

雑談対話からのユーザの興味推定

Estimating User Interest from Chat Dialogues

稲葉 通将^{1*} 高橋 健一¹
Michimasa INABA¹ Kenichi TAKAHASHI¹

¹ 広島市立大学大学院情報科学研究科

¹ Graduate School of Information Sciences, Hiroshima City University

Abstract: This paper proposes a neural network-based model for estimating speakers' levels of interest from dialogues. Our model first converts given utterances into utterance vectors using a word sequence encoder with word attention. Afterward, our novel attention approach, sentence-specific sentence attention extracts useful information for estimating the level of interest. Additionally, we introduce a new pre-training method for our model. Experimental results indicated that it was most effective to use topic-specific sentence attention and proposed pre-training in combination.

1 はじめに

雑談対話システムの研究開発において、対話のパーソナライズは重要な課題の一つである。ユーザの経験やユーザの好みに応じた対話を行うことができれば、ユーザの満足度は大きく向上することが期待できる。対話のパーソナライズを目指した研究としては、雑談対話ログからエンティティ間の関係 [1] やエンティティと述語項構造 [2] を獲得する研究があるが、これらは、過去のユーザの発話を用いてシステム発話を生成することによるパーソナライズの実現を目指したものである。

一方、本研究ではユーザの興味に基づくシステムのパーソナライズを目的とし、トピック（例えば音楽、ファッション、健康など）に対するユーザの興味の度合いを推定する。これにより、システムはユーザが興味のないトピックを避け、興味があるトピックについて対話することが可能になり、ユーザのエンゲージメント・満足度の向上が期待できる。

そこで本研究では、雑談対話中の発話を入力とし、話者の興味を推定するニューラルネットワークベースの手法を提案する。なお、本手法では対話中に出現したトピックだけではなく、それ以外のトピックに対しても同時に推定を行う。これは、たとえユーザが興味を持つトピックであっても、延々と同じトピックについて話せばユーザは飽きてしまう危険性があるためである。対話中に直接出てきていないトピックに対しても興味の有無が推定できれば、ユーザが飽きないうちに話を次のトピックに展開させることが可能となる。

なお、現在の対話システムの性能は十分ではなく、自

然な対話を行うことが困難であることから、本研究では人間同士の対話を対象とする。また、対話システムの発話に依存しない興味推定を実現するため、本研究では対話相手の発話は利用せず、興味推定対象の話者の発話のみを興味推定のために用いる。

2 関連研究

テキストデータからのユーザの興味推定に関する研究では、SNS、特に Twitter を対象とした研究が活発に行われている。例えば Chen らは対象となるユーザのツイート、およびそのユーザのフォロワーのツイートの bag-of-words を用いて興味をモデリングする手法を提案した [3]。また、知識ベースから獲得した単語の上位概念を用いる手法もいくつか提案されている。例えば、Abel らは OpenCalais¹ を用いて取得した人名 (Person)、イベント名 (Events)、音楽グループ名 (Music groups) といった固有表現の出現頻度により Twitter ユーザのモデリングを行った [4]。その他にも単語の上位概念を用いる同様の手法として、Wikipedia のカテゴリーを用いた手法 [5]、DBPedia を用いた手法 [6] などが提案されている。以上の手法はユーザは自分が興味のあることのみを投稿している、という前提に基づいている。この前提は、ユーザが常に主体的に発信する SNS であれば問題ない。しかし、対話では人は必ずしも興味のあることについてのみ発話するとは限らない。対話では、ユーザが現在の話題にそれほど興味がなくても、相手に話を合わせ対話を継続することはそれほど珍しいことではないため、これらの手法を直接適用すること

*連絡先： 広島市立大学大学院情報科学研究科
〒731-3194 広島市安佐南区大塚東 3-4-1
E-mail: inaba@hiroshima-cu.ac.jp

¹<http://www.opencalais.com/>

が難しい．後述する実験では，これらの SNS を対象とした興味推定手法の 1 つを比較手法として用い，対話への適用が困難であることを示す．

対話からのユーザの興味を推定する研究としては，ユーザの興味レベル (Level-of-Interest: LOI) の推定に関する研究が知られている．Schuller らは，ある製品に関して紹介する話者とその聞き手の 2 者間の対話において，聞き手側の製品に対する興味を推定するというタスクを対象とし，音響的素性と言語的素性をを用いた SVM により 3 段階で興味を推定した [7]．また，2010 年には INTERSPEECH Paralinguistic Challenge のサブチャレンジとして LOI 推定タスクが設定された [8]．ただし，これらの研究では興味を推定する対象はその対話中で話される話題 1 つに決められており，オープンドメインではない．また，対話内容もビジネス会話であり，雑談を対象としたものでもない．

3 興味推定手法

本研究で対象とするタスクは，話者 s が他の話者との対話中に行った発話集合 $U_s = (u_1, u_2, \dots, u_n)$ とえられ，興味推定対象のトピック $T = (t_1, t_2, \dots, t_m)$ に対応する話者 s の興味度 $Y_s = (y_1, y_2, \dots, y_m)$ を推定することにある． t_i はそれぞれ具体的なトピック，例えば映画や旅行などに対応する．また， y_i は各トピックに対する話者の興味の度合いを意味する値であり，前節で述べた LOI 推定の研究に合わせ 3 段階とした．すなわち， $y_i = 0$ は「興味なし」を， $y_i = 1$ は「少し興味あり」を， $y_i = 2$ は「興味あり」を意味する．

発話から興味の推定を行うには，以下の 2 点の情報を抽出することが重要である．

- その発話はどのようなトピックに関するものであるか
- 話者はそのトピックについてどの程度興味を持っているか

そのために本研究では，効率的な情報抽出を行うためのニューラルネットワークモデル，およびモデルの効率的な学習のための事前学習手法を提案する．提案モデルの概要を図 1 に示す．図に示したように，提案モデルは単語系列エンコーダ，単語 Attention，トピック別発話 Attention，興味推定層の 4 つが順に処理を行い，興味推定を行う．なお，今回提案する事前学習手法は単語系列エンコーダに対して適用する．以下ではそれぞれの詳細について説明する．

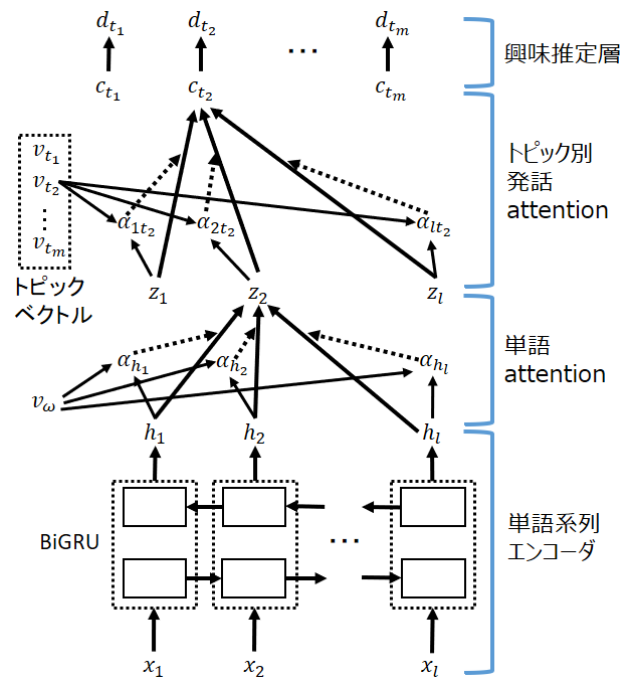


図 1: 提案手法の概要

3.1 単語系列エンコーダ

単語系列エンコーダは発話，すなわち単語の系列を Recurrent Neural Network (RNN) エンコーダを用いて固定長のベクトルにエンコードする．最初に，発話中の単語は学習済みの Word2Vec によりベクトルに変換され，単語ベクトルの系列 $x = (x_1, x_2, \dots, x_l)$ を得る．本研究では RNN の中間層に Bidirectional-GRU (BiGRU) を用いる．BiGRU は系列を x_1 から x_l の順に読み込む順方向 GRU (\overrightarrow{GRU}) と x_l から x_1 の順に読み込む逆方向 GRU (\overleftarrow{GRU}) からなり，順方向 GRU において，時刻 i における隠れ状態 \vec{h}_i は以下の式により得られる．

$$\vec{h}_i = \overrightarrow{GRU}(x_i, \vec{h}_{i-1}) \quad (1)$$

backward GRU についても \overleftarrow{h}_i を同じく計算する．目的となる隠れ状態 h_i は順方向 GRU と逆方向 GRU にそれぞれ x_i を入力した際の出力を用いて以下の式により求められる．

$$h_i = [\vec{h}_i : \overleftarrow{h}_i] \quad (2)$$

ここで $[\cdot]$ はベクトルの連結を意味する．

3.2 トピック分類事前学習

ユーザがどのトピックにどれだけ興味を持つかを推定するためには，まず各発話がどのトピックに関する

ものかを推定する必要がある．すべての発話に人手でアノテーションすることも可能だが，その作業コストは極めて高い．一方，トピックの情報が明示的に与えられない場合，モデルは発話集合とそのユーザの興味度から間接的にトピックを推定する必要があるため，適切な学習は困難である．そこで本モデルでは単語系列エンコーダに対し，事前学習を行う．本モデルの事前学習では，発話中のトピックをより効率的に推定するため，単語系列エンコーダに対し，文のトピック分類を行う学習を実施する．ここで重要な点は，分類クラスを本研究が対象とする興味推定対象トピック T と同一にすることである．このことにより，興味推定の学習時に発話とトピックの対応関係を推定するための学習難易度が下がり，適切な学習が可能となる．

事前学習時には入力文の最後の発話 x_l を BiGRU に入力した際の出力 h_l を用いて，入力文の各トピッククラスへの所属確率 p を以下の式により計算する（ここでは次節で述べる単語 Attention は使用しない）．

$$p = \text{softmax}(W_c h_l + b_c) \quad (3)$$

式中の W_c と b_c は事前学習時にのみ使用されるトピック分類のためのパラメータである．事前学習時の損失関数にはクロスエントロピーを用いる．

3.3 単語 attention

本モデルには Yang らが提案した単語 attention[9] を導入する．単語 attention は，「全ての単語が求めたい結果に平等に貢献するわけではない」という考えに基づき，attention 機構により単語ごとに異なる重みづけを行ったエンコード結果を得るための手法である．例えば「私は野球が好きです」という発話の意味を理解するためには「は」や「です」という単語よりも「野球」や「好き」という単語のほうが重要度が高いと考えられ，単語 attention は後者の単語により大きな重みを与える．

単語 attention よる発話エンコードベクトル z は以下の式により求められる．

$$v_{hi} = \tanh(W_\omega h_i + b_\omega) \quad (4)$$

$$\alpha_{hi} = \frac{\exp(v_{hi}^\top v_\omega)}{\sum_i \exp(v_{hi}^\top v_\omega)} \quad (5)$$

$$z = \sum_i \alpha_{hi} h_i \quad (6)$$

W_ω と b_ω はパラメータである．単語 attention はニューラル翻訳モデルやニューラル対話モデルのようなデコーダを用いて系列を生成する場合の(通常)attention と異なり，各単語において同一の文脈ベクトル v_ω を α_i

を計算に用いる． v_ω は単語の重要性を計算するためのパラメータであり， W_ω と b_ω と同様，学習時に最適化される．

3.4 トピック別発話 attention

単語系列エンコーダは，発話 $U_s = (u_1, u_2, \dots, u_n)$ を発話ベクトル $Z_s = (z_1, z_2, \dots, z_n)$ にエンコードする．次は，この発話ベクトルからユーザのトピックごとの興味推定のための情報を抽出する．

そのために本研究では，単語 attention と同じく，全ての発話が各トピックの興味度の推定に平等に貢献するわけではないという考え方にに基づき，トピック別発話 attention を新たに提案する．なお，Yang らは単語 attention と同じ構造の文 attention[9] を提案しているが，文 attention では文の重要性を計算するためのパラメータは 1 つである(単語 attention における文脈ベクトル v_ω と同等のパラメータ)．しかし，与えられた発話集合において，どの発話が重要であるかはトピックごとに異なると考えられる．例えば「私は毎朝ジョギングをしています」という発話からはスポーツや健康といったトピックの興味推定には役立つと考えられるが，ゲームや車・乗り物に対する興味推定にはおそらく有用ではない．

そこでトピック別発話 attention では，トピック t_i に対する発話の重要性を計算するためのトピックベクトル v_{t_i} を用いて，内容ベクトル c_{t_i} を以下の式により求める．

$$v_j = \tanh(W_r z_j + b_r) \quad (7)$$

$$\alpha_{jt_i} = \frac{\exp(v_j^\top v_{t_i})}{\sum_j \exp(v_j^\top v_{t_i})} \quad (8)$$

$$c_{t_i} = \sum_j \alpha_{jt_i} z_j \quad (9)$$

式中の W_r はトピックによらず共通で用いるパラメータである．また，トピックベクトル v_{t_i} は学習により最適化される．

3.5 興味推定層

トピック別発話 attention により得られた内容ベクトル c_{jt_i} を用いて，トピック t_i に対する興味度 d_{t_i} を以下の式により求める．

$$d_{t_i} = \tanh(W_{t_i} c_{t_i} + b_{t_i}) + 1 \quad (10)$$

ここで，パラメータ W_{t_i} はトピック t_i の興味度推定のためのパラメータであり，対象トピックごとに異なるパラメータを学習する．最後の +1 は興味度 d_{t_i} を正解

表 1: 使用したトピック

| | | |
|----------|--------|--------|
| 旅行 | 映画 | 音楽 |
| アウトドア | 学校 | 家電製品 |
| 芸能人・タレント | PC 関連 | ビデオゲーム |
| 動物・生き物 | 住まい | 家事 |
| 家族・夫婦 | 結婚・恋愛 | 健康 |
| 美容 | ファッション | 料理・グルメ |
| スポーツ・運動 | 読書 | 車・乗り物 |
| 政治・経済・社会 | アニメ・漫画 | 歴史 |

である y_i と同じ 0~2 にスケールリングするための項である。

学習時の損失関数には下式に示すように、正解である y_i と d_{t_i} の平均二乗誤差を用いる。

$$L = \frac{1}{n} \sum_i^n (y_i - d_{t_i})^2 \quad (11)$$

4 評価実験

4.1 使用データ

提案手法の評価のため、初対面の被験者 2 名による日本語のテキストチャットデータと各被験者の興味に関するアンケートを収集した。被験者はクラウドワークス²で募集した。各被験者には対話を行う前にアンケートに回答するよう依頼し、アンケートの内容は表 1 に示す 24 種類のトピックそれぞれに対して、自分の興味の度合いを「興味なし」、「少し興味あり」、「興味あり」の 3 択で選択させた。表 1 のトピックは文献 [2] と同じく、Yahoo!知恵袋³のカテゴリをもとに、2 者間の初対面の対話で取り上げられやすい話題を中心に設定した。

対話は 1 回 1 時間とし、Skype のインスタントメッセージングを介して行った。被験者に対しては「話の中で相手と自分の間で共通して興味のあることを見つけ、それについて話を膨らませていくことを心がけてください」とだけ教示し、話すべき内容・話題についての指示は行わなかった。

収集した全てのチャットデータ中の発話を被験者別に分類し、1 被験者の行った全発話とその被験者のアンケートの結果のペアを 1 データとし、10 分割交差検証により評価する。データの統計情報を表 2 に示す。

表 2: 使用データ

| | |
|------------------|-------|
| ユーザ数 (データ数) | 163 |
| 対話数 | 408 |
| 発話数 | 49029 |
| 「興味あり」の平均トピック数 | 11.48 |
| 「少し興味あり」の平均トピック数 | 7.30 |
| 「興味なし」の平均トピックス数 | 5.21 |

4.2 実験設定

単語ベクトルの獲得には次元数 200、ウインドウサイズは 5、最小出現頻度は 10 とし、約 100GB の Twitter データで学習を行った Word2Vec を用いた。

単語系列エンコーダの入力層の次元は 200、BiGRU の出力の次元は 400 とし、BiGRU は 1 層とした。また、各 attention 層は入出力ともに 400 次元、興味推定層の入力は 400 次元、出力は 1 次元とした。

トピック分類の学習 (事前学習) には、Yahoo!知恵袋データ (第二版)⁴を使用した。各トピックに対応する質問文、および回答文から、77 万文を学習データ、各トピック 100 文ずつの 2400 文をテストデータとした。事前学習の結果、テストデータの正解率は 0.755 であった。

4.3 評価方法

提案手法を対話システムの話題選択に用いることを想定した場合、対話システムがその時点で対応できる話題の中から、より望ましい話題を提案手法が選択するという活用法が有効であると考えられる。そこで、本実験では推定した興味度 d_{t_i} に基づいて各トピックを順位付けし、上位にユーザが興味を持つトピックを、下位に興味を持たないトピックを順位付けできる性能により評価を行う。順位付けの評価尺度として、情報検索の分野で広く使用されている指標である Normalized Discounted Cumulative Gain ($NDCG$) を用いる。 $NDCG$ は 0 から 1 の値を取り、1 に近いほど正しい順位予測結果であることを意味する。 $NDCG$ は以下の式により計算する。

$$NDCG@k = \frac{DCG_k}{IDCG_k} \quad (12)$$

$$DCG_k = rel_i + \sum_{i=2}^k \frac{rel_i}{\log_2 i} \quad (13)$$

ここで、 k は上位何位までを考慮するかを意味する。 rel_i は i 位に順位付けされたターゲットに対する適合度を意味する値であり、本研究では各トピックに対するユーザ s の興味度 Y_s の値 (0, 1, 2) を適合度とする。 $IDCG$

²<https://crowdworks.jp/>

³<https://chiebukuro.yahoo.co.jp/>

⁴http://www.nii.ac.jp/dsc/idr/yahoo/chiebkr2/Y_chiebukuro.html

(Ideal DCG) は順位リストが適合度順に正しく並んだ場合の DCG である。

4.4 比較手法

提案手法の有効性の比較のため、提案手法の一部を変更した以下の3つの設定の実験を行う。

事前学習なし

提案した事前学習を行わず、単語系列エンコーダの初期値を乱数として学習を行う。その他の設定は提案手法と同一である。

トピック別発話 attention なし

トピック別発話 attention による内容ベクトルの獲得の代わりに、発話ベクトルの Max Pooling の結果を用いて内容ベクトルを獲得し、興味推定を行う。その他の設定は提案手法と同一である。

事前学習なし & トピック別発話 attention なし

提案した事前学習を行わず、発話ベクトルの Max Pooling により内容ベクトルを獲得し、興味推定を行う。その他の設定は提案手法と同一である。

また、以下の2つのベースライン手法とも比較を行う。

トピック頻度

1つ目のベースラインとして、Abelらの手法[4]をもとにした興味推定手法を用いる。Abelらの手法はSNSユーザの興味モデリング手法として提案されたものであり、OpenCalais⁵を用いてユーザのツイートに含まれる単語の固有表現を取得し、その頻度ベクトルによりユーザの興味をモデル化した。ただし、本研究では日本語対話を対象としており、英語のみを扱えるOpenCalaisは利用できない。そこで3.4節で述べた事前学習で作成された文のトピック分類器を使用し、発話のトピックを推定する。推定したトピックの出現頻度をトピックへの興味とみなし、頻度順でトピックを順位付けする。

SVR

2つ目のベースラインとして、Support Vector Regression (SVR) を用いてユーザの各トピックの興味度を推定する。素性は unigram と bigram を用い、unigram のみを用いた場合と両方を用いた場合で実験を行う。SVR モデルはトピックごとに個別に学習を行い、各トピックに対する興味度の推定を行う。カーネルは RBF カーネルとした。

⁵<http://www.opencalais.com/>

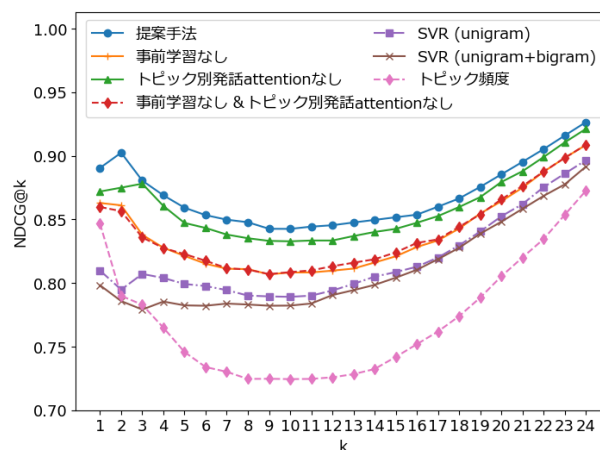


図 2: 実験結果 (NDCG@k)

4.5 実験結果

実験結果を図2に示す。この図はNDCGの計算に用いるトピック数 k を1から24まで変化させた場合のNDCGの値を図示したものである。図より、提案手法が他の手法と比べ、すべての k においてNDCGが高いことがわかる。事前学習を適用した手法(提案手法・トピック別発話 attention なし)と、適用していない手法(事前学習なし・事前学習なし&トピック別発話 attention なし)を比較すると、事前学習を適用した手法が大きく性能を伸ばしていることがわかる。ここから、提案した事前学習の有効性が確認できる。一方、トピック別発話 attention を適用した事前学習なしと事前学習なし&トピック別発話 attention なしの結果がほぼ同じであることがわかる。ただし、わずかな差ではあるが $k \geq 5$ において、事前学習なし&トピック別発話 attention なしの方がNDCGは大きい結果となっている。ここから、事前学習を行わずにトピック別発話 attention を使用することは性能の向上に寄与しないことがわかる。一方、事前学習とトピック別発話 attention を同時に適用した提案手法は、トピック別発話 attention なしと比べ性能が向上している。以上のことから、本研究で提案した事前学習とトピック別発話 attention は同時に適用することが重要であることが確認できた。これは、トピック分類事前学習を行うことで発話ベクトルにトピックの情報が埋め込まれるため、式8のトピックベクトル v_{t_i} が効率的に学習可能になったためであると考えられる。

SVRに関しては、unigram と bigram を用いる場合よりも unigram を単独で用いた方が良い性能を示したものの、提案手法を含むニューラルネットワークベースの手法に劣る結果となった。トピック頻度ベースラインは $k = 1$ の場合はSVR手法よりも高いNDCGを

示した。これは、ユーザが積極的に発言したトピックに関して興味ありと正しく推定できたためである。一方、ニューラルネットワークベースの手法には劣る結果となり、 $k \geq 4$ の場合、すべての手法の間で最低の NDCG となった。トピック頻度ベースラインは SNS の投稿を対象とした手法をベースとしたものであるが、SNS とは異なり、対話では話者は常に興味のあることのみを発言するわけではなく、対話相手に話を合わせるため興味小さいトピックに関して発言していることが低い性能となった原因であると考えられる。

5 おわりに

本研究では、雑談対話におけるユーザの発言を入力とし、事前に設定したトピックに対するそのユーザの興味度を推定するニューラルネットワークベースの手法を提案した。提案手法はまず、BiGRU を用いた単語系列エンコーダと単語 attention により発言をエンコードし、発言ベクトル集合を得る。次に、この発言ベクトル集合からトピック別発言 attention を用いて、推定対象の各トピックに対応したベクトルを生成する。最後に、このベクトルから対応するトピックに対する話者の興味度を推定する。学習時にはトピック分類を行う事前学習を実施することで単語系列エンコーダを学習し、その後、興味度推定の学習を行う。実験の結果、提案手法はベースライン手法と比べ、より適切にトピックに対する興味度を推定できることが確認された。また、提案したトピック別発言 attention とトピック分類事前学習を併用することが性能向上に有効であることが確認された。

今後は、提案手法を対話システムに適用し、ユーザと対話を行う予定である。ユーザが興味を持つトピックが推定できたとしても、具体的な発言をどのように生成するかはまた別の問題である。例えばスポーツに興味があるユーザは、すべての種類のスポーツに興味があるとは限らない。サッカーには興味があるがゴルフには興味がないといったことは十分考えられる。そのため、提案手法の有効な適用方法について検討していきたい。

謝辞

本研究の実施にあたっては、ヤフー株式会社が国立情報学研究所に提供した「Yahoo!知恵袋データ (第2版)」を利用いたしました。また、本研究は JSPS 科研費 16H05880 の助成を受けたものです。

参考文献

- [1] Jeessoo Bang, Hyungjong Noh, Yonghee Kim, and Gary Geunbae Lee. Example-based chat-oriented dialogue system with personalized long-term memory. In *IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 238–243, 2015.
- [2] 平野徹, 小林のぞみ, 東中竜一郎, 牧野俊朗, 松尾義博. パーソナライズ可能な対話システムのためのユーザ情報抽出. 人工知能学会論文誌, Vol. 31, No. 1, pp. DSF-B-1–10, 2016.
- [3] Jilin Chen, Rowan Nairn, Les Nelson, Michael Bernstein, and Ed Chi. Short and tweet: experiments on recommending content from information streams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1185–1194. ACM, 2010.
- [4] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Analyzing user modeling on twitter for personalized news recommendations. *User Modeling, Adaption and Personalization*, pp. 1–12, 2011.
- [5] Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri, Mohsen Kahani, and Weichang Du. Semantics-enabled user interest detection from twitter. In *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pp. 469–476, 2015.
- [6] Pavan Kapanipathi, Fabrizio Orlandi, Amit P Sheth, and Alexandre Passant. Personalized filtering of the twitter stream. pp. 6–13, 2011.
- [7] Björn W Schuller, Niels Köhler, Ronald Müller, and Gerhard Rigoll. Recognition of interest in human conversational speech. In *INTERSPEECH*, 2006.
- [8] Björn Schuller, Stefan Steidl, Anton Batliner, Felix Burkhardt, Laurence Devillers, Christian Müller, and Shrikanth Narayanan. The interspeech 2010 paralinguistic challenge. In *Proc. INTERSPEECH 2010, Makuhari, Japan*, pp. 2794–2797, 2010.
- [9] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J Smola, and Eduard H Hovy. Hierarchical attention networks for document classification. In *HLT-NAACL*, pp. 1480–1489, 2016.