

エデルマンの勇者たち (11)

ソーシャルのOR, ツイッターの場合

伊倉 義郎

1. はじめに

しばらくお休みをしていた本連載ですが、また興味深いケースが何個か出ているので再開します。今回はその一つ SNS (ソーシャル・ネットワーキング・サービス) の Twitter (ツイッター) 社を取り上げます。当社は 2014 年の Edelman 賞コンペティションで優勝は逃したものの、Finalist となり表彰されました。最終プレゼンは、参考文献 [1] を参照ください。

最近急速に利用が広がっている SNS ですが、その実用化技術についてはあまり公にはされていません。その意味で、今回の発表は具体的なシステムの概要や最適化アルゴリズムの中味を解説しており、貴重なケースと考えられます。

Twitter 社が参加した Edelman 賞コンペは、2013 年 10 月から 2014 年 4 月にかけて実施されました。ちょうど 2013 年 11 月に Twitter 社は米国 NASDAQ に上場しており、それ以前に鳴り物入りで IPO した Facebook (フェースブック) との比較が話題になっていた時期でした。特に 10 月の Edelman 賞の予備審査段階では、Twitter 社の IPO 直前だったので、情報公開に制約があるのではという危惧の声もあって、その参加を疑問視する意見もありました。一方で、新しい分野での OR 活用という観点から一部の強い関心や期待もあって、Finalist に残る結果となっています。

Twitter 社のサービスについては、よくご存知の方も多と思うので、詳しくは解説しませんが、無料の会員間のコミュニケーションサービスで、140 文字以内の短文 (Tweet, つぶやき) や画像をやり取りするというものです。他の米国 SNS 大手としては、個人的な趣味や画像中心の Facebook やビジネスに特化した LinkedIn (リンクトイン) などがありますが、Twitter



図 1 Twitter 社のロゴ

の特徴はニュース性、つまり最新情報の提供にあると言えるでしょう。政治、経済、社会のあらゆる出来事や個人的意見や見解をリアルタイムで世界中に届けるというサービスは、現ネット社会の重要なインフラの一つと言えるのではないのでしょうか。ちなみに、2008 年から 2011 年にかけての「アラブの春」という現象も Twitter が重要な役割を果たしたことで記憶に新しいと思います。特に、最近では政治的ニュース、事件・事故の情報、有名人のコメント、企業ニュース、相場の動きなど、Tweet による情報は、他のマスコミよりも一番早く、しかも個人的で直接的であると言えます。また最近では Facebook とともに、Twitter のロゴがあらゆるリンク先に表示されていて、いまやその利用は定番化したという感じさえあります。

今回はそのような Twitter 社で開発された OR 的アルゴリズムを紹介します。その技術的な内容とともに、当アルゴリズムを実装したサービスが会社にとってどんな価値があるかについても論じたいと思います。ここでは Edelman 賞の参加から見た Twitter 社が主な対象なのですが、その裏には「Twitter 社の企業価値はいかほどのもの?」というより大きな課題があり、後ほど技術的な話と絡めて紹介したいと思います。

まずは、簡単な Twitter 社の歴史から振り返ってみますが、他のネット関連の会社と同様に社史といっても極めて短いもので、数人のブロガーが集まって、リアルタイム性の強い新サービスを始めようということ

いくら よしろう

株式会社サイテック・ジャパン

〒113-0033 東京都文京区本郷 2-19-9 田原ビル 2F



図2 Twitterのメイン画面例 (参照: <https://twitter.com> よりログイン後の画面)

で、2006年より開始しています。その後ベンチャーとして支援が強化され、サービスの利用も急速に拡大、日本語の利用も2008年から始まっています。現在では3億人ほどのユーザーがいて、今では1日当たり約5億個のTweetが作られ、35カ国語での利用が可能です。現在ではスマホでの利用が大半(76%程度)を占めており、他のネット・ビジネスと同様に、モバイル戦略が非常に重要になってきています。

さて、Twitterの会員になると、他の会員のTweetを即時に受けとり、閲覧できるサービスが利用できます。これを他の会員を「フォローする」と呼んでいます。たとえば、会員になりタイガーウッズ選手をフォローすると、彼がいつ次の試合にでるのか、体調はどのようなのか、など本人が発表次第、即知ることができます。

この「フォロー」機能をサポートするために、Twitterの大きな仕組みの一つとして、「Who-To-Follow」システムというのがあります。これはTwitterの各会員に対して、他のどの会員をフォローすべきかをお勧めとして知らせる機能です。ちなみに、現在日本語では「Who-To-Follow」は「おすすめユーザー」としてログイン画面の右側に表示されます(図2参照)。

ログインすると3人のおすすめが自動的に表示されますが、更新ボタンを押すとそれらが入れ替えになり、また誰かを新規にフォローしだすと、別のお勧めが出たりします。

普段はなんとなく見過ごしている画面かもしれませんが、この「Who-To-Follow」はTwitterにとって重大な意味をもっています。というのも、この「Who-To-Follow」システムがTwitter社の宣伝広告機能の中核にあり、他のユーザー(特に企業ユーザー)や製品にも適用されているという点です。

「Who-To-Follow」機能を実装する際の難しさは、そのサイズにあると言えるでしょう。つまり、各会員に対して、3億人のユーザーから誰をお勧めとするかを短時間で探さなければならないという点です。全く関連性のないユーザーや興味の湧かない商品をお勧めしても結果は出ませんし、それを続けると逆に会員に嫌われる結果となります。

すべての会員に対して定期的にお勧めを更新する必要から、計算時間もごく限られてきて、一人当たり100msから500msくらいと言われています。

このような「おすすめユーザー」の作成は、2009年ころから始まったということですが、当初は人手でこのリストを作っていたそうです。しかもユーザーごとにカスタマイズされていなかったため、早々に不評を買い、本格ソフトウェア開発の運びとなり、現在ではTwitter内で広範に使われるシステムになっていることです。2009年当時、Stanford大学OR学科のGeol教授がちょうど研究休暇でTwitter社にコンサルタントとして勤務しており、彼を中心に3人で開発グループが結成されました。その後約4か月でプロトタイプが完成し、2010年の6月には最初のリリースが行われました。

2. 「Who-To-Follow」はグラフ理論と方程式

では、「Who-To-Follow」アルゴリズムを解説するために簡単な例を紹介します。今5人のユーザー(Sally, Jin, Kumar, Bob, Alex)がいて、誰が誰をフォローしているかを図3で示します。ノードはユーザーを示し、アーク元がアーク先をフォローしている関係を表しています。

フォロー状況をさらに明確にするために、ユーザー・ノードから二つのグループを作ります。一つのグループは誰かをフォローしているユーザーたちで、消費者と呼ばれ、この例ではSally, Jin, Kumarがそれに当たります。次のグループは、誰かにフォローされているユーザーたちで、生産者と言い、Bob, Kumar, Alexが属することになります。なお、フォローしフォローされているKumarは両方のグループに属することになります。消費者グループのノードを左に置き、生産者グループのノードを右に置くと図4のような2部グラフができます。情報の消費者が、興味のある情報を提供する生産者をフォローしているグラフと解釈できます。

消費者間の比較を行うために、ある消費者Cに対する各消費者の類似度(similarity)という数値を考え、

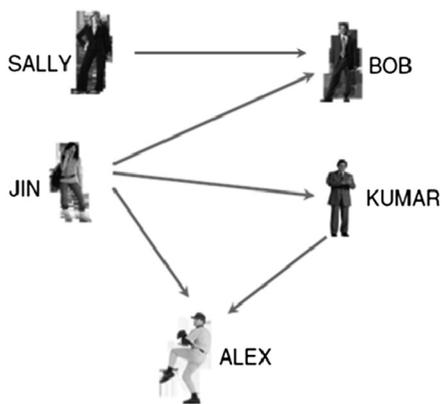


図3 5人のフォロー関係の例 (参照: 文献 [3])

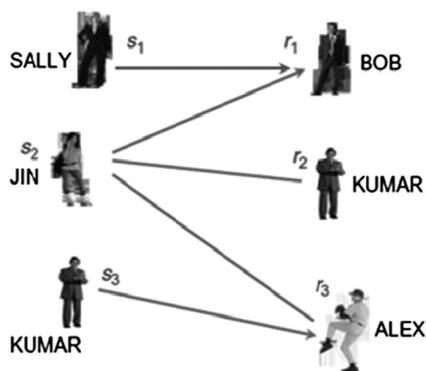


図4 消費者・生産者グラフ (参照: 文献 [3])

全く同じであれば1を、全く違えば0を与えることにします。同じように、Cの生産者への関心度を示すために、関連性 (relevance) という数値を計算することにします。関連性は、非常に興味があれば1、全然興味がなければ0とします。フォロー状況によって与えられたグラフに従い、これらの数値を計算することを考えます。それには、まず元の消費者Cにある数値 (例, 1) を与え、そのノードから出発してアークを伝って右側に数値 (ポイント) をあるロジックによって移動させます。次に右から左にアークを逆向きに辿って数値を減衰伝搬させます。このやり方を繰り返し行うことにより、消費者側の数値と生産者側の数値がある分布に収束するようにし、それを類似度と関連度と定義します。その計算結果から、生産者側ポイントの最も高いものを選び、それによって消費者Cへの推薦リストを作るとします。このような計算方法を、一般的には協調フィルタリング (Collaborative Filtering) と呼びますが、ソーシャルネットワーク上の参加者同志の関連性の数値化にはよく使われる手法とのことです [2]。

実際の協調フィルタリングには細部にわたって種々の方法があり、Twitterで採用されているのは、通称“Money”と呼ばれているやり方です。この手法では消費者から生産者へのポイントの伝搬をする際に、総和が元の値に等しくする (つまり左から右に分割する) という条件を加えます。このやり方は遺産相続と同じで、総和は左と右で同じになります。別のやり方としては“Love”と言われている方法で、左から右へ分割せずにそのまま値を伝搬させるやり方です。ここでどちらが正しいかという議論はなくて、結局どちらが現実のユーザーの行為をより正確に予測するかという判断基準になります。Twitterでは他のユーザーをフォローするには、“Money is better”という結果とのことです [3]。

また別の計算方法としては、“Cosine (コサイン)” というものもあります。これは生産者側での消費者からのフォロー状況を、N次元の0-1ベクターで表現し、二人の生産者間の類似性をそれぞれのN次元ベクターのコサインを取って計算するというやり方です (全く一致すれば1、全く同一のフォローがなければ0)。実はこちらのほうが簡明で対称性もあるという利点がありますが、ユーザーへのお勧めフォロー作成には採用されず、スポンサー企業商品のお勧めに使われているとのことです。

さて、このような協調フィルタリングの実装での困難さは、やはりそのサイズにあります。計算方法を詰めていくと結局一次方程式を解くのに帰着しますが、それを一々解いていたのでは時間が足りなくなります。そこでTwitterでは、各ユーザーに対して類似度が高そうな消費者ノードをある程度 (100から1,000個くらい) 絞り込んで、それらのノードからのポイント伝搬を確率的に行っています。それによって、1ユーザー当たりの計算時間を0.5秒以内に収めることができた、ということです。なお、このような「おすすめユーザー」の計算は、すべてのユーザーに対して定期的に行われる必要があります。一人当たりの計算時間が全体に及ぼす影響は多大なものになります。余分なハードへの投資を抑えるためにも、“Who-To-Follow” アルゴリズムは極力高速化される必要があります。

3. Twitter社の企業価値との関連性

さて、2010年にはめでたく“Who-To-Follow” システムもリリースされ、その成果も徐々に上がり、宣伝収入も急伸びました。それを受けて、2013年秋には無事IPOも遂行されEdelman賞コンペのあった2014年



図5 ツイッターの元 CEO, Dick Costolo 氏
(参照: Twitter 社 HP より)

春の段階では、当時の CEO の Dick Costolo 氏も順風満帆の勢いを感じていたのではと思います。Edelman 賞のプレゼン時でも、2014 年第 1 四半期の売上が \$250 million (内広告収入は 89%) となり、関係者には遂に \$1 billion company という達成感もあったのでは、と思われま。Costolo 氏自身、Edelman 賞参加ビデオでは、“Who-To-Follow” アルゴリズムと OR への最大の賛辞を呈しています (参照 [1])。またプレゼンでは、当時の Twitter 社の時価総額 \$26 billion の 24 分の 1 (約 \$1 billion, 1,200 億円以上) を “Who-To-Follow” システムによる投資効果と位置づけています。

そのまま順調に売り上げが伸びれば、問題は顕在化しなかったのかもしれませんが、今年の第 1 四半期 (1 月～3 月) に異変が起こりました。急に売り上げが減少したのです。アナリストの予想に反して、宣伝収入が 2014 年第 4 四半期に比べて約 10% 以上減ったのです。四半期ごとに 20～30% の高い伸びを示していた宣伝収入だけに、その衝撃は大きかったと推測されます。その影響で、IPO 以来順調に推移してきた株価も 4 割ほどダウン、以前から燻っていた Costolo 氏に対するウォール街からの批判の声も再燃してしまいました。その後 2015 年 6 月に Costolo 氏は退任へ追い込まれ、現時点では創業者の一人 Jack Dorsey 氏が臨時 CEO に就任、ここ数カ月以内に次期 CEO を選ぶという決定がなされています (2015 年 7 月時点)。

さらにウォール街では、そもそも Twitter 社という会社の価値はいかほどのもの? という疑問も上がっています。もちろんニュースを即時伝えるという利便性とそのグローバルなブランド名は誰もが疑問を抱かないとは思いますが、株価を付ける側としては現状に納得がいけないようです。つまり、他の SNS に比べて、技術や知名度の “Monetize (現金化)” がうまくできてい

ないのでは、という疑念です。その裏には、シリコンバレーの創業者とウォール街のファンドマネジャーたちの昔からの確執もあるようです。ファンドマネジャーにしてみれば、いわゆる “ハイテクオタク” に大会社が運営できるのかという疑心感が常にあり、逆にシリコンバレーの創業者たちには、ファンドマネジャーに技術の何がわかるの、という思いがあるようです。どちらもごもつともなのかもしれませんが、Costolo 氏が製品開発畑出身で主に技術的な議論に終始しがちだったことからして、Twitter 社には広告収入戦略に詳しい新しいリーダーが必要なのは確かです。ただし、臨時 CEO の Dorsey 氏は「従来の戦略を延長する方針をとる」と言っているのです。現時点での大幅な方向転換の可能性は低いかもしれません。また別の可能性としては、買収の対象 (特に Google) になるのではとの憶測もありますが、実現性は若干低いようです。

Twitter 社の技術者はどうでしょうか。開発された “Who-To-Follow” アルゴリズムは、他の SNS と比べても、格段に優れているとは思えませんが、逆に劣っているとも思えません。つまり、それなりのものができていると判断されます。この現状で、Twitter 社の技術者には二つの選択があると思います。つまり、そのまま会社に残って事態を見守るか、積極的に転職先を探すかです。一番目を選択すると、新 CEO の元で会社が立ち直るのを期待するか、または買収により他社の経営者の下で働くかになります。前者の場合は、新 CEO の手腕次第でしょうから、自分を含めた会社の行く末も新 CEO に大きく依存することになります。また買収となった場合、買収先の会社でも多数の優秀なエンジニアがいる可能性が高く、現 Twitter 社の技術者が全員生き残れるとは限りません。いずれの選択にしても、早晚技術者の Twitter 社離れが起きる可能性も見てきます。

グラフ理論で最高レベルのアルゴリズム開発ができたにしても、会社トップの経営能力がいまいちであれば環境も安定せず、安閑とはしてられないという始末です。もっとも、技術者たちはストックオプションなどもある恵まれた待遇を受けているはずなので、ハイリスク・ハイリターンキャリアであることには間違いありません。今後の Twitter の動向が注目されます。

参考文献

- [1] “Presentation by Twitter at the 2014 Franz Edelman Competition,” INFORMS Video Learning Center, https://www.pathlms.com/informs/events/217/thumbnail_video_presentations/5905

[2] B. Bahmani, A. Chowdhury and A. Goel, "Fast incremental and personalized pagerank," September 2014, <http://www.vldb.org/pvldb/vol4/p173-bahmani.pdf>

[3] A. Goel, P. Gupta, J. Sirois, D. Wang, A. Sharma and S. Gurumurthy, "The Who-To-Follow System at Twitter: Strategy, algorithm, and revenue impact," *Interfaces*, **45**(1), pp. 98–107, January–February 2015.