

過去会話の整合性に着目したコンピュータ会話における違和感の抽出
 Extraction of Sense of Incompatibility in Computer Conversation
 Paying Attention to Correspondence of Past Conversation

山本 裕司† 吉村 枝里子† 渡部 広一† 河岡 司†
 Yuji Yamamoto Eriko Yoshimura Hirokazu Watabe Tsukasa Kawaoka

1. はじめに

コンピュータと人間の円滑な会話の実現には、過去の会話を記憶として蓄積し、それを利用した会話が必要である。過去の会話を利用することによって、話者の行動の流れを記憶し会話の幅を広げることができると同時に、現在の会話内容と過去の会話内容の矛盾を把握し、会話に活かすことが可能となる。これまでに、現在の会話と過去の会話との内容による矛盾を検出する研究が報告されている^[1]。しかし、過去の会話に対し、行動間隔による違和感を検出することはできなかった。たとえば、「今日、車を買った」という発言が二日連続で行われると、我々は違和感を感じる。これは、「車を買う」という行動に対する行動間隔の常識を保持しているためだと考えられる。そこで、本稿では行動間隔の常識について整理を行う。また、その整理された知識に基づいて、過去の会話と現在の会話の内容と行動間隔を比較し、違和感のある会話を抽出する方法を提案する。

2. 関連技術

2.1 概念ベース^[2]

概念ベースとは、複数の国語辞書や新聞等から機械的に構築した、語（概念）とその意味特徴を単語（属性）の集合のセットを約 12 万語格納した知識ベースである。概念 A に付与される属性 a_n には、その重要性を表す重み w_n が付与されている。

$$A = \{(a_1, w_1), (a_2, w_2), \dots, (a_n, w_n)\}$$

2.2 関連度計算^[2]

関連度計算とは、概念ベースに定義された語と語の関連の強さを定量的に評価する手法である。関連度の値は 0 から 1 で表され、概念同士の関連が強いほど関連度の値が高くなる。

2.3 シソーラス^[3]

本研究で用いるシソーラスとは、一般名詞の意味的用法を表す約 2700 の意味属性(ノード)の上位下位関係、全体部分関係が木構造で示されたものであり、約 13 万語(リーフ)が登録されている。

2.4 過去会話応答システム

過去会話応答システムとは蓄積した過去の会話を利用して応答を行うシステムである。過去の会話情報を整理した過去会話知識ベースを用い、入力文に対して過去の類似の行動や、過去と矛盾のある行動について応答を行

う。例えば、「昨日テニスをしました」という入力に対して、「よくテニスをしますね」や「昨日はサッカーをする予定ではなかったのですか?」といった応答を行う。

3. 違和感判断システム

本稿で提案する違和感判断システムは、入力文と過去の会話を比較し、行動間隔に違和感があるかどうかを判断するシステムである。人間の行動間隔のルールを格納した知識ベースから入力文の行動間隔を取得し、過去会話知識ベースを用いて違和感の判断を行う。出力形式は「違和感あり」、「違和感なし」の 2 択となる。

入力文の処理形態の違いから、文を「挨拶文」、「動作文」の 2 種類に分類する。それぞれの文章には以下のような特徴がある。

1) 挨拶文

「おはようございます」や「久しぶり」といった挨拶に関する文である。挨拶文では挨拶ごとに発言間隔が決まっている。例えば、「おはようございます」という挨拶は一日に 1 度であり、「久しぶり」といった挨拶は数日間会っていない人に対して使われる。挨拶にはこのように規則があり、その規則に反したとき違和感となる。

2) 動作文

「旅行に行きました」や「テニスをしました」といった動作に関する文である。動作によってその間隔が異なる。例えば「旅行に行く」という動作は年 2 回程度であるが、「テニスをやる」は月 2,3 回程度である。この動作間隔は人によって異なるが、アンケートにより求めた動作間隔を一般的な動作間隔とする。動作間隔には、間隔が近すぎる場合の違和感と間隔が遠すぎる場合の違和感がある。

4. 挨拶文の違和感判断

「挨拶文」と判断した入力文について、挨拶文の過去会話履歴を格納した挨拶文過去会話知識ベース(表 1)、挨拶の規則を格納した挨拶間隔知識ベース(表 2)を用いて違和感の有無を判断する。挨拶文過去会話知識ベースには、予め挨拶間隔知識ベースを用いて「種類」を付与しておく。

表 1 挨拶文過去会話知識ベース

会話日	入力文	種類
1/16 10:00	おはようございます	出会い
1/10 13:00	こんにちは	出会い

表 2 挨拶間隔知識ベース (190 文)

挨拶語	挨拶間隔	種類
おはよう	1 日	出会い
ありがとう	—	その他

† 同志社大学院工学研究科

Graduate School of Engineering Doshisha University

挨拶間隔知識ベースから、入力文の「種類」, 「挨拶間隔」を獲得し、挨拶文過去会話知識ベースを用いて過去の「挨拶間隔」以内に同様の「種類」の入力文があるかを検索する。同様の「種類」の挨拶がある場合は「違和感あり」、ない場合は「違和感なし」となる。

5. 動作文の違和感判断

「動作文」と判断した入力文について動作文の過去会話履歴を格納した動作文過去会話知識ベース, 動作文の動作間隔を格納した動作間隔知識ベースを用いて違和感の有無を判断する。

5.1 動作間隔の獲得

入力された動作文に対し、表3の動作間隔知識ベースを参照し、動作間隔の獲得を行う。

表3 動作間隔知識ベース (90語)

用言	目的語 (親ノード)	最小動作 間隔	最大動作 間隔
行く	スポーツ	1日	1月
	行楽	1週	半年
	公共施設	1週	1年

シソーラスから入力文の目的語の全親ノードを取得し、動作間隔知識ベースの目的語項目に親ノードが存在すれば「最小, 最大動作間隔」を獲得する。親ノードが知識ベース内に存在しない場合は、未知語処理^[4]を行う。未知語処理とは、知識ベースに存在しない語に対して関連度計算を用いて意味の近い代替語に置き換える手法である。

入力文の目的語と動作間隔知識ベースの目的語項目の全リーフとの関連度計算を行い、最高関連度となる語を代替目的語として取得する。代替目的語との関連度が閾値以下の場合は「違和感なし」、閾値以上であれば代替目的語の親ノードから「最小, 最大動作間隔」を獲得する。閾値は実験により求められた0.05とする。

5.2 過去会話との違和感判断

獲得した「最小, 最大動作間隔」から表4の動作文過去会話知識ベースを参照し、入力文の違和感の有無を判断する。

表4 動作文過去会話知識ベース

会話日	動作日	主体	場所	目的語	動詞
1/16 19:00	1/15	私	—	テニス	する
1/4 22:00	1/3	私	初詣		行く

入力文と同様の動作が動作文過去会話知識ベース内の「最小, 最大動作間隔」期間外にのみ存在した場合は「違和感あり」、それ以外は「違和感なし」となる。

6. システムの評価

6.1 評価方法

過去会話知識ベースにあらかじめ格納されている400文に対し、アンケートにより取得した200文(挨拶文50文, 情報文150文)について違和感判断システムで違和感の有無を判断し、その結果を3人の人間で目視評価を行った。評価は3人全員が正しいと判断した場合を○, 3人中2人が正しいと判断した場合を△, それ以外を×とする。

6.2 評価結果・考察

評価結果を図1に示す。システムの精度は○・△あわせて72%という精度になった。挨拶文は挨拶間隔知識ベースを参照しており100%の精度を得ることができた。失敗の主な原因は動作文にあり、特に動作間隔獲得での失敗が挙げられる。失敗例を表5に示す。

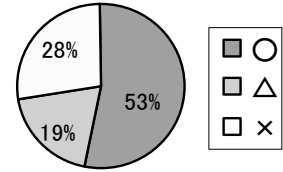


図1 評価結果

表5 動作間隔獲得の失敗例

入力文	代替 目的語	親ノード	最小動作 間隔	最大動作 間隔
スキーに行く	—	スポーツ	1日	1月
努力をする	奮戦	スポーツ	1日	1月

1)シソーラスでの失敗

「スキーに行く」では、目的語(親ノード)が「スポーツ」となり、「スポーツ」の動作間隔「日~月」が獲得された。しかし、「スキー」のように季節が関係する動作は動作期間が限定されており、この結果は正しくないといえる。このような失敗に対しては、目的語の季語を判断するシステムを用いて例外処理を行うことで解決できると考えられる。

2)関連度計算での失敗

「努力をする」では、未知語処理を行い、代替目的語「奮戦」とその親ノードである「スポーツ」が獲得された。しかし、「努力をする」は違和感の生じない文であり、判断することができない。このような目的語が抽象名詞である文は「違和感なし」と考えられるので、目的語が具体名詞であるかを判断することで精度の向上が考えられる。

7. おわりに

本稿では、行動間隔の常識について整理し、過去と現在の会話で行動間隔において違和感のある会話の抽出を行う手法を提案した。システムの精度は72%となった。これにより、行動間隔における違和感を利用した応答方法の作成が可能となる。しかし、「動作文」に関しては、未知語処理、例外処理を用いて精度向上を図ることがこれからの課題となる。

参考文献

- [1]空野皇司, 吉村枝里子, 渡部広一, 河岡司, “常識的なコンピュータ会話のための過去会話知識ベースの自動構築方式”, 電子情報通信学会 2008 総合大会講演論文集, D-5-3 (2008).
- [2]奥村紀之, 土屋誠司, 渡部広一, 河岡司, “概念間の関連度計算のための大規模概念ベースの構築”, 自然言語処理, Vol.14, No.5 pp.41- 64 (2007).
- [3]NTT コミュニケーション科学研究所監修, “日本語語彙体系”, 岩波書店 (1997).
- [4]土屋誠司, 小島一秀, 渡部広一, 河岡司, “常識的判断システムにおける未知語処理方式”, 人工知能学会論文誌, Vol.17, No.6, pp.667-675 (2002).