自然言語処理を用いた極性分析とその応用

鳥取大学　工学部　社会物理学研究室　芦田昇　吉岡朋友

1.はじめに

日本国内におけるSNS 利用者数は年々増加している。株式会社ICT 総研の調査[1]によると一ヶ月平均で約26 万人の利用者が増加している。また、2013 年より、ネット選挙が解禁されたこともあり、インターネット上における政治関連の議論も散見されやすくなった。そこで、本論文ではニコニコ動画[2]を研究対象とし、自然言語処理を用いて極性分析を行う。この結果をオピニオンダイナミクスの実測に応用する。石井、川畑らによるオピニオンダイナミクス理論[3,4,5]は従来からのオピニオンダイナミクスでの合意形成の解析だけでなく、人々の間の意見反発まで考慮を入れて社会のシミュレーションに近い計算結果を示すが[6,7]、議論する人々の意見の極性の実測でその理論の検証が行われていない。そこで、本研究のニコニコ動画に関するコメントの極性の測定をオピニオンダイナミクスの理論計算の検証 に応用する。これに人々の政治に関する動画に対しての意見の推移を視覚化することで、世論や現在の政治に関しての人々の関心を明確にできるであろう。

本研究では、ニコニコ動画より、政治関連のニコニコ動画に対してついたコメントの付けられた時間、ユーザーID を取得しそれらを手動でポジティブ：1、ネガティブ：０に振り分けた。次にそれらをJuman++を利用し、形態素解析を行った。

そして、前処理として、word2vec[8]を用いて単語をID 化し、LSTM へ適用した。その後、オピニオンダイナミクス理論へ応用した。

2.分析方法

本研究では、ニコニコ動画より、政治的内容を含む動画18 個を選定し、コメント15000 件を取得した。コメント内容は、政治関連の動画のため、批判的なものが多く、ポジティブ：ネガティブ=10:1 の不均衡データとなった。このままでは、少数派の精度の低い不均衡データとなってしまうために、アンダーサンプリングを行った。これは、少数派のデータ件数に合うように多数派のデータをランダムに抽出する方法である。次に、Juman++を使用し形態素解析を行った後、word2vec で単語をベクトル化し、LSTM により極性を出力した。最後にオピニオンダイナミクスへ応用した。

　まず、本論文で用いたオピニオンダイナミクス理論の土台となっているHegselmann-Krause[9]の理論より説明する。式は以下の通りである。

$$I_i(t+1) = \sum_{j=1}^{N} D_{ij} I_j(t) \qquad (1)$$

I_iはエージェントi の意見で、D_{ij}はエージェントi がエージェント j からどれだけ影響を受けるかの係数で、Hegselmann-Krause は合意形成だけを考えているので、Dij>0 と正の値が仮定されている。従って、意見Iiの値も正の値となる。そして、計算すると、必ず合意形成に達する。ただ、ある程度意見が離れていると影響を受けないというカットオフが設定されている。しかし、現実社会では、意見が全て合意形成される限りではなく、意見が反発して分裂する場合もある。それらを考慮した石井、川畑らによるオピニオンダイナミクス[3,4]では、Dijに負の値を許し、正の値だと信頼、負の値だと不信とした。さらに、ある程度意見が離れていると影響を受けないというカットオフをHegselmann-Krause[9]では階段関数で入れてるのに対し、石井の理論[4]ではシグモイド関数で、連続的なカットオフとした。以下がその式である。

$$\frac{dI_i(t)}{dt} = c_i A(t) + \sum_{j=1}^{N} D_{ij} \Phi(I_i(t), I_j(t))(I_j(t) - I_i(t)) \qquad (2)$$

$$\Phi(I_i, I_j) = \frac{1}{1 + exp(\beta |I_i - I_j| - b))} \qquad (3)$$

I_i： i さんの意見

c_i：広告量が人に与える強さ表す係数

$A(t)$：時刻 t での広告量

D_{ij}：信頼度　$D_{ij} \geq 0$で i は j を信頼

$D_{ij} \leq 0$ で i は j を不信

である。

これを、自然言語処理の結果と対応させるにあたり、次のように考えた。

I_i：ある地点におけるユーザー i の極性

c_i：ユーザーiの動画影響の受けやすさ

$A(t)$：１秒ごとのユーザー極性の平均値

D_{ij}：コメントの類似度　$D_{ij} \geq 0$で i は j を信頼

$D_{ij} \leq 0$ で i は j を不信

また、コメントの類似度にはWMD を使用した。ここで、信頼度D をコメントの類似度とした理由は、日常生活においても友人や尊敬できる人は自分の意見や考え方と似ている人が多いと考えたためである。また、広告の影響をその動画の1 秒ごとの全極性の平均としたことについては、A は動画自体の面白さに関連していると推察したためである。

3.計算結果

テストデータについて
動画内容：朝日新聞の社説について
動画時間：678秒
コメントしたユーザー：302人

図3.1 テストデータについて

まず、LSTMにより得られたコメントの極性の一部を表に示す。

表１　LSTMによる極性値

テキスト	極性値
既に日本のせいになってるから www	-0.235
頭に花が咲いてる人たち	-0.212
ヘイ、ジャパニーズ！スパイ OK？	-0.220
そらそうだw	0.211
やるやんけ	0.0137
冷静になったな	0.082454

次に、1 秒ごとの全極性の平均を示す。図からも分かるように、全体的にネガティブな意見が多かった。動画の内容は政治的な内容なのでネガティブな意見が多いが、番組冒頭と番組最後は、番組制作者にたいして「お疲れ様」などのコメントがあり、これが冒頭と最後でポジティブな極性になっている理由である。

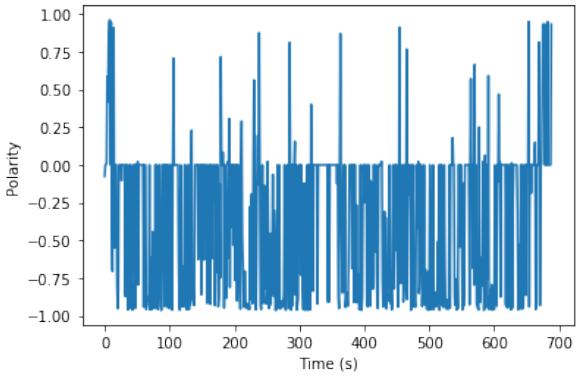

図3.2　1秒ごとの全極性の平均

次にLSTMで学習した極性を用いて、オピニオンダイナミクスへ適用した結果を示す。動画に対してコメントしたユーザーは302人であったので、リアルタイムでコメントしたユーザーを3つのセクションごとに分けた。つまり、セクション1に分類されるユーザーは動画内にコメントが比較的流れていない状態でコメントをしており、セクション３に分類されるユーザーは動画内にコメントが多く流れている状態でコメントをしていることとなる。

まず、全ユーザーの時間順序を入れた全意見に302人全員の各自の意見の時間的推移が、それぞれの人が実際に投稿したコメントから測定した意見の極性と合うようにDijを最適化したオピニオンダイナミクスの計算結果を示す。

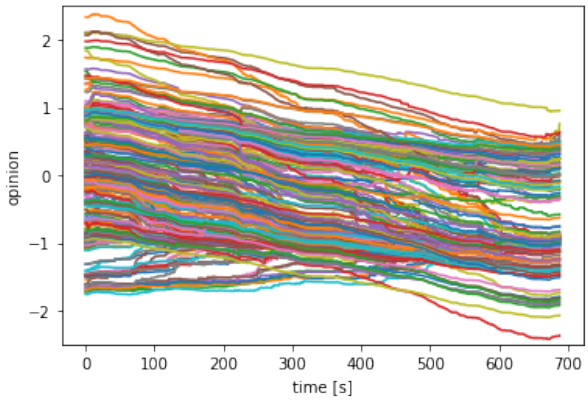

図3.3 全ユーザーの意見推移

全体的に、負の方向へ意見が推移している。次にセクションごとの意見推移の図を以下に示す。

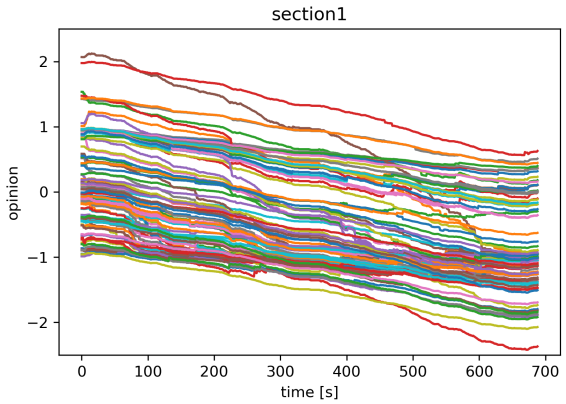

図3.4 セクション1のユーザーの意見推移

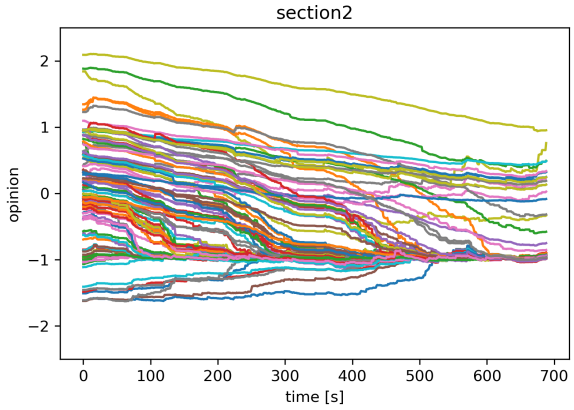

図3.5 セクション２のユーザーの意見推移

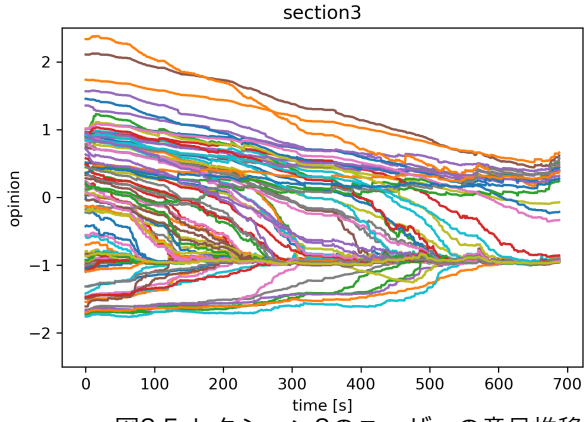

図3.5 セクション3のユーザーの意見推移

セクション３において、動画時間の終盤で意見が二分していることが確認できる。

次にセクション１とセクション３の動画の678秒、すなわち動画の最後で切った意見の分布のヒストグラムを示す。

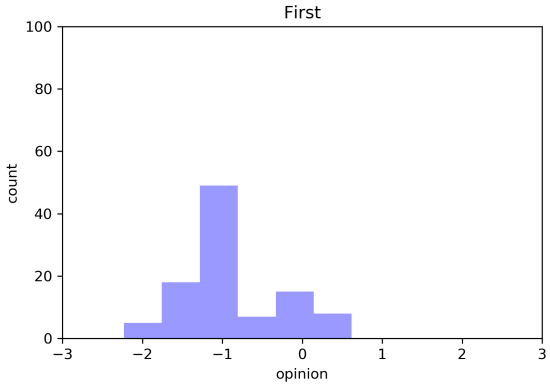

図3.7 セクション１のヒストグラム

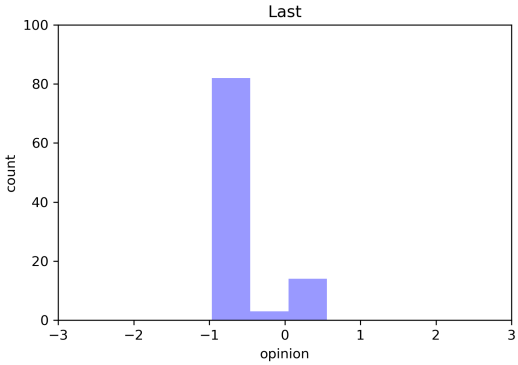

図3.8 セクション１のヒストグラム

意見分布のヒストグラムより、まず、意見に幅があり、全部で合意形成されていないことがわかる。これはHegselmann-Krause の理論の否定とも取れるが、動画の短い時間だけなので、遠い意見は無視するというカットオフでもこれらの分布は説明できる。セクション３では時間の経過で意見＝−１付近に合意形成が見られるものの、正の意見もあり、意見の正負の分裂が見られる。これは図3.6から時間の経過とともに意見が分裂していく様子も見られる。Hegselmann-Krauseの理論だと正負の意見がお互いに無視して別々に合意形成に至るので、この結果は反発を入れて意見の分裂を組み込んだ石井のオピニオンダイナミクスの理論を支持する結果であると思われる。

次にLSTM で出力した極性とオピニオンダイナミクスでシミュレーションした意見推移を示す。

赤点：LSTM による極性
青線：オピニオンダイナミクスによる意見推移

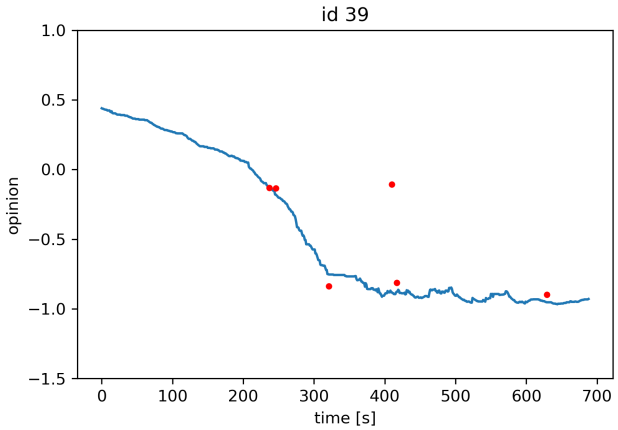

図3.9 ID39の意見分布と推移

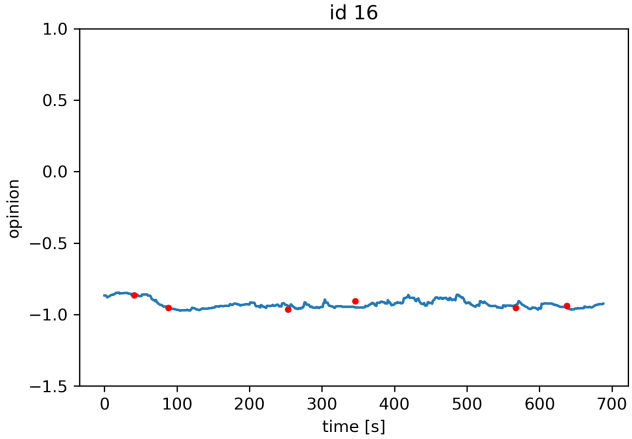

図3.10 ID16の意見推移と極性

4.考察

外れ値について訓練データの不足が考えられる。約14000 件のコメントに対して、実際に使用できたのは約5000 件であった。また、アンダーサンプリング時にネガティブのデータに対してランダムに抽出しただけであったので、特徴量の損失が考えれる。そして、政治関連の動画に絞って訓練データを作成したために今回作成したソフトには汎用性がないことが挙げられる。

5.総括

書き込み日時が古いほど意見が分散しており、新しくなるほど意見が合意されており、コメントによる影響により意見のまとまりができていると考えられる。また、オピニオンダイナミクスによる計算結果と測定した意見のの極性の差より、概ね意見の推移をシミュレーションすることができた。これは、石井のオピニオンダイナミクスをある程度測定から支持した結果と言えるであろう。改善の余地は多いにあるが、SNSや動画に付くコメントを分析することで、社会における人々の意見推移をリアルタイムに視覚化することが可能となると考えられる。具体的には、ライブコマースに応用することで販売の効率化が可能である。

6.展望

今回は、ニコニコ動画を対象としたが、Twitterの呟きに対しての各ユーザーの返信やYoutubeの動画に対してのコメントなどの分析を行い、本研究のニコニコ動画の意見分布と同様の推移が見られるか実験を行いたい。また日本語のコメントに対しての極性分析であったが、英語の文章に対する極性分析などで同様に意見の推移が見られるか試してみたい。

謝辞
本研究は科学研究費補助金基盤研究C(JP19K04881)によって行われた.

参考文献

[1]ICT 総研 SNS 利用者動向に関する調査 <https://ictr.co.jp/report/20181218.html> (2020.1.20 アクセス)

[2]ニコニコ動画 <https://www.nicovideo.jp> (2020.1.20 アクセス)

[3] Akira Ishii and Yasuko Kawahata, "Opinion Dynamics Theory for Analysis of Consensus Formation and Division of Opinion on the Internet", Proceedings of The 22nd Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES2018) 71-76

[4] A. Ishii, "Opinion dynamics theory considering trust and suspicion in human relations" Proceeding of 19th International Conference on Group Decision and Negotiation in 2019 a Joint GDN-EWG/BOR meeting, Springer LNBIP 351, 193-204

[5]石井晃・川畑泰子「不信とメディア効果を考慮したオピニオンダイナミクス理論」第33 回人工知能学会全国大会(2019)セッションID: 2F3-OS-5a-05

[6] N Okano and A Ishii, "Isolated, untrusted people in society and charismatic person using opinion dynamics", Proceedings of ABCSS2019 in Web Intelligence 2019, 1-6

[7] N Okano and A Ishii, "Sociophysics approach of simulation of charismatic person and distrusted people in society using opinion dynamics", Proceedings of the 23rd Asia-Pacific Symposium on Intelligent and Evolutionary Systems" (Springer, 2019). 238-252

[8] Tomas Mikolov, Kai Chen, Greg Corrado Jeffrey Dean" Efficient Estimation of Word Representations in Vector Space" 2013, ICLR