敵対的な攻撃に対するランダムウォークの ロバスト性に関する一検討

0730 椎名 智

関西学院大学 理工学部 情報科学科 大崎研究室
2023 年度 卒業実験および演習 中間審査資料

1 はじめに

近年、未知のグラフにおける目的ノードの探索手法としてランダムウォークが広く用いられている。BFS (Breadth-First Search) や DFS (Depth-First Search) のような決定的なアルゴリズムとは異なり、ランダムウォークに基づくノード探索は確率的なアルゴリズムであることから, 探索を妨害する敵対的な攻撃に対しても比較的堅牢であると期待される。

近年、ランダムウォークに対する敵対的な攻撃として、グラフのトポロジを動的に切り替えるという攻撃に対する特性の分析が行われつつある [1]。このような攻撃手法により、ランダムウォークの探査効率が大幅に低下する可能性があることが示されている。そのため、より現実的な攻撃に対するランダムウォークによるノード探索の堅牢性を明らかにすることも求められる。

そこで本稿では、ランダムウォークに基づくノード探索が、攻撃者によるリンクの張り替え攻撃に対してどの程度堅牢であるかを定量的に分析することを目的とする。攻撃手法や攻撃頻度、グラフのトポロジなどの要因が、ランダムウォークの探査効率に与える影響を実験により調査する。

2 ランダムウォーク攻撃問題

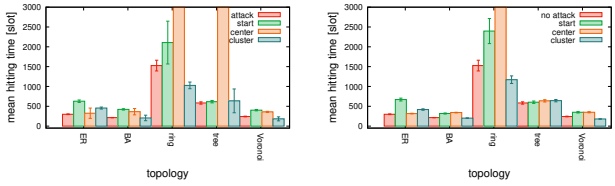
本稿で扱うランダムウォーク攻撃問題は、攻撃者が、ランダムウォークによるグラフ $G = (V, E)$ 上の目的ノード探索をできるだけ妨害する (遅延させる) ために、グラフ上のリンクを張り替えるというものである。

攻撃者は、エージェントの各ステップごとの移動を観測することが可能であり、一定の頻度でリンクの張り替えを行う。時刻 0 から時刻 k までにエージェントが訪問したノードの系列を $P(k) = \{v_0, v_1, ..., v_k\}$、攻撃者によるリンク張り替え攻撃の頻度を A とする。攻撃者は、エージェントの移動開始から A ステップ移動ごとに 1 回、グラフ $G = (V, E)$ 上の任意のリンクを張り替える。ただし、攻撃の事実が検出されることを避けるため、グラフが非連結となるようなリンクの張り替えは行わないものとする。

3 ランダムウォーク攻撃手法

以下では、時刻 k において訪問済みのノード集合からなる部分グラフを S_k とする。本稿では攻撃手法として、現在訪問しているノード v_k に接続されているリンクの 1 つを、 (1) S_k の中心ノードへと張り替える方法 (中心法)、 (2) S_k 中のクラスタの 1 つへ張り替える方法 (クラスタ法)、 (3) エージェントの始点ノード v_0 にできるだけ近いノード へと張り替える方法 (始点法) を対象とする。いずれの手法においても、ノード v_k に接続されているリンク $(v_k, u) \in E(u \notin P(k))$ のうち、ノード v_{k-1} からノード u までのホップ数が最大のリンクを選択する。3 つの手法ともに、リンク (v_k, u) をリンク (v_k, u') へと張り替えるが、張り替え先のノード u' の決定法がそれぞれ異なる。

- 中心法：中心法では、S_k において次数中心性が最大のノードを張り替え先のノード u' として選択する。つまり、ノード v の次数を $d(v)$ とすると、$\mathrm{argmax}_{u' \in P(k),(v_k,u') \notin E}\, d(u')$ となるノード u' を選択する。
- クラスタ：クラスタ法では、S_k 中の最大クラスタを構成するノードの 1 つを張り替え先のノード u' として選択する。つまり、時刻 k において ノード $v \in P(k)$ が属するクラスタを $C_k(v)$ とすると、$\mathrm{argmax}_{u' \in P(k),(v_k,u') \notin E} |C_k(u')|$ となるノード u' を選択する。
- 始点法： 始点法では、エージェントが初期に訪問したノードを張り替え先のノード u' として選択する。つまり、時刻 k におけるノード $v \in P(k)$ の初回訪問時刻を $H_k(v)$ とすると、$\mathrm{argmin}_{u' \in P(k),(v_k,u') \notin E} H_k(u')$ となるノード u' を選択する。



(a) 攻撃頻度 A=50　　(b) 攻撃頻度 A=100

図 1　3 種類の攻撃手法を用いたときの各トポロジにおける探査エージェントの平均初回到着時間

4 実験

5 種類のグラフにおいて、3 種類の攻撃によって単純ランダムウォーク (SRW) の初回到着時間がそれぞれどの程度変化するかをシミュレーション実験により調査する。

トポロジ構造の異なるグラフを生成するため、5 種類のネットワーク生成モデルを用いて 100 ノードの ER (Erdõs-Rényi) グラフ、BA (Barabási–Albert) グラフ、ツリー、リング、ボロノイグラフを生成した。生成したグラフにおいて、エージェントの移動開始から $A = 50$、100 ステップごとに 1 回、リンク張り替え攻撃をした時の、ランダムに選択した始点ノードから、ランダムに選択した目的ノードに到達するまでのステップ数 (初回到着時間) を計測した。同一条件下でのシミュレーションを 100 回試行することで、初回到着時間の平均および 95%信頼区間を計測した。

$A = 50$、100 ステップごとに 1 回攻撃した時の平均初回到着時間を図 1 に示す。これらの結果より、グラフのトポロジによっては、攻撃によって平均初回到着時間が数倍程度に増加することがわかる。

5 今後の課題

今後の課題として、ランダムウォークとそれに対する攻撃が、敵対的学習で設計することができるのかの実験や、複数の攻撃種別の中でもっとも困る攻撃を見つけるなどが挙げられる。

参考文献

[1] O. Denysyuk and L. Rodrigues, “Random Walks on Evolving Graphs with Recurring Topologies,” in *Proceedings of 28th International Symposium on DISTributed Computing (DISC 14)*, pp. 333–345, Oct. 2014.