

# 機械学習を用いた検索エッジ数の推定によるグラフベース近似最近傍探索の高速化 Speeding Up Graph-Based Approximate Nearest Neighbor Search by Estimating the Number of Search Edges Using Machine Learning

菅 寧  
Ning Jian

陸 可鏡  
Kejing Lu

石川 佳治  
Yoshiharu Ishikawa

杉浦 健人  
Kento Sugiura

## 1 はじめに

最近傍探索は幅広い分野で用いられる基礎的な情報処理技術である。この技術は与えられた問合せデータに対して、データベースから問合せデータに最も近いデータ（最近傍点）を見つける。しかし、データの次元が大きくなると「次元の呪い」という問題を引き起こすに加え、検索精度より検索速度が重要であるという要求が増えている。そのため、検索時間を大幅に削減できる近似最近傍探索が盛んに行われている。

近似最近傍探索の一種であるグラフベースの手法は検索時間が短くかつ検索精度も高いため、現在広く研究されている。グラフベースの探索の基本的なアイデアは、データセットから近似最近傍グラフを生成することである。問合せ点の最近傍点を検索するには近傍グラフ上の任意のノードを開始点とし、そのノードの近傍点、つまりエッジの張られている各ノードと問合せ点との距離を計算する。最も問合せ点に近いノードへ移動しこの処理を繰り返すことで、問合せ点の最近傍点となるノードを探す。

本研究ではグラフベースの近似最近傍探索を高速化するために、最先端の手法である hierarchical navigable small world (HNSW) グラフ [1] の改善点について検討する。特に、機械学習によるより高速な近似最近傍探索の実現を図る。

## 2 関連研究

本章では HNSW の基本的なアイデアを紹介する。

近似最近傍グラフ上での検索速度を向上させるために、Malkov らは navigable small world (NSW) グラフを提案した [2]。NSW はノードを 1 つずつグラフに追加し、ビームサーチによって現在のグラフ上で距離の近い  $k$  個のノードへエッジを張り近似最近傍グラフを生成する。つまり、後半に追加されるノードは近傍点へのエッジしか持たないが、早期に追加されたノードは遠くのノードへのエッジを持つ可能性がある。NSW ではそのような長いエッジを問合せ点への近道として利用できるため、検索速度が大幅に向上する。

近道となるエッジをより効果的に構築するために、Malkov らは NSW に階層構造を導入した HNSW を提案した [1]。HNSW はスキップリストのアイデアを参考にしており、1 つ下位の層に存在するノードが現在の層にも存在する確率が  $p$  となるようにグラフを構築する。つまり、最下層を第 0 層とし全ノード

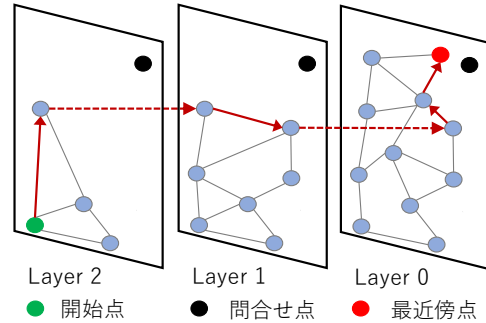


図 1: HNSW のグラフ構造

数を  $N$  とする場合、第  $i$  層にあるノードの数はおよそ  $p^i N$  であり上層になるほどノード数が減少する。したがって、図 1 に示すように上位層のグラフを検索のための近道として利用できる。検索時には最上層から順に問合せ点の最近傍点を見つけ、ある層の最近傍点を次の層の開始点として繰り返すことで、最下層における問合せ点の最近傍を検索する。

HNSW では以下の 3 つのパラメータが使用される。

1.  $M$ : 各ノードの近傍点の最大数。
2.  $efConstruction$ : グラフ作成のためのビームサイズ。
3.  $efSearch$ : 検索のためのビームサイズ。

既存研究では  $efConstruction$  と  $M$  の増加に伴いグラフの作成時間とメモリの使用量も増加するが、検索精度は高くなると報告されている。

## 3 機械学習による検索エッジ数の予測

HNSW を高速化するために、Li らは機械学習によって各問合せ点の終了条件を予測し、検索を早期に終了させる手法を提案した [3]。しかし、この方法は高い検索精度が要求される場合にのみ有効であり、検索精度が求められない場合は早期終了が上手く働かない。また、検索を早期に終了しても、検索中の不要な検索を回避できないという問題点もある。

そこで、本研究では終了条件の予測ではなく各ノードにおける最適な検索エッジ数の予測に着目する。HNSW はビームサーチでの検索時にノード全ての近傍点を走査するため、各ノードの近傍点の最大数  $M$  がそのまま検索エッジ数となる。 $M$  は検索精度と検索時間についてトレードオフの関係を持つため、精度と時間の両方において支配的な  $M$  は見つけられない。そのため、本研究では同じ検索時間で検索精度が最も高い  $M$  を最適値とする。しかし、検索の進行段階によって最適な

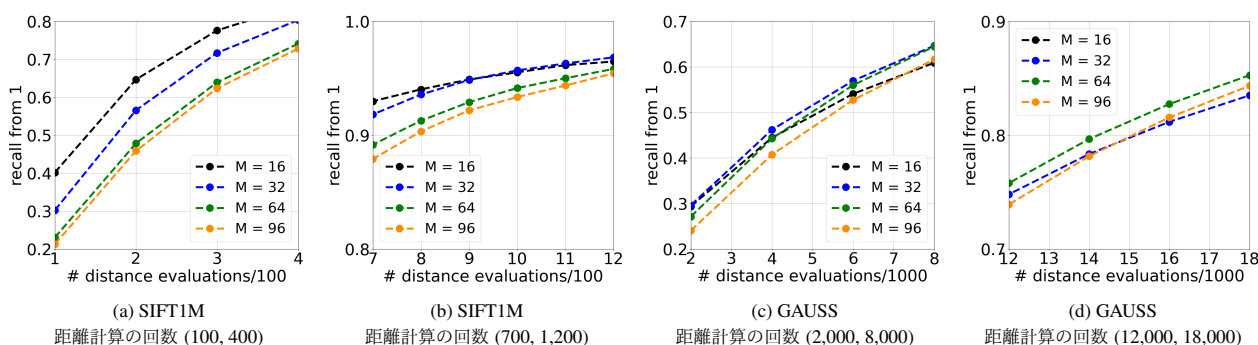


図 2: 探索の各段階（距離の計算回数）において最大エッジ数  $M$  が再現率に与える影響。

エッジ数は変化すると考えられるため、進行段階に応じて最適な値を予測することで検索における不要な距離計算の回避を目指す。なお、HNSW では近傍点全てを検索するためその検索順序は結果に影響を与えないが、提案手法により検索エッジ数を削減した場合は検索順序も重要となる。そのため、近傍点の検索順が固定されるように各ノードの近傍点はそのノードから近い順に格納する。

以下では、予備実験として  $M$  が検索精度に対して与える影響について紹介し、その結果から最適なエッジ数の予測に利用可能な特徴量について検討する。

### 3.1 パラメータ $M$ に対する予備実験

$M$  と検索精度の関係を検討するために、実データである SIFT1M [4] と正規分布に従う人工データ GAUSS を用いて予備実験を実施した。各データセットはいずれも 128 次元のデータを持ち、 $M \in \{16, 32, 64, 96\}$  と  $efConstruction = 1000$  を用いて HNSW グラフを構築した。構築した各グラフに対して近似最近傍探索を実施し、検索の各段階、つまり検索開始時からの距離計算回数の総数に応じた再現率を計測した。

図 2 に実験結果を示す。図 2 (a) と図 2 (c) は距離計算が少ない、つまり検索の初期段階の結果を示しており、 $M = 16$  の再現率が大きい。一方、図 2 (b) と図 2 (d) は検索の後期段階の結果であり、おおむね  $M = 64$  の再現率が大きくなっている。この結果から、検索の初期段階では  $M$  の値が小さいほど検索が効果的に行われるが、高い再現率を達成するには検索が進行するにつれて  $M$  の値を増やす必要があるとわかる。また、SIFT1M と GAUSS の結果の差から、同じ次元数を持つデータセットであっても最適な検索エッジ数が異なるとわかる。

### 3.2 機械学習のための特徴量の検討

以上の予備実験を基に、検索の進行段階に応じて適切な検索エッジ数を予測するための特徴量について検討する。既存手法である Li らの研究では検索中のノードから問合せ点までの距離に関する特徴量を利用して終了条件を予測する [3]。一方、本研究では後述する 3 つの特徴量を使用し、勾配ブースティング決定木を用いて検索エッジ数を予測する。

第一に、現在の最近傍点から問合せ点までの距離である。現在の最近傍点が問合せ点からまだ遠い場合、問合せ点への大まかな方向さえ正しければ距離を縮められる。つまり、確認するエッジ数が少なくても妥当な次の検索ノードが見つけられると

考えられる。一方、探索が進み問合せ点に近づくにつれて真の最近傍点を見つけるには微調整が必要となり、確認するエッジ数も増加すると考えられる。

第二に、データセット自体の intrinsic dimension [5] である。intrinsic dimension はデータセットの検索の難易度を測定するために使用される。例えば、100 次元持つデータセットが 2 つあるとき、次元削減により有効な次元が 80 次元残るものと 20 次元しか残らないものとはその検索の難易度は大きく異なる。つまり、検索難易度が高いほどより多くの検索エッジが必要になると考えられる。

最後に、現在の最近傍点周辺の点密度である。ある点の周囲の点の密度が高い場合、それらは同じ近傍点を多数共有している可能性が高い。したがって、検索エッジ数を減らし周囲の点の一部を検索するだけで次の最近傍点を発見できると考える。

## 4 まとめと今後の課題

本稿では、HNSW の概要を述べ、HNSW に対する改善点を分析した。また、検索エッジ数の予測に適した特徴量を検討した。今後の課題としては、本稿で述べた特徴量を利用して検索中に最適なエッジ数を実際に予測し、その性能を評価することが挙げられる。

## 謝辞

本研究は JSPS 科研費 JP20K19804, JP21H03555, JP22H03594 の助成を受けたものである。

## 参考文献

- [1] Y. A. Malkov and D. A. Yashunin, "Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs," *IEEE TPAMI*, vol. 42, no. 4, pp. 824–836, 2018.
- [2] Y. Malkov, A. Ponomarenko, A. Logvinov, and V. Krylov, "Approximate nearest neighbor algorithm based on navigable small world graphs," *Information Systems*, vol. 45, pp. 61–68, 01 2013.
- [3] C. Li, M. Zhang, D. G. Andersen, and Y. He, "Improving approximate nearest neighbor search through learned adaptive early termination," in *Proc. SIGMOD*, 2020.
- [4] SIFT1M: <http://corpus-texmex.irisa.fr/> (accessed: June 21, 2022).
- [5] L. Amsaleg, O. Chelly, T. Furon, S. Girard, M. E. Houle, K. Kawarabayashi, and M. Nett, "Estimating local intrinsic dimensionality," in *Proc. SIGKDD*, pp. 29–38, 2015.