

D-003

ユーザの位置情報を考慮した 領域内の影響最大化に対する効率的なアプローチ

勝田 健斗[†] 石川 佳治[†] 杉浦 健人[†][†] 名古屋大学大学院情報学研究科

1 はじめに

近年、バイラルマーケティングと呼ばれるソーシャルネットワーク上の口コミを利用したマーケティング戦略が普及している。企業は、ソーシャルネットワーク上の少数のユーザに対して情報の伝達や試供品の提供を行い、情報の発信や商品の宣伝を依頼する。依頼先のユーザを発信源とする口コミ効果により、間接的に大多数のユーザに対して商品やサービスをプロモーションすることが可能である。企業は少数のユーザに対する情報発信の依頼により大多数の人数に宣伝可能なため、費用対効果も高い。ここで重要な問題が、口コミによる情報拡散が大きくなるように発信源として依頼すべき影響力のあるユーザの特定である。

バイラルマーケティングを背景としたソーシャルネットワークにおける影響力のあるユーザの特定は影響最大化問題として多く研究されている。既存研究における影響最大化問題 [2, 5] は主にソーシャルネットワークのグラフとユーザ間の伝搬確率を入力とし、より多くのユーザに情報を伝搬できる k 人のシード (口コミを依頼するユーザ) を選択する最適化問題として定義されている。また、ユーザの位置情報を考慮し、ソーシャルネットワーク上の特定領域内のユーザに対して影響が最大になるようにシードを特定する既存研究 [3] も存在する。これは企業側がイベントや商品を宣伝したい場合、マーケティングのターゲット領域を限定できる点で実用的である。しかし、既存研究では時間経過によるユーザの位置変化を考慮しておらず、領域内のユーザが変化した場合に効率的にシードを再特定する手法は議論されていない。また、ユーザの過去の位置情報よりも直近の位置情報の方が重要であるため、位置情報が生成された時間も考慮すべきである。そこで、本論文ではソーシャルネットワークにおけるユーザのチェックイン情報を考慮した新しい領域内の影響最大化問題を提案し、さらにユーザの変化に対して効率的にシードを更新する手法を検討する。

2 問題定義

2.1 領域内のチェックインユーザに対する影響最大化

バイラルマーケティングにおいて、ソーシャルネットワークの全ユーザをターゲットにして影響を与えた

い場合だけでなく、ある領域内にいる人々といった特定のユーザのみをターゲットにして影響を与えたい場合もある。例として、観光場所を訪れた観光客に観光協会がイベント情報などを宣伝したいといったことが挙げられる。観光場所では訪れている人々は常に変化しているため、影響を与えたいターゲットも常に変化している。そのため、ソーシャルネットワークの特定の人々のみをターゲットにした影響最大化に加えて、ターゲットのユーザが変化した場合に効率的に影響最大化問題を解くことも重要である。

本論文ではユーザのチェックイン情報を考慮し、ある時刻 T から直近 W 時間以内に領域 R でチェックインしたユーザに対して、最大の影響を与える k 人のシードを選択する領域内の影響最大化問題を提案する。また、時刻の変化に対して効率的にシードを保持する手法を提案する。図1は時間の変化とともに影響を与えるターゲットユーザが変化することを示す。

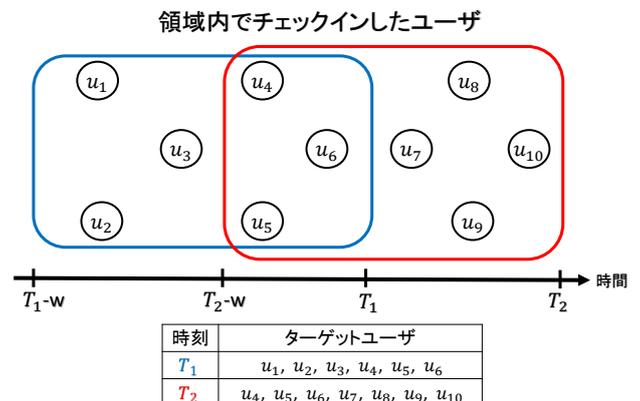


図1: チェックインに対するターゲットユーザの変化

2.2 影響の伝播について

本論文ではソーシャルネットワークにおけるユーザ間の影響伝播はICモデル (Independent Cascade) [2] を適用する。ICモデルではソーシャルネットワーク $G = (V, E)$ (V : 頂点の集合, E : 辺の集合) の各辺 (u, v) はパラメータとして伝搬確率 p_{uv} ($0 < p_{uv} < 1$) を持ち、ユーザ u が影響を受け active 状態になった場合、ユーザ u は p_{uv} の確率でユーザ v に影響を与える。さらにユーザ v が active 状態になった場合、ユーザ v と結ばれているユーザに対して影響を与えようとする。以下、active 状態となるユーザがいなくなるまで影響は伝搬する。本論文ではICモデルにおける影響の伝

Efficient Approach to the Influence Maximization in the Area Considering User's Positions

Kento Katsuda[†], Yoshiharu Ishikawa[†], Kento Sugiura[†]

[†] Graduate School of Informatics, Nagoya University

播をターゲットユーザからの逆伝搬として捉え、影響力のあるユーザを特定する。例として、図2はユーザ u とユーザ v の関係を表すグラフである。ICモデルではユーザ u が影響元となり、 p_{uv} の伝搬確率でユーザ v を active 状態にする。本論文ではユーザ v を領域内のターゲットユーザとし、 p_{uv} の確率でユーザ u がユーザ v の影響元になると考える。

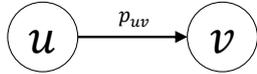


図2: ユーザ u とユーザ v の関係を表すグラフ

2.3 影響最大化の定式化

ユーザのチェックイン集合を $C = \{c_1, c_2, \dots\}$ で表す。各チェックインは $c = (user, time, location)$ であり、ユーザ、チェックイン時間、チェックイン場所の3つの情報を保持する。時刻 t から直近 W 時間以内に領域内でチェックインしたユーザを $U_{R,W}(t) = \{c.user \mid t - W \leq c.time \leq t \wedge c.location \in R\}$ と定義する。またICモデルの下でユーザ v から逆伝搬したとき、ユーザ v の影響元となるユーザ集合を $I^-(v)$ とする。このとき、ユーザ u が影響を及ぼす領域 R 中のユーザの集合 $I_{R,w}(u, t)$ は式(1)のように定義できる。

$$I_{R,W}(u, t) = \{v \mid v \in U_{R,t-W} \wedge u \in I^-(v)\} \quad (1)$$

本論文では影響を与えるユーザ数を最大化することを影響最大化として捉え、式(2)の値が最大になるように k 以下のシード集合 S を見つける。

$$f(S, t) = |\cup_{u \in S} I_{R,W}(u, t)| \quad (2)$$

3 提案手法

式(2)の関数 $f(\cdot)$ は劣モジュラ関数であり、劣モジュラ関数の最大化は一般的にNP困難である[2]。 $f(S \cup u) - f(S)$ の値が最大になるように u をシードとして貪欲的に選択していく手法は、近似比 $1 - 1/e$ を達成する[4]。しかし、時間の経過ごとに1から貪欲的にシードを選択していくことは非効率である。本論文では既存研究[6]のCheckpointの概念を導入し、SSM(Streaming Submodular Maximization)アルゴリズム[1]を用いてシードを選択することにより、ターゲットユーザの変化に対しても効率的にシードを更新する。Checkpointは一部のターゲットユーザに対する部分分解であり、SSMアルゴリズムは要素の追加に対する劣モジュラ関数最大化のストリーミングアルゴリズムである。本論文で考慮すべき時間経過によるターゲットユーザの変化に対して、1ステップ前の時刻における部分分解を用いて

新しいユーザの追加処理(SSMアルゴリズム)を実行することでシードを更新する。

図3はCheckpointを用いたシード更新の例を示す。時刻 T_1 において、ターゲットユーザが u_1, u_2, \dots, u_6 の場合、ターゲットユーザの部分集合に対してCheckpointと呼ばれる部分分解を保持する。また時刻 t_2 において、ターゲットユーザの一部は時刻 t_1 と変化がないため、時刻 t_1 と時刻 t_2 で重複するユーザのCheckpointに対して、時刻 t_1 から時刻 t_2 までに新しくチェックインしたユーザのみの追加処理(SSMアルゴリズム)を実行することでシードを更新する。

T_1		T_2	
ターゲットユーザ (時系列順)	Checkpoint (部分分解)	ターゲットユーザ (時系列順)	Checkpoint (部分分解)
$u_1, u_2, u_3, u_4, u_5, u_6$	S_1^1	$u_4, u_5, u_6, u_7,$ u_8, u_9, u_{10}	S_1^2
u_2, u_3, u_4, u_5, u_6	S_2^1	$u_5, u_6, u_7, u_8, u_9, u_{10}$	S_2^2
u_3, u_4, u_5, u_6	S_3^1
u_4, u_5, u_6	S_4^1
u_5, u_6	S_5^1
...

S_4^1 に対して u_7, u_8, u_9, u_{10} の追加処理を行い、 S_1^2 を計算

図3: Checkpointを用いたシードの更新

4 まとめと今後の課題

本論文では、時間経過及びユーザの位置変化を考慮した特定領域内における影響最大化問題を提案した。また、領域内の影響最大化問題においてユーザの変化に対して効率的にシードを更新するためにCheckpointの概念を導入し、SSMアルゴリズムを適用した。今後の課題としては提案手法の実装及び実験による有効性の評価が挙げられる。

謝辞

本研究の一部は科研費(16H01722)による。

参考文献

- [1] A. Badanidiyuru, B. Mirzasoleiman, A. Karbasi, and A. Krause. Streaming submodular maximization: Massive data summarization on the fly. In *KDD*, pp. 671–680, 2014.
- [2] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD*, pp. 137–146, 2003.
- [3] G. Li, S. Chen, J. Feng, K.-l. Tan, and W.-s. Li. Efficient location-aware influence maximization. In *SIGMOD*, pp. 87–98, 2014.
- [4] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14(1):265–294, 1978.
- [5] N. Ohsaka, T. Akiba, Y. Yoshida, and K.-i. Kawarabayashi. Dynamic influence analysis in evolving networks. *PVLDB*, 9(12):1077–1088, Aug. 2016.
- [6] Y. Wang, Q. Fan, Y. Li, and K.-L. Tan. Real-time influence maximization on dynamic social streams. *PVLDB*, 10(7):805–816, 2017.