

建設的ニュースコメントの順位付けのためのデータセット構築

藤田 綜一郎^{1,a)} 小林 隼人^{2,b)} 奥村 学^{3,c)}

概要：オンラインニュースサイトには、ニュース記事毎に読者の議論の場としてコメント欄が設けられているものがある。そこではニュースに関する建設的なコメントが投稿される一方で、差別的または特定の主義主張に偏ったコメントも投稿されるため、活発な議論を促すコメントが埋もれてしまう可能性があるという問題がある。そこで、本稿では建設的なコメントを優先的に提示することによる議論の活性化を目指し、コメントを建設的なコメント順に並び変えることを試みる。具体的には、コメントの建設的度合いをスコア化したデータセットを構築し、ランキング学習を用いて順位付けを行う。加えて、その応用として、効率的なデータの拡張による順位付けの精度向上を目的に、能動学習を用いてランキングモデルを学習する方法について考察する。

Dataset Creation for Ranking Constructive News Comments

FUJITA SOICHIRO^{1,a)} KOBAYASHI HAYATO^{2,b)} OKUMURA MANABU^{3,c)}

1. はじめに

オンラインニュースサイトには、ニュース記事毎に読者の議論の場としてコメント欄が設けられているものがある。コメント欄の例として、Yahoo!ニュースにおけるある記事の上位3件のコメントを図1に示す。このサービスでは、読者がコメントに対して返信したり、「そう思う/そう思わない」のフィードバック（以下、読者評価とする）をつけてコメントを評価したりすることができる。図中のコメントには、「人として終わっている」という必要以上に攻撃的な内容や、「圧力があつたと思う人はクリック」という読者評価目的のコメントが見られる。この例のように、一部の記事のコメント欄では攻撃的または特定の主義主張に偏ったコメントが投稿されることがあり、活発な議論を促す建設的なコメントが埋もれてしまう可能性があることが問題視されている。



図 1: Yahoo!ニュースのコメント欄の例

そこで本稿では、コメント欄の改善を目的に、コメントを建設的度合いでランク付けするモデルの構築を行う(3.3節参照)。そのためにまず、「建設的」を定義付けする(2.1節参照)。次に、事前調査としてモデル構築の前に読者がどのようなコメントを建設的と感じるのかについて分析を行った(2.2節参照)。そして、建設的度合いに基づいてスコア付けしたデータセットをクラウドソーシングを用いて作成した(2.3節参照)。また、作成したデータセットを用いて、効率的なデータセット拡張の方法についてシミュレーションを行った(4節参照)。

¹ 東京工業大学大学院
² ヤフー株式会社
³ 東京工業大学科学技術創成研究院
a) fujiso@lr.pi.titech.ac.jp
b) hakobaya@yahoo-corp.jp
c) oku@lr.pi.titech.ac.jp

本研究の貢献は、次の通りにまとめられる。

- 建設的なコメントと読者評価順との関係を調査し、読者評価と建設的度合いが必ずしも関係しないことを示した。
- 約 10 万件のコメントに対して建設的度合いのラベル付けを行い、ランキング学習用のデータセットを作成した。
- 効率的なデータ拡張方法について検証し、約 6.25% のデータ作成コストで全データで学習したものと同等の精度を達成した。

2. データ分析及びデータセット構築

本節では、記事のタイトルとコメントを用いてデータ作成、分析を行う。まず、2.1 節で「建設的」の定義を行う。次に、2.2 節でどの程度のコメントが建設的と判断されるのか、建設的度合いと読者評価には関係性があるのかについて調査する。2.3 節でクラウドソーシングを用いてデータセットを構築する。

2.1 「建設的」の定義

「建設的」という単語は一般に、物事の成立や進行をおし進めようとするさまのことを指す [10] が、これだけでは具体性に欠けるため、建設的かどうかの判断に無視できない個人差が現れる可能性がある。そのため、データセットを構築する前に、何をもって「建設的」と呼ぶかについて具体的に定義する必要がある。「建設的」の定義については、Kolhatkar らの survey monkey^{*1} で 100 人にアンケートを取り、投稿コメントが建設的となる条件を複数決定した例 [3] がある。我々は、彼らの定義を参考に（一部簡略化して）、建設的コメントの定義を表 1 に示す前提条件と主条件により設定した。前提条件は最低限のコメントの質を担保する条件で、主条件は建設的となりうるシチュエーションを示す条件である。本稿では、前提条件を満たし、かつ主条件の項目を 1 つ以上満たすものを建設的なコメントと定義する。

表 1: 建設的コメントとなる条件

前提条件	● 記事に関係しており、誹謗中傷ではない
主条件	<ul style="list-style-type: none"> ● 自分の意見をもとに議論を引き起こそうとしている ● 客観的で、必要であれば証拠や根拠を提示している ● 新たな考え方、解決案、洞察を提供する発言 ● 記事に関係する珍しい経験談

2.2 事前調査

本稿では、データとして、Yahoo!ニュース記事のコメントを扱う。その中でも、返信でないコメントが 100 件以上ある記事の、10 文字以上 125 文字以下のコメントを使用した。データ作成には、Yahoo!クラウドソーシング^{*2} を利用

^{*1} <https://www.surveymonkey.com/>

^{*2} <https://crowdsourcing.yahoo.co.jp/>

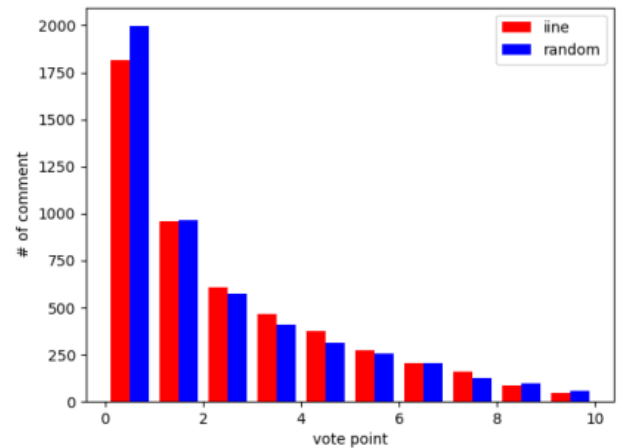


図 2: 建設的スコアの分布

した。

我々は建設的なコメントと読者評価の関係性を調べるために、ニュース記事 1000 件を対象に事前調査を行った。具体的には、各記事から読者評価スコアが高い 5 コメントとランダムに選んだ 5 コメントを抽出し、作業者に各コメントが建設的であるかどうかを判定してもらう。読者評価にはプラス評価（「そう思う」）とマイナス評価（「そう思わない」）の二つが付与されている。今回使用する 5000 件の記事についてのコメント全ての「そう思う」と「そう思わない」の数を集計すると、約 5:1 の割合で付与されていた。そのため、読者評価スコアを次式で定義して並び替えを行った、

$$\text{読者評価スコア} = \text{「そう思う」数} - 5 \times \text{「そう思わない」数} \quad (1)$$

クラウドソーシングのタスクでは、記事のタイトルと 5 件のコメントを提示し、それぞれのコメントに対し建設的と感じるか否かを判定してもらう。各コメントにつき 10 人に判定してもらい、建設的として判定された票数を集計する。以後、この票数を建設的スコアと呼ぶ。例えば、10 人中 8 人が建設的と判断したコメントの建設的スコアは 8 となる。仮に、建設的なコメントが読者によりプラス評価されているのであれば、読者評価順はランダムに選択するより建設的スコアの平均が高くなるはずである。

建設的スコアの分布を図 2 に示す。横軸が建設的スコア、縦軸がそのスコアがついたコメント数となっている。平均スコアは読者評価が高い 5 件選択 (iine) が 1.981 で、ランダム 5 件選択 (random) が 1.857 となった。また、読者評価と建設的スコア間の相関係数は -0.0036 となり、両者に相関関係は見られなかった。このことから、読者評価を建設的の指標として用いることは難しいという結論が得られた。また、建設的スコアの高いコメントについて分析してみると、全体のコメントの文字数の平均が 34.38 文字であるのに対し、建設的スコアが 7 以上の平均は 54.70 文字となり、20 文字程度長くなるという特徴が得られた。

表 2: 作成したデータセットの統計情報

	記事数	コメント数	コメント/記事	スコア範囲
deep	650	65,000	100	0 ~ 40
shallow	8,858	44,290	5	0 ~ 10
test	200	20,000	100	0 ~ 40

2.3 データセット構築

2.2 節より、読者評価をスコアとすることは不適切であると判断したため、コメントに建設的スコアを付与し、データセットを構築する。スコア付与は、2.1 節と同様にクラウドソーシングで行う。作業者は、まずタスクを行う前に建設的コメントの定義と、その判別例を読む。その後、提示された記事タイトルとコメントを読み、そのコメントが建設的かどうかを判別する。ここで、回答の信頼性を担保するために、我々が正解を用意したチェック質問によるフィルタリングを行った。チェック質問は、建設的でないコメントの中で、明らかに建設的なコメントが一つだけ含まれており、タスク中にランダムに複数回出題される。建設的スコアは、そのチェック質問に正解した作業者の結果のみを集計した。

多数の記事の全てのコメントに対してスコア付与することは、膨大なコストがかかり、現実的でない。そのため、我々は 2 種類の学習用データとして記事の数を重視したものと、記事あたりのコメント数を重視したものを作成した。テスト用データを加えると 3 種類のデータセットとなる。3 つのデータセットの詳細は、次のようにまとめられる。

- shallow データ
記事の網羅性を重視して、記事数を増やした学習用データ。各記事ランダムに選んだ 5 件のコメントに対して、0 ~ 10 の建設的スコアを付与した（各コメントに作業者 10 人）。
- deep データ
記事内のコメントの順位付けを重視して、記事あたりのコメント数を増やした学習用データ。各記事ランダムに選んだ 100 件のコメントに対して、0 ~ 40 の建設的スコアを付与した（各コメントに作業者 40 人）。
- test データ
deep データと同様の条件で作成した、テスト用データ。実際のサービスの環境を想定している。

構築したデータの統計情報を表 2 に示す。また、表 3 に、作成された deep データの例を示す。例 1 ~ 3 のコメントを比較すると、例 1 と例 2 は自分の意見を述べているが、例 1 は意見に理由をつけているため、より建設的であると捉えられている。例 3 は攻撃的な内容のため、建設的でないとして捉えられていることが分かる。

表 3: 「酒・タバコの 18 歳解禁」についてのニュースに対するコメントと建設的スコアの例

	コメント	スコア
例 1	将来的には酒やタバコを必要としない社会を目指すべき。どちらも健康被害、事故や事件の原因になるから。	30
例 2	自由を与えるなら刑罰も厳しく与えるべき	16
例 3	煙草を吸うから馬鹿になるのか、馬鹿だから吸うのか。	1

3. 建設的コメントの順位付け実験

3.1 実験設定

ランキング学習では、記事ごとにコメントを各データセットは訓練データと検証用データに 9:1 の割合で分割して実験を行った。比較するランキング手法を次に示す。

- Good
式 (1) によって算出された読者評価スコアでランキング。
- Random
ランダムにランキング。10 回の結果を平均。
- Length
コメントの文字数でランキング。2.2 節のコメントが長い方が建設的と取られやすいという知見を元にして。
- RankSVM
2 つのコメントの出力を比較し、建設的スコアの高い方が出力値も高くなるように相対的に学習。出力値の大きい順にランキング。deep データと shallow データのそれぞれで学習する。
- SVR
コメントの建設的スコアを学習。予測スコアの高さでランキング。deep データと shallow データのそれぞれで学習する。

RankSVM と SVR の実装は、公開されている liblinear-rankSVM^{*3} [5] を利用した。実験では、線形カーネルを用い、分類器のコストパラメータは、 $\{2^{-13}, 2^{-12}, \dots, 2^0\}$ の中から検証用データを用いて調整した。

また、分類器の素性には次のものを使用した。

- コメントの Bag-of-Words (BoW)
- コメント内の出現語彙数
- コメントと記事タイトルの cos 類似度
- コメントと記事タイトルの共起単語 BoW

3.2 評価方法

ランキングの有効性の評価として、情報検索などのランキング問題の評価指標として広く用いられている Normalized Discounted Cumulative Gain (NDCG) [1] を用いた。NDCG@ k は、ランキング結果上位 k 件において、正解のラ

*3 <https://github.com/FurongPeng/liblinear-ranksvm>

表 4: 建設的コメントの順位付け実験における NDCG@k の結果 .

	NDCG@1	@3	@5	@10	@20	@50
Good	0.0004	0.0041	0.0189	0.0663	0.1057	0.1674
Random	0.0032	0.0258	0.0281	0.0558	0.0747	0.1279
Length	0.1217	0.1806	0.2064	0.2799	0.3642	0.4205
RankSVM (shallow)	0.3036	0.3519	0.3980	0.4677	0.5269	0.5668
SVR (shallow)	0.2729	0.3380	0.4079	0.4758	0.5265	0.5670
RankSVM (deep)	0.2658	0.3661[†]	0.4200[†]	0.4929[†]	0.5517[†]	0.5822[†]
SVR (deep)	0.2145	0.2858	0.3382	0.4236	0.4951	0.5299

ランキングとの近さを示しており、値が大きいほど正解に近い。NDCG@k は次式で表される。

$$\text{NDCG}@k = Z_k \sum_{i=1}^k \frac{r_i + 1}{\log_2(1 + i)} \quad (2)$$

ここで、 r_i はモデルによって順位付けられた i 番目のコメントの建設的スコアを示している。また、 Z_k は正規化項であり、正解と同じ順序で並べたときに最大値 1 をとるように設定されている。

3.3 結果と考察

表 4 に、deep データと shallow データそれぞれを SVR, RankSVM によりランキングを行い、NDCG@k を算出した結果を示す*4。結果を見ると、次の 3 点の特徴が読み取れる。

- 読者評価と建設的スコアの関係

Good と Random を比較すると、NDCG@k は $k \leq 5$ のとき Random の方が高く、 $k \geq 10$ の時 Good の方が高い。ここから、建設的なコメントは読者評価上位ではなく $10 \leq k \leq 50$ あたりの中位に多く分布しており、読者評価は建設的スコア上位のランキングと関連性が薄い事が分かる。

- 手法の比較

RankSVM と SVR について比較すると、shallow データでは両者の NDCG@k に有意差なし、deep データでは、NDCG@k に有意差ありという結果が得られた。このことから、ランキング手法として、SVR よりも RankSVM の方が適当といえる。

- データ構築方法の比較

shallow データと deep データで学習した RankSVM, SVR について NDCG@k を比較すると、RankSVM では deep データ、SVR については shallow データが有意に高いという結果が得られた。

総じて、deep データを用いて RankSVM で学習する方法が最も優れている事が確認できた。

*4 †は有意水準 5%とした時、他の手法との差異が、統計的に有意である事を示す。検定は記事ごとの NDCG@k の対応あり t 検定を行った。

4. データ作成のコスト削減に関する考察

本節では、構築した deep データを静的なデータプールとして、様々なデータ構築手法を試すことで、学習データ作成コスト削減の方法について考察する。4.1 節では、建設的スコアの削減について検証する。4.3 節では、能動学習を用いた 1 記事あたりのコメント数の削減法について検証する。

4.1 建設的スコアの削減

deep データは現在、0 ~ 40 の建設的スコアが付与されているが、例えば建設的スコアを半分にした 0 ~ 20 でも、同等の精度が得られるのであればスコア付与コストは半分に削減できる。しかし、スコアが減ると、同率順位となるコメントが増加し、詳細なランキングが難しくなることが予想される。

我々は、建設的スコアの最大値を 1 から 40 まで変化させた時、RankSVM の NDCG@k がどのように変化するかを調査した。建設的スコアの 40 から n ($n < 40$) への圧縮は、建設的か否かの 40 人の判定結果の中から n 件をサンプリングすることによって実現した。結果については、4.3 節でまとめて報告する。

4.2 コメント数の削減

ここでは、能動学習を用いて、既存のデータにラベル付きコメントを逐次的に追加する手法について述べる。記事集合 A 中の記事 a について、これまでにスコアが付けられたコメント集合を D_a 、まだスコアが付いていないコメント集合を X_a とする。能動学習の目的は、 D_a に基づく順位付けモデルや D_a そのものを用いて、最終的な順位付けモデルを改善する可能性の高いコメント \tilde{x} を抽出してスコア付けすることである。本稿では、そのようなコメントを下記の方法により抽出する。ただし、関数 f はスコア付きデータ D で学習した RankSVM であり、 x^* は、コメント x から抽出した特徴量である。

- random

ランダムに追加するコメントを選択する。

- best

表 5: 建設的スコアと NDCG@k の値

	NDCG@1	@3	@5	@10	@20	@50
0~1	0.2396	0.3259	0.3805	0.4502	0.5217	0.5543
0~3	0.2537	0.3322	0.3935	0.4683	0.5426	0.5745
0~5	0.2657	0.3433	0.4082	0.4825	0.5426	0.5745
0~10	0.2984	0.3656	0.4260	0.4998	0.5579	0.5880
0~20	0.2996	0.3827	0.4311	0.5058	0.5646	0.5923
0~40	0.2933	0.3749	0.4302	0.5005	0.5583	0.5886

上位のランキングを重視した手法．RankSVM が最大値をとる，つまりスコアなしコメント中でランキング 1 位に来るコメントを選択する．高い順位に来るコメントに優先してスコアをつけることで，上位のランキング性能をより向上させる狙いがある，

$$\tilde{x} = \operatorname{argmax}_{x \in X_a} f(x^*) \quad (3)$$

• cosine

コメントの内容の網羅性を重視した手法．スコア付きコメント $d \in D_a$ と x の \cos 類似度を取り，平均値が最も低いコメントを選択する． \cos 類似度が低い，スコア付きデータ D に見られない加えることで般化性能を上げる狙いがある．

$$\tilde{x} = \operatorname{argmin}_{x \in X_a} \sum_{d \in D_a} \cos(d^*, x^*) \quad (4)$$

• uncertainty

RankSVM のランキングの不確実性に注目した手法．スコア付きコメント $d \in D_a$ と x の出力が近いコメントを選択する．近い順位にランキングされるコメントにスコア付けすることで，出力が曖昧なペアの順位を明確にする狙いがある．

$$\tilde{x} = \operatorname{argmin}_{x \in X_a} \sum_{d \in D_a} \|f(x^* - d^*)\| \quad (5)$$

• local and global

Qian らの手法 [8]．local uncertainty と global uncertainty の二つの指標を合わせて判定する．global uncertainty で X 中のコメント全てを考慮しながら式 (5) と同様に出力が曖昧なペアを探している．ここで p はハイパーパラメータであり，実験では $p = 1$ とした．

$$local(x) = \sum_{d \in D_a} |f(x^* - d^*)|^{-1} \quad (6)$$

$$G(x) = \frac{f(x^* - d^*) - \min_d f(x^* - d^*)}{\max_d f(x^* - d^*) - \min_d f(x^* - d^*)} \quad (7)$$

$$global(x) = - \sum_{x \in X_a} G(x) \log G(x) \quad (8)$$

$$\tilde{x} = \operatorname{argmax}_{x \in X_a} global(x) \times local(x)^p \quad (9)$$

選択された \tilde{x} はラベルをつけて D_a に追加する．この，

各記事 1 コメントずつ選んでラベルを追加する処理を繰り返すことによって．データセットを拡張してゆく．実験では，最初に deep データから各記事 2 コメントをランダムで取り出してラベル付きコメント D ，それ以外をラベル無しコメント X と設定する．そこから，能動学習により，各記事 1 コメントずつラベル付きコメントを追加してゆく．それを 1 記事 50 コメントになるまで，同様の操作を繰り返した．

4.3 結果と考察

図 3 に建設的スコアの最大値と NDCG@k の変化のグラフを示す．横軸が建設的スコアの最大値，縦軸が NDCG@k の値を示す．グラフより，1 ~ 10 までは建設的スコアとともに NDCG@k も上昇しているが，それ以降は横ばいで NDCG@k の向上が見込めないことが読み取れる．また，表 5 に代表的な値の例を示す．最大スコア 10 以上とそれ未満を比較すると，有意な差が見られたが，最大スコア 10 と 20, 40 を比較すると，有意な差が見られなかった．ここから，学習データの建設的スコアは，0 ~ 10 で十分であることが分かった．

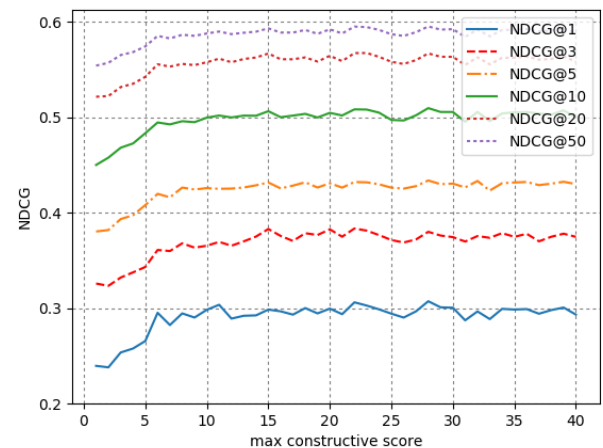


図 3: 建設的スコアの最大値による変化と NDCG@k の関係

この結果を受けて，学習データの建設的スコアを 0 ~ 10 にして，能動学習を行った．図 4 に能動学習を用いて記事毎にコメントを一つずつ増やした際の NDCG@3 の変遷を示す．横軸は 1 記事あたりのコメント数，縦軸は NDCG@3 の値となっている．ここで，NDCG@3 に焦点を当てた評

価を行っているのは Yahoo!ニュースの記事ページにおける表示がコメント 3 件だからである。学習初期では、データの数が少ないため、random に追加しても NDCG の向上が見られるが、10 コメント/記事を越えると、random と他の手法では大きく差がついた。手法としては、best が最もよく、25 コメント/記事で 100 コメント/記事で学習した時と同等の結果が得られた。よって、1 コメントあたりのコストを 75%、カットした上でコメント数を 75%削減して同等の精度を得ることができ、作成した deep データの約 6.25%のコストで、NDCG@3 において同等の精度を得ることができた。

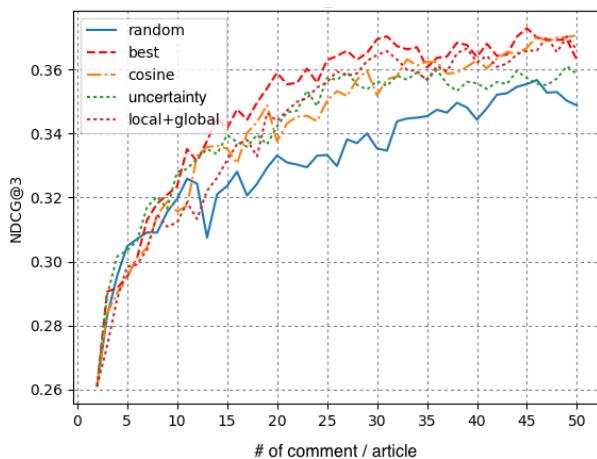


図 4: 能動学習手法毎の NDCG@3 の値の遷移

5. 関連研究

5.1 オンラインフォーラムに関する研究

ニュースのコメント欄をはじめとするオンラインフォーラムについては、近年多くの研究がなされてきた。中でもコメントのランキングについては、読者評価を予測しランク付けする研究 [2] や、説得力のあるコメントをランク付けする研究 [9] など様々な観点から順位付けをする試みがされている。しかし、「建設的」という観点でのランク付けは未だ行われていない。また、これらは正解ラベルに読者評価スコアを用いている点で、2.2 節で示した通り、読者評価と相関がない建設的に関するランキングとは大きく異なる。

「建設的」という観点は、議論分析のフレームワークでしばしば導入される。オンラインフォーラムにおける議論分析のデータセット YNACC [7] では、Yahoo News の返信スレッド (親コメントと、それに対する複数の返信コメント) に対して、建設的な議論であるか否かのラベルが付与されている。本稿にもっとも関係する研究は、Yahoo News のコメントについて、建設的なコメントか否かの分類を行った研究 [4] である。彼らは、YNACC のスレッドレベルの建設的ラベルがついているコメントを全て建設的コメントとみなし、二値分類器を学習している。

5.2 能動学習

能動学習は、教師データを追加で作成する際に、分類器への影響が大きいデータを選択して追加してゆく方法である。大量のデータ全てに教師ラベルを付けると非常にコストがかかる。そこで、精度が向上するデータだけをラベル付けしてデータに追加することで、少ないラベル付けでの精度向上を実現出来る。ランキング学習の為のデータ選択手法も多数考案されている。例えば、期待損失が小さくなるように選択する手法 [6] や、出力が曖昧なペアを選択する手法 [8] がある。後者の手法はコメント間の曖昧さと、ベクトル空間上でのエントロピーを合わせて、選択している。本研究では、ペアをデータに加える代わりに、既存のデータに対していいペアが多く作れるようなコメントの一つ取得するように改良している。

6. まとめ

本研究では、ニュースのコメントについて建設的度合いをスコア付けしたデータセットの構築を行った。また、ランキング学習により、建設的コメントを順位付けするモデルを構築した。応用として、能動学習を用いて効率的なデータセット拡張のシミュレーションを行い、約 6.25%のデータ作成コストで全データで学習したものと同等の精度を達成できることを確認した。

今後の展望として、次の 3 つが考えられる。第一に、今回構築したモデルを実際にサービスに組み込むことである。建設的コメントの提示が読者評価にどのような影響を与えるのか、分析を行いたい。第二に、能動学習により得られた知見を用いて shallow データを拡張し、さらに精度の良いモデルの構築を行うことである。第三に、1 章で述べたコメント欄に対するアプローチとして、建設的かつ内容が異なるコメントを提示することである。今回のデータセットを利用して、要約モデルを構築したい。

参考文献

- [1] Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N. and Hullender, G.: Learning to rank using gradient descent, *Proceedings of the 22nd international conference on Machine learning*, ACM, pp. 89–96 (2005).
- [2] Hsu, C.-F., Khabiri, E. and Caverlee, J.: Ranking comments on the social web, *Computational Science and Engineering, 2009. CSE'09. International Conference on*, Vol. 4, IEEE, pp. 90–97 (2009).
- [3] Kolhatkar, V. and Taboada, M.: Constructive Language in News Comments, *Proceedings of the First Workshop on Abusive Language Online*, pp. 11–17 (2017).
- [4] Kolhatkar, V. and Taboada, M.: Constructive Language in News Comments, *Proceedings of the First Workshop on Abusive Language Online*, pp. 11–17 (2017).
- [5] Lee, C.-P. and Lin, C.-J.: Large-scale linear ranksvm, *Neural computation*, Vol. 26, No. 4, pp. 781–817 (2014).
- [6] Long, B., Chapelle, O., Zhang, Y., Chang, Y., Zheng,

- Z. and Tseng, B.: Active learning for ranking through expected loss optimization, *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, ACM, pp. 267–274 (2010).
- [7] Napoles, C., Pappu, A. and Tetreault, J. R.: Automatically Identifying Good Conversations Online (Yes, They Do Exist!)., *ICWSM*, pp. 628–631 (2017).
- [8] Qian, B., Li, H., Wang, J., Wang, X. and Davidson, I.: Active learning to rank using pairwise supervision, *Proceedings of the 2013 SIAM International Conference on Data Mining*, SIAM, pp. 297–305 (2013).
- [9] Wei, Z., Liu, Y. and Li, Y.: Is this post persuasive? Ranking argumentative comments in online forum, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vol. 2, pp. 195–200 (2016).
- [10] 松村明ほか：大辞林，第3版，三省堂，p. 11 (2006).