

SVM を用いた感情抽出に基づくフォロワー推薦

11371002 秋山 和寛 (灘本研究室)

あらまし：これまで我々の提案してきた話題と感情を考慮したフォロワー推薦手法では、ツイートの感情の抽出の精度が十分であるとは言い難い。そこで、本研究では機械学習による感情の決定を行い、コンテンツの類似性だけでなくそのコンテンツに対する感情も考慮したフォロワー推薦手法を提案する。

1. はじめに

マイクロブログを代表する Twitter の機能の一つにフォロー機能がある。ユーザがどのような人物をフォローするかには友人、共通の趣味、情報収集、有名人など様々な基準があると考えられる。しかしながら、膨大な数のユーザの中から自分がフォローしたいと思うユーザを探し出すのは困難であるといえる。そこでこの問題を解決するために様々なフォロワー推薦の研究が行われている。フォロワー推薦は大きく分けると、ユーザ間のフォロー・フォロワー関係を考慮したネットワークに基づく推薦手法と、ユーザ同士のツイートした内容の類似性を考慮したコンテンツの類似性に基づく推薦手法の 2 つに分類される。

山本ら[1]は後者に着目し、話題の類似性だけでなく感情の類似性も考慮したフォロワー推薦手法を提案している。これまでの提案手法ではツイートの感情抽出に我々の作成した感情語辞書を用いている。具体的には感情語辞書を用いて、ツイート文中の単語毎の感情値を求め、それらの和によりツイートの感情値を決定している。しかしながら、実験よりその精度が十分であるとは言い難いことが判明した。そこで、本研究では感情の分類に SVM を用いる。

本研究では感情に基づくフォロワー推薦として、まず話題の類似性を求め、次に、類似している話題に対する感情を機械学習により分類し感情の類似性を求める。そして、話題及び感情が類似しているユーザを推薦する。

2. 提案手法

2.1. 話題の類似性の抽出

話題の類似性の抽出には、これまで山本らが提案してきた手法を用いる。本論文では、フォロワー推薦を行いたいユーザを「閲覧ユーザ」と呼び、閲覧ユーザにフォロワーとして推薦するユーザは閲覧ユーザのフォロワーのフォロワー且つ、閲覧ユーザのフォロワーでないユーザを対象とする。フォロワーのフォロワーを本論文では「ff ユーザ」と呼ぶ。

まず、閲覧ユーザと ff ユーザの過去 200 件のツイートを対象とし、両者のツイートを収集

する。次に、閲覧ユーザと各 ff ユーザのツイートを併せてクラスタリングを行う。この際、クラスタリング手法には短文のクラスタリングに優れているといわれる [2] Repeated Bisection 法 [3] を用いる。クラスタリング対象は名詞のみとする。各クラスタには話題を示す名詞であるトピックが複数存在するが、本研究ではそのクラスタの中心ベクトルに最も近いトピックをそのクラスタの話題とする。クラスタの分割数については予備実験より 10 とする。また、共通の話題を抽出する際に以下の 2 点を考慮する。

●話題の非共通度

あるクラスタ内に閲覧ユーザと ff ユーザのツイートが混在している場合、そのクラスタは両者にとって共通の話題となる可能性がある。しかし、両者のツイート数の比率がどちらかに偏っている場合、そのクラスタを両者に共通の話題とすることは適切とはいえない。そこで、以下の式を用いて両者のツイート数の比 R_i を求め、その値が閾値 0.25 以下のクラスタの話題を両者に共通の話題とする。

$$R_i = \frac{|X_i - Y_i|}{X_i + Y_i}$$

ここで、 X_i は i 番目のクラスタに属するユーザ X のツイート数を示し、 Y_i は i 番目のクラスタに属するユーザ Y のツイート数を示す。

●クラスタ内のツイートの凝集性

クラスタリングに用いた Repeated Bisection 法はハードクラスタリングであるため、相互に関連性のない話題を含むツイートで構成されたクラスタが生じる場合がある。本研究ではそのようなクラスタをガベージクラスタと呼ぶ。ガベージクラスタの話題は、たとえ両者のツイート数が同じだとしても共通の話題とはいえないため、取り除く必要がある。この際、ガベージクラスタ内のツイート同士の話題の関連性が低いという点を利用する。具体的には i 番目のクラスタ C_i のセントロイド c_i とそのクラスタに含まれるツイート x_i のコサイン類似度をツイート毎に求め、その平方和をクラスタ C_i の凝集性 A_i と定義し、その値が閾値 0.6 未満のクラスタをガベージクラスタとして取り除く。

以上より、2 つの閾値を満たすクラスタを「共

通話話題クラスタ」と呼び、共通話題クラスタの話題を両者に共通の話題として扱う。

2.2. 感情の類似性の抽出

本研究では感情の分類に機械学習を用いる。そして分類した各ツイートの感情からユーザ間の話題ごとの類似度を求め、フォロワー候補の推薦を行う。機械学習には SVM を用いる。また、SVM による感情分類の際には感情軸は山本らの提案する Twitter 用の 8 軸の感情軸「喜・好」、「哀」、「怒・厭」、「安」、「怖」、「恥」、「昂」、「驚」を対象とする。

2.2.1. SVM を用いた感情の類似性の分類

抽出された共通の話題毎に共通話題クラスタ内の各ツイートの感情を SVM により分類する。そして、両ユーザのその話題に対する感情の類似度を求める。学習データと素性は以下の通りである。

●学習データ

SVM の学習データには、アンケートにより感情毎に分類されたツイートを用いる。具体的には、1 セットにつき、無作為に抽出した 25 件のツイートに対し、10 人の被験者による各軸 5 段階評価を 58 セット、計 1450 件のツイートに対し 10 人の被験者による評価を行い、学習データを決定した。

●素性

素性は全学習データに形態素解析を行い、得られた名詞・動詞・形容詞全てを対象とし、各々の単語がツイート内に含まれるか否かを素性としました。

2.2.2. 感情の割合に基づく類似度算出

類似度の算出には共通話題クラスタ内の両者の感情の割合を用いる。類似度の計算には、2 つの確率分布間の距離尺度を測ることができる JS ダイバージェンスを用いる。

$$JS_i = \frac{1}{2} KL(P||M) + \frac{1}{2} KL(Q||M)$$

$$KL(P||Q) = \sum_{i=1}^8 P(i) \log_2 \frac{P(i)}{Q(i)}$$

$$M = \frac{1}{2} (P + Q)$$

ここで、P は共通話題クラスタ i 内における閲覧ユーザの 8 軸の感情の割合を示し、Q は推薦候補ユーザの 8 軸の感情の割合を示す。上記の式より、共通話題クラスタ毎に感情の類似度を求め、その和を両者の感情の類似度 DS とする。

DS の値が、ff ユーザ 200 人それぞれの類似度の平均値未満であるとき、そのときの ff ユーザをフォロワー候補として推薦する。

3. 評価実験

提案手法の有用性を測るために、評価実験を行った。具体的には、4 名の被験者に自分の ff

ユーザ 200 名分のツイートを見てもらい、実際にフォローしたいと評価した ff ユーザを正解データとし、適合率、再現率、F 値を求める。

実験結果を表 1 に表す。4 人のユーザの特徴は以下の通りである。

●ユーザ 1

ツイート内容は日常生活に関するものやフォロワーへのリプライが多く、フォロワーは現実世界での友人・知人が多い。

●ユーザ 2

ツイート内容は日常生活に関するものが多く、フォロワーは現実世界での友人・知人と情報収集目的のユーザが多い。

●ユーザ 3

ツイート内容は趣味に関するものが多く、フォロワーもその趣味に関係するユーザで占められている。

●ユーザ 4

ツイート内容は趣味に関するツイートが多く、フォロワーはそれとは別の趣味に関するユーザで占められている。

表 1: 実験結果

ユーザ	適合率	再現率	F 値
ユーザ 1	0.072	0.45	0.125
ユーザ 2	0.106	0.667	0.183
ユーザ 3	0.129	0.653	0.25
ユーザ 4	—	—	—

表 1 より、ユーザ 1-3 に対しては、全般に適合率が低いが、これは機械学習の分類精度が低いためと考えられる。

ユーザ 1 の値が特に低くなっているのは、ユーザ 1 が様々なタイプのユーザと交流をしており、話題の偏りが少なく、個々のユーザとの類似性が低くなっているためだと考えられる。

ユーザ 4 については、本人がつぶやいている内容と ff ユーザのつぶやいている内容が全く異なっており、共通話題クラスタが抽出されなかったため、類似度の算出が行われなかった。

4. まとめと今後の課題

本研究では、機械学習を用いたフォロワー推薦手法の提案を行った。結果として、精度は良いとは言えなかった。今後は機械学習の精度の向上及び、ユーザのツイート・フォロー傾向を考慮していきたい。

参考文献

- [1] Yuki Yamamoto, Tadahiko Kumamoto, Akiyo Nadamoto, "Followee Recommendation Based on Topic Extraction and Sentiment Analysis from Tweets," iiWAS2015, pp. 215-225, 2015.
- [2] 花井俊介, 灘本明代. 酷似レシピ抽出のためのクラスタリング手法の提案, DEIM2014, F8-6, 2014.
- [3] Y. Zhao and G. Karypis, "Comparison of Agglomerative and Partitional Document Clustering Algorithms", University of Minnesota, pp.2-14, 2002.