Lesson 4

Finding Hidden Patterns

Collapsing Infinity to a Point

Wholeness of the Lesson

Human brain is a pattern matching machine. Therefore our problem solving techniques are based on finding hidden patterns buried deeper in Nature. In this context, boils down to finding hidden patterns buried deeper in Big Data.

A successful action results from a deeper dive into silence, into pure intelligence, just as, in archery, the arrow flies truer and hits its mark more consistently if it is pulled back farther on the bow.

PAIRS AND STRIPES

One common approach for synchronization in MapReduce is to construct complex keys and values in such a way that data necessary for a computation are naturally brought together by the execution framework.

This section introduces two common design patterns known as "pairs" and "stripes".

Example - Building co-occurrence matrix

Event. Any thing we are interested in can be thought of as the event. We can represent an event using integer.

Example: A customer bought the following books.

Gone With the Wind   // 145 be its ID

Titanic   // 329 be its ID

Life of Pi   // 278 be its ID

…

This can be represented as

145 329 278 …

Example - Building co-occurrence matrix

Point 1:

Thus sequence of events associated with one customer can be represented in one record.

Point 2: Two different records belong to two different customers.

Example:

 5 7 8 2 9   // belongs to one customer
 8 9 7 8 2   // belongs to another customer

Example - Building co-occurrence matrix

Concept of a window

Given a event, you can define a window based on the problem at hand.

Examples

 8 9 7 8 7 2

Example 1. Window is just two events following the event different from the current event.

W(8) = {9, 7}, W(9) = {7, 8}, W(7) = {8, 2}, W(8) = {7, 2}, W(7) = {2}, W(2) = {}

Example - Building co-occurrence matrix

Examples (cont'd)

 8 9 7 8 7 2

Example 2.

W(x) = {y | y appear after x and before another x or the end of the list}

W(8) = {9, 7}, W(9) = {7, 8, 7, 2}, W(7) = {8}, W(8) = {7, 2}, W(7) = {2}, W(2) = {}

More examples of window definition

Window(x) = {y | y appear after x in the same sentence}

Window(x) = {y | y is the next two items after x}

Window(x) = {y | y appear after x in the same chapter}

Window(x) = {y | y happened within two weeks after x}

Example - Building co-occurrence matrix

Formally, the co-occurrence matrix is a square $n \times n$ matrix where n is the number of events.

A cell m_{ij} contains the number of times j appear in all windows of i.

Example - Building co-occurence matrix

Let W(x) = {y | y appear after x and before another x or the end of the list}

Example

8 9 7 8 7 2

7 2 8 4 2 1

m_{72} is 3. That is, 2 appear in all windows of 7, 3 times.