

Big Data

Big Data in Action

Big Data

❖ New FLU Virus Starts in the U.S.!

- H1N1 flu virus (which has combined virus elements of the bird and swine (pig) flu) started to spread in the U.S. in 2009
- U.S. CDC (Centers for Disease Control and Prevention) was only collecting diagnostic data of Medical Doctors once a week
- Using the CDC information to find how the flu was spreading would have an approximate 2 week lag, which is far too slow compared to the speed of the virus spreading

Big Data

❖ New FLU Virus Starts in the U.S.!

- What vaccine was needed?
- How much vaccine was needed?
- Where was the vaccine needed?
- Vaccine preparation and delivery plans could not be setup fast enough to safely prevent the virus from spreading out of control

Big Data

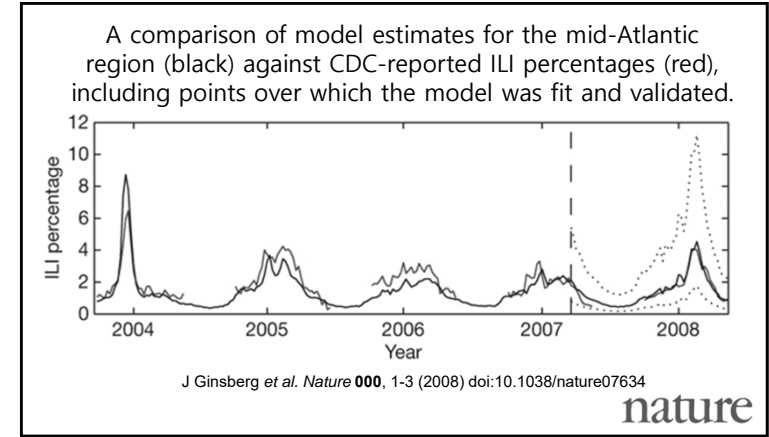
❖ New FLU Virus Starts in the U.S.!

- Fortunately, Google published a paper about how they could predict the spread of the winter flu in the U.S. accurately down to specific regions and states
- This paper was published in the journal *Nature* a few weeks before the H1N1 virus made the headline news

Big Data

Detecting Influenza Epidemics using Search Engine Query Data

Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. *Nature*, 457, pp. 1012–1014, February 19, 2009.

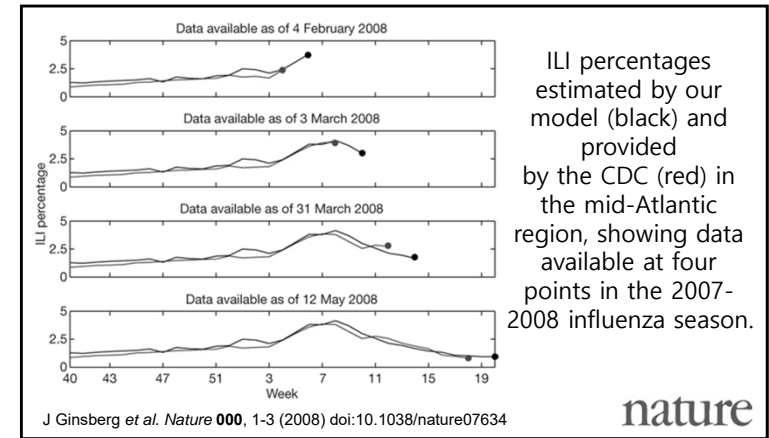


Reference : <https://www.nature.com/articles/nature07634t>

Big Data

Detecting Influenza Epidemics using Search Engine Query Data

Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. *Nature*, 457, pp. 1012–1014, February 19, 2009.



Reference : <https://www.nature.com/articles/nature07634t>

Big Data

❖ New FLU Virus Starts in the U.S.!

- Millions of the most common search terms and Millions of different mathematical models were tested on Google's database
 - Google receives more than 3 billion search queries a day
- Analysis system was set to look for correlation between the frequency of certain search queues and the spread of the flu over time and space

Big Data

❖ New FLU Virus Starts in the U.S.!

- Google's method of analysis did not use data provided from hospitals or Medical Doctors
- Google used Big Data analysis on the most common search terms people use
 - Unstructured Data, Semi-Structured Data
- Google's system proved to be more accurate and faster than analyzing government statistics

Big Data

Big Data Characteristics

Big Data Characteristics

❖ Big Data Analysis

- Analysis on the Whole Dataset
- Highly Scalable
- Structured, Semi-structured & Unstructured data
- Clusters, Highly Distributed
- Commodity Computers/Servers
- General type fast network

❖ Traditional Data Analysis

- Analysis on Sampled Data (Error Margin, Reliability Range)
- Limited Scalability
- Structured Data
- Centralized
- Custom-made Servers
- SAN (Storage Area Network)

Big Data References

References

- V. Mayer-Schönberger, and K. Cukier, Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt, 2013.
- T. White, Hadoop: The Definitive Guide. O'Reilly Media, 2012.
- J. Venner, Pro Hadoop. Apress, 2009.
- S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, "Big Data, Analytics and the Path From Insights to Value," MIT Sloan Management Review, vol. 52, no. 2, Winter 2011.
- B. Randal, R. H. Katz, and E. D. Lazowska, "Big-data Computing: Creating revolutionary breakthroughs in commerce, science and society," Computing Community Consortium, pp. 1-15, Dec. 2008.
- G. Linden, B. Smith, and J. York. "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Computing, vol. 7, no. 1, pp. 76-80, Jan/Feb. 2003.

References

- J. R. Galbraith, "Organizational Design Challenges Resulting From Big Data," *Journal of Organization Design*, vol. 3, no. 1, pp. 2-13, Apr. 2014.
- S. Sagioglu and D. Sinanc, "Big data: A review," *Proc. IEEE International Conference on Collaboration Technologies and Systems*, pp. 42-47, May 2013.
- M. Chen, S. Mao, and Y. Liu, "Big Data: A Survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171-209, Jan. 2014.
- X. Wu, X. Zhu, G. Q. Wu, and W. Ding, "Data Mining with Big Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 97-107, Jan. 2014.
- Z. Zheng, J. Zhu, and M. R. Lyu, "Service-Generated Big Data and Big Data-as-a-Service: An Overview," *Proc. IEEE International Congress on Big Data*, pp. 403-410, Jun/Jul. 2013.

References

- I. Palit and C.K. Reddy, "Scalable and Parallel Boosting with MapReduce," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 10, pp. 1904-1916, 2012.
- M.-Y Choi, E.-A. Cho, D.-H. Park, C.-J Moon, and D.-K. Baik, "A Database Synchronization Algorithm for Mobile Devices," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 392-398, May 2010.
- IBM, What is big data?, <http://www.ibm.com/software/data/bigdata/what-is-big-data.html> [Accessed June 1, 2015]
- Hadoop Apache, <http://hadoop.apache.org>
- Wikipedia, <http://www.wikipedia.org>

Image sources

- Walmart Logo, By Walmart [Public domain], via Wikimedia Commons
- Amazon Logo, By Balajimuthazhagan (Own work) [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)], via Wikimedia Commons