

DISTINGUISHED LECTURE SERIES 2022/23

CARDINALITY ESTIMATION OF APPROXIMATE SUBSTRING QUERIES USING DEEP LEARNING

18 JUL 2023 (TUE)
4:00-5:00 PM (HKT)

Mr. and Mrs. Lee Siu Lun Lecture Theatre
WLB205, Shaw Campus, HKBU /

Online via Zoom

PROF. KYUSEOK SHIM

Department of Electrical and Computer Engineering
Seoul National University

Abstract:

Cardinality estimation of an approximate substring query is an important problem in database systems. Traditional approaches build a summary from the text data and estimate the cardinality using the summary with some statistical assumptions. Since deep learning models can learn underlying complex data patterns effectively, they have been successfully applied and shown to outperform traditional methods for cardinality estimations of queries in database systems. However, since they are not yet applied to approximate substring queries, we investigate a deep learning approach for cardinality estimation of such queries. Although the accuracy of deep learning models tends to improve as the train data size increases, producing a large train data is computationally expensive for cardinality estimation of approximate substring queries. Thus, we develop efficient train data generation algorithms by avoiding unnecessary computations and sharing common computations. We also propose a deep learning model as well as a novel learning method to quickly obtain an accurate deep learning-based estimator. Extensive experiments confirm the superiority of our data generation algorithms and deep learning model with the novel learning method.

Enquiries: +852 3411 2385 / comp@comp.hkbu.edu.hk

