雑談対話を利用したニュース記事推薦システム

武田 悠† 熊本 忠彦‡

千葉工業大学情報科学部情報ネットワーク学科 〒 275-0016 千葉県習志野市津田沼 2-17-1
E-mail: †s1432092hr@s.chibakoudai.jp, ‡kumamoto@net.it-chiba.ac.jp

あらまし 本論文では，ユーザが興味を有する語（興味語）とその興味語に抱いている印象をユーザとの雑談対話を通して学習する手法を提案するとともに，その興味語を含み，かつ似た印象を有するニュース記事を推薦するシステムを提案する．興味語が抽出されなかった場合は，ユーザとの対話履歴の中から頻出語を抽出し，その頻出語を含むニュース記事の中からユーザ発話の印象に最も近いニュース記事を推薦する．なお，本研究では，印象として「楽しい⇔悲しい，うれしい⇔怒り，のどか⇔緊迫」という 3 つの印象軸を採用している．

キーワード 雑談対話，記事印象，記事推薦

1. はじめに

Web 上の情報は日々更新されており，毎日膨大な量のデータが流通している．しかしながら，その多くは無用なデータであり，そういったデータの中から自分の嗜好に合うデータを探し出すのは容易でない．そのため，ユーザの嗜好に合った情報を効率的に推薦するためのシステム[1][2]が数多く提案されている．

ユーザの嗜好に合った情報を推薦するためのシステムは，内容に基づくフィルタリング方式に基づくもの[1]と協調フィルタリング方式に基づくもの[2]とに大きく分けられる．内容に基づくフィルタリング方式では，単語の出現頻度を利用しており，過去にユーザが好んだ情報に含まれている単語の出現頻度からユーザが好むであろう情報を予想し，推薦する．しかしながら，ユーザが過去に興味を示した情報ばかりを推薦することとなり，ユーザの潜在的な嗜好や表明されていない嗜好に合った情報を推薦することは困難である．一方，協調フィルタリング方式では，様々な情報に対して自分と似たような興味を示す他のユーザの評価（嗜好）を利用する．例えば，あるユーザに対し，そのユーザと類似した興味を持つ他のユーザらを探し出し，そのユーザら（の多く）が興味を持っている情報は当該ユーザも興味を持つという仮定の下，ユーザらが高評価を与えた情報を推薦するというものである．この場合，推薦される情報は当該ユーザが過去に興味を示した情報のみに限定されないため，ユーザの潜在的な嗜好や表明されていない嗜好に合った情報を推薦することができる可能性がある．しかしながら，その一方で，数多くのユーザがいなければ，任意のユーザと似たような興味を持つユーザらを十分に探し出すことができないかもしれないという欠点や誰も評価していないような新規な情報は推薦対象にならないという欠点がある．

そこで本論文では，ニュース記事推薦に焦点を当て，内容に基づくフィルタリング方式に雑談対話能力と印象マイニング手法[3]を導入することで，ユーザの興味を有する語（本論文では「興味語」と呼ぶ）とその興味語に抱いている印象をユーザとの雑談対話から学習し，その興味語を含み，かつ似たような印象を有するニュース記事を推薦する雑談対話システムを提案する．なお，印象マイニング手法とは，入力文章を読んだ人が感じるであろう印象（本研究では「楽しい⇔悲しい，うれしい⇔怒り，のどか⇔緊迫」という 3 種類の印象軸に限定している）の強さを数値的に求めるための手法であり，印象値が 1 に近いほど「楽しい，うれしい，のどか」という印象に近いことを表し，7 に近いほど「悲しい，怒り，緊迫」という印象に近いことを表している．

2. 関連研究

2.1 ニュース記事推薦

ニュース記事推薦に関する先行研究として早川らの研究[4]がある．早川らは，ユーザが Twitter 上でフォローしている友人らに対して自分の興味のあることをツイートしているかどうかに応じて 3 段階の重要度（hi, mid, low）を設定すると，この重要度を重みとして友人らのツイート内容とニュース記事との類似度を求め，類似度が高い記事を興味のある記事としてユーザに推薦する手法を提案しており，ユーザと友人らとの間で盛り上がっている話題に関するニュース記事が推薦されることを確認している．しかしながら，この方法では友人らとの共通の話題しか推薦の対象にならないため，ユーザの突発的な興味には対応できない．これに対し，提案手法では，ユーザの興味の変化を雑談対話を通して得ることができるので，ユーザの突発的な興味にも対応することができる．

一方，対話を介してニュース記事を推薦する研究として，斉藤らの研究[5]が挙げられる．斉藤らの研究では，Web 上に存在するニュース記事や天候に関する情報などの実世界情報を利用して，システムからの発話文を生成している．例えば，「スポーツの話題は？」というユーザ発話文に対し，「スポーツ」カテゴリに分類されたニュース記事の中からランダムに選んだ記事を対話テンプレートに当てはめ，システム発話文を生成している．また，システムからの話題提供に対

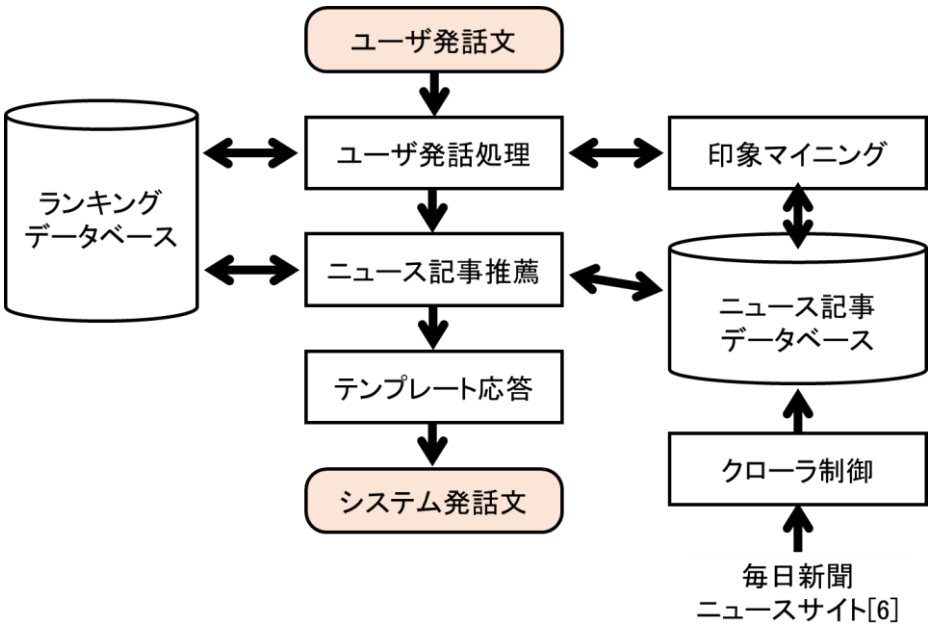

図1 システム構成

し肯定的な反応を示したか否かといった情報やユーザの情報要求が多いカテゴリに関する情報，特定のカテゴリに対する「興味があるか？」という問いかけへの肯定的な返答の有無などからユーザの興味のある話題を抽出し，システム発話文の生成に利用している．しかしながら，彼らの手法ではカテゴリレベルで興味のあるなしを判定しており，単語レベルでの興味のあるなしを考慮していない．加えて，興味のあり方（興味語に対してユーザがどのような印象を抱いているかといった心象的な側面）については全く考慮していない．これに対し，提案手法では，単語レベルで興味の有無を判断するとともに，抽出された興味語に対し，どのような印象を抱いてるかを考慮した記事推薦を実現している．

2.2 印象マイニング[3]

印象マイニング手法は，「楽しい⇔悲しい」，「うれしい⇔怒り」，「のどか⇔緊迫」という 3 種類の印象軸のそれぞれに対し，印象の強さを表す印象値として 1.0～7.0 の実数値を出力する．この実数値は，「（楽しい，うれしい，のどかを）感じる（1 点），わりと感じる（2 点），やや感じる（3 点），（どちらの印象も）感じない（4 点），（悲しい，怒り，緊迫を）やや感じる（5 点），わりと感じる（6 点），感じる（7 点）」という 7 段階評価スケールに対応しており，例えば，「楽しい⇔悲しい」に対して5.52595という実数値が出力された場合は，その記事の悲しさに関する印象が「やや悲しい」と「わりと悲しい」のほぼ中間であることを意味している．

この印象マイニング手法は，印象の強さを数値化するために，事前に構築されている印象辞書（各単語の記事印象への影響力を数値化したもの）を用いる．この印象辞書は，新聞記事データベースから抽出された各単語と特定の印象語群との記事内共起関係に基づいて印象の種類ごとに構築されており，この印象辞書を用いることで，入力文章の印象を数値化することができる．さらに，この印象辞書を用いて算出された記事の印象値とその記事を読んだ人々が感じた印象の強さとの対応関係を 3 次関数あるいは 5 次関数を用いた高次の回帰分析により定式化することで，記事の印象値をより高精度に求めることができるようになっている．

なお，記事を読んだ人々が感じた印象を数値化するにあたり，900 人（男女 450 人ずつ）が参加するアンケート調査を行っている．具体的には，回答者 900 人を年齢や性別が均等になるよう 9 つのグループ（男女 50 人ずつ，計 100 人からなるグループ）に分け，各グループに 10 個の記事を読んでもらい，各記事の印象を 3 つの印象軸のそれぞれにおいて 7 段階評価してもらっている．その結果得られたデータの平均値を記事ごと・印象ごとに求めたものを，その記事の当該印象における強さとして用いている．

3. 雑談対話システムの構成

図1に提案する雑談対話システムの構成を示す．

まず，クローラ制御モジュールは，ユーザとの対話を開始する前に毎日新聞のニュースサイト[6]から最新のニュース記事を 100 件取得し，各記事のタイトルや本文，URL を抽出するとともに，これらのデータをニュース記事データベース（記事DB）に登録する．また，このとき，取得したニュース記事の本文に対して印象マイニング手法[3]を適用することで，各記事の印象値を 3 種類の印象軸（楽しい⇔悲しい，うれしい⇔怒り，のどか⇔緊迫）のそれぞれにおいて算出し，これらの印象値もニュース記事DBに登録する．

ユーザ発話処理モジュールは，汎用日本語形態素解析

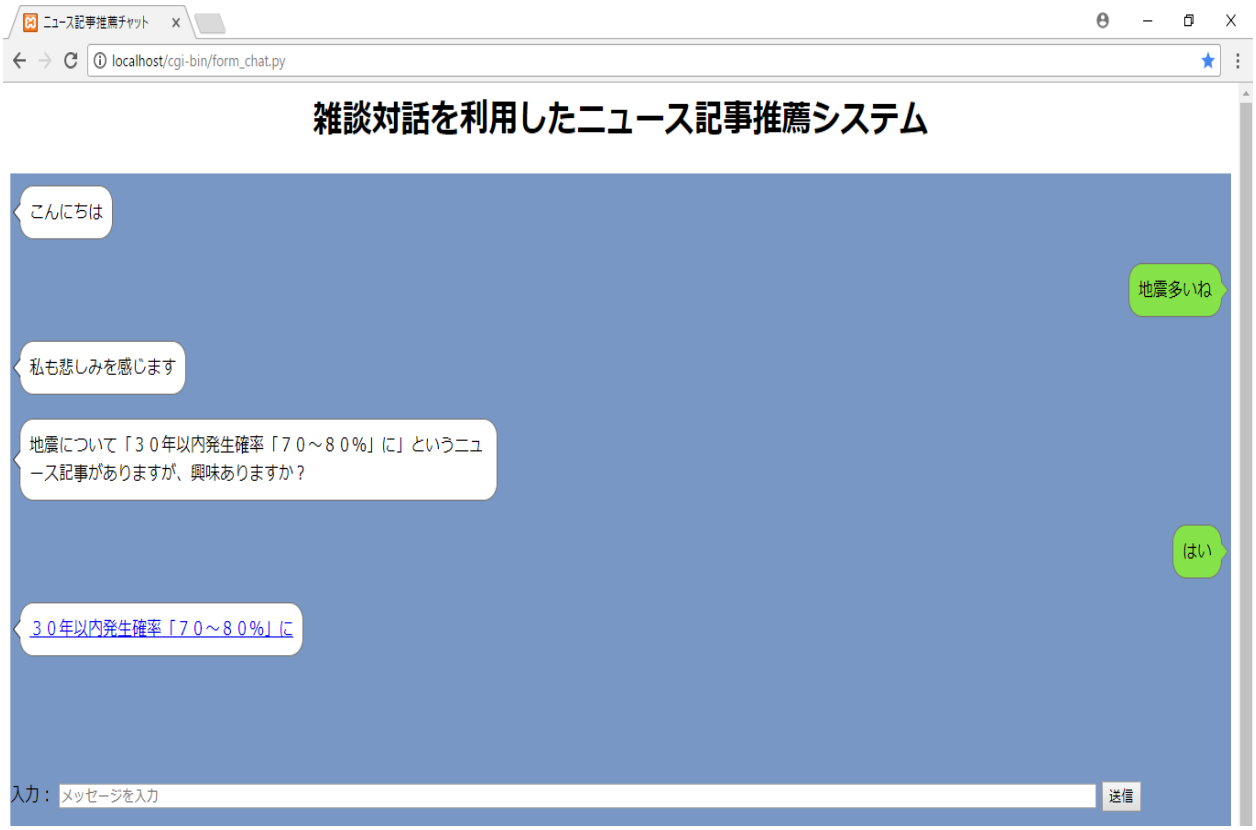

図２　雑談対話・記事推薦の実行例

システムである Juman[7]を用いてユーザ発話文を形態素に分解し，名詞（形式名詞・副詞的名詞を除く）を抽出するとともに，抽出した名詞をその出現頻度とともにランキングデータベース（ランキングDB）に登録する．これらの名詞のうち，最初に出現した名詞が興味語として扱われる．また，ユーザ発話文に対して印象マイニング手法を適用することで求められる 3 つの印象値がその興味語に対してユーザが抱く印象として扱われる．

ニュース記事推薦モジュールは，興味語が抽出された場合と抽出されなかった場合とで処理内容が異なっている．興味語が抽出された場合は，その興味語を含むニュース記事を記事DB内で検索し，結果得られる記事のうち，ユーザ発話文の印象に最も近いニュース記事を推薦する．一方，興味語が抽出されなかった場合は，ランキングDB内のデータを用いて各単語（形式名詞・副詞的名詞以外の名詞）の TF 値を求め，TF 値の高い上位 3 単語を含むニュース記事を記事 DB 内で OR 検索した結果の中から，ユーザ発話文の印象に最も近いニュース記事を選出し，推薦する．なお，単語 w の TF 値は次の式(1)により算出する．

$$TF(w) = \frac{単語w\text{の出現回数}}{全単語\text{の総出現回数}} \quad (1)$$

テンプレート応答モジュールは，まず初めに，ユーザ発話文から得た 3 つの印象値の中で最も近い印象をテンプレートに当てはめ，システム発話文を生成する．テンプレートは「私も○○を感じます」という形式であり，例えば，最も近い印象として「うれしい」が選択された場合は，「私もうれしさを感じます」というシステム発話文が生成される．なお，各印象値は 4.0 が中間であり，「（どちらの印象も）感じない」に対応しているため，すべての印象値が 3.5～4.5 の範囲になった場合は，「印象なし」としてこのテンプレートは利用しない．次に，ニュース記事推薦モジュールから推薦された記事をテンプレートに当てはめ，システム発話文を生成する．このとき，ユーザ発話文から興味語が抽出されている場合は，「[興味語]について[ニュース記事のタイトル]というニュース記事がありますが，興味ありますか？」というテンプレートを用いて，システム発話文を生成する．興味語が抽出されていなかった場合は，「[最も近い印象]といえば，[ニュース記事のタイトル]というニュース記事がありますが，興味ありますか？」というテンプレートを用いて，システム発話文を生成する．最も近い印象が「印象なし」と判定されている場合は，「[最も近い印象]といえば，」というフレーズを省略して，「[ニュース記事のタイトル]というニュース記事がありますが，興味ありますか？」というテンプレートを用いて，システム発話文を生成する．いずれの場合も，ユーザが「はい」のような肯定的

な反応を示した場合は，その記事の URL を辿り，別画面に当該記事の掲示サイトを表示する．

4. システムの実装と実行例

提案システムは，PYTHON を用いて実装した Web アプリケーションであり，ブラウザ上で動作する．雑談対話のためのインタラクション画面は図2に示したようにLINE風になっている．左側の吹き出しがシステムからの発話文であり，システム起動時には「こんにちは」と表示される．ユーザ発話文の入力は画面下部のテキストボックスへの入力により行われるが，入力した内容は右側の吹き出しとして表示される．

図2に示したように，ユーザが「地震多いね」というユーザ発話文を入力した場合，システムは，ユーザ発話文に対して形態素解析を行い，普通名詞「地震」を興味語として抽出するとともに，印象マイニング手法[3]を適用して，ユーザ発話文から表1に示したような 3 つの印象値を算出する．各印象値は，1.0～7.0 の実数値をとり，1 に近いほど「楽しい，うれしい，のどか」という印象を表し，7 に近いほど「悲しい，怒り，緊迫」という印象を表している．ユーザ発話文「地震多いね」に対する印象値は，「うれしい⇔悲しい」が 5.52595，「喜び⇔怒り」が 3.78311，「のどか⇔緊迫」が 4.80353 であった．それぞれの印象値と中間値である 4.0 との差をとると，「楽しい⇔悲しい」が 1.52595，「うれしい⇔怒り」が-0.21689，「のどか⇔緊迫」が 0.80353 となることから，このユーザ発話文の印象は「悲しい」と判定される．その結果，テンプレートである「私も○○を感じます」の○○の部分に，「悲しい」に対応する名詞「悲しみ」が挿入され，「私も悲しみを感じます」というシステム発話文が出力される．

さらに続けてシステム発話文を生成するために，抽出した興味語「地震」を含むニュース記事を記事DB内で検索し，得られたニュース記事の中から最も「悲しい」に近い印象を持つ記事を選択する．その結果，抽出した興味語と選択した記事のタイトルを用いて『地震について「30年以内発生確率「70～80％」に」というニュース記事がありますが，興味ありますか？』というシステム発話文を生成し，ユーザに提示している．このシステム発話文に対し，ユーザが「はい」のような肯定的な応答を返した場合は，その記事の URL を用いて，別画面にニュース記事サイト内の記事を提示する．

表1 ユーザ発話文「地震多いね」の印象値

印象軸		
楽しい ⇔悲しい	うれしい ⇔怒り	のどか ⇔緊迫
5.52595	3.78311	4.80353

5. まとめ

本論文では，ユーザが興味を有する語（興味語）とその興味語に抱いている印象をユーザとの雑談対話を通して学習するとともに，これらの情報をもとにユーザの嗜好に合ったニュース記事を推薦するシステムを提案した．

具体的には，ユーザ発話文に興味語（形式名詞・副詞的名詞以外の名詞）が含まれていた場合は，その興味語を含み，かつユーザ発話文と似た印象を有するニュース記事を推薦し，含まれていなかった場合は，ユーザとの対話履歴の中から出現頻度の高い単語を抽出し，その単語を含むニュース記事の中からユーザ発話文の印象に最も近いニュース記事を推薦する．なお，ユーザ発話文から抽出された名詞はいずれもその頻度情報とともにランキングDBに登録され，雑談対話を通してこのような情報を蓄えていくことで，ユーザの興味のあるニュース記事をより高精度に提示することができるようになると考えられる．

今後の課題としては，まずユーザ発話文の意味理解性能の向上や対話文脈の利用などが挙げられる．現段階のシステムでは，対話文脈を保持していないため，ユーザが「地震に関するニュースある？」といった質問に対し，関連するニュース記事を推薦・提示した後，ユーザが続けて「他には？」と質問しても，地震関連の別のニュース記事を検索することはできなず，名詞「他」を興味語として扱い，ニュース記事を検索してしまう．また，対話の多様性を確保するために，テンプレートを増やしていく必要がある．さらに，提案システムの対話性能や記事推薦の精度を評価するためのユーザ評価実験も行っていきたい．

謝辞 本研究の一部は，福田将治奨学寄附金により実施された．ここに記し，感謝の意を表する．

参 考 文 献

[1] 土方嘉徳，嗜好抽出と情報推薦技術，情報処理，Vol.48，No.9，pp.957-965 (2007)．

[2] 小原恭介，山田剛一，絹川博之，中川裕志，Blogger の嗜好を利用した協調フィルタリングによる Web 情報推薦システム，第 19 回人工知能学会全国大会講演論文集，2C2-02 (2005)．

[3] 熊本忠彦，河合由起子，田中克己，新聞記事を対象とするテキスト印象マイニング手法の設計と評価，電子情報通信学会論文誌，Vol.J94-D，No.3，pp.540-548 (2011)．

[4] 早川豪，岡部誠，尾内理紀夫，Twitter を利用したソーシャルニュース記事推薦システム，情報処理学会研究報告，Vol.2011-DBS-153，No.16，pp.1-4 (2011)．

[5] 斉藤哲也，広田健一，星野准一，Web 情報を用いたキャラクタの発話・世間話モデル，情報処理学会研究報告，Vol.2007-NL-181，No.9，pp.53-58 (2007)．

[6] 毎日新聞，https://mainichi.jp/

[7] 形態素解析 Juman，http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN