Utility-preserving database and datastream summarization system

Stanford researchers have developed a data sketching method that leverages neural networks to perform queries on large datasets. As datasets grow larger and more complex, they must be compacted (sketched) in ways such that they are easily stored and processed. Performing analyses on these large datasets requires extensive computing power and conventional methods use ad-hoc, randomized algorithms to develop sketches. This technology uses neural networks, a machine learning algorithm to develop sketches and facilitate queries and other data analyses. This neural network method better captures the properties of the data and preserves their utility, which reduces the computation power required and increases the accuracy in downstream applications.

Stage of Development
Proof of concept

Applications

- Performing queries on large datasets, such as genomic data
- Performing traditional data analyses (k-means, PCA) using only data summary
- Biobanks
- Financial data
- Genomic companies (genomic data)

Advantages

- **Faster** than conventional data summarization methods
- **Less computing power** required than conventional data summarization methods

- **Increased accuracy** in downstream analyses

Innovators

- Alexander Ioannidis
- Daniel Montserrat

Licensing Contact

Jon Gortat

Licensing & Strategic Alliances Director for Physical Science

[Email](#)