

評価視点別レビュー要約のための重要文候補抽出

小池 惇爾[†]松吉 俊[‡]福本 文代[‡][†] 山梨大学大学院医学工学総合教育部[‡] 山梨大学大学院医学工学総合研究部

{g11mk018, sugurum, fukumoto}@yamanashi.ac.jp

1 はじめに

近年、インターネットの普及によって、web上には大量の文書が存在している。大量の情報の中からユーザーが短時間で必要な情報を探すことは容易ではないため、それを支援するための研究が盛んに行われている。それらの研究の1つとして、web文書集合や新聞記事集合などを対象とした複数文書要約がある。本研究では、レビュー文書集合を対象とした複数文書要約を取り扱う。レビュー文書集合から重要である少数の文を抽出し、要約を生成する。

レビュー対象には、予め評価視点が定められていることが多い。例えば、レビュー対象「宿泊施設」には、“部屋”、“立地”、“食事”などの評価視点が設定される。一般に、レビューはこのような評価視点を念頭において作成される。レビュー文書集合を要約する場合、評価視点を全く考慮せず、レビュー対象に対して1つの要約を生成する方法と、評価視点ごとにそれぞれ1つの要約を生成する方法が考えられる。レビューを利用するユーザーは、必ずしも全ての評価視点の情報を含む要約を求めているわけではないため、後者の要約を自動生成する技術は有用である。このような理由により、我々は、評価視点ごとにレビュー文書集合から重要である少数の文を抽出し、要約を生成する。

本論文では、我々が現在構築している要約コーパスと、重要文抽出の前処理である重要文候補抽出の実験について報告する。

2 関連研究

複数文書要約の手法としては、“文の位置”や“TF・IDF”などの統計的尺度やSupport Vector Machine (SVM)などの学習器を用いる手法、グラフベースの要約手法など、様々な手法が提案されている[10, 8, 2]。

Consumer Generated Media (CGM)の発展により、レビューや口コミといった評判情報をユーザーが投稿する機会が増えており、近年、それらを対象とした研究も行われている。次章以降で述べるように、我々は、まず重要文候補抽出を対象とするが、それに関する先

行研究として、レビュー対象の評価視点に基づいてレビューの各文を分類するものがある[9, 6]。

3 レビュー重要文抽出における問題点

我々は、楽天データ¹における楽天トラベル:レビューデータを用いて、評価視点ごとに重要文にラベルを付与した評価コーパスを作成し、6章で説明するベースラインのようなシステムを実装し、レビュー文書集合から重要文を抽出し、要約を生成する予備実験を行った。システムのF値は評価視点の平均で0.248と低かった。コーパスを観察し、レビュー文書集合から重要である少数の文を抽出する際には、次のようなことを考慮すべきことが分かった。

- (i) 相対的な重要度 レビュー文書集合に含まれる、重要であると見なされるレビュー文の数により、ある文を最終的に出力すべきかどうかは相対的に決まることもある。(4.1節参照)
- (ii) 言い換え・含意 ある文(A)と同じ内容を持つ文(B)、もしくは、文(A)の内容を含む文(c)が存在する。要約においては、冗長性は最大限排除すべきであるので、文(A)と文(B)の両方、文(A)と文(C)の両方を出力すべきでない。
- (iii) 評価視点のカバー範囲 各評価視点のカバー範囲が自明ではないため、例えば、部屋のドライヤーについてのレビュー文を、“部屋”、“設備・アメニティ”、その両方のどれに分類すべきか曖昧である。

上記のコーパス構築においては、“サービス”、“立地”、“部屋”、“設備・アメニティ”、“風呂”、“食事”の6種類の各評価視点ごとに3人の作業者を置き、レビュー文数の5%までの重要文を選択してもらった。作業員間の一致率を測定するため、2人ずつF値を計算し、その平均をとった。評価視点6種類の一致率の平均は0.58であった。一致率がそれほど高くないのは、上記(ii)と(iii)が原因であると考えられる。

¹<http://travel.rakuten.co.jp/>

4 要約コーパス

前章で述べた問題点に対処するため、要約コーパスの設計を改めた。

4.1 段階的ラベル付け

重要文抽出により要約を作成する場合、レビュー文数の5%以内や10%以内の文を選択する。各レビュー文のみからその文に関する素性ベクトルを作成し、機械学習手法を用いて重要文を抽出することを考える。このとき、宿泊施設 H_1 と宿泊施設 H_2 のレビュー文書に、「(☆) レストランの食事がおいしいです。」という、内容がほぼ等しいレビュー文が存在すると問題となる。例えば、評価視点“食事”に関する H_1 のレビュー文書には、「レストランでおいしい郷土料理を食べられます。」や「レストランの朝ご飯はバイキングで、味も良いです。」などの文が多数存在し、 H_2 のレビュー文書にそのようなものが一切存在しない場合、文(☆)は、 H_2 の重要文の正例であるが、 H_1 の重要文の負例となる。このような矛盾する学習例が訓練コーパスに少なくない数存在すると、学習が困難になる。

そこで、我々は、次のように重要文候補という概念を定義する。

重要文候補 評価視点に関して重要な内容を含み、この文より重要度が高い文が存在しない場合、この文を最終出力としてもよいと判断できる。

ラベル付けにおいて、重要文であるかどうかは他の文も考慮して相対的に判断する必要があるのに対して、重要文候補であるかはその文単体で絶対的に判断できる。これは、機械学習手法においても同様である。

前章で述べた (iii) の問題に対処するため、各評価視点のカバー範囲を明確に規定した²。また、ラベル付けにおける判断の揺れを軽減するため、次のように関連文という概念を定義した。

関連文 評価視点に関連する文であると判断できる。重要であるかどうか、その宿泊施設についての記述であるかどうかを問わない。

以上をまとめると、各レビュー文に、次のように段階的にラベルを付与することになる。

関連文か? → 重要文候補か? → 重要文か?

4.2 コーパス構築

楽天トラベル:レビューデータを用いて、6種類の評価視点ごとに各レビュー文にラベルを付与し、要約コーパスを構築する。

対象の宿泊施設を決定するにあたり、楽天トラベル:レビューデータにおけるレビュー数の分布を調査した。調査の結果、90%以上の宿泊施設に対して、レビュー数は1から58の範囲にあることが分かった。レビュー数が1桁と少ない場合、評価視点ごとに要約を作る必要はないと考える。以上の理由により、レビュー数が10から58の範囲の宿泊施設の全体から、ランダムに40の宿泊施設を抽出し、ラベル付与の対象とした。

これらの宿泊施設に対するレビュー文書集合に対して、文分割を行った。レビュー文書は記法が統制されているわけではないので、単純に句点、「!」、「?」で区切るだけでは不十分である。我々は独自に文分割規則³を作成し、それに基づいて半自動的に文分割を行った。これにより、5,111のレビュー文を得た。この各レビュー文に対してラベル付けを行う。

4.3 コーパスの現状

コーパスは独自のXML形式で表現されており、人間が見やすいHTML形式に自動変換可能である。現在までに、関連文と重要文候補のラベル付けが一通り完了している。今後は、前章の(ii)の問題に対処しながら、重要文のラベル付けに取り組む予定である。

5 重要文候補抽出手法

前章でラベル付けした重要文候補を抽出する手法を提案する。本研究では、各レビュー文に対して素性ベクトルを構築し、SVMsを用いて重要文候補か否かを判定する。用いた素性は以下の5種類である：

評価視点の類義語、その共起表現、類義語の分布類似語、分布類似語の共起表現、機能表現

5.1 評価視点の類義語

日本語 WordNet[1] と分類語彙表[4] から抽出した、各評価視点の類義語(評価視点自身の“風呂”などを含む)の有無を素性として用いる。

5.2 評価視点の類義語の共起表現

前節の類義語名詞 n とそれと共起する述語 p の組 (n, p) のうち、後述する閾値を超えるものの有無を素性として用いる。本研究では、Hindle らの手法[3]を参考に、項を考慮した次式の共起スコアを用いる。

$$C_a(n, p) = \log \frac{\frac{f_a(n, p)}{N}}{\frac{f_a(n)}{N} \frac{f(p)}{N}} * DF_{a, n, p} \quad (1)$$

ここで、 a は項を表す変数であり、本研究では、ガ格、ヲ格、ニ格のいずれかをとる。 N はコーパス内の述語の数、 $f(p)$ は述語 p の出現頻度、 $f_a(n)$ は項 a に出現する名詞 n の頻度、 $f_a(n, p)$ は、述語 p の項 a に出現す

²この基準書は次のサイトで公開予定である。

<http://cl.cs.yamanashi.ac.jp/nldata/review-summary/>

³文分割規則に関しても前記のサイトで公開予定である。

表 1: 重要文候補抽出精度

評価視点	適合率		再現率		F 値	
	ベースライン	本手法	ベースライン	本手法	ベースライン	本手法
サービス	0.766	0.673	0.611	0.489	0.680	0.567
立地	0.794	0.734	0.587	0.484	0.675	0.584
部屋	0.776	0.803	0.662	0.670	0.714	0.730
設備・アメニティ	0.678	0.564	0.410	0.341	0.511	0.425
風呂	0.846	0.868	0.734	0.756	0.786	0.808
食事	0.859	0.819	0.720	0.810	0.784	0.814
平均	0.787	0.744	0.621	0.592	0.692	0.655

る名詞 n の頻度である。 $DF_{a,n,p}$ は、低頻度の語に対するバイアスを軽減するための discounting factor[5] である。形態素解析には JUMAN⁴を、構文解析と格解析には KNP⁵を使用した。訓練コーパスから、前節の類義語を含まない共起表現の組をランダムに 50 組抽出し、予備実験により、素性として利用する共起表現のスコアの閾値を定めた。

5.3 評価視点の類義語の分布類似語

5.1 節の類義語名詞 n との分布類似度が高い名詞 n' の有無を素性として用いる。次式により、2つの名詞の分布類似度を計算する。

$$SIM(n, n') = \sum_{p \in P} \sum_{a \in A} SIM_a(p, n, n') \quad (2)$$

ここで、 P はコーパス内の述語の集合、 $A = \{ \text{ガ格, ヲ格, ニ格} \}$ 、 $SIM_a(p, n, n')$ は、述語 p の項 a に関して、2つの名詞 n と n' がどれほど類似しているかを表す関数であり、 $C_a(n, p)$ と $C_a(n', p)$ が異符号の場合、0 を、同符号の場合、それらの絶対値の小さいほうを返す。訓練コーパスから、5.1 節の類義語を含まない共起表現の組をランダムに 50 組抽出し、予備実験により、分布類似度の閾値を定めた。

5.4 分布類似語の共起表現

前節の分布類似語名詞 n' とそれと共起する述語 p の組 (n', p) のうち、5.1 節で述べた閾値を超えるものの有無を素性として用いる。

5.5 機能表現

日本語機能表現辞書「つつじ」[7] に収録される文字列長 4 以上の機能表現が、単純な文字列一致で文中に存在する場合、その意味的等価クラスを素性として用いる。重要文候補の判断に有効であろうと考え、前節までに導入した内容語に加え、機能表現を素性として導入した。語義曖昧性解消を行わない場合、文字列が短い機能表現はノイズとなる可能性が高いため、それらを除外した。

⁴<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

⁵<http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

6 実験

6.1 実験方法

4 章で構築した要約コーパスを用いて、評価視点ごとに重要文候補抽出実験を行った。40 交差検定を行い、適合率、再現率、F 値で評価した。比較対象として、文に出現するすべての語（記号を含む）の出現頻度を素性として、SVMs を用いて重要文候補か否かを判定するベースライン手法を用意した。

6.2 実験結果

結果を表 1 に示す。本手法とベースラインの F 値はそれぞれ 0.655、0.692 であるため、本手法の有効性は示せなかった。評価視点別に見ると、“風呂”、“食事”、“部屋”についてはベースラインを上回り、特に、“風呂”と“食事”については F 値が 8 割を超えているため、本手法で用いた素性はある程度有効であると言える。しかし、残り 3つの評価視点では著しく精度が悪く、特に、再現率に関してはいずれも 0.5 を下回った。

そこで、各素性の有効性を検証するため、平尾らの手法 [10] を用いて、SVMs の学習結果から素性の重みを算出した。この値が大きい素性は、重要文候補の判断に有効な素性である。重み上位の素性 5 つを降順に表 2 に示す。表において、「辞書-」は 5.1 節、「ガ-」や「ヲ-」、「ニ-」は 5.2 節か 5.4 節、「類-」は 5.3 節、「機能-」は 5.5 節の素性であることを表す。この表より、ベースラインを上回った“風呂”、“食事”、“部屋”に関しては、「温泉」や「焼き」、「ベット」など、その評価視点の関連語が上位に現れている。しかし、残りの評価視点については、「面」や「後」など、その評価視点に関連しない語が上位に現れている。この原因として、辞書から抽出した類義語を直接使用したことに問題があると考えられる。例えば、“立地”の場合、“配置”や“サイト”など評価視点に関連しない語が存在していた。従って、より評価視点に関係のある語の抽出法を検討する必要がある。共起表現については、“設備・アメニティ”や“立地”、“サービス”では上位に出現しなかった。これに関しては、利用した係り受け

表 2: 素性の重みを降順にソートした結果

サービス		立地		部屋	
ベースライン	本手法	ベースライン	本手法	ベースライン	本手法
安い	類-“用意”	立地	類-“湖”	部屋	類-“ツイン”
!	類-“精算”	海	類-“朝市”	景色	ガ-“窓 開く”
サービス	類-“後”	が	類-“前”	が	類-“ベット”
の	機能-“不許可”	景色	類-“市”	夜景	類-“布団”
スタッフ	類-“箱”	夜景	類-“見晴らし”	サイコー	類-“見晴らし”
設備・アメニティ		風呂		食事	
ベースライン	本手法	ベースライン	本手法	ベースライン	本手法
アメニティ	類-“用意”	風呂	辞書-“サウナ”	朝食	類-“焼き”
少なめ	類-“コテージ”	温泉	ヲ-“湯 楽しめる”	美味しい	類-“コース”
が	類-“面”	露天風呂	ヲ-“湯 塗る”	おいしい	類-“質”
トイレ	類-“ショップ”	は	類-“風呂”	料理	類-“量”
ベッド	類-“ベッド”	だらけ	類-“温泉”	食事	類-“酒 飲める”

表 3: 機能表現に関する素性上位 3 つとその重み

サービス		立地		部屋		設備・アメニティ		風呂		食事	
クラス	重み	クラス	重み	クラス	重み	クラス	重み	クラス	重み	クラス	重み
不許可	0.286	否定	0.033	否定	0.004	逆接	0.261	逆接	0.022	否定	0.004
授受	0.286	不必要	-0.000	添加	0.004	判断	0.058	許可	0.004	許可	0.000
逆接	0.163	可能	-0.000	許可	0.001	逆接	0.017	判断	0.000	添加	-0.000

の種類をガ格, ヲ格, ニ格に限定したためと考えられる。例えば, “立地” の場合, 「駅から近くて便利でした。」や「空港からのアクセスは満足できます。」など, カラ格も考慮した方が良いと考えられる。今後は, 利用する格についても検討する必要がある。

機能表現の素性上位 3 つとその重みを表 3 に示す。この表において, “サービス” に対する重みが他の評価視点よりも高い。このことから, 機能表現は “サービス” に対して有効であると言える。この理由は, 例えば, “授受” の場合, “サービス” に関する文は, 「従業員の方が～してくださいました。」のように「てくれる」形で書かれることが多いためである。また, “逆接” の場合, 「予定より遅く到着したにも関わらず, 丁寧な対応でした。」のように, 「にも関わらず」などの表現を伴う文が “サービス” に多いためである。他の評価視点では重みは 0 に近いため, このことから, 機能表現が “サービス” の重要文候補抽出に特に有効であると言える。

7 おわりに

本論文では, 現在構築中の要約コーパスについて報告し, 重要文抽出の前処理である重要文候補抽出についての手法を提案した。いくつかの評価視点において, ベースライン手法を上回ることができなかったため, 今後は, 利用する素性について検討する必要がある。

構築中のコーパスは, 楽天データとの差分形式で, 前記のサイトにて無償で一般公開する予定である。

謝辞: 本研究の一部は, 科研費若手研究 (B) 「高精度モダリティ解析のための言語資源構築に関する研究」(課題番号: 23700176, 代表: 松吉俊) の支援を受けている。

参考文献

- [1] Francis Bond, Hitoshi Isahara, Kyoko Kanzaki, and Kiyotaka Uchimoto. Boot-strapping a WordNet using multiple existing WordNets. In *Proceedings of the 6th International Language Resources and Evaluation (LREC2008)*, 2008.
- [2] G. Erkan, D. Radev. Lexpagerank:prestige in multi-document text summarization. In *Proc. of the Conference on Empirical Method in Natural Language Processing*, pp. 365–371, 2004.
- [3] Hindle, D. Noun classification from predicate-argument structures. *ACL-90*, pp. 268–275, 1990.
- [4] 国立国語研究所. 分類語彙表 [増補改訂版]. 大日本図書, 2004.
- [5] Patrick Pantel and Deepak Ravichandran. Automatically labeling semantic classes. In *Proc. of HLT/NAACL-04*, pp. 321–328, 2004.
- [6] 波多野匡, 嶋田和孝, 遠藤勉. クラスタリングを利用した評価文のアスペクト推定. 情報科学技術フォーラム講演論文集, Vol. 9, No. 2, pp. 103–134, 2010.
- [7] 松吉俊, 佐藤理史, 宇津呂武仁. 日本語機能表現辞書の編纂. 自然言語処理, Vol. 14, No. 5, pp. 123–146, 2007.
- [8] 阪野慎司, 松原茂樹, 吉川正俊. 機械学習に基づく判決文の重要箇所特定. 言語処理学会第 12 回年次大会発表論文集, pp. 1075–1078, 2006.
- [9] 新井智也, 佐藤哲司. 評価視点別の言及度を用いた意見文の分類手法の提案. In *DEIM Forum, A2-2*, 2010.
- [10] 平尾努. 機械学習による複数文書からの重要文抽出. 自然言語処理, Vol. 10, No. 1, pp. 81–108, 2003.