关于论文的提问及回答

介绍一下文章：

首先，我们有两个出发点：1.想要引入一些非传统方法，用来防御日益增强的攻击。2.传统的加密方法已经落后，因为对图片、视频等多媒体数据的特征没有进行过多的考虑。

基于这样的目的我们进行了研究。引入元胞自动机的模型，对图像进行加密。最后在检测中发现当若干次加密轮次，图像会满足全局雪崩的条件，说明我们的算法敏感性满足要求。

元胞自动机的模型是冯诺依曼提出来的，模拟生物学中细胞的自我复制和繁殖的行为。现在很多领域都有利用大量元胞模拟复杂的动态的宏观现象。比如说，研究人口问题时一些群体性行为；生物学中癌细胞的生长，交通密度一定情况下车辆的运行规律等等…

创新点有三点

1. 使用了元胞自动机的模型
2. 在图像加密的过程中，只使用替换和置乱的操作，因此保证了算法具有很高的效率。
3. 该算法支持并行计算。由于每个格点的下一状态依赖于相邻格点的当前状态，所以难以实现并行加密。解决并行是的关键在于领域的构造，把多媒体数据分成互不相交的子集，分成很多个元胞机，实现全体格点同时加密。

我的任务是什么？

那么根据我的本身这种擅长编程的特点，老师给我布置的任务是编程去实现染色模型的元胞自动机，用每个粒子的一个单元分成8部分，每个部分包含n位。一方面这些数据可以和不同颜色深度的图像之间的关系，另一方面可以设置表示图像的精度。因此，加密模型可以灵活地对数字图像进行加密。

由于每一个格点会影响其相邻的格点，多次应用之后，局部微小的变化会扩散到全局。相当于每个人去将身边的邻居染色，这样一个一个扩散出去，最后变成几百个，最终实现整个图像的变化。一开始只做了十几个元胞进行实验，最后是做了一个300的元胞机。

难点在于：

1. 运行速度。主要是我一开始用Python写，优化代码后还是没有达不到老师的要求。 那段时间一直在熬夜，脑子都是在想怎么进行代码优化的问题。最后解决办法是用C++来编写，因为像C++、汇编这种底层语言运行效率要更高一些。

广告中违法关键词的识别项目概述

首先我们根据新修订的广告法提出违法的种子词汇，然后利用同义词林（使用哈工大的扩展版本）、语义相似度过滤扩充词汇，从而得到关键词库。基于这个词库，在字符串匹配基础上，考虑了上下文词性序列的影响，最终识别出违法关键字。我的工作是对于大量数据的收集，我用Python写了一个爬虫，……

在广告中违法词的识别做了什么

用Python做爬虫，利用request+BeautifulSoup的库和正则表达去抓取，再把数据导入到Excel中使用。对于开发自己的JS框架的一些动态网站而言，需要去模拟浏览器环境，不过效率比较低。所以我们在实验过程找的是没有开发自己的JS框架的一些网站，比如中国化妆品网、美妆网等。

这两个实验室是做什么的，有什么区别

这两个实验室分别隶属于经管学院和软件学院，负责院内和校外委托的一些项目的开发，信管工作室偏向软件这一块，而数字媒体实验室的规模更大一些，包括软件硬件都有在研究。我后面加入树莓的原因是因为，在软件范围内，我们学校有两个最厉害的实验室，树莓是其一，所以我想进去多学习，多交流。

说说你的毕业设计？

网络广告中违法文本自动识别研究，这次主要是想借助谷歌的词向量word2evc、以及模式识别的一些方法来进行研究识别。与之前不一样的是，搭建词库的方法更自动。但是由于复习复试，所以毕设的进展比较慢。

雪碧图？利用css的背景定位来显示需要显示的图片部分。主要是为了提高加载速度。但是后来的http2对请求数量没有太大考虑，所以已经不怎么使用了。

artTemplate.js？是一种很轻量的模板引擎，用来分离页面和业务数据。

Compass？ 是SASS的一个工具库。

Handlebars？是一个js的模块库，主要是用于分离进行页面渲染。在加载时就已经预编译了，而不是到了客户端执行代码时在进行编译，可以提高速度。

对不起老师，我对这个方面没有深入了解，希望以后可以跟老师讨教。

为什么不报计算机？是，现在的计算机非常的火，编程已经变成一种基础工具了，我想要利用编程数据挖掘去结合我的兴趣点，解决一些社会上的一些管理方面的问题。但是我的兴趣点是管理科学当中的运筹优化方向，比如博弈和决策，我打算未来结合自己的编程基础对优化决策领域进行研究，用数据帮助决策。其实人工智能、文本分析等都是先进行训练，最后求解优化问题。我觉得正好是对上了。

管理方面的：

读过罗宾斯的《管理学》、谷再秋《客户关系管理》、胡运权的《运筹学》、曼昆的《微观经济学》

马斯洛需求层次理论：：555生理需求、安全需求、社交需求、尊重需求和自我实现需求。

数学：

正态分布--中心极限定理：多个独立统计量的和的平均值，符合正太。一个东西受多个因素影响，最后多个因素加总后就是正太分布

数据挖掘：

相关分析（correlation analysis）

研究两个或两个以上随机变量之间相互依存关系的方向和密切程度的方法。

图表、协方差及矩阵、相关系数、回归分析

分类：

SVM支持向量机（三八线或平面进行分类，保持离两边的元素最远）、朴素贝叶斯（按照哪个类下这种类型的数量最多-给他）、神经网络

神经网络

有输出输入，多个神经元组成一个网络，已经各个权重和阈值。最后得到数值，期望预测与真实之间的差距（损失）减小。 通过已知的正确类号去训练模型，得到权重和阈值。

即一个分类的过程，想人体的神经元一样，越多层越准确（声音视觉触觉），每一个神经元化一层，最终得到它的类号。为了保证分类的准确，之前要先进行反向传播的训练，已知一个正确的，往前改。

最近我看的比较多的是文本分析。比如说

分词：中科院的分词系统、结巴中文分词等等，调用它的接口，有些是基于字符串（词典）匹配，快，但是要看词典好不好；基于统计/机器学习等（SVM、hmm、神经网络）。现在都是结合。我最近学习了一种利用自由度进行分词的方法，因为分词常常是以词语出现的频率判断是否为一个完整词语，但是很有可能出现诸如白日—但实际上应为白日梦。这里的原理是白日他不是自由，他是常常被梦绑住，一起出现，那么梦出现的频率就会很高。说明白日的自由度不够高，很有可能白日梦是一个词。

最近看了一篇比较有趣的文章在聊对红楼梦进行文本分析，比较用词习惯从而分析出后四十回不是曹雪芹携写的。

还看了一些决策序列和量子博弈的文章

1.决策序列：（供应链中序列决策优化方法的妍究）主要是考虑到未来的工作的随机不确定性和模糊不确定性。随机环境：用了一个双层随机值模型，先处理下层的规划问题，通过下层的最优解去求解上层最优，最后得到整体最优。

2.量子博弈均衡：风险者的推出困境。量子

3.基于元胞自动机的股票市场行为模拟研究——因为政府、羊群效应、异象效应之类的股票投资者不能对未来的股市波动做出比较准确的预测，所以作者用这三个作为影响因素去引入元胞自动机进行股票市场的模拟。分两种，一种是短期的，短期内主要受政府的影响，如果政策偏向利空，那异象影响会比较大，投资者的利润比较少；另一种是长期的，这时投资者的决策起主导作用。

数据挖掘：

1.回归分析（regression analysis)是确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法。多用于预测

2.分类：按照共同特点划分：K均值，svm支持向量机，神经网络

3.聚类：相似度和差异性划分

比如给你一堆数据，你怎么去预测未来的天气？

我比较喜欢用分类算法中的神经网络，比如我去用三个权重，一个是昨天的天气，第二个是去年今天的天气，以及这一周的一个 平均走势之类的，利用给出来的已有的天气数据，和反向传播的一个训练方法去得到模型，再用这个模型做出预测。

php的好处：入门门槛很低，免费，有很多优秀的开源项目。

我认为一门语言能否被广泛应用要看开源项目多不多，以及有没有被一些主流公司大佬运用。

Python-matplotlib 画图

请你简单说说你的毕业论文(毕业设计)。

我的毕设题目是网络广告中违法文本自动识别研究。由于近年来，网络广告业发展迅速。网络广告违规的问题随之日益突出。针对网络广告数量巨大，难以采用人工方式监管与审查的问题，课题的目标是设计并实现了一种智能识别网络广告中违规词的方法。可以有效辅助网络广告的监管。初步设想是利用模式识别以及Google词向量和句向量的相关方法。

请你简单说说你的毕业论文(毕业设计)。

我的毕设题目是网络广告中违法文本自动识别研究。由于近年来，网络广告业发展迅速。网络广告违规的问题随之日益突出。针对网络广告数量巨大，难以采用人工方式监管与审查的问题，课题的目标是设计并实现了一种智能识别网络广告中违规词的方法。可以有效辅助网络广告的监管。初步设想是利用模式识别以及Google词向量和句向量的相关方法。

1. 研究如何从网络页码中获取相应的广告文本内容。 2. 研究广告法的相关规定，为判断广告词违法提供支持。 3. 基于自然语言的研究匹配识别算法，能够发现广告词中违法的用词和信息。 研究过程中可使用如下方法，但不限于： 1. 文献法 2. 模式识别中的方法 3. 自然语音中的相关方法 4. Google词向量和句向量的相关方法。

通过 ｗｏｒｄ２ｖｅｃ计算语义相似度

ｗｏｒｄ２ｖｅｃ是 Ｇｏｏｇｌｅ在２０１３年开 源 的 一 款 将 词 表 示 为实数值向量的高效工具。通过训练可以把对文本内容的处 理简化为 Ｋ 维向量运算，而向量空间上的相似度可以用来表 示文本语义 上 的 相 似 度。ｗｏｒｄ２ｖｅｃ输出的词向量可以用来 做很多 ＮＬＰ相关的研究，比如词聚类、找同近义词、词性分析 等。

生成词向量之后，利用生成的词向量，ｗｏｒｄ２ｖｅｃ采用 ｋ－ｍｅａｎｓ聚类算法可以 实 现大数据集的文本分类

1. 自己今后研究的优势在哪里?缺点在哪?

我的优势第一个在于我对新知识很感兴趣，总是充满激情去学习新的东西。举个小例子，市面上的手机开发平台有三种，除了最常见的苹果和安卓，还有微软的windows phone。学习学校开设的安卓开发课程之后，我还自己跟同学一起学习了市场占有率只有1%的Winphone开发，（visual studio 2013）虽然只写了一个计算机，不过也挺开心的。第二个就是我的团队协作能力了。

缺点的话，我会给自己定过高的要求，出现压力过大的情况。不过目前我发现，酣畅淋漓的一场跑步可以很好的疏解我的压力，同时也在慢慢的调整自己的心态。

1. 研究生毕业后打算如何——求职抑或深造?

如果可能的话，我希望深造，继续从事研究。我相信 我的专业值得我努力，而且我也愿意为之花费更多的时间和精力。

1. 你为什么考研？为什么考我们学校?

因为我很想要看看更优秀的人是怎么学习生活的。其实本科之初，我的目标是当程序员，那时候常常听到code改变世界，但是随着我去比赛去交流，接触到更多更优秀的人，才知道自己的狭隘。代码是科研中的工具，科学才改变世界。那个时候我慢慢对科研产生兴趣，也期待能到更大的平台，与更多优秀的人交流。武大底蕴深厚，端博弘毅，武大人身上的气度和自信非常吸引我，所以我想成为这里的一员。

2. 为什么会报考这个专业，以及今后的打算。

因为我本科就是信息管理与商务智能专业，用科学手段去解决管理、决策上面的问题非常有趣，所以我希望可以进行利用大数据进行决策博弈等方向的研究。

关于未来的打算，我是这样思考的。在研究生期间，我会阅读大量的专业文献书籍来巩固基础知识，并且主动承担一些任务去提高自己的解决实际问题的能力，

1. 如何评价你的大学生活？

我的大学生活充实而愉快。在本科学习期间，我从各方面锻炼自己，培养扎实的基础知识和综合能力。课余时间，我还参加了许多活动，第一次做中超联赛的志愿者时，内心非常兴奋。科研上，我主动找老师参与课题，培养自己运用知识解决问题的能力。这些不仅给我留下了愉快的回忆，也从各个方面培养了自己。总的来说，我感谢我的母校和学校的老师们。