

数据挖掘讲座报告

黄勇 201521210022

IIKI2015 年会从 2015 年 10 月 22 日到 2015 年 10 月 23 日正式在北京师范大学举办，这次会议紧跟当前大数据和物联网的热潮，在 knowledge engineering, big data and cloud computing 这些和大数据、知识工程等相关的项目上进行了交流和探讨。除此之外，会议也紧跟无线穿戴设备，无线网络安全，移动网络，deep web，物联网用户体验等方面进行了深入有效地探讨和交流，从而为我们提供了一个绝佳的学习和接触前沿科学知识，紧跟信息爆炸和移动互联网时代潮流的机会。

在这两天的会议中，我看到了计算机和电子通讯专业各个方面的精英人才，有来自美国各个顶尖大学的 IEEE fellow，教授等知名专家学者，也有国内其他大学相关领域内有相关经验的专家和科研人员。这些研究人员对自己相关领域的最新研究和自己的研究成果都做了相应的介绍的讲述，同时相关工业界的人才也会参与这个学术会议，从而保证学术界和工业界相结合，使得学术研究成果能够被工业界快速应用，从而加速产学研用相结合。首先，22 日上午很多学术和工业界大牛都在大会开幕式上进行了介绍和演讲，最让我印象深刻的就是富士通集团的 IAN THOMAS，他是富士通的 strategist，演讲的题目是 web of everything，介绍了富士通对于当前广泛应用的移动设备和物联网设备的思考，这涉及到这家大公司对于自己未来面对复杂而激烈的物联网市场而需要做出的反应和应对策略，我们通过学习一家世界 500 强企业面对激烈的市场竞争而做出自身对于未来的决策和反应，可以让我们能早日接触到公司思考的方式和方法，让我们对于自己未来找工作对自身的定位有了很多清晰的认识。同时，富士通公司介绍了在物联网环境下整个网络安全的思考和认识，这让我对自己以前狭隘的网络思考有了下一步的认识和理解，能让我更好的学习从前没有接触的 web 安全知识，让我对自己以后学习的新方向有了一个新的了解，这不仅有利于我学习和巩固自己以往的计算机相关知识，又能学习到企业相关的思路 and 想法，让我对自己的今后学习思路 and 想法有了更深一步的认识。

除了 Thomas 介绍的和企业相关的知识，另外一个让我感兴趣的知识点就是 deep web，维基百科上对于 deep web 的定义是使用搜索引擎难以抓取的 web 页面。Deep web 不仅在数量和内容上远超 surface web。在形式上也包括动态网页、PDF、word 等形式，这些 web 站点和页面在内容质量上也远超传统的 surface web，如何能科学利用和获取这些网页的内容，也让我深深的为之着迷。作者从 web 信息的爬取、采样、分析和建表这一维度和搜索关键字

的获取、搜索缺陷、搜索图形化等维度进行了深入的分析 and 介绍。并介绍了作者自己的基于隐马尔可夫模型 MCMC 等。除了作者介绍的这些工作，作者也对自己对于 deep web 中人们的隐私问题进行了探讨、并提出介绍了很多关于 deep web 中关于如何建立爬虫模型、获取 web 信息等多方面思考和认识。这些思考和认识让我深深思考起自己的工作 and 学习。我自己的研究方向和 web 搜索有一些交叉关系，对于如何构建爬虫，获取 web 数据我都有自己的一些实际项目基础，但是如果让我接触到 deep web 中数据的爬取和采样、构建聚类、分类模型，并为公众提供方便易用的 web 搜索接口，对我当前的实战工程能力来说还是一个很大的挑战。有幸得是，我也抓取过部分的微博、researchgate 等网站的数据。对这些数据进行过分析和整理。这样的学术研讨会让我能深入思考自己今后的工程实践方向。除了抓取部分方便可获取的 surface web 页面，也要深入考虑 deep web 中信息知识的不同，不断提高和提升自己的工程实践能力。

当前对于大数据和知识工程等方面的知识都有很多新的探索 and 发现，上面的两个重点报告让我明确了自己的在学习相关方向时要结合企业和科研的热点，努力提升自己的工作中的工程实践能力，只有这样，我自己的学习效果才能越来越好。除了这些给我影响深刻的报告外，其他嘉宾和演讲者也都谈论了自己的研究领域内的一些新成果。这些成果为我提供了很好的知识拓宽途径，让我明白互联网+和物联网都有哪些研究的重点和研究的热点，这些研究方向让我深深思考自己的研究领域能否和其他研究领域相结合，从而加速自己科研的创新创造力。在无线可穿戴设备、物联网、互联网+等新的知识快速普及 and 应用的今天，通过这样的快速学习了解其他专业知识，让我受益匪浅。

这次会议也让我有一个很大的遗憾。自己的英语水平一直不高，在听国外高水平学者的研究报告是，自己有些不能完全理解他们的英语意思。这对我在很多报告中的知识细节都有很多的疑问。不过这样的会议和知识报告通过实践给我深深的上了一课，要努力提高自己的英语水平，只有学好英语，不断提高自己英语方面的听说读写能力，才能在这样的国际化的会议中游刃有余，自由的学习更多更开阔的知识。英语这样的通用型语言对于我们这些科研人员努力了解国外先进的研究思路和研究方法，是在是太重要了，只有将自己的英语水平提高，才能更好的了解外面的科研状况，而不是闭门造车，固步自封。这场高水平的学术研讨会让我对自己的学习思路和学习方法有了很多新的认识，对于加强自己英语听说读写能力有了迫切的要求。同时，我也希望自己以后能在研究生期间，做出好的研究成果，让自己的研究成果也能发表在同样的高水平的会议上，为此我要努力学习除了英语外的数理统计和数据挖掘知识。这些知识在未来物联网的大环境下有着极为重要的基础型作用。