MSc thesis defense presentation

Ioannis Psaros defends his MSc thesis

Date:	Thursday, 23 Apr 2015
Time:	13:00
Location:	Department of Informatics, University of Athens, A56
Thesis title:	Low-quality dimension reduction and high-dimensional Approximate Nearest Neighbor"
Committee:	• Ioannis Emiris • Stavros Kolliopoulos • Aristeidis T. Pagourtzis

---

Thesis abstract

The approximate nearest neighbor problem (ANN) in Euclidean settings is a fundamental question, which has been addressed by two main approaches: Data-dependent space parti- tioning techniques perform well when the dimension is relatively low, but are affected by the curse of dimensionality. On the other hand, locality sensitive hashing has polynomial dependence in the dimension, sublinear query time with an exponent inversely proportional to the error factor ■■, and subquadratic space requirement. We generalize the Johnson-Lindenstrauss lemma to define “low-quality” mappings to a Euclidean space of significantly lower dimension, such that they satisfy a requirement weaker than approximately preserving all distances or even preserving the nearest neighbor. This mapping guarantees, with arbitrarily high probability, that an approximate nearest neighbor lies among the ■■ approximate nearest neighbors in the projected space. This leads to a randomized tree based data structure that avoids the curse of dimensionality for (1 + ■■)-ANN. Our algorithm, given ■■ points in dimension ■■, achieves space usage in ■■(■■■■), preprocessing time in ■■(■■■■ log ■■), and query time in ■■(■■■■^■■ log ■■), where ■■ is proportional to 1 − 1/ln ln ■■, for fixed ■■ ∈ (0, 1). It employs a data structure, such as BBD-trees, that efficiently finds ■■ approximate nearest neighbors. The dimension reduction is larger if one assumes that pointsets possess some structure, namely bounded expansion rate.