

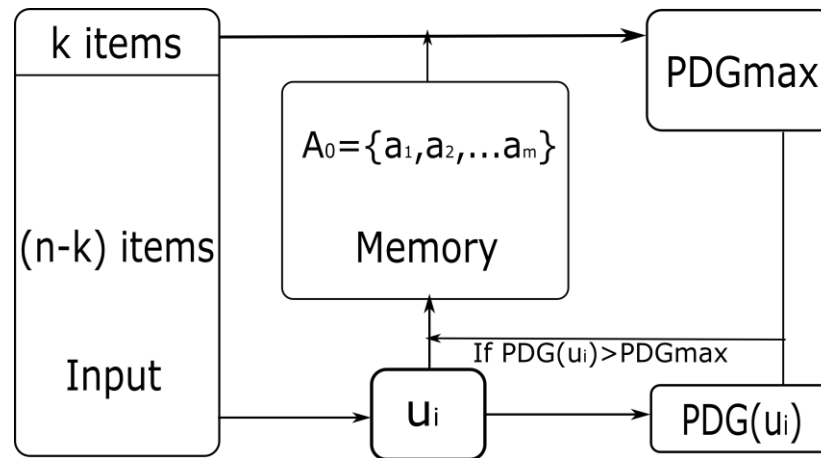
Diversification on big data in query processing

**Meifan ZHANG, Hongzhi WANG, Jianzhong LI,
Hong GAO**

Frontiers of Computer Science, DOI: [10.1007/s11704-019-8324-9](https://doi.org/10.1007/s11704-019-8324-9)

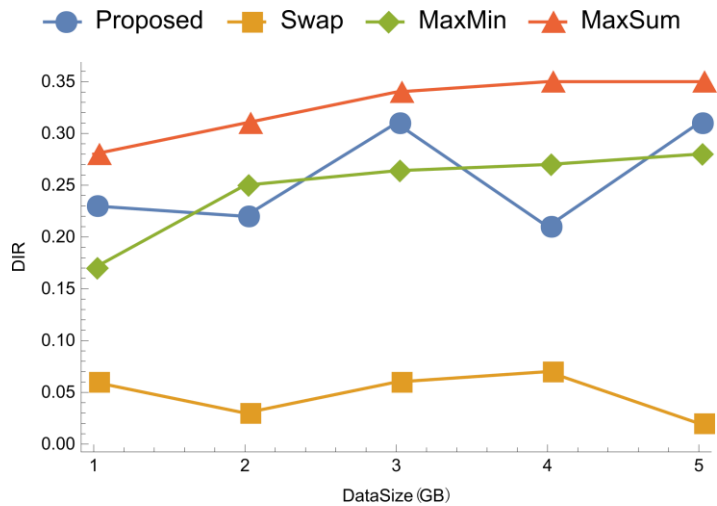
Problems & Ideas

- Problems of improving diversification of query results in large scale:
 - Limitation of Memory
 - Requirement of fast response
- Ideas: Improve the diversification with one-pass data scan.
 - Find the maximum Possible Diversity Gain (PDG_{max}) from k items.
 - Find the item u_i with $PDG(u_i) > PDG_{max}$ from the rest items.



Main Contributions

- **Performance comparison (Numerical data)**



- **Performance comparison (String data)**

