

# 关于新建计算机专业实习基地及开设前沿 技术研习班的请示报告

教务处领导：

为不断深化人才培养模式改革，加强实践育人，提高实践创新解决问题能力，拓宽大学生校外实践渠道，在前期与产业界多家单位进行交流沟通的基础上，遴选“北京蓝点数据科技有限公司”进行合作，建立校外实习基地，实习基地协议见附件 1。

北京蓝点数据科技有限公司位于北京大数据基地，是上市公司宏福集团控股子公司。北京大数据基地是北京市唯一的大数据产业聚集平台，定位为促进大数据产业发展，由北京市科委授牌宏福孵化集团运营（国家级）。大数据基地通过与中国科学院及北京大数据研究院等国内著名科研机构合作，探索产学研结合的新模式。

为进一步探索产学合作协同育人，培养一批在大数据/人工智能领域有所建树的学生，拟计划，在实习基地建设的基础上与“北京蓝点数据科技有限公司”进一步深度合作，面向我院计算机专业二、三年级在校生，为学有余力、希望

在大数据/人工智能领域有所发展的学生，免费开设为期三个学期的前沿技术研习班，使学生在校期间就接触行业先进技术，缩短学生就业通道，夯实学生实践能力。教学计划安排见附件 2。

特此申请，请批示。

信息与管理科学学院

2018-3-3

附件 2：教学安排表

|                                        |    |    |
|----------------------------------------|----|----|
| 第一学期：基础巩固与提高/人工智能理论与实践                 | 学时 | 说明 |
| 模块一：数学基础与程序设计提高                        | 20 |    |
| 1.1 概率论回顾                              | 4  |    |
| 1.1.1 概率论实训作业：随机数的生成和不同分布的变换           |    |    |
| 1.2 随机模拟                               | 2  |    |
| 1.2.1 随机模拟实训作业：蒙特卡洛算法计算圆周率、Ising Model |    |    |
| 1.3 信息论                                | 2  |    |
| 1.3.1 信息论实训作业：哈夫曼编码和信息熵的计算             |    |    |
| 1.4 统计学基础                              | 8  |    |
| 1.4.1 统计学实训 1：利用微博数据抽样验证中心极限定理         |    |    |
| 1.4.2 统计学实训 2：利用 R 语言计算统计量             |    |    |
| 1.5 Java 程序设计提高                        | 4  |    |
| 模块二：人工智能理论与实践                          | 40 |    |
| 2.1 概述                                 | 2  |    |
| 2.2 回归                                 | 2  |    |
| 2.2.1 回归实训作业：利用股票数据进行回归，理解过拟合与欠拟合      |    |    |
| 2.3 聚类                                 | 2  |    |
| 2.3.1 聚类实训作业：利用微博数据进行人群划分              |    |    |
| 2.4 逻辑回归与线性分类                          | 2  |    |
| 2.4.1 逻辑回归实训：用 Java 实现逻辑回归             |    |    |
| 2.5 SVM                                | 2  |    |
| 2.6 贝叶斯分类                              | 2  |    |
| 2.7 神经网络                               | 2  |    |
| 2.8 特征工程                               | 4  |    |
| 2.8.1 特征工程实训：利用特征提取技巧对设备工况进行分类         |    |    |
| 2.9 关联规则                               | 2  |    |
| 2.9.1 关联规则实训：利用关联规则处理                  |    |    |
| 2.10 异常检测                              | 2  |    |
| 2.10.1 异常检测实训：利用回归和动态阈值找到异常数据          |    |    |
| 2.11 模型的评价                             | 4  |    |
| 2.11.1 模型评价实训：逻辑回归的准确率、召回率与 PR 曲线的计算   |    |    |
| 2.12 深度学习导论                            | 2  |    |
| 2.13 自然语言处理简介                          | 6  |    |
| 2.14 图像识别简介                            | 4  |    |
| 2.15 总结                                | 2  |    |

|                         |     |    |
|-------------------------|-----|----|
| 第二学期：大数据并行计算平台          | 学时  | 说明 |
| 模块一：大数据与并行计算简介          | 6   |    |
| 1.1 大数据简介与行业应用          | 2   |    |
| 1.2 并行计算理论              | 4   |    |
| 模块二：MPI 平台简介            | 6   |    |
| 2.1 MPI 简介              | 2   |    |
| 2.2 数值计算与 MPI 实践        | 4   |    |
| 模块三：MapReduce 框架        | 18  |    |
| 3.1 大数据时代 MPI 存在的问题     | 0.5 |    |
| 3.2 MapReduce 模型        | 1   |    |
| 3.3 MPI 与 MapReduce 的对比 | 0.5 |    |
| 3.4 Hadoop 安装           | 2   |    |
| 3.5 HDFS 简介             | 2   |    |
| 3.6 MapReduce 程序设计      | 8   |    |
| 3.7 集群管理                | 2   |    |
| 3.8 实训                  | 2   |    |
| 模块四：Spark 框架            | 16  |    |
| 4.1 MapReduce 在迭代计算的缺陷  | 1   |    |
| 4.2 内存计算与 Spark         | 1   |    |
| 4.3 Spark 安装            | 2   |    |
| 4.4 RDD 与 RDD 编程        | 8   |    |
| 4.5 集群管理                | 2   |    |
| 4.6 实训                  | 2   |    |
| 模块五：大数据上的 SQL           | 4   |    |
| 5.1 HIVE                | 2   |    |
| 5.2 SparkSQL            | 2   |    |
| 模块六：NoSQL 简介            | 6   |    |
| 6.1 结构化和非结构化数据          | 2   |    |
| 6.2 四种常见的 NoSQL 数据库     | 4   |    |
| 模块七：深度学习矿建简介            | 4   |    |
| 7.1 TensorFlow 与 Caffe  | 4   |    |
| 总计                      | 60  |    |

|                            |    |    |
|----------------------------|----|----|
| 第三学期：大数据在实际领域中的应用          | 学时 | 说明 |
| 模块一：自然语言处理深入               | 8  |    |
| 1.1 分词与词性标注                | 2  |    |
| 1.2 词向量                    | 2  |    |
| 1.3 情感识别                   | 2  |    |
| 1.4 LDA 主题提取               | 2  |    |
| 模块二：图计算与社交网络分析             | 4  |    |
| 2.1 图计算                    | 2  |    |
| 2.2 社交网络计算与挖掘              | 2  |    |
| 模块三：数据可视化工具简介              | 2  |    |
| 模块四：实训项目，构建基于微博数据的社交媒体分析平台 | 46 |    |
| 总计                         | 60 |    |



前沿技术研习班面试选拔（1）



前沿技术研习班面试选拔（2）



前沿技术研习班面试选拔（3）



前沿技术研习班面试选拔（4）

| 大数据前沿研习班学习总结 |                               |
|--------------|-------------------------------|
| 学号           | 1510101049                    |
| 姓名           | 马志华                           |
| 专业           | 计算机科学与技术                      |
| 班级           | 15-2                          |
| 联系方式         | 13598051171 1493878432@qq.com |

## 1. 学习总结及收获

上个学期很荣幸能参加大数据前沿研习班的学习，通过学习，一定程度上提升了自己的能力，同时也发现了自己的很多不足，现做出总结：

陶老师上课很有热情，通过多种途径活跃课堂气氛，让大家全神贯注到课堂中去，陶老师的专业知识也非常的扎实，对各种问题都非常了解，重视引导我们的解题方法和解题思路，不刻板教学。比如有一次留的作业是分析微博数据，大家都在忙着写程序，想各种方法对数据进行处理，得到想要的结果，得到的结果也是五花八门，然而上课讲解这个题的时候老师把数据导入 SQL Server，用了几条命令结果就出来了。但有时陶老师的思路太快，以至于没有办法跟上，这是我的不足之一。

就这次研习班给我带来的收获来讲，我觉得视野的开阔要大过自己真正学到的东西的价值。坦白地讲，我以前的学习很封闭，老师上课讲的内容，我学会了，题知道怎么做了就放在一边了，应对期末考试足矣。但是到底这些知识离开了课本上和卷子上的习题，遇到实际问题怎么运用这些知识与解决，没有过多的想过，自然掌握的技能也很少，结果是这门课讲完了一个学期之后，就忘得差不多了。我现在发现这个问题，开始重视而且一定会影响我今后的学习，但是对大学本科阶段的我来说，作用已经非常有限。据我观察，不仅是我一个人存在这样的问题，但是在 16 级，17 级中已经有所改善，希望咱们院能够继续加强引导。

我计划考取本专业的硕士，继续发展，纠正本科阶段一直犯的错误，用更高的视野看待问题。

## 2. 对研习班的建议

通过本次研习班的开办，有以下个人的建议：

首先是在报名，据我所知有很多同学对研习班的具体内容并不了解，是冲着

目前大数据的热度和来自北京的老师去的，在后来开始上课，发现跟自己想象的不一样，以后就再不去上课了。建议以后再开办研习班的时候能够让同学们更充分的了解研习班的课程安排，课程要求。

其次是选拔，建议在选拔阶段改善选拔模式，由外聘任课老师和熟悉同学们的本学院老师共同选拔确定名单，尽可能避免出现因为外聘任课老师对同学们不熟悉，参考材料少造成的选拔失误。

再次是上课，在上课过程中出现有同学因为各种原因造成的缺勤，希望能和行政班级上课一样，有严格的课堂管理制度，保证教学的正常开展。

最后一点，也是我认为最重要的一点，我们的同学们习惯于老师在课堂上讲，自己在下面被动地接受，不习惯于研习班这种互问互答，交流讨论的学习方式和氛围，同时同学们普遍缺乏自信心，面对问题不相信自己能够做到，质疑自己的学习能力，希望咱们院在以后的教学中多加入一些高质量的可持续的实践环节，改善这种状况。

### 3. 个人对我系教学方面提出的建议

以下内容仅代表我个人的片面的看法

第一、课堂和考试需要改革。课堂需要靠有价值的内容留住同学们，而不是通过考勤，存在一些课程的教学方式枯燥，教学内容陈旧，无法吸引同学们去听课。有一个很现实的问题摆在我们面前：互联网发展迅速的今天，反正我上课跟老师也没有交流，我在宿舍通过网络也可以免费听到名校名师老师讲课，或者我看看课件就跟老师讲课差不多，我为什么还要去上课？课堂需要改革，在我们现有教学资源的基础上尽可能的让课堂变得有价值，有吸引力。考试需要改革，考试要真正能起到检验学习情况的作用，不能盲目变成检验老师教学能力的工具，曾有老师讲要把期末的挂科率控制在 25%以内，不然就是教学事故，再看成绩单上好多的 60 分,61 分，期末考试已经变了味道。所以又有一个现实的问题摆在我们面前，期末考试又不难，到最后的时候看看课件，做做往年的卷子我就能及格，运气好还能拿个高分，在身边充满各种诱惑的情况下，我凭什么抵制诱惑好好学习？

第二、有一句话，老师们讲，同学们讲，“计科软件的学生喜欢打游戏，缺勤率高，拿不到学位证的人多”。现实可能就是这样，但是这背后深层次的原因

却没怎么听人讲。来上大学就是为了玩儿，为了打游戏的人很少，但是是什么原因让他们变成这样，这是需要思考的。我个人认为原因主要有两点，第一点，专业和学习认知。我的专业是一个怎么样的专业？应该怎么学？我可以通过什么样的途径学习？通过什么样的途径提升自己？我的大学应该怎么度过才算有价值？这是摆在大一新生面前的问题，现在新生已经有了专业认知讲座，但我觉得，我们要以这个为起点，做的更多。第二点，机会。很多人包括老师会说机会是需要自己去把握的，不能等别人送上门，这句话很对，但是对刚入学的新生来说他们并不知道哪些机会需要把握，并不知道如何把握，需要有人去引导。比如蓝桥杯，如果有人反复强调重要性，有老师负责做一些培训和鼓励，增强同学们的自信心，参加的人数会不会多？取得的成绩会不会更优秀呢？如果我们有更多的参加比赛的机会，会不会更多调动同学们的积极性？除了已有的机会，我们能不能创造更多的机会呢。

以上所有都只代表个人片面之看法，如有不恰当的地方，请老师见谅。

| 大数据前沿研习班学习总结 |                                     |
|--------------|-------------------------------------|
| 学号           | 1510101004                          |
| 姓名           | 杨柳青                                 |
| 专业           | 计算机科学与技术                            |
| 班级           | 15-1                                |
| 联系方式         | 手机号：13598060090 邮箱：308779049@qq.com |

### 学习总结及收获

首先很荣幸能够在大三下学期拥有这个难得的机会，通过筛选加入学院组织的大数据前沿研习班进行学习，非常感谢各位老师和学院为我们的付出。虽然大江老师所讲授的课程只有短短几节，却令我们得到了一个充分的学习，收获良多。老师带给我们的收获远远超出了课程本身的讲授，更多的是在学习态度和方法上，甚至是每周四往返郑州及北京两地，尽职尽责为我们授课的态度。

大数据时代已经悄然到来，如何应对大数据时代带来的挑战与机遇，是我们当代大学生特别是计算机专业人才必须面对的一个严峻课题。大数据的学习对我们的数学知识、编程能力以及知识储备都有一定的要求，课程的开始我们大量的知识欠缺就暴露出来了，要能有所突破，首先要肯下功夫，付出努力。在大数据研习班我们学习了一些模型求最优解的方法和策略，了解了最小二乘法、贪心算法、熵值法在求解模型系数时的应用原理，通过学习对模型底层算法有了一定了解。通过老师对大数据相关概念的解析，这个近几年在网络上十分火热的名词在我们的脑海里有了更为生动的概念。

经过在大数据前沿研习班的学习，我了解到自己在这方面的道路还很长，不管是数学基础还是编程能力都还迫切地需要提升，我要时刻告诫自己，解放思想，更新观念，确立创新意识，善于动脑，勤于思考，开拓进取，争取站在时代的前沿，不断研究新情况，解决新问题，使自己的学习走上一个新台阶。我还要结合自己的实际情况，学习其他同学的态度，求知精神，协作能力，加强课外的学习、充电。对整个体系的全面掌控、深度学习等方面都可以成为自己未来发展的方向，在后续学习过程中，我会根据需要确定优先深入学习的方向，努力增强实力为今后的职业发展打下坚实基础。

在老师的帮助下，我们发现，自己和老师所要求达到的水平，无论是知识储

备，还是学习能力都存在着很大的差距。为了学弟学妹们今后能有更好的发展，还是希望学院能够继续组织更多优秀的研习班，这不仅能够学习前沿知识，更可以有效的激发学习兴趣。同时建议授课过程可以考虑网络直播的形式，这样对老师授课、同学学习都增添了更多的可能性，这样可以惠及到更多有兴趣、想学习的同学，外地的老师也可以免于奔波赶路，同时授课时长也能得到保证。最后，感谢老师和学院为我们的学习和成长提供了这么多优秀的平台和学习资源，希望我们信管的计算机专业在大家的共同努力下可以越来越好。