LUT
Lappeenranta
University of Technology

BIG DATA – CHAPTER 4

Marc Seidel 0448380

Correlation

Amazon - the early days

In the early days Amazon employed a dozen book critics and editors to write reviews and suggest new titles. Their work was considered to be the company's crown jewels and a source of competitive advantage.

From the start Amazon had captured what customers had purchased, what books they looked at and how long they looked at them.

The first analysis was conduced by taking a sample and analysing it to find similarities among customers. However the results were crude as they offered only a tiny variation to the previous purchase.

Correlation

Amazon - the early days

A solution to the problem was found by detecting associations among products instead of customer to customer relationships.

Later it also worked across product categories

In a test between critics and computer-generated content data-derived material generated much more sales.

Amazons recommender system has been adopted by almost everyone in ecommerce.

Amazon's innovative recommendation system teased out valuable correlations without knowing the underlying causes. Knowing *what*, not *why*, is good enough.

Data set

From a limited data set to n = all

Before big data, correlations usefulness was limited. Because data was scarce and collecting it was expensive, statisticians often chose a proxy, then collected the relevant data and ran the correlation analysis to find out how good that proxy was.

In the big-data age, it is no longer efficient to make decisions about what variables to examine by relying on hypotheses alone. The data sets are far too big and the area under consideration is probably far too complex.

Predictions

Shops personalize their offers depending on customers situation in life

Correlation quantifies the statistical relationship between two data values. A strong correlation means that when one of the data values changes, the other is highly likely to change as well.

Correlations cannot foretell the future, they can only predict it with a certain likelihood. But that ability is extremely valuable.

Prediction

Earlier identification of problems leads to cost saving / security improvements

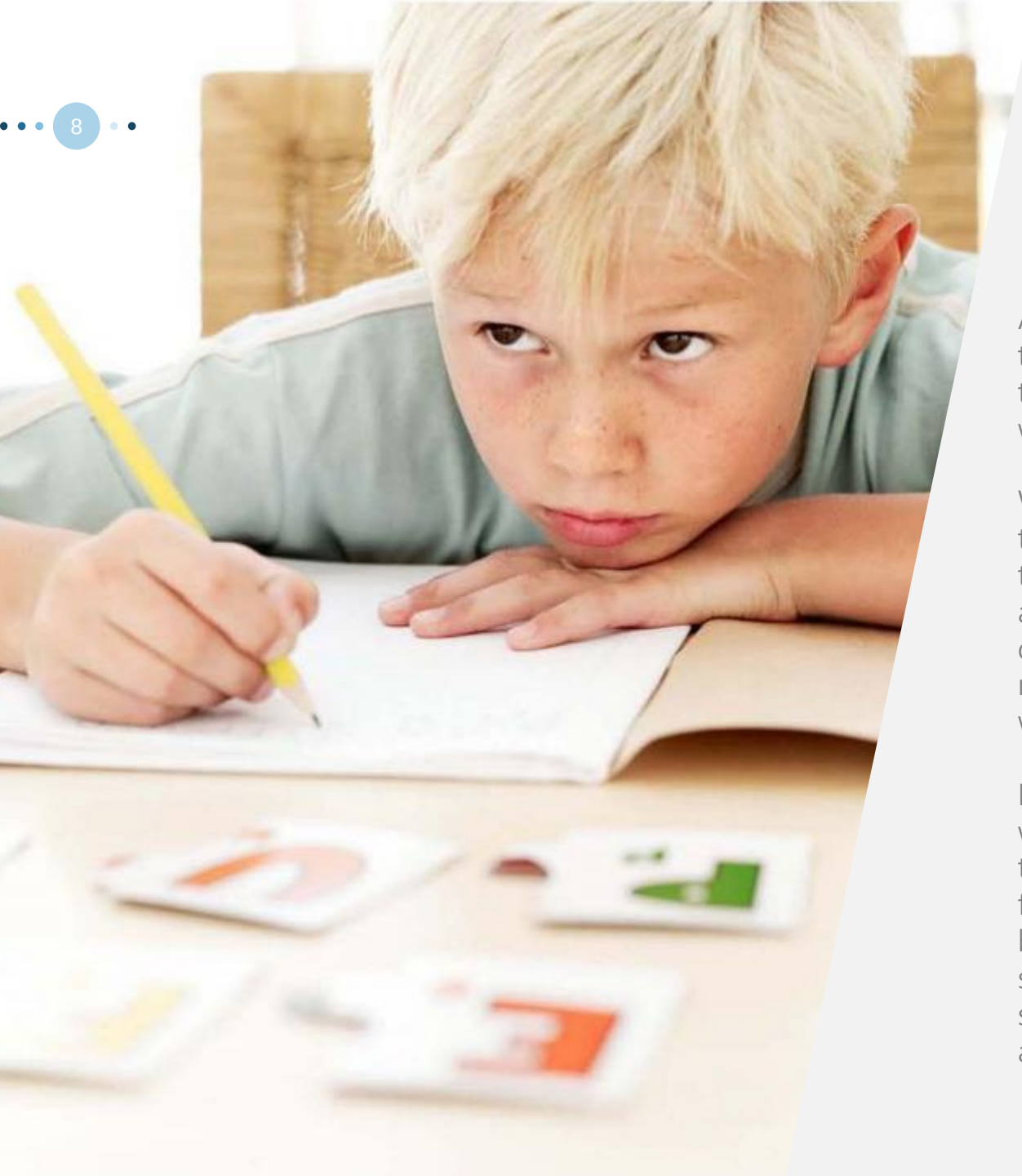
Note that predictive analytics may not explain the cause of a problem; it only indicates that a problem exists. It will alert you that an engine is overheating, but it may not tell you whether the overheating is due to a frayed fan belt or a poorly screwed cap.

Non-linear

Before big data, partly because of inadequate computing power, most correlational analysis using large data sets was limited to looking for linear relationships. In reality, of course, many relationships are far more complex. With more sophisticated analyses, we can identify non-linear relationships among data.

Experts are just now developing the necessary tools to identify and compare non-linear correlations.

Ultimately, in the age of big data, new types of analyses will lead to a wave of novel insights and helpful predictions. We will see links we never saw before.

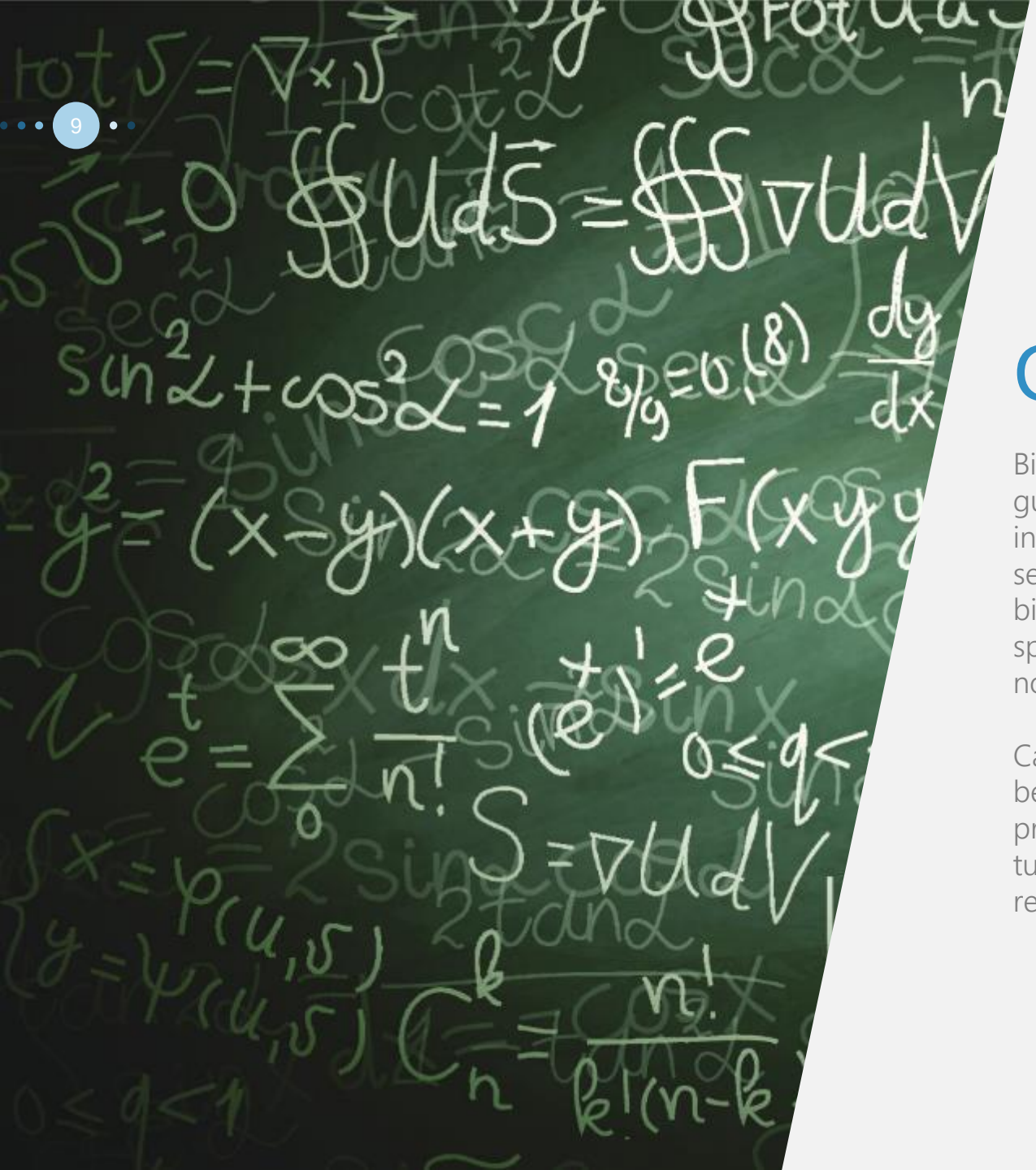


Non-Causal

As humans, we desire to make sense of the world through causal links; we want to believe that every effect has a cause, if we only look closely enough.

When we say that human sees the world through causalities, we're referring to two fundamental ways humans explain and understand the world: through quick, illusory causality; and via slow, methodical causal experiments. Big data will transform the roles of both.

In a small-data world, showing how wrong causal intuitions were took a long time. This is going to change. In the future, big-data correlations will routinely be used to disprove our causal intuitions, showing that often there is little if any statistical connection between the effect and its supposed cause.



Causality

Big data itself aids causal inquiries as it guides experts toward likely causes to investigate. In many cases, the deeper search for causality will take place after big data has done its work, when we specifically want to investigate the why, not just appreciate the what.

Causality won't be discarded, but it is being knocked off its pedestal as the primary fountain of meaning. Big data turbocharges non-causal analyses, often replacing causal investigations.

Big data may not spell the “end of theory” but it does fundamentally transform the way we make sense of the world. This change will take a lot of getting used to. It challenges many institutions. Yet the tremendous value that it unleashes will make it not only a worthwhile trade-off, but an inevitable one.

Appendix - Images

Cover
http://i.huffpost.com/gen/1258120/thumbs/o-BIG-DATA-facebook.jpg

Amazon Headquarter
http://thenextweb.com/wp-content/blogs.dir/1/files/2011/06/Amazon.jpeg

Greg Linden
http://lists10.com/wp-content/uploads/2014/05/door-desks-2.jpg

Mother & Child
http://www.canadianpainsummit2012.ca/wp-content/uploads/2014/06/mother-and-baby.jpg

Medication
http://www.mhsi.org/image/jpeg/MEDICATION.jpg

Plane
http://www.multiflight.com/uploads/images/0K8A4796.jpg

Homework
http://www.camdencountylibrary.org/sites/default/files/images/HomeworkHelp1-08-24-11(1).jpg

Mathematics
http://i.huffpost.com/gen/1344524/thumbs/o-IS-MATHEMATICS-INVENTED-OR-DISCOVERED-facebook.jpg