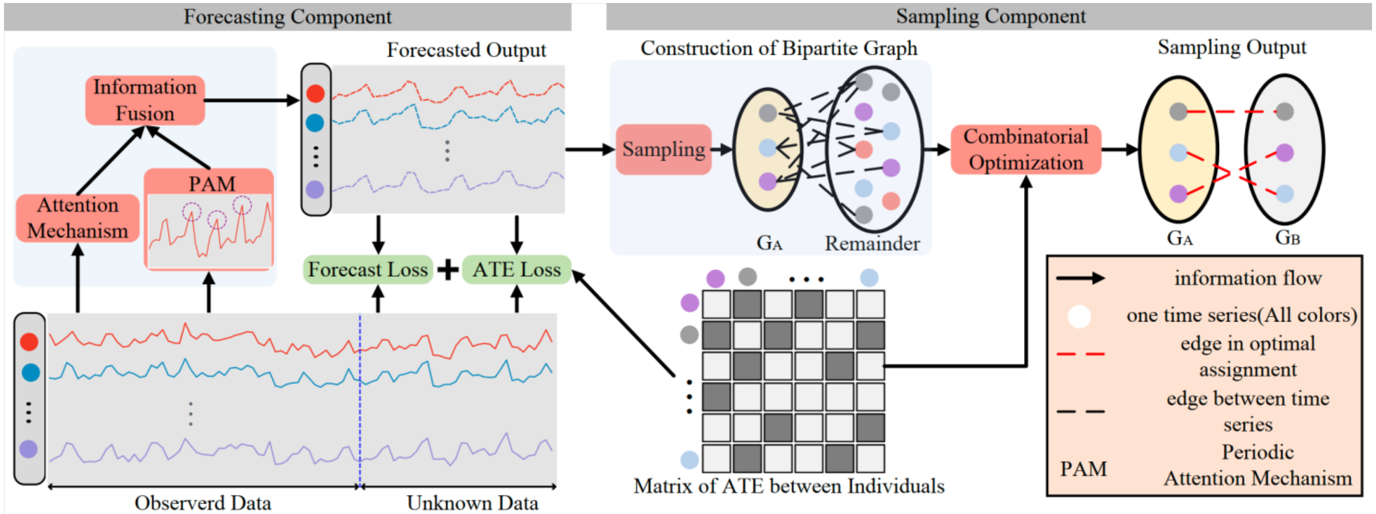
# A Sampling Method based on Forecasting and Combinatorial Optimization for High Performance A/B testing

**Fan LI, Tiancheng ZHANG, Shengjia CUI, Hengyu LIU, Zhibin REN, Donglin DI, Xiao WANG, Po ZHANG, Ge YU**

Frontiers of Computer Science, DOI: 10.1007/s11704-023-3460-7

# Problems & Ideas

- Problems of conventional sampling approaches in A/B testing:
  - The features of individuals being sampled change over time and the sampling results used for evaluation are valid only for a period of time.
  - The sampling methods in A/B testing require determining whether the two groups of samples have an identical distribution while Simple Random Sampling(SRS) has a different purpose, which only concerns whether the sample distribution is identical to the population.
- Ideas: We propose a forecasting and combinatorial optimization-based sampling method.




An overview of our Framework. The Forecasting Component aims to predict future data and the Sampling component is a Combinatorial Optimization-based sampling method, which is designed for A/B testing

# Main Contributions

- Contributions:
  - We propose a general framework, SFCO, to address the problems in sampling time-varying data. The results suggest that forecasting-based sampling is a promising new line of research for sampling in A/B testing.
  - We present a combinatorial optimization-based sampling method designed to promote both the similarity between the distributions of G_A and G_B, as well as the similarity between the distribution of samples and that of the entire population.
  - We present a systematic approach for evaluating sampling methods.

**Table 1**: SFCO compare with Simple Random Sampling(SRS) in Average Treatment Effect(ATE) on different datasets

| Dataset | Feature | SRS | SFCO | $R$ |
|---|---|---|---|---|
| Baidu | $f_1$ | 114.37 | 78.70 | 0.31 |
| | $f_2$ | 79.15 | 57.71 | 0.27 |
| | $f_3$ | 94.46 | 72.01 | 0.24 |
| | $f_4$ | 545611.34 | 398264.60 | 0.27 |
| | $f_5$ | 447381.69 | 361849.30 | 0.19 |
| | $f_6$ | 301198.07 | 270563.53 | 0.10 |
| | $f_7$ | 6708.25 | 5126.37 | 0.24 |
| | $f_8$ | 878.37 | 760.60 | 0.13 |
| | $f_9$ | 855.94 | 719.76 | 0.16 |
| | $f_{10}$ | 13668.47 | 9625.19 | 0.30 |
| | Average | - | - | 0.22 |
| Electricity | - | 2440.23 | 257.57 | 0.89 |
| Traffic | - | 0.004745 | 0.002992 | 0.37 |

**Table 2**: The $D_{n,m}$ between $G_A$ and $G_B$(DAB) results on different datasets. The best-performing method is boldfaced

| Dataset | SRS | SFCO |
|---|---|---|
| Baidu | 0.2910 | **0.2880** |
| Traffic | 0.3695 | **0.3418** |
| Electricity | 0.5944 | **0.4395** |

**Table 3**: The $D_{n,m}$ between $(G_A, G_B)$ and Population(DABP) results on different datasets. The best-performing method is boldfaced

| Dataset | SRS | SFCO |
|---|---|---|
| Baidu | 0.3168 | **0.3094** |
| Traffic | 0.3296 | **0.3135** |
| Electricity | 0.7479 | **0.7017** |