

MMRを用いた文書ランキングのための能動学習

エルゲン瑛夏[†] 鈴木 優[†]

[†] 岐阜大学工学部電気電子・情報工学科 〒501-1193 岐阜県岐阜市柳戸1番1

E-mail: †y3033024@edu.gifu-u.ac.jp, †ysuzuki@gifu-u.ac.jp

あらまし 本研究ではレビュー文を有用性の高い順に並び替えることを目的としている。有用性とは商品の購入を決定付ける際に参考になる度合いのことを指す。本論文ではその一環としてランキング推定を行うモデル構築に使用するデータを削減することを目的とする。通販サイトは良いイメージを持たせるために出品者側が自演的に投稿したレビューがあること、投稿されているレビュー文の数が多いという理由から商品購入者が有用なレビューを見つけることは難しい。そこで商品購入者が有用なレビューを見つけやすくするためにレビュー文を有用性の高い順に並び替えることを考えた。モデルの学習はランキング結果の上位の予測に重きを置くことができるランキング学習を用いて行った。学習の際に用いるデータのラベル付けには一対比較法というアンケート方法を採用した。一対比較法では対象評価物中の総当たりペアについてアンケートを行う必要があり、作業数が膨大になってしまう。そこで小データでも予測精度を高められる手法が必要であると考えた。本論文では文書群に情報の多様性を持たせて並び替える MMR と能動学習の一手法である Uncertainty Sampling を組み合わせて訓練データに追加するデータの選択を行いモデルを学習した。評価実験の結果 MMR と Uncertainty Sampling を組み合わせたものは MMR 単体、ランダムと比べて実験設定のほとんど全データ区間において一致率が上回ったが、Uncertainty Sampling 単体のものと比べると実験設定の全データ区間において一致率が下回った。この結果から本論文の実験設定において提案手法は有効であるとは言えなかった。

キーワード 機械学習, 自然言語処理, ランク学習, 一対比較法, MMR, 能動学習

1 はじめに

通販サイトの利用者の商品購入を決定づける要素はいくつかあり、その中でも大きな要素として商品に投稿されたレビューがある。通販サイトには出品者側がイメージを良くするために投稿した自演的なレビューや、キャンペーン目当てにレビューを寄せている人が多くいる事による質の低いレビューが多くあること、投稿されているレビューが多いという現状がある。そのため購入を決定するのに参考になるレビューは見つけにくい。この問題を解消するために通販サイト上のレビューのページにはレビューに投稿された「役に立った」の投票数順のソートが存在するが、「役に立った」の投票が上位の数レビューに集中しているという現状があり不便と感ずることが多い。そこで本研究ではレビューを有用性の高い順に並び替えることを目指す。

通販サイト利用者は、有用性の高い順に並び替えされたレビューを上から順に見ていくと考えられる。そこで本研究では予測ランキングの上位の正確さに重きを置くために、ランキング学習を用いてモデルの学習を行った。ランキング学習とはモデルによる予測スコアではなく、予測ランキングの正確性を高めることを目的とした学習のことである。

学習の前段階として対象レビューそれぞれに有用性についてのラベルを付ける必要があり、その際のラベル付けの手法はいくつか考えられる。一つ目に数値評価によるラベル付けがある。その中でも5段階による数値評価を例に挙げる。レビューの有用性という曖昧な評価軸で評価するときに最低評価の1と最高

評価の5は判断が簡単だが、2,3,4の違いをを厳密に評価することは難しい。また、アンケート中に評価軸がブレてしまうことも考えられる。二つ目に、有用・非有用の2値によるラベル付けがある。この方法では各評価対象を有用であるか非有用であるかで評価するため、アンケートとしては簡単である。しかしレビューに対して有用性の度合いを評価することはできない。そこで本研究では一対比較法というアンケート方法を採用した。

一対比較法とは作業者に二つの評価対象を提示してその優劣を問うという形式のアンケート手法である。これによりレビューの有用性という曖昧な評価軸でも、評価軸がブレることがなく簡単にアンケート結果を得られると考えた。また、数値評価や有用・非有用の2値ラベルでは似た文書に同じラベルが付くことが多いが、一対比較法では評価対象群中で優劣の結果を集計することで、レビューの有用性の度合いをより詳細な段階で評価することができる。評価対象の総数が100個の場合を例に考える。評価対象中で総当たりのペアを作って評価対象それぞれについて、各ペアの中で優れているとされた数を集計する。これにより各評価対象の勝利数が算出され、それをもとに全体をソートすることで100段階で有用な度合いを求めることができる。しかし一対比較法では評価対象群中の総当たりペアについてアンケートを行う必要があり、作業数が膨大になってしまう。そこで、入力する学習用データを減らしても評価指標の精度が下がらないような効率の良い学習を行うことができれば、アンケート作業数の削減につながると考えた。

学習の早い段階で多様性のあるデータを入力することによる学習効率の向上を期待し、MMR(Maximal Marginal Rele-

vance) [1] という手法を用いてアンケートペアの並び替えを行った。MMR とは周辺関連性を最大化することを目的に文書群を並び替える手法であり、情報の冗長性が少ないように文書群を並び替えることができる手法である。これを用いて対象レビュー群の総当たりペアの並び替えを行い、上位のアンケートペアから作業を行うことで学習に用いる小データの区間で学習効率を高めることができると考えた。

また、Uncertainty Sampling によるデータ選択も併せて行った。Uncertainty Sampling とは機械学習の一手法である。ラベル付けのコストが高い問題設定に対して、有効な能動学習の手法のことであり、現状のモデルによる予測が不確かなものから学習データに追加していく方法である。これと MMR を組み合わせることでデータの多様性を考慮しながら能動学習を行った。これにより学習効率が向上すると考えた。

評価実験としてモデル構築に使用するデータ数を変更しながらモデルを学習し、レビューの有用性予測の精度を評価した。ランダムによる削減、MMR と Uncertainty Sampling 単体による削減、提案手法による削減それぞれにおいて評価指標の精度の推移を比較した。その結果、MMR と Uncertainty Sampling を組み合わせたものは MMR 単体、ランダムと比べてほとんど全データ区間で一致率が上回ったが、Uncertainty Sampling 単体と比べると全データ区間において一致率が下回った。この結果から、本論文の実験設定において提案手法はモデル構築の際の使用データ削減に有効であるとは言えなかった。

本論文では MMR を用いた文書並び替えを考慮した能動学習手法を提案した。本手法では MMR で並び替えた上位 20% を対象に Uncertainty Sampling を用いているため、その際の対象とする割合を変えることで精度が向上する可能性がある。本論文ではこの割合の設定による精度の変化を比較していない。また、Uncertainty Sampling の後に MMR を適用するという方法も考えられる。今後はこのような方法でも実験を行い精度の向上に努めたいと考える。

本論文による貢献は以下の通りである。

- MMR によるランキング下位の小データ削減時に、ランダムによるデータ削減と比べて精度の向上を確認した。
- ランキング学習による予測スコアに基づいた Uncertainty Sampling の有用性の確認をした。

2 関連研究

レビューの有用性推定についての研究はすでいくつか行われている。山澤ら [2] の研究では形態素情報を用いてレビューの有用性を判別する SVM を作成した。その際に、有用なレビューとそうでないレビューでのすべての品詞の出現頻度を用いて有効性の評価を行っている。柴田ら [3] の研究ではレビューの有用性スコアに基づいてランキング推定を行う。レビューに投稿されている「役に立った」、「役に立たなかった」を有用性のスコアにしており、その評価数が多いほど信頼性が高いと考え、EAP 推定値を用いてスコアの信頼性を考慮できるように有用性のスコアを再定義することを行った。

一対比較法についての研究も行われている。阿部ら [4] の研究では物質の質感について一対比較法でアンケートを取り、ランキング関数を得ている。ランキング関数の学習には SVM を用いて行う。高濱ら [5] の研究では一対比較法を行う際に評価対象の提示の順序を工夫する事で作業者が評価対象を評価する回数を削減する漸進比較法を提案している。また、レビューペアに対して期待分布変化量、期待確率変化量という二つの変化量を定義して能動学習を行っている。実験では実際に並び替えたデータ集合の中の二つのデータ間の優劣情報を増やしている。ランキング推定の精度を確認するというをしている。

文書の特性に着目した能動学習によって未知の文書を分類する研究も行われている。佐々木ら [6] の研究では SVM を用いて文書分類をする際の学習過程で、分離平面に近い N 個のデータを学習用データに追加する。この時にそれらのデータのクラスタリングを行い、各クラスターの分離平面に近いものを学習データに追加するというを行っている。これにより多様性を考慮した能動学習を行っている。

このように先行研究ではレビューの有用性のスコアとしてレビューに投稿されている「役に立った」、「役に立たなかった」の投票数を採用しているものが多い。しかしこれらの評価が集まっているレビューが学習に十分な数投稿されている商品は多くないと考える。そこで本研究ではこれらの評価ではなくアンケートによってレビューに評価を付けて学習することを考える。

文書の特徴を用いた能動学習の研究としては、SVM を学習する際に分離平面からの距離が近く、予測モデルによる予測が不確かであるデータから学習に追加するというをしている。本研究では予測モデルによる未使用アンケートペア中の二つのレビューの予測スコアの差の絶対値が小さいものを不確かなデータとして Uncertainty Sampling を行う。予測スコアが大きく離れている二つのレビューは現状のモデルで違いを認識できていると言えるが、予測スコアが近い値をとるものは現状の予測モデルでは同等の有用性があると認識していると考え。そのようなペアの中には文書が類似しているものもあると考え。違う内容の文書の場合にはその正解ラベルを学習データに追加することで効率の良いモデルの改善ができると考える。

3 提案手法

本研究ではレビューを有用性の高い順に並び替えることを目指す。その一環として本論文では、MMR を用いた能動学習を行いレビューの有用性のランキング推定を行うモデル構築に使用するデータを削減することを目的とする。

数値評価の難しい評価物や、似ている評価物の評価に適しているアンケート手法として一対比較法がある。しかし一対比較法では全データ中の二つのデータの総当たり組について優劣を問うというアンケートを行う必要があり、データ数が増えるにつれて行う必要のあるアンケート数は膨大となる。この問題を解決するために少ないアンケート数で評価精度を高められる方法があると良いと考えた。

本論文では MMR という文書群中の文書を検索クエリとの関

連性を考慮しながら内容の冗長性の少ない順に並び替えるという手法を応用して、アンケートペアを冗長性の少ない順に並び替え、上位のデータに多様性を持たせる。その後上位のペアからアンケートを実施していくことによって少ない作業数でも精度を高められるのではないかと考えた。

本研究の概要を図1に示す。はじめに対象とするレビューにアンケートを行いデータセットを作成する。このデータセットから提案手法によるデータ選択を行い予測モデルを構築する。最後に構築後のモデルによる予測結果を出力する。

3.1 MMRの適用

MMRとは文書群を周辺関連性を最大にすることを試みながら並び替える手法である。周辺関連性とは並び替えたい文書群中のとある文書と検索クエリとの関連性と、並び替え後の文書との類似度の二つの変数を線形結合したもののものである。

MMRは下式で表される。中括弧中の左項で検索クエリ Q と並び替え候補文書群 R 中のある文書 D_i との類似度 $Sim(D_i, Q)$ を算出し、右項では並び替え済み文書群 S 中のある文書 D_j と、並び替え候補文書群 R 中のある文書 D_i との類似度 $Sim(D_i, D_j)$ を算出する。0から1の値をとる可変数である λ を用いて各項に重み付けを行い、括弧中が最も大きい際の文書 D_i を並び替え済み文書群 S の末尾に追加し並び替え候補文書群 R 中から削除する。 λ が1に近いと検索クエリとの関連度の高い順で文書が並び替えられ、 λ が0に近いと並び替えたい文書群中の文書間の関連度が低い順に文書が並び替えられる。

$$MMR = \operatorname{argmax}_{D_i \in R} [\lambda Sim(D_i, Q) - (1 - \lambda) \max_{D_j \in S} Sim(D_i, D_j)] \quad (1)$$

本論文ではコサイン類似度を用いて文書間の類似度を算出している。コサイン類似度を下式に示す。 a, b は文書を表し、 \mathbf{a}, \mathbf{b} は a, b の分散表現を表す。対象とする二つの文書の分散表現の内積を、二つの分散表現の大きさの積で割ることで算出される。 -1 から 1 までの値をとり、1に近いほど類似度が高く、 -1 に近づくほど類似度は低くなる。

$$\cos(a, b) = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| \times |\mathbf{b}|} \quad (2)$$

本論文ではアンケートペアを冗長性が少ないように並び替

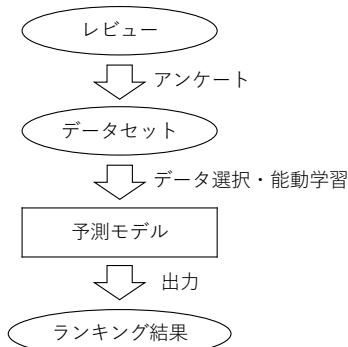


図1 本研究の概要

えることを目的としており、検索クエリを必要としないため(2.4.1)式中の λ の値を0とした。

はじめに訓練データに用いるレビュー全てをMMRを用いて並び替える。その後並び替え済みレビュー群の上から順に選択してペアの一方を固定する。ペアのもう一方は訓練データに用いるレビューをランダムに並び替えた順序の上から選択した。

アンケートペア作成の様子を図2に示す。図2では例としてデータ数が4の場合を示しているが、実際には100データで行っている。左の数字がMMRによる並び替え後のレビュー番号を表しており、上部の数字が対象レビュー全体をランダムに並び替えたもののレビュー番号を表している。表の上の行から着目して左から右に見ていき、交わった列とペアを作成する。上の2行を例にあげる。1行目では既存のペアがなく重複するペアがないため、同じレビューが交わる4-4以外でペアが作られる。2行目では4-1が既存で、1-1で同じレビューが交わるため1-4と1-1以外でペアが作られる。このようにして総当たり表に沿って、対象レビュー中の異なる二つのレビューの全組み合わせでペアを作る。この時ペア中の二つのレビューの順序は考えないため、作られるペアの数は ${}_4C_2$ となる。作られたペア番号の昇順でアンケートペアは並び替えられる。

このようにしてアンケートの総当たりペアに順序を付け、上位から作業を行うことで情報の冗長性が少なく多様性があり、学習する価値の大きい順番でアンケートを実施することができると考えた。下位のアンケートペアは既出の情報が多く学習する価値の小さいアンケートペアだと考え、下位のアンケートペアを削り、上位のアンケートペアのみのアンケートを取ることによりアンケート数を減らしても精度が下がらないと考えた。

MMRの使用にあたって対象文書間の類似度を算出する必要がある。本論文ではWord2Vecを用いて対象文書群それぞれの一つのレビュー中の名詞、動詞、形容詞、形容動詞の単語の相加平均をとることで分散表現を得た。Word2Vecのモデルは対象としたデータセットから学習した。

3.2 データセット作成

一対比較法とはアンケート手法の一つである。アンケート作業員に対して評価対象を二つ提示し、評価軸に対してどちらが優れているかを問う手法である。提示する二つの評価対象のペアは、評価対象群から総当たりの組み合わせで作成する。

一対比較法の優れている点は二点ある。一つ目は数値評価の

		ランダム順			
		2	4	3	1
M M R 順	4	ペア1		ペア2	ペア3
	1	ペア4		ペア5	
	2			ペア6	
	3				

図2 アンケートペア作成の様子

難しい曖昧な評価軸についてのアンケートの場合においても簡単に評価を行えることである。アンケート手法としてよく用いられている数値評価では、曖昧な評価軸に対して厳密に評価することは難しいため、アンケート途中で評価軸がブレてしまうことが考えられる。一対比較法では提示された評価対象の優劣を答えるのみでよく、評価軸のブレが生じにくいと考える。

二つ目は似ている評価対象に対して優劣の情報を得られることである。数値評価の場合では似ている評価対象同士に同じラベルが付くことが多くある。一方、一対比較法では評価対象間の総当たりによる優劣情報を集計することで、より詳細な段階で評価することが可能になる。

アンケート作業は著者一人で行い、対象レビュー群中の総当たりペアの優劣をつけることでデータセットを作成した。

3.3 モデル構築

使用するモデルはランク学習を用いて学習した。ランク学習とは予測スコアの最適化ではなく、予測順位の最適化を目標にモデルを学習する手法のことである。検索エンジンによる検索結果の最適化などに用いられる。ランク学習は損失関数の定義の仕方によって大きく3種類に分類される。本稿ではその中でも pairwise 手法を用いてモデルの構築を行った。

二つ目は Pairwise 手法である。Pairwise 手法とは任意の二つのデータからなるペアの優劣関係を正確に予測できれば文書群を正しく並び替えられるという発想に基づいて学習を行い、重みを最適化する手法である。損失関数は下式のように表される。クロスエントロピーなどを用いて、モデルによるクエリ q 中の i 番目のデータ $x_{q,i}$ についての予測スコア $f(x_{q,i})$ と、 $x_{q,j}$ についての予測スコア $f(x_{q,j})$ を用いた損失関数 $L_{Pairwise}$ を最適化する2値分類問題を解いて学習を行う。 m_q がクエリ q 中のデータ数を表し、 $y_{q,i}, y_{q,j}$ がクエリ q 中の i 番目と j 番目のデータのラベルを表す。

$$L_{Pairwise} = \sum_q \sum_{i,j: y_{q,i} > y_{q,j}}^{m_q} L(f(x_{q,i}), f(x_{q,j})) \quad (3)$$

実装には LightGBM [7] を用いて、LambdaMART [8] という手法を用いて行った。LambdaMART とはニューラルネットワークを用いて Pairwise 法を考える RankNet [9] を拡張した LambdaRank [10] と MART [11] を組み合わせたもので 2010 年の Yahoo! Learning To Rank Challenge [12] で優勝したモデルである。

学習の際は検証データの予測について評価指標の精度が 100 回の Iteration の間に向上しなかった場合に学習を中断し、精度が最も高かった時点のモデルを採用するというを行った。入力データには前述したデータセットを用いた。ここで入力に用いる訓練データの数を減らす手法を四つ行いモデルを構築した。三、四つ目は能動学習によりデータの削減を行った。

一つ目はランダムに減らす方法である。全アンケートペア中から減らしたい割合のデータをランダムに選択し、学習用のデータから取り除く。

二つ目は前述の MMR を適用して優先順位を付けて作成した

アンケート順の下位のデータから取り除く方法である。

三、四つ目は能動学習によりデータの削減を行った。能動学習とはモデルの学習に有効なデータからラベルをつけていく学習方法のことである。

能動学習の手法の一つに uncertainty sampling というものがある。uncertainty sampling とは現段階の予測モデルによる予測が不確かなものからラベルをつけて学習すれば予測の不確かさが最も軽減するだろうという期待に基づいた手法である。

三つ目は Uncertainty Sampling を用いる方法である。未使用アンケートペアすべての中でその段階のモデルによるペア中の二つのレビューについての予測スコアの絶対値の差が小さいものから N 個を学習データに追加し再度学習する。学習後のモデルを用いて再度この工程を行うということを未使用データがなくなるまで繰り返す。予測スコアの差の絶対値が小さいということは予測モデルが二つのレビューを同程度有用であると認識しているということである。その中にはレビューが類似しているものもあると考えるが、レビューの内容が異なっているような場合にはその優劣の正解データを学習データに追加することでモデルは効率よく改善されると考えた。

四つ目は MMR と Uncertainty Sampling を用いる方法である。前述の MMR を適用して優先順位を付けて作成したアンケートペアの未使用データ上位 N 個の中で、その段階のモデルによるペア中の二つのレビューについての予測スコアの差の絶対値が小さいものから $N/2$ 個を学習データに追加し再度学習する。学習後のモデルを用いて再度この工程を行うということを未使用データがなくなるまで繰り返す。これにより MMR によるアンケートの優先順位を考慮しつつ、スコア差が小さく不確かな予測をしている可能性の高いペアから優先的にラベルを付けることができると考えた。

手法 3、4 の様子を下に示す。これら 4 つの方法でモデルを構築した。これら四つの方法で構築したモデルにテストアンケートペアを入力し、ペア中の二つのレビューのスコアを予測することで、二つのレビューの優劣の予測を出力した。

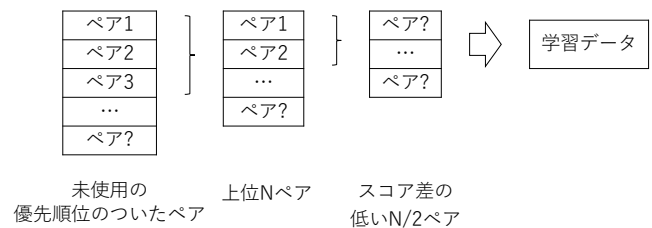


図 3 手法 3 の様子

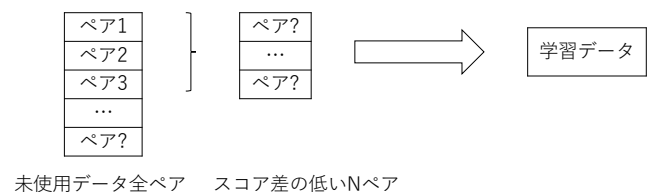


図 4 手法 4 の様子

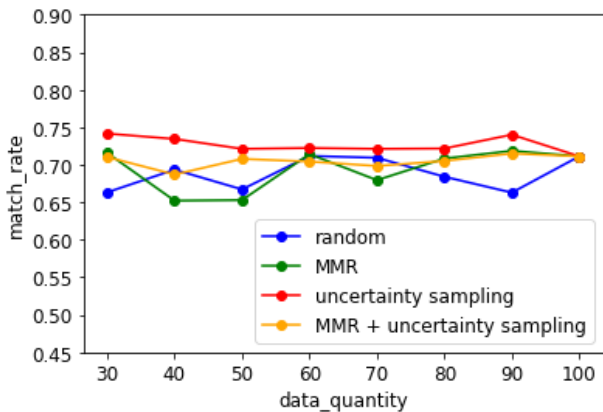


図5 実験1から4の一致率の推移

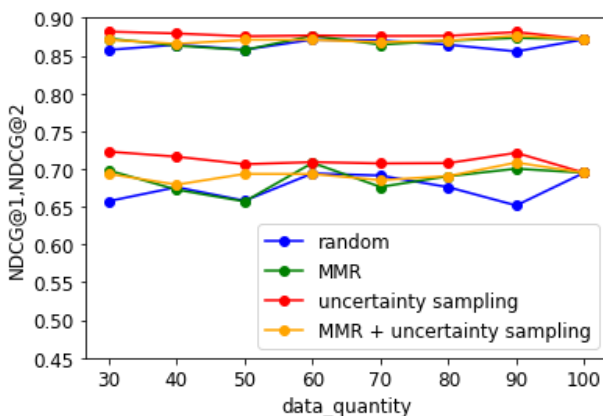


図6 実験1から4のNDCGの推移

4 評価実験

提案手法について、以下の評価実験を行った。

実験1 ランダムによるデータ削減

無作為にデータを取り除いた際の精度指標の推移を見ることで他手法と比較するためのベースラインを作る。

実験2 MMRによるデータ削減

MMRを適用した冗長性が少なく多様性があるデータを優先的に学習に用いることで、他手法と比較して小データの区間で精度に変化が見られるかを確認する。

実験3 Uncertainty Samplingによるデータ削減

MMRによる並び替えを考慮せず、Uncertainty Samplingのみを行った際に、他手法と比較して小データの区間で精度に変化が見られるかを確認する。

実験4 MMRとUncertainty Samplingによるデータ削減

Uncertainty SamplingとMMRを組み合わせることでMMR、Uncertainty Sampling単体の場合と精度に変化が見られるかを確認する。

本論文では国立情報学研究所から提供されている楽天市場の

表1 ランダムデータ削減による精度の推移

	データの割合						
	90	80	70	60	50	40	30
一致率	0.663	0.684	0.709	0.712	0.667	0.693	0.663
NDCG@1	0.652	0.676	0.691	0.694	0.658	0.676	0.657
NDCG@2	0.855	0.864	0.870	0.871	0.858	0.864	0.858

表2 MMR適用後のデータ削減による精度の推移

	データの割合						
	90	80	70	60	50	40	30
一致率	0.719	0.708	0.679	0.714	0.653	0.652	0.716
NDCG@1	0.700	0.690	0.676	0.708	0.657	0.673	0.698
NDCG@2	0.873	0.870	0.864	0.876	0.857	0.863	0.873

表3 Uncertainty Samplingを用いた能動学習による精度の推移

	データの割合						
	90	80	70	60	50	40	30
一致率	0.740	0.722	0.721	0.722	0.721	0.734	0.742
NDCG@1	0.721	0.708	0.707	0.709	0.706	0.716	0.723
NDCG@2	0.881	0.876	0.876	0.877	0.876	0.879	0.882

表4 MMRとUncertainty Samplingを用いた能動学習による精度の推移

	データの割合						
	90	80	70	60	50	40	30
一致率	0.715	0.705	0.698	0.704	0.708	0.687	0.710
NDCG@1	0.708	0.691	0.685	0.693	0.693	0.679	0.693
NDCG@2	0.876	0.870	0.868	0.871	0.871	0.866	0.871

商品レビューデータのデータセット¹のうち、レビュー本文を使用した。また、懐中電灯が商品名に入る商品の中で2019年の中で再頻出の商品を対象にし、その商品に投稿されていたレビューを用いて研究を行った。多ジャンルや、同ジャンル間の複数商品のレビューを対象にすると有用であるレビューの傾向に差が生じる可能性があると考え、一種類の商品を対象に行った。

対象とするレビューについて3.2節で示したアンケート方法によりデータセットを作成した。本論文ではアンケート作業は著者一人で、訓練データ、検証データ、テストデータについてそれぞれ100件、36件、36件のレビューを対象にして行い、4950件、630件、630件からなるデータセットを作成した。

データの割合が全体の100%である場合は四つの手法で精度が等しくなるため各実験の結果の表から省略した。実験1から4の評価指標の精度の推移を図5、6に示す。図5の縦軸が一致率を表し、図6の縦軸がNDCGのスコアを表す。図6の上部はNDCG@2の、下部はNDCG@1の値を表す。横軸はモデル構築に使用するデータの割合を示す。

4.1 評価指標

本稿では評価指標としてNDCG@kと一致率を用いる。それぞれ評価指標の定義を下に示す。

1: 楽天グループ株式会社 (2020): 楽天市場データ. 国立情報学研究所情報学研究所データリポジトリ. (データセット). <https://doi.org/10.32130/idr.2.1>

4.1.1 NDCG@k

NDCG とは順位に依存したスコア減損を考慮した評価指標である $DCG(DCG_{pred})$ を、理想の $DCG(DCG_{ideal})$ で割って正規化したものである。理想の DCG とは正解データのスコアの降順通りに順位を予測できている場合の DCG である。NDCG@k を下式に示す。NDCG@k は上位 k 番目までを対象とした NDCG のことである。0 から 1 の値を取り、1 に近い方が精度が良い。

$DCG@k$ は下式のように表され上位 k 番目までのデータを考慮した DCG のことである。順位を r 番と予測されたデータの予測スコア $g(r)$ を順位に応じたスコア減損である $\log_2(r+1)$ で割ることで与えている。本研究の目的であるレビューを有用性で並び替える際に予測結果の上位の正しさを評価するため、下位の順位にスコア減損を与えるこの評価指標を採用した。

$$NDCG@k = \frac{DCG_{pred}@k}{DCG_{ideal}@k} \quad (4)$$

$$DCG@k = \sum_{r=1}^k \frac{g(r)}{\log_2(r+1)} \quad (5)$$

4.1.2 一致率

一致率とは、対象とする全アンケートペア All_pairs の中でモデルによるペア中の二つのレビューに対する予測スコアの優劣関係が、正解ラベルの優劣関係と一致しているペア $match_pairs$ の割合のことである。一致率を下式に示す。あるペアについて予測結果と正解ラベルが一致している割合が高ければ、予測結果の優劣関係に基づいて各レビューを並び替えた際のランキングが正しいものになるという考えからこの評価指標を採用した。

$$Match_rate = \frac{Match_pairs}{All_pairs} \quad (6)$$

4.2 ランダムによるデータ削減

本節では訓練データの全アンケートペアの中から無作為にデータを取り除いてモデルを構築した際の評価指標の精度の推移を確認し、他手法と比較するためのベースラインを作成した。

4.2.1 実験内容

訓練データから無作為にデータを取り除き評価指標の精度を確認する。データを取り除く際には訓練データ中の全ペアをランダムに並び替え、末尾のデータをアンケートにより作成した訓練データの全データ数に対して 10% ずつ取り除くことを、訓練データの数が全体の 30% になるまで行う。モデルの構築は 3.3 節で述べた方法で行った。レビューの並び順の偏りによる結果への影響を減らすため、ランダムによる並び替えを 10 回行い評価指標の平均を算出した。

4.2.2 結果・考察

モデルによる予測結果の評価指標の精度を表 1 に示す。上部がモデル構築に使用したデータ割合を示し左側が評価指標を示す。交わった点に評価指標の値が書かれている。

データ数を減らすにつれて評価指標の精度は下がる場合もあれば上がる場合もあった。データ数を減らすことで、評価指標の精度は単調に下がると考えていたがそうはならなかった。その理由として、学習に用いる正解ラベルに矛盾が含まれていることが考えられる。アンケート段階において、あるペアでレビュー A < レビュー B、あるペアでレビュー B < レビュー C となる場合にレビュー A < レビュー C と予測するように学習した場合でも実際のアンケートではレビュー A < レビュー C となるとは限らず、レビュー A > レビュー C となることもある。このような場合にモデルが間違えて予測することで学習が妨げられることがあると考える。このような学習の妨げとなるアンケートペアが取り除かれた場合はデータ数が減ることで評価指標の精度が向上することがあると考える。その他には検証用データ中に多く含まれるレビューに似た特徴のレビューを重点的に学習できた場合に、小データでも精度が向上することがあると考える。

4.3 MMR によるデータ削減

本節では訓練データの全アンケートペアに MMR を適用し、冗長性の少ない順序に並び替えを行う。その後下位のアンケートペアからデータを取り除いてモデルを構築した際の評価指標の精度をランダムによるデータ削減結果と比較することで、MMR を用いて並び替えを行うことが小データでのモデルの学習に有効であるかを確認する。

4.3.1 実験内容

MMR を適用してアンケートペアを並び替え、下位ペアを取り除いてモデルを構築し評価指標の精度を確認する。

MMR の適用は 3.1 節で述べた方法で行いアンケートペアを並び替えた。MMR を用いてレビューを並び替える際に並び替え後の 1 番上にくるレビューを設定する必要がある。この一つのレビューにより並び替え結果が偏ることによる学習結果への影響を減らすために、一つのレビューをランダムに 10 回選択し、評価指標の精度の平均を算出した。

4.3.2 結果・考察

モデルによる予測結果の評価指標の精度を表 2 に示す。上部がモデル構築に使用したデータ割合を示し左側が評価指標を示す。交わった点に評価指標の値が書かれている。

ランダムによるデータ削減と比較して大データの区間で精度が高く、小データの区間で精度が低くなった。大データの区間で精度が向上する理由として、MMR で並び替えた下位のデータは既出の情報が占めているため学習に役立たないものが多いのに対して、ランダムの場合は学習に役立つものも取り除いている場合があるためだと考える。小データの区間で精度が下がる理由として、MMR ではアンケートペアの片方のレビューが固定されているため、いくつかのレビューのみを多く学習することになり、かえって学習データの多様性が減っているためだと考える。

4.4 Uncertainty Sampling によるデータ削減

本節では能動学習の一手法である Uncertainty Sampling を用

いて能動学習を行い、MMR による並び替えを考慮しなかった場合の小データの区間での精度を確認し、Uncertainty Sampling の有効性を確認する。

4.4.1 実験内容

3.3 節で述べた方法で Uncertainty Sampling を行い、学習に用いるデータ数の各段階でのモデルによる予測結果について評価指標の精度を確認する。

最初に全訓練データのうちの 30% をランダムに選び、モデルを構築する。次に構築後のモデルを用いて未使用の訓練用アンケートペア中の二つのレビューについてスコアを予測し、その差の絶対値が小さいペア順に並び替える。その後上位の 495 ペア (訓練データ全体の 10%) を学習用のデータに追加し、モデルを再び構築する。この工程を未使用データがなくなるまで繰り返し行い、各段階のモデルによってテストデータの予測を行い評価指標の精度を算出した。最初にモデルを学習する際の 30% のデータの偏りによる学習結果への影響を減らすために 30% のデータをランダムに 10 通りで繰り返し、評価指標の精度の平均を算出した。

4.4.2 結果・考察

モデルによる予測結果の評価指標の精度を表 3 に示す。上部がモデル構築に使用したデータ割合を示し左側が評価指標を示す。交わった点に評価指標の値が書かれている。

ランダム、MMR 単体によるデータ削減に比べて、実験設定とする全データ区間において評価指標の精度が高くなった。このことから、モデルによるアンケートペア中の二つのレビューの予測スコアの差の絶対値が小さいペアを不確かなデータとして Uncertainty Sampling を行うことは有効であった。よって、モデルが近いスコアを予測する文書ペアには二つのレビューが類似しているもの以外も存在しており、その優劣の判別を優先的に可能にさせることが学習効率の向上につながったと考える。また、MMR を用いて単に学習データに多様性を持たせることよりも、各段階のモデルによる予測が不確かな学習データを与えることの方が学習効率が高くなった。このことから、多様性を持たせたデータでモデルを構築しても、モデルが近いスコアを予測する類似していないレビューのペアを判別できないと学習精度は向上しにくいと考える。

4.5 Uncertainty Sampling と MMR によるデータ削減

本節では能動学習の一手法である Uncertainty Sampling と MMR を組み合わせて、小データの区間で評価指標の精度を確認し、Uncertainty Sampling, MMR それぞれ単体の場合と比較し、精度に変化が見られるかを確認する。

4.5.1 実験内容

3.3 節で述べた方法でレビューを MMR を用いて並び替えたのちに順序を考慮して Uncertainty Sampling を行い、各段階でのモデルによる予測結果の評価指標の精度を確認する。

最初に 3.1 節で述べた方法で MMR を用いてアンケートペアを並び替える。その後上位の 30% のデータを用いてモデルを構築する。次に構築後のモデルを用いて未使用のアンケートペアのうち上位 990 ペア (訓練データ全体の 20%) を対象に、ペア

の中の二つのレビューについてスコアを予測し、その差の絶対値が小さい順にペアを並び替える。その後上位の 495 ペア (訓練データ全体の 10%) を学習用のデータに追加し、モデルを再び構築する。この工程を未使用データがなくなるまで繰り返し行い、各段階のモデルによってテストデータの予測を行い評価指標の精度を算出した。実験 2 と同様に、MMR で並び替えを行う際の並び替え後の 1 番上とするレビューをランダムに 10 回選択し、評価指標の精度の平均を算出した。

4.5.2 結果・考察

モデルによる予測結果の評価指標の精度を表 4 に示す。上部がモデル構築に使用したデータ割合を示し左側が評価指標を示す。交わった点に評価指標の値が書かれている。

ランダム、MMR 単体によるデータ削減と比べてそれぞれの場合で、ほとんど全データ区間において評価指標の精度は向上している。一方で Uncertainty Sampling 単体の場合に比べて全データ区間での精度は下回った。このことからランダム、MMR 単体によるデータ削減に対しては 4.4.2 節と同様の考察をする。Uncertainty Sampling 単体と比べて全データ区間で精度が下回ったことから、MMR による多様性を持った順序よりも、各段階のモデルによる予測の不確かさのみを考慮して学習データを選択することが学習効率を高めると考える。また、データ割合が 80,90% の段階において MMR 単体の場合と精度がほとんど同じになった。このことから、これら二つの手法で学習に使われていないデータは似たものとなっていると考えた。つまり MMR によるランキングの上位 80, 90% で既出なデータは Uncertainty Sampling 時においても不確かなデータと識別されず、最後まで学習に用いられなかったと考える。よって、学習データに多く用いられた情報の優劣はモデルが比較的はっきり判別できていると考える。

5 おわりに

本研究ではレビューを有用性の高い順に並び替えることを目的とした。その一環として本論文では MMR を用いた能動学習を行い、レビューの有用性のランキング推定を行うモデル構築に使用するデータを削減することを目的とした。

通販サイトには出品者側がイメージを良くするために投稿した自演的なレビューや、キャンペーン目当てにレビューを寄せている人が多くいる事、さらには投稿されているレビューが多いという現状がある。そのため利用者が有用なレビューを見つけることが難しいという問題がある。通販サイト上のレビューのページにはレビューに集まった「役に立った」の投票数順のソートが存在するが、「役に立った」の投票が上位の数レビューに集中しているという現状があり不便と感じる人が多い。そこでレビューを有用性で並び替えることを考えた。利用者は有用性の高い順に並び替えられたレビューを上から順に見ていくと考えられるため上位の予測に重きを置くことのできるランキング学習を用いて学習を行った。

学習の際にはレビューに有用性についてのラベルを付ける必要がある。ラベル付けのアンケート手法としては 5 段階の数値

評価が考えられるが、レビューの有用性という評価軸で厳密な数値評価をアンケートによって行うことは難しいと考えた。また、有用、非有用の2値のラベルを付けるというアンケート方法も考えられるが、レビューに対して有用性の度合いを評価することはできない。そこで一対比較法というアンケート方法を採用した。

一対比較法では提示された二つの評価対象の優劣を問う形式のアンケートであるため、有用性という曖昧な評価軸に対しても簡単に正確なアンケート結果を得られると考えた。また、数値評価や有用・非有用の2値ラベルでは似たレビューに同じラベルがつくことが多いが、この手法では優劣の結果を集計することで、レビューの有用性の度合いをより詳細な段階で評価することができる。しかし一対比較法では評価対象群中の総当たりペアについてアンケートを行う必要があり、作業数が膨大になってしまう。そこで、入力する学習用データを減らしても評価指標の精度が下がらないような効率の良い学習を行うことができればアンケート作業数の削減につながると考えた。

学習の早い段階で多様性のあるデータを入力とすることによる学習効率の向上を期待し、MMR という手法を用いて対象レビュー群中の総当たりペアを冗長性の少なくなるように並び替えを行った。その後上位のアンケートペアから作業を行い学習に用いることで、小データの区間で学習効率を高めることができると考えた。また、能動学習の一手法である Uncertainty Sampling によるデータ選択も併せて行った。これと MMR を組み合わせることでデータの多様性を考慮しながら能動学習を行った。これによりそれぞれ単体の場合に比べて学習効率が向上すると考えた。

評価実験として提案手法を用いて使用データ数を変更しながらモデルを構築し、各データ数におけるレビューの有用性予測の精度を評価した。ランダムによる削減、MMR、Uncertainty Sampling 単体による削減、MMR と Uncertainty Sampling を組み合わせた削減それぞれにおいて評価指標の精度の推移を比較した。その結果、MMR と Uncertainty Sampling を組み合わせたものは MMR 単体、ランダムと比べて実験設定のほとんど全データ区間で一致率が上回ったが、Uncertainty Sampling 単体のものと比べると実験設定のほとんど全データ区間において一致率が下回った。この結果から、本論文の実験設定において MMR による多様性を持った順序を考慮して能動学習することはモデル構築の際の使用データ削減に有効であるとは言えなかった。

本論文では MMR を用いた文書並び替えを考慮した能動学習手法を提案した。アンケート作成の面では作業者が一人であることからアンケート結果に偏りが生まれている可能性が考えられるため、作業者を増やしたいと考えている。またアンケートは、提示した二つの評価対象の優劣のみを問う形式であり、ペア中の二つのデータが類似した場合の評価が難しくなっているためアンケート結果のミスを生じさせた可能性があると考えられる。そこで、アンケート時に「どちらでもない」という評価項目を追加することで作業者の負担、アンケート結果のミスを減らすのが良いと考える。提案手法の面では、MMR で並び替え

た上位 20%を対象に Uncertainty Sampling を用いており、その際の対象とする割合の設定が精度に影響するが、本論文ではこの割合の設定による精度の変化を比較していない。また、Uncertainty Sampling の後に MMR を適用するという方法も考えられる。今後はこのような方法でも実験を行うことで手法を改善し、精度の向上に努めたいと考えている。

謝辞 本研究の一部は JSPS 科研費 19H04218 および越山科学技術振興財団の助成を受けたものです。本研究では、国立情報学研究所の IDR データセット提供サービスにより楽天グループ株式会社から提供を受けた「楽天データセット」(https://rit.rakuten.com/data_release/) を利用した。

文 献

- [1] Jaime Carbonell and Jade Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 335–336, 1998.
- [2] 山澤美由起, 吉村宏樹, 増市博ほか. Amazon レビュー文の有用性判別実験. 情報処理学会研究報告自然言語処理 (NL), Vol. 2006, No. 53 (2006-NL-173), pp. 15–20, 2006.
- [3] 柴田知親, 伊東栄典. 商品レビューの有用性ランキング推定. 人工知能学会研究会資料 知識ベースシステム研究会 118 回, p. 05. 一般社団法人 人工知能学会, 2019.
- [4] 阿部 敏, 岡谷貴之, 出口光一郎ほか. 自然画像を対象とするランキング学習に基づく質感認識の一手法. 研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2012, No. 7, pp. 1–7, 2012.
- [5] 高濱隆輔, 神島敏弘, 鹿島久嗣. 漸進比較法によるランキング推定. 人工知能学会全国大会論文集 第 29 回 (2015), pp. 3K3OS20a4i–3K3OS20a4i. 一般社団法人 人工知能学会, 2015.
- [6] 佐々木寛, 高村大也, 松本裕治. Svm とクラスタリングを用いた文書分類のための能動学習. 第 1 回 情報科学技術 フォーラム (FIT 2002) 講演論文集, Vol. 2, pp. 99–100, 2002.
- [7] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, Vol. 30, , 2017.
- [8] Christopher JC Burges. From ranknet to lambdarank to lambdamart: An overview. *Learning*, Vol. 11, No. 23-581, p. 81, 2010.
- [9] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning*, pp. 89–96, 2005.
- [10] Christopher Burges, Robert Ragno, and Quoc Le. Learning to rank with nonsmooth cost functions. *Advances in neural information processing systems*, Vol. 19, , 2006.
- [11] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232, 2001.
- [12] Olivier Chapelle and Yi Chang. Yahoo! learning to rank challenge overview. In *Proceedings of the learning to rank challenge*, pp. 1–24. PMLR, 2011.