

SeeDB: Archaeologist Role

Alice Yeh

Influenced By: Multi-Armed Bandit (MAB) Strategy [VM05]

- Multi-armed bandit problem (k -armed slot machine)
 - Maximizing gain via selections over time steps w/ increased knowledge about choices over allocations
- POKER strategy
 - **Pricing uncertainty**: assign price to knowledge gained by pulling a particular arm
 - **Exploit lever distribution**: can estimate properties of unexplored arms by looking at ones already observed
 - **Take into account the horizon**: num rounds remaining will influence exploration
- SeeDB → multi-armed bandit pruning
 - Goal: find the visualizations (arms) with the highest utility (reward) → adapted Successive Accepts and Rejects algorithm



Influenced: Data Storytelling in Notebooks [LYZ+23]

- SeeDB → data fact recommendations
- Notable
 - On-the-fly data storytelling assistance in notebooks
 - Data fact recommendation algorithms to infer facts that users might be interested in

Notable

a Year

b Category

c Model

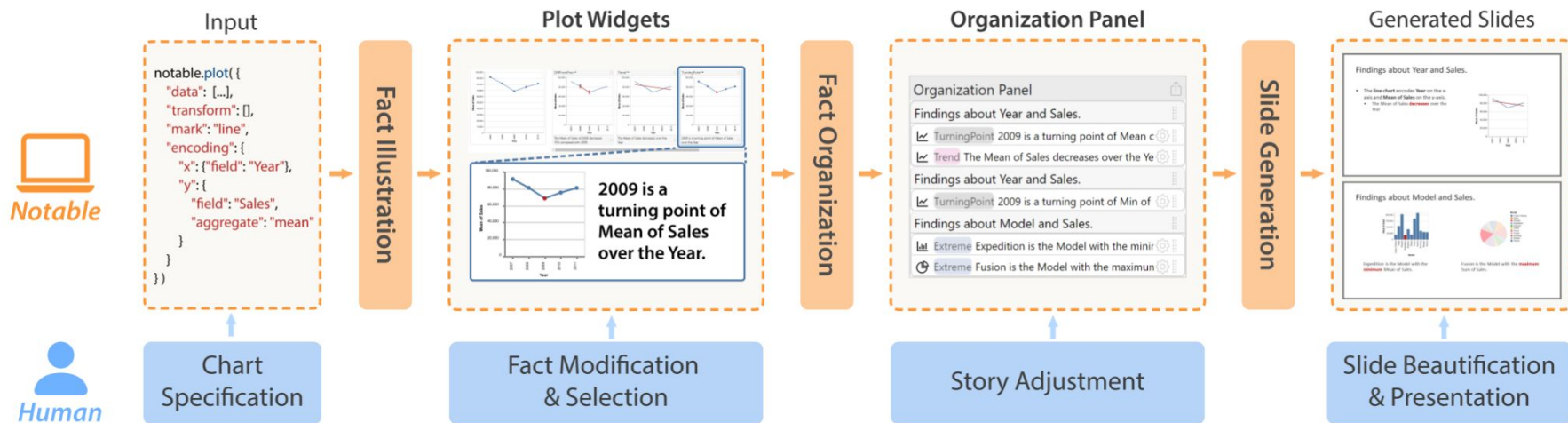
d Slides

Influenced: Data Storytelling in Notebooks [LYZ+23]

The image displays a Jupyter Notebook interface with the following components:

- Code Cell (a):** Contains a `notable.plot()` function call with parameters for data, transform, mark, and encoding. The code is enclosed in a blue dashed box.
- Plot Widgets (b):** Four plots are generated from the code:
 - Main Plot:** A line plot showing sales from 2007 to 2011. The y-axis is labeled 'Sales' and ranges from 0 to 70,000. The x-axis is labeled 'Year' and ranges from 2007 to 2011. The data points are approximately: (2007, 22000), (2008, 23000), (2009, 23000), (2010, 24000), (2011, 68000).
 - DiffFromPrev:** A line plot showing the difference in sales from the previous year. The y-axis is labeled 'Sales' and ranges from 0 to 70,000. The x-axis is labeled 'Year' and ranges from 2007 to 2011. The data points are approximately: (2007, 0), (2008, 1000), (2009, 1000), (2010, 1000), (2011, 44000).
 - Extreme:** A line plot showing the maximum sales for each year. The y-axis is labeled 'Sales' and ranges from 0 to 70,000. The x-axis is labeled 'Year' and ranges from 2007 to 2011. The data points are approximately: (2007, 22000), (2008, 23000), (2009, 23000), (2010, 24000), (2011, 68000).
 - Trend:** A line plot showing the trend of sales. The y-axis is labeled 'Sales' and ranges from 0 to 70,000. The x-axis is labeled 'Year' and ranges from 2007 to 2011. The data points are approximately: (2007, 22000), (2008, 23000), (2009, 23000), (2010, 24000), (2011, 68000).
- Organization Panel (c):** A panel on the right side of the notebook, enclosed in an orange dashed box, containing a list of findings about sales and year. The findings are:
 - Findings about Sales and Year:
 - Trend The Mean of Sales decreases over the Year
 - DiffFromPrev The Mean of Sales of 2009 decreases 1%
 - TurningPoint 2009 is a turning point of Mean of Sales
 - Findings about Fusion:
 - Extreme Fusion is the Model with the maximum Mean
 - Trend The Sales increases over the Year
 - Findings about Fiesta:
 - Extreme Fiesta is the Model with the minimum Mean
 - Extreme Fiesta is the Model with the minimum Sum c
 - Findings about Fiesta:
 - DiffFromPrev The Sales of 2011 increases 195% comp

Influenced: Data Storytelling in Notebooks [LYZ+23]



Sources

<https://cs.nyu.edu/~mohri/pub/bandit.pdf>

https://en.wikipedia.org/wiki/Multi-armed_bandit

<https://arxiv.org/pdf/2303.04059.pdf>