빅데이터 기술 개론 강의 노트

Slide 1: 강의 소개

안녕하세요!《빅데이터 기술 개론》강좌에 오신 것을 환영합니다. 저는 리양위안 선생님입니다. 여러분은 저를 "양위안 선생님"이라고 불러도 되고 "데이터 왕자님"이라고 불러도 됩니다 😉.
오늘 우리는 제1장 내용인 "빅데이터 기술 개론"을 시작합니다.
제 연락처와 홈페이지는 위에 있으니, 밤에 잠이 안 오시면 제 홈페이지를 방문해 보세요. 어쩌면 수면에 도움이 될지도 모릅니다 💤.

Slide 2: 강의 개요

이번 장에서 주로 다룰 내용:

- 빅데이터란 무엇인가?
- 빅데이터는 어떤 영향을 가져왔는가?
- 핵심 기술에는 어떤 것들이 있는가?
- 빅데이터의 컴퓨팅 방식과 대표적인 기술은 무엇인가?

간단히 말하면, 이번 장은 "빅데이터의 과거와 현재"입니다.

Slide 1:

大家好！欢迎来到我们的《大数据技术概述》课程。我是李阳源老师，你们可以叫我"阳源老师"也可以叫我"数据小王子"😊。

我们今天进入的是第一章内容——"大数据技术概述"。
我的联系方式和主页都在上面，如果你晚上睡不着觉，欢迎去我主页看看，也许能帮你催眠 💤。

Slide 2:

我们这章主要讲：

- 什么是大数据？
- 大数据带来了什么影响？
- 核心技术有哪些？
- 它的计算方式和代表性技术又是什么？

简单来说，这一章就是"大数据的前世今生"。

Slide 3-4: 제3절 제목

안녕하세요, 오늘 우리는 계속해서 《빅데이터 기술 개론》의 제3절인 '빅데이터의 영향'을 다루겠습니다.

이 절을 얕보지 마세요. 이것은 우리가 어떻게 생각하고, 어떻게 연구하며, 심지어 어떻게 의사결정을 하는지와 관련이 있습니다. 📊

Slide 3:章节标题

大家好，今天我们继续讲《大数据技术概述》的第三小节：大数据的影响。

别小看这个章节，它可关系到我们怎么思考、怎么做科研，甚至怎么决策。📊

Slide 5: 강의 목차

오늘 우리는 주로 두 가지 측면을 다룹니다:

1. 빅데이터가 과학 연구를 어떻게 변화시키는가
2. 빅데이터가 우리의 사고방식에 미치는 영향

여러분은 빅데이터가 단지 "기술적인 일"이라고 생각할 수 있지만, 사실 빅데이터는 이미 우리의 머리가 어떻게 돌아가는지에 영향을 미치고 있습니다!

Slide 5:提纲

今天我们主要讲两个方面：

1. 大数据对科学研究的改变
2. 大数据对我们思维方式的影响

你可能以为大数据只是"技术活"，
其实它已经影响我们的脑子是怎么转的！

Slide 6: 빅데이터 시대의 개막

빅데이터가 도래했습니다. 이것은 마치 "신석기 시대"와 같이 완전히 새로운 문명을 열었습니다.

과거에는 인간의 두뇌와 주판에 의존했지만, 이제는 데이터와 알고리즘에 의존합니다.

예전에는 날씨에 맞춰 농사를 지었지만, 이제는 모델을 통해 태풍이 언제 올지 예측합니다 🌪️!

Slide 7: 과학 연구의 새로운 단계

과학 연구도 빅데이터에 의해 재정의되고 있습니다.
과거에는 이론 + 실험에 의존했지만, 이제 두 가지 핵심 키워드가 추가되었습니다:
📌 데이터
📌 계산
이들은 이론 및 실험과 나란히 새로운 과학 연구 체계를 구성합니다!

Slide 8:

Slide 9:

Slide 6:大数据时代的开启

大数据来了，它就像"新石器时代"一样，开辟了全新的文明。

从前我们靠人脑和算盘，现在我们靠数据和算法。

以前看天吃饭，现在靠模型预测台风什么时候来🌪️！

Slide 7:科学研究的新阶段

科学研究也被大数据重新(chóng xīn) 定义。
以前靠理论 + 实验，现在多了两个关键词：
📌 数据
📌 计算
它们和理论与实验并列，
组成新的科学研究体系！

Slide 8:

科学研究包含实验、理论、计算和数据四大领域。

Slide 9:

数据库专家Jim Gray 博士
进入大数据时代以后
是以数据为中心的研究范式.

Slide 10: 제1 패러다임 -

우리에게 익숙한 첫 번째 과학 연구 패러다임입니다:

👨‍🔬 실험, 예를 들어 갈릴레오가 피사의 사탑에서 두 개의 쇠공을 떨어뜨린 것과 같습니다.

이것은 "관찰을 통해 가설을 검증하는" 전형적인 방법입니다.

Slide 10:第一范式：实验

这是我们熟悉的第一种科研范式：

👨‍🔬 实验，比如伽利略在比萨斜塔上扔两个铁球。

这是"用观察验证假设"的典型方法。

Slide 11: 제2 패러다임 - 이론

다음은 이론 시대입니다.

예를 들어 우리가 배운 📐기하학, 뉴턴의 법칙은 모두 "공식에서 출발하는" 이론 연구에 속합니다.

이때 강조되는 것은 "머리로 세계를 추론하는 것"입니다!

Slide 11:第二范式：理论

接下来是理论时代。

比如我们学过的📐几何、牛顿定律，都属于"从公式出发"的理论研究。

这时候强调的是"用脑子推演(**tuī yǎn**)世界"！

Slide 12:

과학의 이론 단계에서는 우리는 천재적인 재능을 가진 사람이 필요합니다.

뉴턴의 세 가지 법칙

1. **뉴턴 제1법칙 (관성의 법칙)**
2. **뉴턴 제2법칙 (운동 법칙)**
3. **뉴턴 제3법칙 (작용-반작용의 법칙)**

Slide 12:

在科学理论阶段，我们需要拥有天才智慧的人才。

牛顿三大定律

1. **牛顿第一定律（惯性定律）**
2. **牛顿第二定律（运动定律）**
3. **牛顿第三定律（作用与反作用定律）**

Slide 13: 제3 패러다임 - 계산

인터넷과 슈퍼컴퓨터(예: 톈허 2호) 시대에 이르러, 우리는 세 번째 방식을 갖게 되었습니다:

🧠 이론 → 🧪 실험 → 💻 시뮬레이션!

이것이 바로 "계산"을 중심으로 하는 새로운 연구 방식이며, 특히 물리학, 기후, 생물학 모델에서 광범위하게 사용됩니다.

Slide 13: 第三范式：计算

到了互联网和超级计算机（比如天河二号）时代，我们有了第三种方式：

🧠 理论 → 🧪 实验 → 💻 模拟！

这就是以“计算”为中心的新研究方式，尤其在物理、气候、生物模型中广泛使用。

Slide 14: 제4 패러다임 - 데이터 중심

이제 우리는 제4 패러다임에 진입합니다:

💡 "데이터 주도"의 과학 연구!

우리는 더 이상 가설에서만 출발하는 것이 아니라, 데이터에서 "새로운 현상을 발견"하고 그것을 설명합니다.

이것은 마치 이런 것과 같습니다:

"내가 무엇을 알고 싶은가가 아니라, 데이터가 내가 무엇을 알아야 하는지 알려준다."

Slide 14:第四范式：以数据为中心

现在我们进入第四范式：

💡 “数据驱动(**qū dòng**)”的科学研究！

我们不再只是从假设出发，而是从数据中去“发现新现象”，再去解释它。

这有点像：

“不是我想知道什么，

而是数据告诉我该知道什么”。

Slide 15:

계산 중심 시대는 "어떻게 문제를 해결할 것인가"에 관한 것이었다면, 데이터 중심 시대는 "데이터에서 문제와 그 해답을 발견하는 것"에 관한 것입니다.

Slide 15:

简单来说，以计算为中心的时代是“如何解决问题”，而以数据为中心的时代是“从数据中发现问题及其答案”。

Slide 16-19: 실제 사례 - 상품 판매 감소

예를 들어 이 그래프는 한 상품의 1년 내 판매량 하락을 보여줍니다 📉.

전통적인 방법은 마케팅 부서를 불러 회의를 하는 것일 수 있습니다.
하지만 이제 우리는 빅데이터 분석을 사용합니다:

- 어느 달에 급격히 떨어졌는가?
- 어느 지역의 판매량이 가장 나쁜가?
- 이 달에 경쟁 제품이 할인을 했는가?

👉 모두 데이터로 말합니다!

Slide 20-21: 사고방식에 미치는 영향

이제 중점적으로 빅데이터가 우리의 사고방식을 어떻게 변화시키는지 살펴보겠습니다.

답은 유명한 책에 나와 있습니다 — 📖 《빅데이터 시대》 by 빅터 마이어-쇤베르거

Slide 16:实际案例——商品销售(**xiāo shòu**)下降

比如这张图，一个商品一年内销量下滑📉。

传统做法可能找市场部开会讨论。

但现在我们用大数据分析：

- 哪个月掉得快？
- 哪个地区销量最差？
- 是不是有竞品在这个月打折？

👉 全部靠数据说话！

Slide 17, 18.19 그냥 보면서 설명

Slide 20:对思维方式的影响

我们接下来重点看，大数据到底怎么改变我们思维方式？

Slide 21 :

答案出现在一本著名的书里——

📖 《大数据时代》by 维克托·迈尔·舍恩伯格

Slide 22-27: 전수조사, 표본조사가 아님

전통 통계학에서는 표본 조사를 사용했습니다: 예를 들어 단지 1,000명만 조사하여 전체 도시 인구를 추정하는 것입니다.

하지만 빅데이터 시대에는 우리가 할 수 있는 것이 —

📦 전수 표본 분석!

왜냐하면:

- 저장 장치가 저렴해졌고
- 컴퓨팅 능력이 강력해졌으며
- 데이터 획득이 더 편리해졌기 때문입니다

마치 여러분이 밀크티를 주문할 때 시스템이 묻지 않아도 됩니다: "최근에 무엇을 좋아하세요?" 시스템은 이미 여러분이 매번 설탕을 몇 스푼 넣는지 알고 있습니다 😆

Slide 22-27:全样而非抽样

제목만 이야기하고 다음으로

Slide 23:

在传统统计学中，我们用抽样调查：

比如只调查1000人就推断整个城市人口。

Slide 24: 슬라이드 보면서 이유 설명

Slide 25:26 그냥보면서 넘어감

Slide 27:

但在大数据时代，我们可以做到——

📦 全样本分析！

因为：

- 存储设备便宜了
- 算力变强了
- 数据获取更方便

就像你点奶茶时系统不用问你：

“你最近爱喝什么？”

它已经知道了你每次加了几分糖 😆

我们学校北区食堂有一家蜜雪冰城。

午饭后我去的时候，都不用我开口，打工的学生就会给我做一杯“咖啡拿铁，不加糖”。

我只需要付7块钱就行。

Slide 28-32: 효율성, 정확성이 아님

과거에는 알고리즘이 99.9%까지 정확하기를 추구했지만, 이제는 반응 속도와 전체적인 추세를 더 중시합니다.

다시 말하면:
📌 "빠름 + 대략적으로 정확함" > "느림 + 매우 정확함"
마치 논문을 쓸 때 표절률 93% vs 94%는 사실 비슷합니다. 하지만 10분 안에 한 번 검사를 완료할 수 있다면, 그것이 진짜 중요한 것입니다!

Slide 28-32:效率而非精确
제목만 얘기하고 다음으로

Slide 29:
以前我们追求算法精确到99.9%，现在更注重反应速度和整体趋势**(qū shì)**。
换句话说：
📌 “快+大致正确” > “慢+超级精确”

就像你写论文，查重率**(chá chóng lǜ**)93% vs 94%，其实差不多。
但如果你10分钟能查完一次，才是真正重要！

Slide 30, 31,32：
슬라이드 보면서 내용 설명

Slide 33-37: 상관관계, 인과관계가 아님

전통적인 연구는 인과관계를 강조합니다 (예: 과식 → 비만)

빅데이터는 상관성을 강조합니다:
🍰 디저트를 좋아하는 사람은 자주 콜라를 산다
📚 추리 소설을 보는 사람은 자주 스릴러 영화 티켓을 산다

👉 시스템은 "왜 그런지" 알 필요가 없고, 단지 "이것을 클릭하면 저것도 클릭할 가능성이 있다"는 것만 알면 됩니다.

이것이 바로 추천 시스템의 원리입니다!

Slide 33:相关而非因果
传统研究强调因果关系（比如：吃太多 → 变胖）

大数据则强调相关性：
🍰 喜欢甜品**(tián pǐn)**的人常买可乐

📚 看推理小说的人常买惊悚**(jīng song)**电影票

👉 系统不需要知道“为什么”，只要“你点了这个，可能也会点那个”。

这就是推荐系统的原理！

Slide 38-46: 정답 모음

- **Slide 38**: ✅ 정답: D
- **Slide 39**: ✅ 정답: C
- **Slide 40**: ✅ 정답: C
- **Slide 41**: ✅ 정답: D
- **Slide 42**: ✅ 정답: B
- **Slide 43**: ✅ 정답: C
- **Slide 44**: ✅ 정답: C
- **Slide 45**: ✅ 정답: C **rǒng yú 중복**
- **Slide 46**: ✅ 정답: D
- **Slide 47**: ✅ 정답: B

Slide 38-46: 정답 모음

- Slide 38:✅ 正确答案：D
- Slide 39:✅ 正确答案：C
- Slide 40:✅ 正确答案：C
- Slide 41:✅ 正确答案：D
- Slide 42:✅ 正确答案：B
- Slide 43:✅ 正确答案：C
- Slide 44:✅ 正确答案：C
- Slide 45:✅ 正确答案：C
- **데이터 중복**
- Slide 46:✅ 正确答案：D **废弃(fèiqì)**
- Slide 47:✅ 正确答案：B

Slide 48: 마무리

좋습니다, 오늘 수업은 여기서 마치겠습니다!

오늘 우리는 "빅데이터 시대"의 전체 맥락을 알게 되었습니다. 정보화 물결, 세 가지 지원 기술, 데이터가 어떻게 생성되고 어떻게 진화하는지를 배웠습니다.

다음 수업에서는 계속해서 "빅데이터의 핵심 기술"을 탐구하겠습니다.

마지막으로 질문 하나: 여러분은 오늘 매일 자신도 "데이터를 생산"하고 있다는 것을 발견하셨나요? 😉

Slide 48:

好了，我们本节课到这里就结束啦！

今天我们认识了"大数据时代"的来龙去脉，从信息化浪潮、三大支撑技术，到数据如何产生、如何演变。

下节课我们会继续探讨"大数据的关键技术"。

最后提问一下：你今天有没有发现，自己每天其实也在"生产数据"呢？😉

谢谢大家，下课！👏👏👏