Unveiling the Power of Big Data

Transforming Insights through Data Science on Big Data Platforms

Zhanbo Hua Xiangrui Zhao Haozhe Zhou Haopo Zhang

School of Computer Science & Technology
China University of Mining and Technology

September 20, 2023

Table of Contents

Introduction

- Big Data is transforming industries worldwide.
- In this presentation, we explore its significance and applications.
- Let's begin our journey into the world of Big Data.

Big Data Definition:

- Big Data refers to extremely large and complex datasets that cannot be easily managed, processed, or analyzed using traditional data processing tools.
- It is characterized by the four Vs: Volume, Velocity, Variety, and Veracity.

Figure: Big Data

Figure: Data Generation

Data is generated all the time:

- The Internet of Things, cloud computing, mobile Internet, mobile phones and tablets, PCs, and various sensors everywhere are all sources or carriers of big data.
- It can be said that big data is all around us, producing and carrying large amounts of data every day.

- Data Science is the key to unlocking insights from Big Data.
- It involves data collection, analysis, and interpretation.
- Data Science helps organizations make data-driven decisions.

Figure: Data Science

Case Studies

- Let's explore real-world examples to understand the impact of Big Data and Data Science.
- Case Study 1: E-commerce Personalization
- Case Study 2: Healthcare Analytics

Personalization features on business-to-consumer e-commerce:

- The rapid growth of technology and social media enables easy real-time information exchange and interaction among users, impacting online personalized advertising performance. However, it also poses challenges in terms of consumer trust and privacy concerns for researchers.

Commercial Personalization Features	Description	Ref
Rewards	A feature that offers rewards to consumers based on their purchasing or loyalty performance.	[16]
Online Advertisement	An online advertisement feature that adjusts advertisement content based on user preferences.	[33] [32]
Pricing	Personalized pricing based on the users' characters, purchase history or purchase power.	[34]
Bundle-Based Pricing	Offering bundle-based pricing according to user preferences.	[15]
Best Offer	A feature that offers the best deals of a product according to user preference.	[23] [35]
Promotion Deal	A personalized feature that promotes a product based on user behavior or purchase history.	[15] [36] [37] [23]
Discount	A feature that enables personalization of discount price based on user purchase power, loyalty and user purchase history.	[38] [39]

Figure: Commercial personalization features

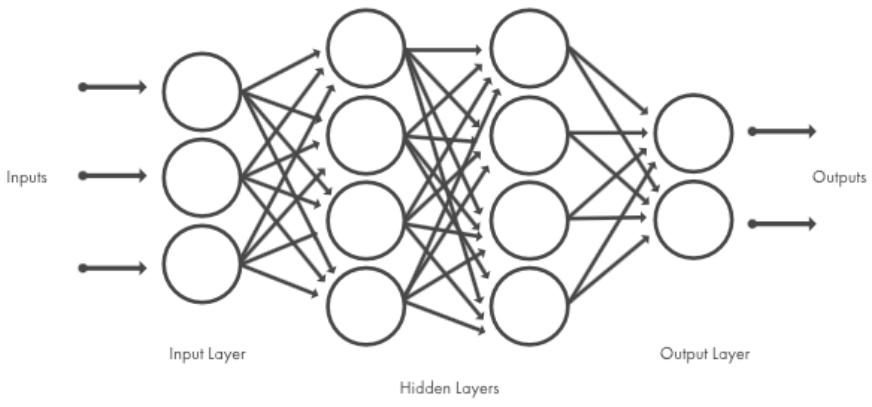

Figure: Neural networks

Applying data science to healthcare data:

- Coronavirus disease 2019 (COVID-19) data are notable healthcare data. The disease broke out in 2019, was declared as a pandemic on 11 March 2020, and is still prevailing more than two years later (in 2022). Healthcare informatics on COVID-19 data helps get a better understanding of the disease and find ways to prevent or combat it.

Challenges and Limitations

- While Big Data offers immense potential, it comes with challenges and limitations:
- Challenge 1: Data Privacy and Security
- Challenge 2: Scalability
- Limitation 1: Representativeness: a complete sample is never possible
- Limitation 2: Timeliness: Second-level value exists
- We must address these challenges to harness the full power of Big Data.

Future Trends

- The world of Big Data is constantly evolving.
- Future trends include:
- Trend 1: Edge Computing
- Trend 2: AI and Machine Learning Integration
- Trend 3: Advancements in Big Data Platforms
- Stay updated to stay competitive in the era of Big Data.

Figure: Future Trends

Conclusion

- Big Data, Data Science, and Data Mining are revolutionizing industries.
- With the right strategies, organizations can harness Big Data's power.
- Embrace the future of data for informed decision-making.

References

[1] Chen, J., Song, L., Liu, X., & Wang, L. (2014). Big Data Deep Learning: Challenges and Perspectives. IEEE Access, 2, 514-525.

[2] Provost, F., & Fawcett, T. (2013). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media.

[3] Zikopoulos, P., Eaton, C., deRoos, D., Deutsch, T., Lapis, G., & Deroos, D. (2012). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill Osborne Media.

[4] Ghemawat, S., Gobioff, H., & Leung, S. T. (2003). The Google file system. ACM SIGOPS Operating Systems Review, 37(5), 29-43.

[5] White, T. (2015). Hadoop: The Definitive Guide. O'Reilly Media.

[6] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute.