

## 写给数据挖掘/时空数据挖掘方向研究生新生的一封信

各位 20、21 级研究生新生：

你们好！首先欢迎你们加入数据智能实验室！进入实验室进行科研是一项辛苦但却非常有趣的事情，能够做出很有意思的工作也是非常有成就感的。希望你们在接下来的 3 年里能够享受这段痛并快乐的时光。下面我就你们关于数据挖掘与机器学习的学习路径参考。

首先，实验室的相关研究内容可以参考实验室主页 ([di.ustc.edu.cn](http://di.ustc.edu.cn)) 和我的主页，其中我的 <http://home.ustc.edu.cn/~zzy0929/Home/> slides for speech 中有三个 ppt，请依次阅读。前两个讨论研究内容，第三个是关于如何阅读文献和进行写作的。

你们可以通过李宏毅“机器学习”相关课程和周志华“机器学习”西瓜书和李航“统计学习方法”来了解和学习机器学习相关的思想和算法。关于代码，可以在 github 和机器学习实战中找到相应练习的代码。

在论文阅读方面，我们现在比较流行的数据挖掘与深度学习技术有 CNN（卷积神经网络）、GNN（图神经网络）、GAN（对抗生成学习）、LDL（标记分布学习）等。在应用层面，数据挖掘方向有图节点分类、图的边预测问题（Link prediction）、时空数据预测/补全问题、推荐系统与人类活动（包括人的出行预测、人的 POI 签到活动预测），计算机视觉方向有目标检测、动作识别、跟踪与追踪等。你们可以通过搜索论文、CSDN 博客去大致了解他们。进而可以去搜索相应的论文。我们推荐阅读的数据挖掘类的会议和期刊，大致如下：

**会议：**

**NIPS, AAAI, IJCAI（人工智能），ICML, SIGKDD, WWW, ICDE, SIGIR（信息检索与推荐系统）、**

**CIKM, ICDM, WSDM（B 类）**

**期刊：IEEE TKDE, TPAMI**

视觉方向：CVPR, ICCV, AAAI, NIPS, IJCAI.

一般而言，研一同学可以先通过阅读 AAAI, IJCAI 上面的文章了解近期人工智能领域的主流方向和技术，然后阅读 CIKM, SIGIR, ICDM, UBICOMP 等数据挖掘相关领域的专题文章，由浅入深。在具有一定基础后，可以阅读 KDD、WWW 等更为顶级的会议。人工智能方面，可以阅读 NIPS, ICML 等顶级会议。事实上，我们方向比较主要的阅读文献来源就是 **AAAI, IJCAI, CIKM, KDD, WWW, SIGIR, IEEE TKDE 等。**

附 CCF 推荐会议列表：

[https://www.ccf.org.cn/Academic\\_Evaluation/AI/](https://www.ccf.org.cn/Academic_Evaluation/AI/) （人工智能领域）

[https://www.ccf.org.cn/Academic\\_Evaluation/DM\\_CS/](https://www.ccf.org.cn/Academic_Evaluation/DM_CS/) （数据挖掘领域）

此外，还要学会刷会议，以上标红的会议，可以经常在 **dblp（主流论文检索）** 里 <https://dblp.uni-trier.de/> 进行检索，并对一整年该会议的论文进行整理，顺序是这样：**先看标题，选择感兴趣的阅读摘要，再阅读模型图和 introduction，最后阅读全文。**此外，还要勤做笔记，记录好的 idea 和表达好的句子、公式。

阅读论文的顺序如下：

标题=> 摘要=> 模型图 => introduction => 方法（Methods）=> 实验，具体

在任何一步如果你觉得不感兴趣或者和自己的研究相关度不高，都可以放弃，但是你需要保证每周 3 篇左右完整论文的阅读量。我们也会推荐一些有价值或者有意思的论文发到群里，你们可以选择性阅读。

好了，就写到这儿了，有想法随时补充！

哈哈哈，祝大家开心科研，顺利毕业！