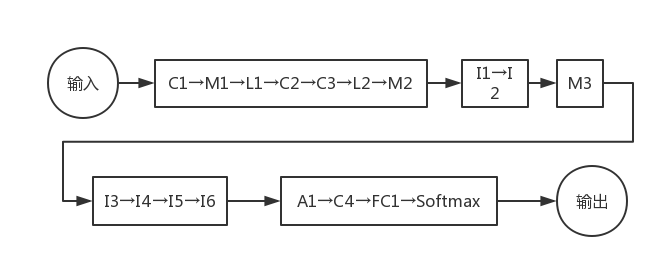


图5.17 卷积神经网络模型

图中，以C开头的代表卷积层，M开头代表最大池化层，L开头代表局部响应归一化层(Local Response Normalization)，I开头代表Inception模块，A 开头代表平均池化层。可以看出，如果只计算含有待学习参数的层数，网络模型共有17层，分别是16层卷积神经网络和1层全连接层。所设计的网络使用120×120像素的RGB图像作为输入，网络输入图像后，将图片的每个通道单独处理。经过一系列卷积池化等操作后，再经过Inception模块的处理以及全连接网络，全连接层之前的网络是将原始的图片数据映射在隐藏层的特征空间表达上，而全连接层则起到将学习到的特征表示映射到样本标记空间的作用。尽管全连接层参数冗余，但过多的参数并不是一无是处的。全连接层的使用可以让网络保持较强表达能力，捕捉到尽可能多的图像特征。最后使用Softmax分类器输出不同标签的预测结果。

本算法所设计的网络参数总数约为250万，而CaffeNet参数大约是本算法的30倍，这意味着在保证性能的前提下，可以在同等数量数据集的条件下更不易过拟合，准确率更高，并且所需要的计算量也大幅降低，占用计算机资源更少，有效降低测试时间。

## 本章小结

本章主要介绍了深度神经网络在不良图片检测识别中的应用，并以此展开了对神经网络的原理以及结构的介绍，介绍了神经网络的工作原理。然后介绍了卷积神经网络算法，详细讲述了AlexNet、VGGNet、ResNet、DenseNet和InceptionNet这五种经典的卷积网络结构，并重点介绍了InceptionNet在不良图片检测中的应用。

# 未来技术展望



本章内容结合目前已有的技术，给出了未来在不良信息发现领域可能会用上生成式对抗网络的大胆猜测，并对生成式对抗网络进行了介绍，分析其优缺点。

## 生成式对抗网络介绍

生成式对抗网络（GAN）是2014年由Goodfellow等人[Goodfellow I J , Pouget-Abadie J , Mirza M , et al. Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.]提出的一种生成式深度学习模型，该模型一经提出就成为了计算机视觉研究领域热点研究方向之一。它能够自动生成目标数据集，以弥补训练数据不足的缺陷，因此对深度学习意义重大。此外GAN在场景生成、图像翻译、文本与图像的相互生成、视频预测等领域都发挥了独特的作用。

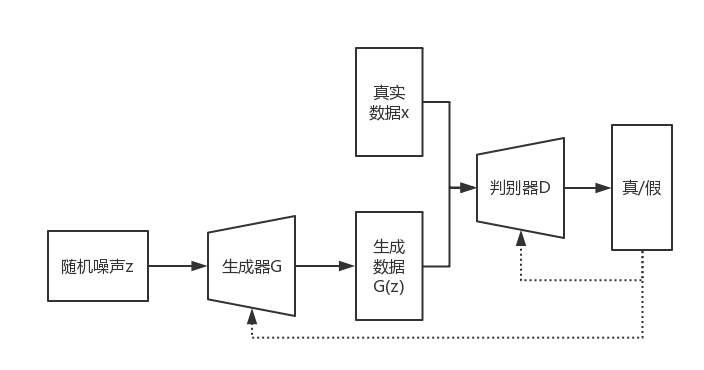


图6.1 生成式对抗网络模型结构

GAN包括两个模型，生成模型（G）和判别模型（D）。生成器G和判别器D本质上都是函数，它们通常用深层神经网络来实现。GAN模型结构如图6.1所示，G从真实样本中捕获数据分布映射到某个新的数据空间，输出生成的数据记作G(z)，其分布记作Pg(z)，并尽量使其看上去和训练集中样本Pr(x)一样。D的输入包括真实数据x与生成数据G(z)，输出是一个概率值或一个标量值，表示D认定输入是真实分布的概率，数值越大，是真实数据的概率越大，反之认为输入的是生成样本，D根据输出反馈给G，使G生成的数据与真实数据逐渐一致。理想状态下D无法分辨输入的是真实数据还是生成数据，即满足Pr(x)/(Pr(x) + Pg(z))最小，此时认为G已经学到真实数据的分布，模型达到最优。

GAN对应于极小极大的双玩家游戏，又叫二人零和博弈[51]，在包含两个神经网络的零和游戏框架中相互竞争。系统中两名游戏者由两个函数表示，这两个函数对于它们各自的输入和参数都是可微的。判别器的函数用D表示，它的输入为x，参数为θ(D)。生成器的函数用G表示，它的输入为z，参数为θ(G)。两个玩家都有各自定义的损失函数。D需要通过更新θ(D)极小化J(D)(θ(D)，θ(G))。G需要通过更新θ(G)极小化J(G)(θ(D)，θ(G))。两个玩家的损失函数都依赖于对方的参数，但是却不能更新对方的参数，经过训练达到一个纳什均衡[52]。

因此，在生成式对抗网络中最核心的就是生成网络和判别网络。



### 生成网络

生成器用可微函数G表示，输入z是一个随机变量或者隐空间的随机变量，一般使用高斯变量或噪声，G生成假样本分布G(z)。G网络只要求少量限制条件，对于输入变量，既可以把它输入到第一层，也可以输入到最后一层；还可以对隐层增加噪声，增加的方式可以是求和、乘积或做拼接。GAN对输入变量z的维度没有限制，它通常是一个100维的随机编码向量。但需要注意G必须是可微的，因为经过判别器的“判断”会将它的梯度传回G、D来更新参数，否则误差无法传递。

### 判别网络

在GAN中，判别器D的主要目标是判断输入是否为真实样本并提供反馈机制，其与生成网络构成一个零和游戏。这个游戏由两个场景构成，在第一个场景中，从真实训练数据中采样x，作为D的输入，D输出的是一个0到1之间的数，表示x属于真实样本的概率。通常还会假设真实样本和伪造样本的先验比例是1:1。在第1个场景下，D(x)被训练地尽量输出接近1的概率值。在第2个场景中，从一个先验分布中采样出变量z，将G(z)作为D的输入，在这个场景中，两名玩家都要参与，D的目标是使得输出D(G(z))接近0，而G的目标是使得它输出接近1。两个玩家的模型经过足够的训练，游戏最终会达到一个纳什均衡，此时G(z)与从真实样本中采样出的一样，而D(x)对所有输入x的函数值都是1/2，无法判断真假。

## 优劣分析

GAN自出现以来，针对不同领域的许多变体被提出，它们或在结构上有所改进，或在理论有所发展，或在应用上有所创新。在Goodfellow等人[62]提出的原始GAN中，先验假设很少，对于数据没有做任何假设，它可以是任何分布，最终目标使GAN具有无限的建模能力，可以拟合一切分布。另外，GAN模型设计简单，不必预先设计复杂函数模型，使用反向传播算法（BP）训练网络，生成器和判别器就能正常工作；GAN为创建无监督学习模型提供了强有力的算法框架，它颠覆了传统人工智能算法，不是用人的思维去限定机器，而是用机器来“对话”机器，通过自身的不断对抗博弈，经过足够的数据训练，能够学到现实世界内在规律。

然而GAN无限建模能力的目标背后隐藏一系列问题，由于生成过程过于自由，训练过程的稳定性和收敛性难以保证，容易发生模式崩塌，进而出现无法继续训练的情况；原始GAN存在如梯度消失，训练困难，生成器和判别器的损失无法指示训练进程，生成样本缺乏多样性，容易过拟合等问题；在由于GAN本身的局限性，它很难学习生成离散的分布，比如文本。

## 不良信息发现中的应用

本文认为，在不良信息发现中，生成式对抗网络（GAN）的作用主要体现在通过改善数据的分类边界，对少量类型数据的提升上。对于基于机器学习的物体识别来说，数据是越多越好，但对于有些特定类型的数据，不是数据量高就可以训练到好的效果。此时GAN就是改善模型性能的一个较好方法。

但是算法本身的迭代优化速度没有数据迭代快，而且各种图片识别算法差距并不是多大，所以目前物体识别技术的优化差异依然是基于数据本身，而数据的难点是数据多样化，而不是数据的纯数量。拥有更多数据类型，并且能针对不同数据应用场景提出最合适的数据模型，才是物体识别技术的关键所在。而生成式对抗网络 GAN 在图片处理方面的优势，将来也许能使其在不良信息发现方面的应用越来越广泛。

## 本章小结

生成式对抗网络作为一种生成模型，对于解决样本不足、生成质量差、提取特征难度大等问题提供了一种较好的解决方案。对基于深度学习的生成对抗网络在计算机视觉方面的应用进行了分析总结，总结了其优缺点，并对其未来在不良信息发现中的应用进行了可行性分析。

参考文献

1. 2017年互联网发展现状探讨分析
2. 中国互联网络发展状况统计报告
3. Sood S, Antin J, Churchill E. Profanity use in online communities[C]/lProceedings of the 2012 ACM annual conference on Human Factors in Computing Systems. ACM, 2012: 1481一1490.
4. Lee W, Lee S S, Chung S, et al. Harmful contents classification using the harmful word filtering and SVM[M]//Computational Science-ICCS 2007. Springer Berlin Heidelberg, 2007: 18-25.
5. 周学广，张焕国一种柔性中文字符串匹配算法[C][C]//第二十七届中国控制会议论文集.2008.
6. 周学广，张焕国.抗中文主动干扰的柔性中文串匹配算法闭.武汉大学学报:理学版，2009, 55(1): 101-104.
7. 帅正化，周学广.基于柔性匹配的中文文本特征提取方法明.计算机工程2010, 36(16):63-64.
8. Yoon T, Park S Y, Cho H G. A smart filtering system for newly coined profanities by using approximate string alignment[C]//Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on. IEEE, 2010:643-650.
9. Su G, Li J, Ma Y, et al. Improving the precision of the keyword-matching pornographic text filtering method using a hybrid model[J].Journal of Zhejiang University Science, 2004, 5(9): 1106-1113.
10. Lee P Y, Hui S C, Fong A C M. An intelligent categorization engine for bilingual web content filtering[J]. Multimedia, IEEE Transactions on, 2005, 7(6): 1183-1190
11. Ahmadi A, Fotouhi M, Khaleghi M. Intelligent classification of web pages using contextual and visual features[J].Applied Soft Computing, 2011, 11(2): 1638-1647.
12. Caulkins J P, Ding W, Duncan G, et al. A method for managing access to web pages: filtering by statistical classification (FSC) applied to text[J]. Decision Support Systems, 2006, 42(1): 144-161.
13. 唐坚刚，熊国萍.自适应不良网页过滤模式的研究与实践[[J].计算机工程与设计，2008, 29(20): 5324-5326.
14. Yin D, Xue Z, Hong L, et aI. Detection of harassment on web 2.0[J].Proceedings of the Content Analysis in the WEB, 2009, 2.
15. Rubenstein H, Lewis S S, Rubenstein M A. Evidence for phonemic recoding in visual word recognition[J].Journal of Verbal Learning and Verbal Behavior, 1971, 10(6): 645-657.
16. Gough P B. Word recognition[J].Handbook of reading research, 1984, 1: 225-253.
17. Meyer D E, Schvaneveldt R W, Ruddy M G. Functions of graphemic and phonemic codes in visual word-recognition[J].Memory&Cognition, 1974, 2(2): 309-321
18. Carroll D W. Psychology of language[M]. Wadsworth Publishing Company, 2007.
19. 林学民，王炜.集合和字符串的相似度查询叨.计算机学报，2011, 34(10): 1853-1852.
20. 孙海霞，成颖.信息集成中的字符串匹配技术研究田.现代图书情报技术2007, 7: 22-26.
21. Theobald M, Siddharth J, Paepcke A. Spotsigs: robust and efficient near duplicate detection in large web collections[C]//Proceedings of the 31 st annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2008: 563-570.
22. Wang J, Li G, Fe J. Fast join: An efficient method for fuzzy token matching based string similarity join[C]!/Data Engineering (ICDE), 2011 IEEE 27th International Conference on. IEEE, 2011:458-469.
23. Melnick S, Molina-Garcia H, Rahm E. Similarity flooding: a versatile graph matching algorithm[C]//proc. of the International Conference on Data Engineering-ICDE, San Jose. 2002: 117-128.
24. Levenshtein V I. Binary codes capable of correcting deletions, insertions and reversals[C]//Soviet physics doklady. 1966, 10: 707.
25. Zobel J, Dart P. Phonetic string matching: Lessons from information retrieval[C]//Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 1996: 166-172.
26. Wu S, Manber U. A fast algorithm for multi-pattern searching[R]. Technical Report TR-94-17, University of Arizona, 1994.
27. Sood S O, Anon J, Churchill E F. Using Crowdsourcing to Improve Profanity Detection[C]//AAAI Spring Symposium Series. 2012: 69-74.
28. 手机色情[EB/OL]. [http://baike.sogou.com/v6314199.htm. 2014-04-05](http://baike.sogou.com/v6314199.htm.%202014-04-05).
29. 王璐，赵永建.浅析手机上网不良信息监测实现方案[[J].邮电设计技术，2011(9): 68-70.
30. 李航.统计学习方法[M].北京:清华大学出版社，2012.
31. 思科推新工具:可有效阻止员工浏览色情网站[EB/OL].

<http://security.zdnet.com.cn/security_zone/2009/1013/1479271.shtm1.2014-04-05>.

1. 共青团与人大代表、政协委员面对面探讨净化手机网络环境[[EB/OL] .

http://zjnews.zj ol.com.cn/system/2014/01 /02/019790645. shtml. 2014-04-05.

1. Surhone L M, Tennoe M T, Henssonow S F. Activation Function[M]. 2010.
2. Saul L K , Jaakkola T , Jordan M I . Mean Field Theory for Sigmoid Belief Networks[J]. Journal of Artificial Intelligence Research, 1996, 4(1):61--76.
3. Glorot X, Bordes A, Bengio Y. Deep Sparse Rectifier Neural Networks[J]. Jmlr W & Cp, 2011, 15:315-323.
4. Hu J , Zhang J , Zhang C , et al. A new deep neural network based on a stack of single-hidden-layer feedforward neural networks with randomly fixed hidden neurons[J]. NEUROCOMPUTING, 2016, 171(C):63-72.
5. Kalchbrenner N, Grefenstette E, Blunsom P. A Convolutional Neural Network for Modelling Sentences[J]. Eprint Arxiv, 2014, 1.
6. He K , Zhang X , Ren S , et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9):1904-16.
7. Huang G B , Bai Z , Kasun L L C , et al. Local Receptive Fields Based Extreme Learning Machine[J]. IEEE Computational Intelligence Magazine, 2015, 10(2):18-29.
8. Lecun Y . LeNet-5, convolutional neural networks[J].
9. Lecun Y, Cortes C. The mnist database of handwritten digits[J]. 2010.
10. Krizhevsky A , Sutskever I , Hinton G E . ImageNet Classification with Deep Convolutional Neural Networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012.
11. imonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
12. Szegedy C , Liu W , Jia Y , et al. Going Deeper with Convolutions[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.
13. He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[J]. 2015.
14. Huang G , Liu Z , Laurens V D M , et al. Densely Connected Convolutional Networks[J]. 2016.
15. Kawano Y, Yanai K. ILSVRC on a Smartphone[J]. Ipsj Transactions on Computer Vision & Applications, 2014, 6:83-87.
16. Srivastava N , Hinton G , Krizhevsky A , et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.
17. Ioffe S , Szegedy C . Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// International Conference on International Conference on Machine Learning. JMLR.org, 2015.
18. Szegedy C , Vanhoucke V , Ioffe S , et al. [IEEE 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Las Vegas, NV, USA (2016.6.27-2016.6.30)] 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Rethinking the Inception Architecture for Computer Vision[J]. 2016:2818-2826.
19. He D , Chen W , Wang L , et al. A Game-theoretic Machine Learning Approach for Revenue Maximization in, Sponsored Search[C]// International Joint Conference on Artificial Intelligence. AAAI Press, 2013.
20. Ratliff L J , Burden S A , Sastry S S . Characterization and Computation of Local Nash Equilibria in Continuous Games[J]. IEEE Transactions on Automatic Control, 2016, 61(8):2301-2307.