

グラフマイニングとテキストマイニング技術の統合によるソーシャルセンサからの実社会イベント予測手法の開発

研究代表者

津川 翔

筑波大学 システム情報系 助教

1 はじめに

電気通信技術の発展により、多くの人が日常的に様々なネットワークサービスを利用している。本研究では、ネットワークサービスのユーザを、実社会を観測するセンサであるソーシャルセンサとして捉える。ソーシャルセンサから得られる様々な情報、すなわちユーザの検索履歴や Web 閲覧の履歴、ソーシャルメディアにおける投稿などの情報は、実社会におけるトレンドを反映していると考えられている。そのため、ソーシャルセンサから得られる情報を用いて、病気の流行や経済状況の変動などの実社会におけるイベントを予測する技術の実現に期待が高まっている [1-6]。

ソーシャルセンサを用いて実社会イベントを予測する試み自体は既に国内外の学会において報告されている。これまでに、Web 検索の履歴を用いた景気動向の予測 [2] やインフルエンザの流行の予測 [6]、選挙結果の予測 [7] などが試みられている。このような技術によって、例えば、病気の流行を予測することができれば、国や自治体がワクチンの備蓄や伝染病の対策などを行う上で有用である。ただし、このような研究分野は未だ発展途上の段階であるため、現在も研究開発が進められている。

本研究は、従来研究を発展させ、ソーシャルセンサから実社会イベントを予測する技術を確立することを目指し、研究期間内において、特に以下の 2 つの研究課題に取り組んだ。

- A) 検索行動量を用いて自動車販売台数を予測するモデルの研究
- B) ソーシャルメディアの投稿からノイズを除去するためのユーザの影響力推定手法の研究

研究課題 (A) においては、Google Trends という Web 検索ユーザの行動を反映したデータと、Wikipedia のページの閲覧数のデータを用いて、自動車の販売台数を予測するという問題に取り組んだ。自動車の過去の販売台数とその車に関する検索行動量 (Google Trends 値と Wikipedia のページ閲覧数) を用いて、将来の販売台数を予測するモデルを構築した。さらに、実験によって予測の精度を評価した。その結果、検索行動量が自動車の販売台数を予測する上で有用な情報であることを示した。

一方、研究課題 (B) では、ソーシャルメディアにおけるユーザの投稿から実社会イベントを予測する際にノイズとなる投稿を除去するための技術の研究を行った。研究代表者はこれまで、Twitter における投稿のテキストデータから、車の販売台数を予測する研究を行っている [8]。この研究において、予測精度を向上させる上で、ソーシャルメディアの投稿の中に実社会イベントを予測する上で有用でないノイズとなる投稿が多数存在することが問題となっていた。そこで、本課題では、グラフマイニングの技術を用いて、ソーシャルメディアにおけるユーザの中から、予測に有用な投稿を行うと考えられる重要なユーザを推定する手法の研究を行った。従来のグラフマイニングの技術を用いて、重要なユーザを推定するためには、ソーシャルメディアのユーザ間の関係を表現したソーシャルネットワーク全体の構造が必要であった。しかしながら、ソーシャルメディアにおいては、ソーシャルネットワークの規模が非常に大きいため、ネットワーク全体の構造を得ることは容易ではない。そこで、本課題では、部分的なソーシャルネットワークの構造のみから重要なユーザを推定するための手法を提案した。

以降、2 章では研究課題 (A) について、3 章では、研究課題 (B) について報告する。4 章において全体のまとめと今後の課題について述べる。

2 検索行動量を用いた自動車販売台数の予測

2-1 研究の背景と目的

本研究では、Google Trends という Web 検索ユーザの行動を反映したデータと、Wikipedia のページの閲覧数のデータを用いて、自動車の販売台数を予測するという問題に取り組む。自動車は高額商品であるため、自動車の購入に先だって、多くの消費者が Web 検索を行うと考えられる。そのため、検索行動量は自動車の販売台数を予測するのに有用な情報ではないかと期待される。

まず、実際の車種ごとの自動車販売台数と、検索行動量を表す Google Trends 値との関係、およびその自動車の Wikipedia ページの閲覧数との関係を明らかにする。次に、過去の自動車販売台数、Google Trends 値および Wikipedia ページ閲覧数から、将来の自動車の販売台数を予測するモデルを構築し、その有効性を検証する。

2-2 自動車販売台数と検索行動量との関係

自動車販売台数と、検索行動量との関係を明らかにするため、以下の手順によって、データを取得した。各月の自動車販売台数については、日本自動車販売協会連合会（普通及び小型乗用車）及び全国軽自動車協会連合会（軽自動車）が公開している毎月の新車販売台数データを利用した。このうち、2015年2月時点でも発売されており、かつ上記の期間内で販売台数が比較的多い22車種を分析の対象とする。また、Google Trends については、車名にメーカー名を追加したキーワード（例えば「トヨタプリウス」など）に対する人気度値（Interest）を利用した。Wikipedia 閲覧数については、日本語 Wikipedia の該当する項目（例えば「トヨタ・プリウス」など）に対する毎月の閲覧数を Wikimedia 閲覧数統計データから集計した。

データの得られた22車種について、それぞれ対応する Google Trends 値及び Wikipedia 閲覧数と販売台数の間の相関係数を求めた。ここで、検索行動は購買行動に先立って行われることを考慮し、各検索行動量の系列を1ヶ月及び2ヶ月先行させた場合についても計算を行った。

相関係数を計算した結果、Google Trends の場合は7車種、Wikipedia 閲覧数の場合は6車種について、相関係数が0.4以上であった。更に、Google Trends を1、2ヶ月先行させた場合はそれぞれ10、6車種、Wikipedia 閲覧数の場合は4、2車種について相関係数が0.4以上であった。これらの車種については販売台数と検索行動量との相関がある程度高いことから、販売台数の予測に検索行動量を用いることで予測精度を向上させられる可能性がある。

また、先行させる期間に注目すると、多くの車種について Google Trends に対しては1ヶ月先行させた場合、Wikipedia 閲覧数に対しては先行させない場合が最も相関が高いが、一部の車種については異なる期間の場合の方が高い相関を持つ。このため、予測モデルでは予め定めた期間だけ先行させるのではなく、複数の先行期間について最良の場合を判断できることが望ましいと考えられる。

具体的な例として、ホンダ・フリードの販売台数と各検索行動量の推移の比較を図1に示す。この例からも Google Trends 値および Wikipedia 閲覧数と販売台数の間に相関が存在することが分かる。

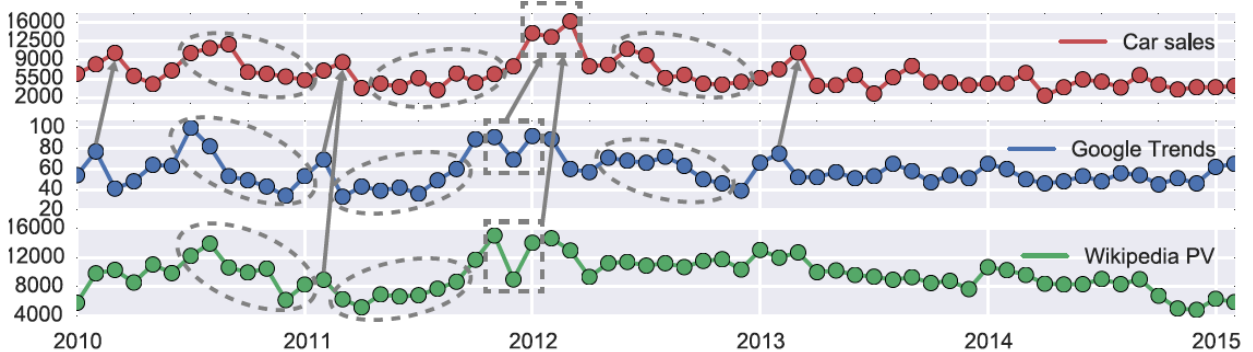


図1: ホンダ・フリードの販売台数、Google Trends 値、Wikipedia 閲覧数の推移の比較 [9]

2-3 自動車販売台数の予測モデル

状態空間モデルを用いて、自動車販売台数を予測するモデルを構築する。状態空間モデルは時系列変化する変数の予測に広く用いられるモデルである。状態空間モデルを用いて、自動車販売台数を予測する単純な方法として、過去の自動車販売台数の時系列データのみから、将来の販売台数を予測するという方法が存在する。本研究では、このような予測モデルをベースラインモデルと呼ぶ。

本研究では、過去の自動車販売台数の時系列データに加えて、検索行動量の時系列データを用いた予測モデルを提案する。提案するモデルでは、検索行動量の時系列から、トレンド成分を抽出し、抽出した検索行動量のトレンドを変数としてベースラインモデルに加える。提案モデルの詳細については、文献 [9] を参照されたい。

2-4 実験

提案するモデルを用いて予測した自動車販売台数と、実際の自動車販売台数がどの程度一致するかを調査することにより、提案するモデルの有効性を検証する。予測精度を評価するテスト期間は 2014 年 1 月～2014 年 12 月の 1 年間とした。モデルのパラメータを推定する学習期間は、2013 年 12 月以前の期間で学習期間の長さを変化させた。評価指標には、実際の自動車販売台数と予測した自動車販売台数の RMS (Root Mean Square; 誤差の二乗和の平均) を用いた。但し、販売台数は車種によって大きく異なることから、誤差には真の値に対する予測値の比率に基づく相対的な値を用いた。なお、RMS は低いほど予測精度が高いことを意味する。

図 2 に学習期間の長さを変化させた時の予測精度を表す RMS の値を示す。図中の Base はベースラインモデルの、GT は Google Trends を用いたモデルの、WP は Wikipedia ページ閲覧数を用いたモデルの RMS 値を示している。この結果より、提案する検索行動量を用いたモデルがベースラインモデルよりも低い RMS 値を示していることが分かる。この結果は、検索行動量は自動車販売台数の推定に有用であることを示唆している。また、予測に用いる学習期間の長さに着目すると、おおよそ 24 ヶ月程度の期間を学習期間とすれば十分であり、それ以上学習期間を長くしても予測精度はそれほど向上しないことも分かる。

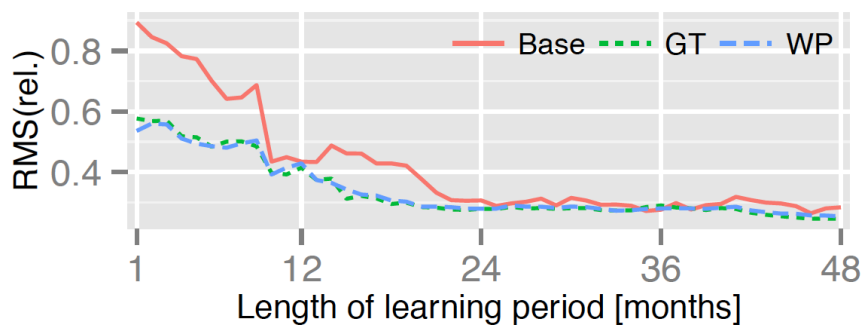


図 2: 学習期間の長さ と 推定精度を表す RMS との関係 [10]

2-5 まとめ

本研究課題では、ソーシャルセンサから得られる情報として検索行動量に着目し、検索行動量を用いて自動車販売台数を予測するモデルを提案した。さらに本研究課題では、実験により提案するモデルの有効性を検証した。実験の結果、検索行動量は自動車販売台数の推定に有効であることを示した。

3 ソーシャルメディアにおけるユーザの影響力推定

3-1 研究の背景と目的

ソーシャルメディアの流行により、ソーシャルメディアにおけるユーザの影響力を推定する指標の研究が活発に行われている [11, 12]。このような指標は、実社会イベントを予測する際に有用な情報を投稿するユーザを推定するのに有用ではないかと期待している。多くの研究では、ユーザの影響力を推定するために、ユ

ユーザ間の関係を表現したソーシャルネットワークの構造を用いる。ソーシャルメディアにおけるユーザ間の関係は、友人関係やフォロー関係と呼ばれるユーザ間のつながりをリンクとするネットワークによって表すことができる。このようなネットワークの構造をユーザの影響力推定に用いる。具体的には、ソーシャルネットワークにおける各ノード（ソーシャルメディアにおけるユーザに相当する）の中心性（次数中心性、近接中心性、媒介中心性） [13]、やPageRank [14]、k-core 指標 [15]などを求めることにより、影響力を推定する。

これまでの研究では、ソーシャルネットワーク全体の構造からユーザの影響力をどの程度推定できるかが評価されている。ただし、現実にはソーシャルメディアにおけるユーザ間の関係を正確に取得することは容易ではない。ユーザ間の関係を表現したソーシャルネットワークには通常、データの欠損や誤り、ノードやリンクのサンプリングなどに起因して様々な誤りが含まれている [16]。しかし、このようなソーシャルネットワークの不完全性が、ユーザの影響力推定にどのような影響を与えるかは明らかにされていない。

そこで本研究では、ソーシャルネットワークにおける不完全性の中でも特に、ソーシャルネットワークからのノードのサンプリングが、ノードの影響力推定指標の有効性にどのような影響を与えるか調査する。Twitter や Facebook に代表されるソーシャルメディアのネットワークは非常に大規模であり、一般的にサンプリングによってしかその構造を得ることができない。本研究ではこのような状況を想定し、様々なサンプリング手法によって得られたソーシャルネットワークから、ユーザの影響力をどの程度推定できるか調査する。

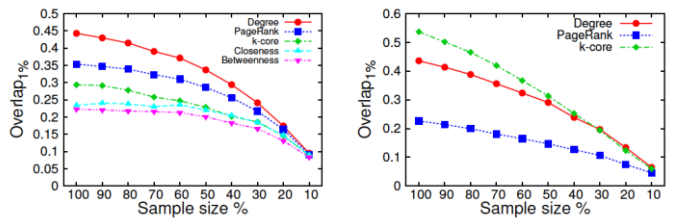
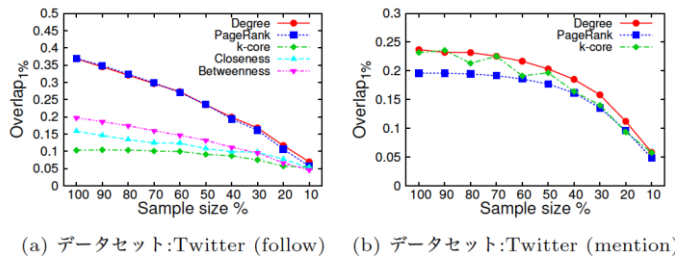


図 3: 各データセットにおいて、ランダムにサンプリングしたネットワークから推定したランキングと情報拡散数に基づくランキングの Overlap1% [18]

3-2 データを用いた評価

サンプリングによって得られた不完全なソーシャルネットワークから推定したユーザの影響力と、各ユーザの実際の影響力の関係を調査する。情報の欠落がないソーシャルネットワークからノードをサンプリングすることにより、不完全なソーシャルネットワークを得る。得られた不完全なソーシャルネットワークから、次数中心性・近接中心性・媒介中心性・PageRank・k-core 指標を用いて、各ユーザの影響力を推定する。

不完全なソーシャルネットワークにおいて推定した各ユーザの影響力と、各ユーザの実際の影響力に相当する情報拡散数を比較する。特に影響力の強いユーザをどの程度推定できるかに着目して評価を行う。推定したユーザの影響力と、実際の情報拡散数それぞれに基づくユーザランキングを作成し、ランキング上位のユーザの一致度を求める。ランキングの一致度を測る指標として、Overlap1% [16]を用いる。これは、不完全なネットワークから推定したランキングと実際の情報拡散数に基づくランキングの上位 1% のユーザの一致率である。

本実験では、我々の収集した Twitter (follow) ならびに公開データセットである、Twitter (mention)、Facebook、API Journals の 4 種類のデータセットを用いる。

本課題では、代表的なネットワークのサンプリング手法である BFS (Breadth First Search : 幅優先探索)、DFS (Depth First Search : 深さ優先探索)、SEC (Sample Edge Count) [17]、およびランダムサンプリングを用いる。なお、BFS、DFS、SEC ではランダムサンプリングと比較して、次数の高いノードがサンプリングされることが知られている [17]。影響力の強いノードは次数が高い場合が多いため、これらのサンプリング手法を用いることで、サンプリングするノードの割合が小さい場合でもサンプリングされた部分ネットワークから影響力の強いユーザを推定できると予想される。本実験では、各サンプリング手法を用いた場合、サンプリングするノードの割合がユーザの影響力を推定する指標にどのような影響を与えるかを明らかにする。

まず、ランダムサンプリングがノードの影響力を推定する指標にどのような影響を与えるかを調査した。サンプリングされたソーシャルネットワークを用いてノードの影響力のランキングを作成し、それを現実の情報拡散数に基づくノードランキングと比較した時の Overlap1% を求めた。各データセットにおけるサンプリングするノードの割合と Overlap1% の関係を図 3 に示す。ただし、計算量の問題で Twitter (mention) と APS Journals のデータセットでは近接中心性、媒介中心性の計算を省略した。

この結果から、どのデータセットにおいても、またどの影響力を推定する指標を用いた場合でも、サンプリングするノードの割合に対して、Overlap1% が徐々に低下していることがわかる。当然ながらランダムにノードをサンプリングする場合、サンプリングされた部分ネットワークに影響力の強いノードが含まれる確率は、サンプリングするノードの割合に対して線形に低下する。そのため、Overlap1% の値が、サンプリングするノードの割合に対して徐々に低下するのは自然な結果である。サンプリングするノードの割合が 10% の時に着目すると、どのデータセットにおいても、またどの指標を用いた場合においても、Overlap1% の値は 0.05 程度である。ネットワークの構造を用いずに、ランダムにネットワーク中からノードを 1% 選んだ時の Overlap1% の値は、0.01 であるため、この結果はランダムサンプリングでネットワークの情報の 10% 程度しか得られない場合には、ネットワークの構造からノードの影響力を推定する効果が非常に小さくなってしまっていることを意味している。

次に、SEC 法を用いてネットワークからノードをサンプリングした場合の結果を図 4 に示す。ランダムサンプリングと比較すると、どの指標を用いた場合でも、サンプリングの割合が 10% 程度であっても、Overlap1% の値が、ネットワーク全体の構造を用いた場合と同程度、もしくはそれ以上であることが分かる。これは、SEC 法が貪欲的に次数の期待値が高いノードをサンプリングする手法であるため、影響力の強いノードをサンプリングできていることが原因であると考えられる。また、サンプリングされたノードの次数が高いということは、元のネットワークに含まれるリンクのうち、サンプリングによって得られた部分ネットワークに含まれるリンクの割合が高いことを意味している。そのため、SEC 法によって得られた部分ネットワークは、ランダムサンプリングによって得られた部分ネットワークと比べて、元のネットワークの構造的な特徴を有していると考えられ、その結果として、Overlap が高い値になっているのではないかと考えられる。なお、BFS、DFS 法を用いてノードをサンプリングした場合にも、同様の結果が得られた。

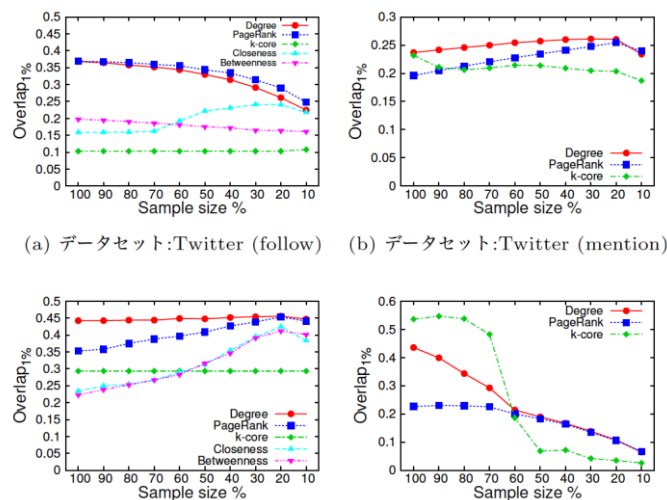


図 4: 各データセットにおいて、SEC 法によりサンプリングしたネットワークから推定したランキングと情報拡散数に基づくランキングの Overlap1% [18]

これらの結果は、ソーシャルメディアにおけるユーザの影響力推定において、ソーシャルネットワークがサンプリングによってしか得られない場合は、ランダムサンプリングではなく、SEC、BFS、DFS 法などのサンプリング手法を用いた方が影響力の強いユーザを推定できることを示している。ただし、APS Journals の結果を見ると、ランダムサンプリングの場合に比べ、SEC 法を用いた場合の方が、サンプリングの割合の低下に伴う Overlap の低下が早いことが分かる。この原因を明らかにするためにはさらなる調査が必要であるが、研究者のソーシャルネットワークは影響力の強いノードが広範囲に分散しているのに対して、BFS、DFS、SEC がネットワーク中の特定の狭い範囲をサンプリングしてしまうことが原因であると考えている。しかしながら、Twitter と Facebook のようなソーシャルメディアにおけるネットワークでは同様の結果であるため、本研究で対象としている影響力の強いノードの抽出には SEC サンプリングが有効であると考えられる。

3-3 モデルを用いた評価

さらに本研究では、サンプリングされた部分ネットワークの構造のみからどの程度影響力の強いノードが推定できるかを明らかにするため、影響伝搬のモデルを用いた評価を行う。

まず、部分的なネットワークから影響力の強いノードを特定する問題を未知のグラフに対する影響最大化問題として定式化する。従来の研究では、グラフの構造が既知であるという条件の下で、影響伝搬を最大化するシードノードの集合を選択するという問題が考えられているが、本研究では、グラフの構造が未知であるという条件の下で、同様の問題を考える。初期状態においては、グラフに関するノード数のみが分かっている状況を想定する。ただし、グラフの探査を行うことで部分的なグラフの構造を把握することが可能であり、部分的な構造の情報からシードノードを選択できるものとする。ここでのグラフの探査は、ソーシャルメディアが提供する Application Programming Interface (API) を利用してユーザの友人関係のリストを得るようなことを想定している。さらに、複数回の影響伝搬が可能であり、過去の影響伝搬の結果も、シードノードの選択に利用できるものとする。

以降では、具体的に問題の定式化を行う。本研究では重みなし無向グラフ $G = (V, E)$ 上の影響伝搬を考えるが、初期状態においては、 V のみが既知であり、グラフ G の構造、すなわち E は未知であるとする。グラフ G における各ノードはソーシャルメディアのユーザを、各リンクはユーザ間の関係をモデル化している。 R 回のラウンドが存在し、各ラウンドにおいて、自由に選択した単一のノードの探査と、自由に選択した単一のノードからの影響伝搬が可能であるとする。ノード v の探査により、ノード v の隣接ノードのリストを取得できる。また、シードノードからの影響伝搬は IC (Independent Cascade) モデル[19] に従うとする。IC モデルでは、シードノードは、確率 p (影響伝搬確率) で各隣接ノードに影響を伝搬させる。隣接ノードに影響が伝搬した場合、そのノードは「影響を受けた」状態となる。影響を受けたノードはさらにその隣接ノードの中で影響を受けていないノードに対して再帰的に確率 p で影響を伝搬させる。ただし、IC モデルにおける影響伝搬確率は未知であるとする。シードノードからの影響伝搬の結果として、影響を受けたノードのリストが得られるものとする。各ラウンドにおいて単一の探査と単一のシードノードからの影響伝搬が可能であるため、あわせて R 回の探査と R 回の影響伝搬が可能であるものとする。ただし各ラウンドの影響伝搬は独立であり、過去のラウンドで影響を受けたノードが再びそのラウンドで影響を受けることもあり得る。つまり、各ラウンドでは、各ノードが過去のラウンドで影響を受けたかどうかとは独立に、隣接ノードに対して確率的に影響伝搬を行うものとする。

R 回のラウンド中に影響を受けた(重複を除く)ノード数の合計を最大化することを目的とし、ラウンドごとに、探査するノードとシードノードを選択する。各ラウンドにおける探査ノードおよびシードノードの選択には、過去のラウンドの探査と影響伝搬の結果を利用することができる。

本課題では、未知のグラフに対する影響最大化アルゴリズム IMUG (Influence Maximization for Unknown Graphs) を提案する。IMUG の基本的なアイデアは、(1) 過去のラウンドにおけるノードの探査とノードからの影響伝搬の結果から、各ノードの次数を推定し、(2) 次のラウンドでは、次数の期待値が最も大きく、なおかつ過去に影響を受けていないノードをシードノードとして選択する、というものである。IMUG では、ネットワーク全体の構造を利用することができないため、各ノードの次数の期待値を利用してシードノードの選択を行う。ただし、本研究では各ノードの次数も既知ではないため、過去の探査および影響伝搬の結果から、各ノードの次数の期待値を推定する。まず初期状態においては、各ノードの次数に関する知識がないため、各ノードの次数を k と推定する。 k は IMUG のパラメータであり、事前知識からグラフの平均次数がおおよそ推定できる場合には、その値を用いることができる。そして、各ラウンドの探査によって、各ノード

ドの次数の期待値を更新していく。ノード vp を探査した場合、 vp の真の次数が確定するため、 vp の次数の期待値を vp の隣接ノード数とする。さらに vp の隣接ノードは、少なくとも vp とのリンクを 1 本有していることが判明するため、 vp の隣接ノードの次数の期待値を 1 増加させる。探査ノードの選択には、前節で紹介した SEC 法を用いる。

シミュレーション実験により、提案する IMUG が、未知のグラフに対する影響最大化アルゴリズムとしての程度有効であるかを評価する。モデルで生成したグラフ上で、IMUG を用いた影響伝搬のシミュレーション実験を行い、各ラウンドにおいて影響を受けた（重複を除く）ノード数を調査する。

図 5 に NetHept と呼ばれるソーシャルネットワークにおいて、IMUG を用いて選択したシードノードの数と、そのノードの影響力の強さの関係を示す。ここでは比較のため、ランダムにシードノードを選択した場合の結果（random）、ランダムサンプリングによってネットワークの探査を行い、次数の期待値の高いノードをシードノードに選んだ場合の結果（random-degree）、およびネットワーク全体の情報を用いてシードノードを選択した結果（DegreeDiscountIC）も掲載している。探査するノード数はラウンドごとにノード数の 0.1% としている。この結果より、IMUG を用いることにより、ネットワーク全体の情報を用いた場合と同程度の影響力の強いノードを、部分的なソーシャルネットワークの構造のみから推定できていることが分かる。

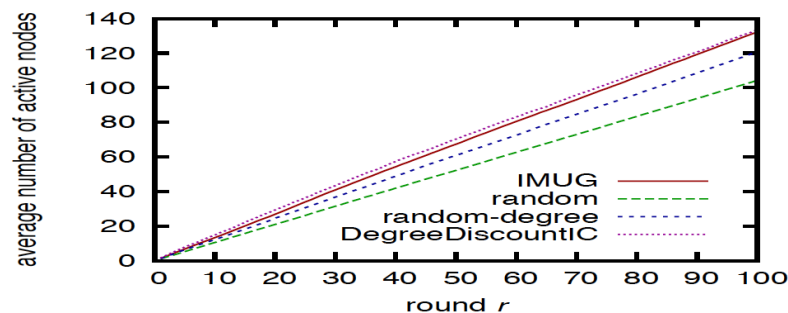


図 5: ラウンドごとの影響力を受けたノード数（縦軸は各手法が選択したノードの影響力の強さを表す） [20]

3-4 まとめ

本研究課題では、ソーシャルメディアにおいて影響力の強いユーザを部分的なソーシャルネットワークの構造のみから推定する手法を検討した。SEC 法と呼ばれるサンプリング手法を用いることにより、ネットワーク全体の 10% 程度の構造のみを用いて、十分に影響力の強いノードを発見できることを示した。本課題では、実社会イベント予測そのものには取り組んでいないものの、ここで開発したユーザの影響力を推定する技術は、ソーシャルメディアの投稿から実社会イベント予測を行う際に、重要な投稿を抽出するのに利用可能であると考えている。

4 まとめと今後の課題

研究期間内において、(A) 検索行動量を用いて自動車販売台数を予測するモデルの研究、(B) ソーシャルメディアの投稿からノイズを除去するためのユーザの影響力推定手法の研究、に取り組んだ。研究課題 (A) では、状態空間モデルを用いて、Google Trends と Wikipedia のページ閲覧数から自動車販売台数を予測するモデルを構築した。研究課題 (B) では、部分的なソーシャルネットワークの構造のみから重要なユーザを推定する手法を提案した。

今後は、研究課題 (B) で開発した手法を用いて、ソーシャルメディアにおける重要なユーザを推定し、重要なユーザの投稿のみを用いて実社会イベントを予測するモデルの構築に取り組む予定である。また、ソーシャルメディアにおける投稿と、研究課題 (A) で用いた検索行動量を合わせて用いて高精度な予測モデルを開発することも重要な今後の課題である。

【参考文献】

- [1] S. Asur, B. Huberman et al., “Predicting the future with social media,” in Proc. of WI-IAT, 2010, pp. 492–499.
- [2] H. Choi and H. Varian, “Predicting the present with Google trends,” *Economic Record*, vol. 88, no. s1, pp. 2–9, 2012.
- [3] E. Aramaki, S. Maskawa, and M. Morita, “Twitter catches the flu: detecting influenza epidemics using Twitter,” in Proc. of EMNLP, 2011, pp. 1568–1576.
- [4] W. Fan and M. D. Gordon, “The power of social media analytics,” *Communications of the ACM*, vol. 57, no. 6, pp. 74–81, 2014.
- [5] D. Gayo-Avello, “A meta-analysis of state-of-the-art electoral prediction from Twitter data,” *Social Science Computer Review*, vol. 31, no. 6, pp. 649–679, 2013.
- [6] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, “Detecting influenza epidemics using search engine query data,” *Nature*, vol. 457, no. 7232, pp. 1012–1014, 2009.
- [7] L. Granka, “Using online search traffic to predict us presidential elections,” *PS: Political Science & Politics*, vol. 46, no. 02, pp. 271–279, 2013.
- [8] 中島 光夫, 津川 翔, 山本 幹雄, “トピックモデルを用いたソーシャルメディアからの市場シェア予測,” 第13回情報科学技術フォーラム (FIT 2014) 予稿集 第2分冊, pp. 221-222, Sep. 2014.
- [9] 角田 孝昭, 吉田 光男, 津川 翔, 山本 幹雄, “状態空間モデルを用いた検索トレンドとページビューからの自動車販売台数の予測,” 人工知能学会全国大会論文集, pp.3C-3:1-3C-3:4, Jun. 2015.
- [10] 山口 太一, 角田 孝昭, 津川 翔, 山本 幹雄, “検索行動量を用いた自動車販売台数予測に必要な学習期間の長さについての分析,” 電子情報通信学会総合大会講演論文集 (D-4-8), Mar. 2016.
- [11] J.Weng, E.-P.Lim, J.Jiang, and Q.He, “Twitterrank: finding topic-sensitive influential Twitterers,” in Proceedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM’10), 2010, pp.261--270.
- [12] A. Pal and S. Counts, “Identifying topical authorities in microblogs,” in Proc. Of WSDM’11, pp. 45–54
- [13] L.C.Freeman, “Centrality in social networks: conceptual clarification,” *Social Networks*, vol.1, no.3, pp.215--239,1979.
- [14] S.Brin andL. Page, “The anatomy of a large-scale hyper-textual web search engine,” *Computer Networks and ISDN Systems*, vol.30, no.1, pp.107--117, 1998.
- [15] S. B. Seidman, “Network structure and minimum degree,” *Social Networks*, vol.5, no.3, pp.269--287,1983.
- [16] S. P. Borgatti, K. M. Carley, and D. Krackhardt, “On the robustness of centrality measures under conditions of imperfect data,” *Social Networks*, vol. 28, no. 2, pp. 124–136, 2006.
- [17] A. S. Maiya and T. Y. Berger-Wolf, “Benefits of bias : Towards better characterization of network sampling,” in Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’11), Sep. 2011, pp. 105–113.
- [18] 木邑 和馬, 津川 翔, 山本 幹雄, “サンプリングされたソーシャルネットワークからのユーザの影響力推定,” 電子情報通信学会技術研究報告 (CQ2015-97), pp.25-30, Jan. 2016.
- [19] D. Kempe, J. M. Kleinberg, and E. Tardos, “Maximizing the spread of influence through a social network,” in Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’03), Aug. 2003, pp. 137–146.
- [20] S. Mihara, S. Tsugawa, and H. Ohsaki, "Influence maximization problem for unknown social networks," in Proceedings of the 1st International Workshop on Dynamics in Networks (DyNo 2015) in conjunction with ASONAM’15, pp. 1539–1546, Aug. 2015.

〈発表資料〉

題名	掲載誌・学会名等	発表年月
未知のグラフに対する影響最大化アルゴリズムにおけるランダムジャンプの効果に関する一検討	電子情報通信学会総合大会論文集	2016年3月
検索行動量を用いた自動車販売台数予測に必要な学習期間の長さについての分析	電子情報通信学会総合大会論文集	2016年3月
通信行動データを利用したコミュニティ計測に関する研究の動向	電子情報通信学会技術研究報告	2016年1月
サンプリングされたソーシャルネットワークからのユーザの影響力推定	電子情報通信学会技術研究報告	2016年1月
Influence maximization problem for unknown social networks	Proceedings of the 1st International Workshop on Dynamics in Networks (DyNo 2015)	2015年8月
状態空間モデルを用いた検索トレンドとページビューからの自動車販売台数の予測	人工知能学会全国大会論文集	2015年6月