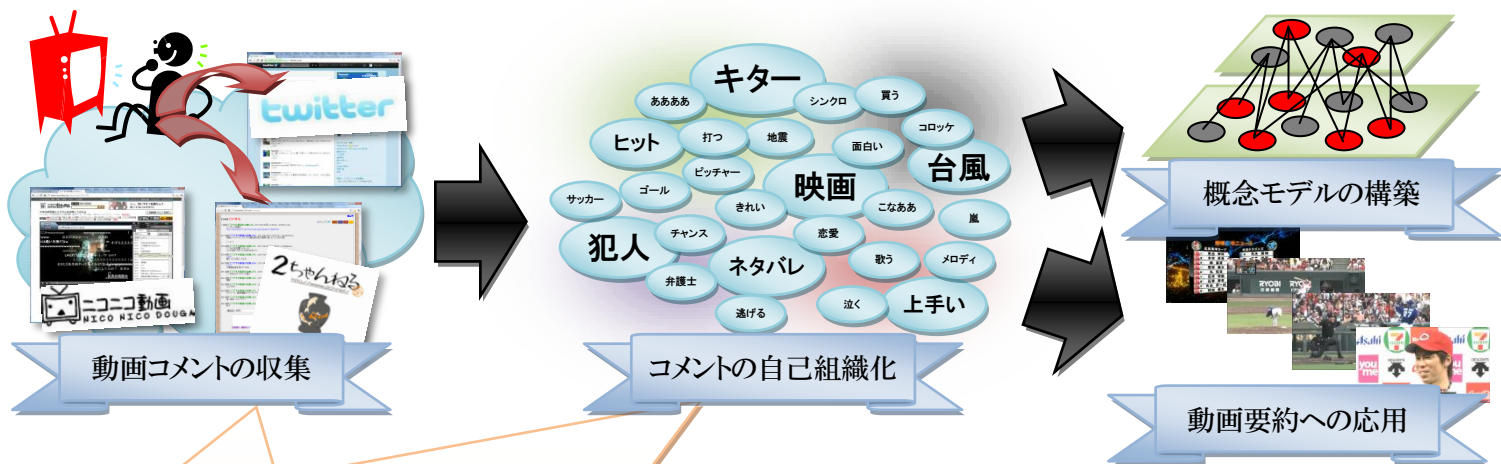


動画コンテンツに対する コメントの自己組織化と 要約・概念学習への応用

安田女子短期大学 坂口 琢哉

Abstract

BBS や Twitter の用途の一つに、動画コンテンツに対するコメントの書き込みと共有が挙げられる。web 上の動画や TV 番組などを視聴している複数のユーザが、その時々感じた意見や感想を同期／非同期的に書き込むことで、視聴体験をバーチャルに共有するものである。本研究ではこうしたコメントを収集し、それらを書き込み時刻や類似度に基づいて収束・自己組織化していく手法を提案する。また、得られた結果を元に動画コンテンツの要約や、概念モデル構築への応用を目指す。



(1) BBS「2ch 実況板」を対象

- TV 番組や中継に対して、リアルタイムに意見や感想を書き込む場
- レス ID・氏名・書込日時などのメタデータと、コメント本文で構成
- 短文、直観的、短絡的な内容が多い

116 名前: どうですか解説の名無しさん : 2011/09/20(火) 18:45:55.32 ID:egDLL81g
キター(´▽`)!!!!

117 名前: どうですか解説の名無しさん : 2011/09/20(火) 18:46:01.34 ID:mm8UrF8R
ホームベース手前で高めの球受けてるんだからなあ

118 名前: どうですか解説の名無しさん : 2011/09/20(火) 18:46:34.50 ID:8Ps78mEi
巨人は青木山井似に弱いんだな

119 名前: どうですか解説の名無しさん : 2011/09/20(火) 18:47:07.97 ID:PPK3dePz
>>115
ええか(げん理解して欲しいのう...
まあ頑張れ大鳥

120 名前: どうですか解説の名無しさん : 2011/09/20(火) 18:47:26.71 ID:PzdVU61L
1点どまりか

(2) コメント間の類似度を計算

- 一定時間枠(±60sec / ±180sec / ±600sec)内に存在する任意のコメント M, N を対象
- 3-Gram 法に基づき類似度 S_{MN} を計算

「私」と広島カープ」の 3-Gram
「私」と広」「と広島」「広島カ」「島カー」「カープ」

「広島カープ」と私」の 3-Gram
「広島カ」「島カー」「カープ」「ープと」「プと私」

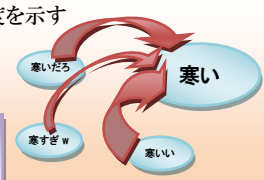
$$S_{MN} = \frac{\sum_{i=1}^{m-2} f(M, M_{i+1}, M_{i+2}, N)}{(m-2) + (n-2) - \sum_{i=1}^{m-2} f(M, M_{i+1}, M_{i+2}, N)}$$

ただし、 $f(A, B) = \begin{cases} 1(A \in B) \\ 0(A \notin B) \end{cases}$

(3) 活性化値の変化による自己組織化

- コメント M に対し、活性化値 a_M の概念を導入: コメントの重要度を示す
- 類似のコメント同士で値をやりとり、活性化値を変化
- 値は、低い方から高い方へと移動→値が集約→自己組織化
- 値の移動量は、コメント間の類似度に依存

$$\begin{cases} a_M + = da_N S_{MN} \\ a_N - = da_N S_{MN} \end{cases} \text{ ただし、} a_M \geq a_N, d \text{ は変化量を表す係数}$$



(4) 提案モデルの流れ

- step0: 対象となるスレッドの dat ファイルから、メタデータとコメントを獲得
- step1: 各コメントの活性化値に初期ゆらぎ $rand * \alpha$ を加える
- step2: 一定時間枠内に存在するコメント $M \sim N$ 間の類似度 S_{MN} を計算
- step3: S_{MN} の大きさに応じて、活性化値 $a_M \sim a_N$ 間で値を移動
- step4: step2~3 を、すべてのコメントについて実行
- step5: step2~4 を、一定回数 (k 回) 繰り返す
- step6: 各コメントの書込時刻と活性化値をプロット

(6) まとめと今後の展望

- 動画コメントに対する自己組織化手法を提案した
- プロ野球の試合に対する 2ch 実況板の書き込みを対象に実験を行った
- 重要場面におけるコメントが活性化され、提案手法の有効性が示された
- ▽コメント単位から単語単位のモデルへ: 形態素解析の利用
- ▽様々な種類の動画コメントに対応: web サービスによる差異性を考慮
- ▽動画内容や書き込み感情とのリンク
- ▽概念モデルの構築、動画要約への応用

(5) 実験結果と考察

- 対象動画: プロ野球広島戦(2011/9/4)の試合
- 対象コメント: 2ch 実況板(野球 ch 板)の広島応援スレッド(計 11 スレ、9673 レス)
- 各種パラメタ: $d=0.1, \alpha=0.1, k=10$
- 各重要場面において、ピンポイントでコメントの活性化が認められた
- 活性化されたコメントは、ある程度その場面の状況に沿ったものであった
→更なる最適化には、単語単位で自己組織化を行うことが有効?
- コメントのバリエーションが少ない
→時間枠を拡大することで、広範囲の自己組織化が実施された結果、多様性が向上

