

異文化間コミュニケーションにおける重要度を考慮した文化差検出手法の提案

Proposal of Cultural Difference Detection Method Considering Importance

of Cross-Cultural Communication

諏訪 智大[†] 宮部 真衣[‡] 吉野 孝^{††}
Tomohiro Suwa Mai Miyabe Takashi Yoshino

1. はじめに

近年、情報技術の発展により、世界中の人と簡単にやりとりすることが可能になり、文化の異なる人とのコミュニケーションの重要性が増してきている。しかし異文化間コミュニケーションにおいては、同一の単語を用いて会話をしている場合でも、相手の国や文化について十分に理解していないために、誤解が生じる可能性がある [1]。また相手の国の言葉を使用して互いに円滑にコミュニケーションできていたとしても、相手の国や文化についての知識が足りないことで、互いに思い浮かべているものが違ったりする場合がある。このように異文化間のコミュニケーションでは、知らないうちにどこかでコミュニケーションにすれ違いや誤解が生まれ、大きなトラブルにつながる可能性がある。

異文化間の摩擦を和らげるために、お互いの価値観などに違いがあることを認識する必要がある。これまでに、遠隔チャット中や対面コミュニケーション中に、画像等のアノテーションを付与する手法を用いて、誤解を減らす工夫がなされている [2][3]。しかし、アノテーションの付与が必要となる語句の選択は、利用者自身が行う必要があった。つまり、文化差の有無の判断は、利用者自身で行う必要がある。しかし、その判断は、相手の文化に関する十分な知識が必要であり、容易ではない。そのため、文化差が存在することを自動的に検出する仕組みが求められている。

我々はこれまでにいくつかの文化差検出手法を提案してきた [4]。しかし、語句の文化差には、コミュニケーションの内容によって、重要な文化差とそうでない文化差が存在するはずである。これまでの手法は、語句に対して文化差が存在するかどうかの判定に留まっており、実際にその語句には、どのような文化差が存在し、どの程度の違いがあるのかわからない。そこで本稿では、コミュニケーションにおける重要度を考慮した文化差検出手法を提案する。

2. 関連研究

関連研究として、まず、異文化間コミュニケーションにおける、文化差に関する研究を示す。Cho らは、異文化話者らがコンピュータとネットワークを介してコミュニケーションを行う際に用いる絵文字に着目した。絵文字は、異文化間で普遍的に解釈されないという問題がある。そこで、その問題を解決するために、解釈に文化差のある絵文字の検出における工学的な手法の適用可能性について検討した [5]。検討の結果、従来の工学的な手法では、人の文化差判定を近似することは困難であることを示した。Koda らは、アバターを介したコミュニケーションにおける、異文化間での表情の解釈に着目した。アバ

ターの表情に関するユーザの解釈について実験を行った結果、表情の解釈が文化によって大きく異なることを示した [6]。文化差に関しては、これまでにいくつかの検討が行われているが、文化差判定は容易ではない。

また、Wikipedia を利用した文化差に関する研究がある。Ulrike らは、Hofstede が明らかにした、国ごとに存在する 4 つの文化的多様性 [7] を Wikipedia に適用し、それぞれの国ごとの Wikipedia の編集操作は、それぞれの国の文化的特徴と関連することを明らかにした [8]。中崎らは、ある同一のトピックについて書かれた日英ブログエントリを選別することで、互いの関心・意見の共通点・差異がどの程度存在するかを発見することを試みている [9]。具体的には、トピックの語句に関して、日英 Wikipedia エントリから抽出した関連語を用いて、あるトピックについて書かれたエントリを選別する。そして、エントリに出現する語句の頻度を見ることで実現している。

このように、Wikipedia は知識抽出分野で資源として注目を集めており、様々な利用が検討されている。しかし、これまでに、Wikipedia のデータを利用した、コミュニケーション支援のための語句の文化差判定・抽出に関する試みは行われていない。

3. 文化差と重要度の定義

3.1 文化差

文化差を定義するためには、まず「文化」の定義が必要である。「文化」(Culture) の定義は、日本と欧米では異なり、一概に定義することは困難である [10]。例えば、今日欧米で用いられる「文化」は、「知識、信仰、芸術、道徳、慣習、その他社会の一員としての人間によって獲得される能力や習慣を包含する複合体である」と定義づけられている [10]。このような「文化」を単純に「測る」ことは困難であるが、コミュニケーション支援に文化差検出手法を適用するためには、何らかの尺度を考える必要がある。そこで本稿では、特に「知識」の面から「文化」を捉えることとし、まず、文化差検出手法の第一歩として、形式知化された知識の違いで文化差を測ることとした。

これまでの文化差検出手法では、「第 1 種の文化差」(一方の文化圏で発生したり、存在したりしている「もの」や「こと」と「第 2 種の文化差」(どちらの文化圏にも存在するが、それぞれの文化圏で意味の異なるもの) という 2 つの文化差を定義していた [4]。

次に、本稿で扱う文化差について考えるために、誤解の発生原因を考える。「第 1 種の文化差」と「第 2 種の文化差」では、誤解の原因が異なる。「第 1 種の文化差」である語句、例えば、日本生まれの「忍者」という語句であれば、これは、日本人と外国人がコミュニケーションする場合、どちらも日本における「忍者」を想像してコミュニケーションしようとする。つまり、外国人が、日本の「忍者」に対して、事実とは異なる知識を有していた場合、そのことによる誤解が発生する可能性がある。一方、「第 2 種の文化差」である語句、例えば、「饅頭」という語句は、日本と中国では大きな違いがある食べ物である。互

[†] 和歌山大学大学院システム工学研究科, Graduate School of Systems Engineering, Wakayama University

[‡] 京都大学学際融合教育研究推進センター, Center for the Promotion of Interdisciplinary Education and Research, Kyoto University

^{††} 和歌山大学システム工学部, Faculty of Systems Engineering, Wakayama University

いに文化差があると知らず「饅頭」を用いてコミュニケーションする場合は、日本人は日本の「饅頭」、中国人は中国の「饅頭」の知識でコミュニケーションしようとするので、誤解が発生する可能性がある。したがって、誤解の発生原因は、以下のようによまとめられる。

- 第1種の文化差による誤解
事実とは異なる誤った知識により発生する
- 第2種の文化差による誤解
互いの国にとっては正しい知識であるが、両国では異なることにより発生する

このように、「第1種の文化差」と「第2種の文化差」では、ともにコミュニケーションにおいて誤解が発生する場合があるが、誤解の原因が異なり、解決すべき問題と方法も大きく異なるはずである。これまでに検討してきた文化差検出手法は、Wikipediaの記事内容をもとに考えられている。つまり、事実に基づく、正しい知識による文化差検出手法を検討してきた。本稿でも、Wikipedia記事を用いた文化差検出手法を、「第2種の文化差」に絞った手法を考えていく。

3.2 コミュニケーションにおける重要度

本節では、本稿で扱う重要度について説明する。文化差のある語句を使用することで、互いにすれ違いや誤解が起きる可能性がある。すぐに誤解に気づくことができれば、その場で誤解を解消し、話を進めることができるが、誤解に気づかないまま話が進んでしまうと、誤解発生時点からの対話が無駄になる可能性がある。このような、その場で気づくことが難しい文化差のある語句は、システムによって注意を促すことが重要な語句（コミュニケーションにおける重要度の高い語句）であると考えられる。そこで、「対話において誤解が起きているにもかかわらず、そのままになってしまう可能性の高さ」を「コミュニケーションにおける重要度」として考える。

例えば、「饅頭はおいしい」という発言をした場合、互いに「饅頭」が具体的にどういったものであるかの同意がないため、何らかの誤解が起きる可能性が高く、コミュニケーションにおける重要度が高いと考える。なぜなら、この場合、「饅頭」が食べ物であるという共通の認識でコミュニケーションが成り立っているが、本当は、日本人と中国人の想像する「饅頭」は異なっており、それに気づかないままにコミュニケーションが進んでしまう可能性が高いからである。

一方「饅頭は蒸したパンである」という内容であれば、これは、「饅頭」がどういったものであるかを表しており、明確に文化差を表している内容である、この場合は、互いの認識のズレを認知しやすく、「饅頭」が日本と中国では何か異なるものなのではないかということに気づく可能性が高い。そして、これをきっかけに、自らの力で誤解が解消できる可能性が高いため、コミュニケーションにおける重要度が低いと考える。

4. 文化差検出手法と重要度の計算方法

4.1 文化差検出手法

日本語版 Wikipedia 記事から、いくつかの抽出条件に従って、文化差を表す文章を抽出していく。最終的に、1つでも文章が抽出できたかによって、「検索語句」に文化差が存在していたかを判断する。そのため、語句は、Wikipedia に記事として必ず存在している必要がある。図1に具体的な文化差検出手法の流れを示す。

- 条件 A
条件 A が当てはまっていた場合、それは中国における

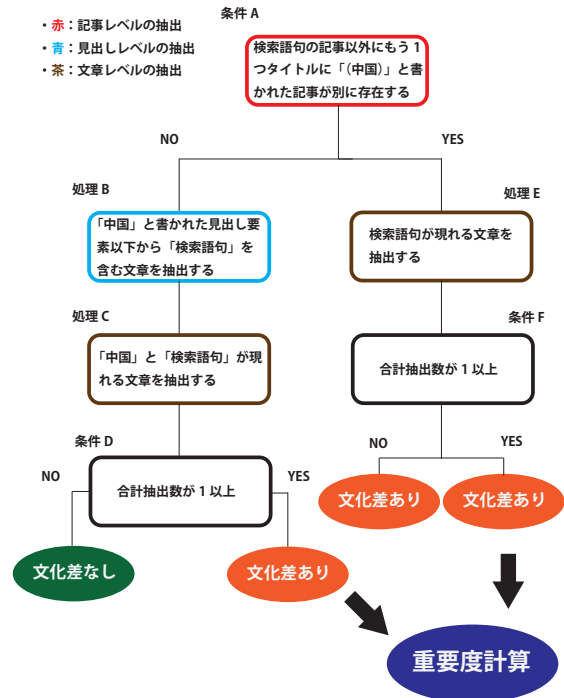


図1: 文化差検出・重要度計算の流れ

「検索語句」について説明している記事が存在することがわかる。

- 処理 B
図2は「ちまき」の記事の一部である。このように、見出し（heading）要素に「中国」と書かれていた場合、見出し要素以下に書かれている「検索語句」についての説明は、中国に関することであることがわかる。
- 処理 C
見出し要素以外にも、記事全体を見て、1文内に「中国」と「検索語句」が書かれていれば、中国に関することであることがわかる。
- 条件 D
合計抽出数が1以上の場合「文化差あり」とし、重要度の計算を行う。そうでない場合「文化差なし」となる。
- 処理 E
記事全体が、中国に関することであることがわかっていいため、「検索語句」について書かれている文章を抽出する。
- 条件 F
合計抽出数の数によらず「文化差あり」となるが、1以上の場合のみ、重要度の計算を行う。

今回の手法では、それぞれの判断で、最終的に1つでも文化差を表す文章を抽出できていれば、文化差が存在していると判断する。しかし、文化差を表す文章が抽出できていなくても、条件 A が当てはまっていたら、その記事全体は、中国用に使われているものであり、日本のものと違うことは明らかであるため、文化差があると判断する。そして、この手法によって獲得した、文化差のある語句と、その語句についての具体的な文化差を表す文章を用いて、重要度の計算を行う。

4.2 重要度の計算方法

重要度の計算は、文化差検出手法において、文化差があると判断された語句に対してのみ行う。文化差検出手法において抽

中国のちまき [編集]

中国において、ちまきは、水分を吸わせたもち米を直接葉の葉で包み、茹でる、もしくは蒸す方法で加熱して、作る方法が主流である。米と一緒に、味付けした肉、塩漬け卵、蕪（なつめ）、栗などの具や、小豆餡などを加えることが多い。特別なものでは、アワビやチャーシューを包んだものもある。形は正四面体が多いが、直方体、円筒形のものもある。中国北部では甘いちまき、南部では塩辛い味のちまきが好まれるが、そうした違いは南北との交流が盛んになった現在では少なくなっている。

図 2: 見出し要素の例 (ちまき)

出した文章と Google ngram[11] を用いて、コミュニケーションにおける重要度順にランキングを作成する。以下に、具体的な重要度の計算手順を示す。

- (1) 文化差検出手法において抽出された、文化差を表す文章を形態素解析する。
- (2) 形態素解析された結果から「名詞」「動詞」「形容詞」「形容動詞」の見出し語を抽出し、一意にする。
- (3) それぞれの「抽出した語句」と「検索語句」でトークンの出現順序を考慮せずに Google ngram の 1 から 7gram、全てのデータで AND 検索し、上位 10 件を取得し、頻度を合計する。
- (4) それぞれの「抽出した語句」について、頻度合計の多い順で「抽出した語句」のランキングを作成する。

検索語句と共に頻度が低い語句であれば、日常的に用いられていない表現や語句であり、コミュニケーションに違和感を感じる可能性が高く、コミュニケーションにおける重要度が低いと考えた。また「名詞」「動詞」「形容詞」「形容動詞」を抽出することとしているのは、用言と体言以外は、語句の文化差には関係がないと判断したためである。このランキングから、文化差のある語句を使用する際に注意すべき語句がわかる可能性がある。

5. 実験

5.1 文化差検出精度

今回、まず重要度の計算方法を適用させる語句を適切に検出できるのか、日本と中国間で調査を行った。用いた語句は、従来の文化差検出手法 [4] の検討において用いられた 114 個の語句である。これらの語句に対し、今回の文化差検出手法を適用することで、データセットの「第 2 種の文化差」である語句を適切に「文化差あり」と判定できるかを検証した。

今回、提案した手法の評価結果と従来手法との比較を表 1 に示す。データセットの 114 個の語句のうち「第 2 種の文化差」である語句は 41 語存在している。提案手法は、適合率 0.745、再現率 0.610、F 値 0.671 という結果となり、従来手法における「第 2 種の文化差」の実験結果よりも少し劣る結果となった。しかし、従来手法は「文化差なし」「第 1 種の文化差」「第 2 種の文化差」すべてを 1 つのフローチャートで判断しようとするものであり、今回の手法は、従来手法よりも比較的簡単かつ「第 2 種の文化差」に絞った手法である。このことから、結果としては少し劣ることにはなったものの、今回の手法は、これまでに検討してきた手法とは異なり、重要度を計算するために、具体的にどういった文化差が存在しているのかを表す文章を獲得しつつ、従来手法と同程度の精度が得られたため、メリットは大きいと考えられる。

表 1: 従来手法と提案手法の比較

	適合率	再現率	F 値
従来手法	0.788	0.634	0.702
提案手法	0.745	0.610	0.671

5.2 重要度の計算

適切な重要度の計算を行えるのか実験を行った。用いた語句は、5.1 節において、適切に「文化差あり」と判断された語句である。

表 2 において、文化差検出手法によって抽出された文章を示す。表 3 に、実験結果のなかでも良い傾向を示した「煎餅」の実験結果を示す。表 2 の「煎餅」の文化差として抽出された文章である「中国の煎餅は、いずれも柔らかくて、巻いたり、曲げたりして食べることが可能なものであり、形状はクレープやドーサに似ている」という文章から、中国の「煎餅」は、日本の「煎餅」とは大きく違うということが分かる。そこで表 2 を見ると、「柔らかい」「巻く」「曲げる」「食べる」「形状」「クレープ」「似る」が存在している。頻度の低い順から、「巻く」「曲げる」「形状」の頻度合計が 0 であり、日本人から考えると、「煎餅」を巻いたり、曲げたりすることは考えられないので、中国人の考える「煎餅」に違和感を覚えるはずである。違和感のある語句が用いられた場合、コミュニケーションの途中で、互いに想像している「煎餅」の認識がずれていることがわかるので、誤解に気づく可能性が高く、自らで誤解を解消しようとするはずであり、重要度は低い。また「クレープ」も 20 と低い値であり、日本人にとっても「煎餅」と「クレープ」が同時に語られること違和感を覚えるはずであり、妥当な結果であると言える。

他に「豆腐」の結果の一部を表 4 に示す。表 2 より、抽出された文化差として、「現在の日本の豆腐は柔らかくて淡泊な食感を特徴とする独特の物で、これに対する中国や韓国の豆腐は炒めたり揚げたりして調理されることが多かったため、日本の豆腐に比べると水分が少なく堅いものとなっている」が存在しており、中国との調理法の違いによる「豆腐」の違いが述べられており、表 4 から「揚げる」「炒める」「堅い」という語句の頻度が少ないことがわかる。

また、抽出された文章に、適切でないパターンがいくつか存在した。例えば、語句の起源や歴史の説明において、「中国」という単語が出現することがあり、これが文化差として、誤って抽出されてしまっている場合が多かった。これは「豆腐」のほかにも「干支」などで見られた。こういった文章は、出典が記されている場合が多く、抽出しないようにする必要がある。他にも全体的に「検索語句」で抽出したものの「検索語句」そのものではなく、「検索語句」周辺に関する文化差が抽出されることが多かった。こういった文章は、文化差を表す文章として抽出すべきではない。

また、今回の手法では、抽出された文化差を表す文章を形態素解析したが、そうすることで、文章としての意味が失われてしまっている。例えば「煎餅」において抽出された要素に、「似る」があったが、「似る」だけでは情報量が少なく、何に似ているのかということが重要なはずである。つまり「煎餅は、クレープに似ている」で検索するほうが、良い結果となった可能性がある。しかし、実験結果から、Google ngram で検索することは、重要度の計算において有用であることがわかったため、今後、語句ではなく、文章レベルで意味を失わずに検索する手法を検討する必要がある。

表 2: 抽出された文章

語句	抽出された文章
煎餅	「そのため、煎餅は餅米を搗いて作る餅を基本的に使っていないが、「餅」という字が入っている」「つまり、中国の「煎餅」は小麦粉などの粉を水で練って、鉄板で焼いたものである」「中国においては、現在も山東煎餅、天津煎餅など、一銭洋食に似た作り方の軽食が作られている」「日本のお好み焼きもこれらの「煎餅」の一種として「日式雑菜煎餅」などと呼ばれることがある」「逆に、日本の揚げ煎餅の類を「仙貝 シエンペイ」と音訳で呼ぶ場合がある」「中国の煎餅は、いずれも軟らかくて、巻いたり、曲げたりして食べることが可能なものであり、形状はクレープやドーサに似ている」「煎餅 山東煎餅 天津煎餅 甘味煎餅に似たもの 鶏蛋仔。」
豆腐	「豆腐は中国から伝来した可能性の高い食品であり、中国でもこの名前で呼ばれているのに対して、納豆は日本独自の言葉であり、取り違えられることはあり得ない」「中国では伝統的には豆腐は生で食べるのではなく発酵豆腐などとして食べていたとされる」「また、中国の伝統的な豆腐は日本の豆腐よりも堅いとされるが、これは油による調理法が多かったため水分が少ないほうが都合がよかったためとされる [1]」

文章の形式は、抽出元のままである。また「豆腐」は、一部の抽出結果を載せている。

表 3: 「煎餅」の共起語ランキング

頻度順	語句	頻度合計	頻度順	語句	頻度合計
1	ため	36130	27	山東	337
2	揚げ	8825	28	可能	334
3	類	5933	29	一種	330
4	ある	4121	30	軟らかい	242
5	いる	3855	31	式	242
6	食べる	3843	32	作り方	236
7	もの	3823	33	呼ぶ	235
8	お好み焼き	3265	34	一銭	220
9	日	2740	35	これら	220
10	米	2635	36	いずれ	214
11	する	2503	37	洋食	132
12	仙	2236	38	クレープ	20
13	れる	2014	39	逆	0
14	日本	1208	40	字	0
15	焼く	1193	41	使う	0
16	小麦粉	888	42	貝	0
17	餅	866	43	入る	0
18	こと	778	44	水	0
19	中国	747	45	巻く	0
20	作る	710	46	曲げる	0
21	現在	638	47	練る	0
22	的	572	48	搗く	0
23	甘味	554	49	鉄板	0
24	粉	542	50	形状	0
25	基本	459	51	似る	0
26	天津	408	52	軽食	0

表 4: 「豆腐」の共起語ランキングの一部

頻度順	語句	頻度合計
1	の	1732100
2	よう	93520
3	納豆	79580
73	揚げる	1215
75	炒める	1084
79	堅い	907

- [2] 藤井薫和, 吉野 孝: 異文化間コミュニケーション支援のためのアノテーション自動獲得システムの開発, 情報処理学会研究報告, グループウェアとネットワークサービス研究会, 2008-GN-66, pp.141-146 (2008) .
- [3] 岡本健吾, 吉野 孝: 会話中の名詞の関連情報を用いた対面型異文化間コミュニケーション支援システムの構築と評価, 情報処理学会論文誌, Vol.52, No.3, pp.1213-1223 (2011) .
- [4] 諏訪智大, 宮部真衣, 吉野 孝: 日本語版・中国語版 Wikipedia を用いた文化差検出手法の提案, 情報処理学会論文誌, Vol.55, No.1, pp.257-266 (2014) .
- [5] Cho Heeryon, 石田 享, 山下直美, 稲葉利江子, 高崎俊之, 神田智子: 絵文字解釈における人間の文化差判定, ヒューマンインタフェース学会論文誌, Vol.10, No.4, pp.427-434 (2008) .
- [6] Tomoko Koda and Toru Ishida : Cross-cultural study of avatar expression interpretations , SAINT 2006 , pp.130-136 (2006) .
- [7] Hofstede G : Cultures and Organizations : Software of the Mind. London : McGraw-Hill (1991) .
- [8] Ulrike Pfeil , Panayiotis Zaphiris , Chee Siang Ang : Cultural Differences in Collaborative Authoring of Wikipedia , Journal of Computer-Mediated Communication , 12 , pp.88-113 (2006) .
- [9] Hiroyuki Nakasaki , Mariko Kawaba , Takehito Utsuro , and Tomohiro Fukuhara : Mining Cross-Lingual/Cross-Cultural Differences in Concerns and Opinions in Blog , Computer Processing of Oriental Languages: Language Technology for the Knowledge-Based Economy: 22nd International Conference, ICCPOL 2009, Lecture Notes in Computer Science/Lecture Notes in Artificial Intelligence , Vol.5459 , Springer , pp.213-224 , March 2009 .
- [10] 西田ひろ子: 異文化間コミュニケーション, 創元社 (2000) .
- [11] 工藤拓, 賀沢秀人著, 「Web 日本語 N グラム第 1 版」, 言語資源協会発行 .

6. おわりに

本稿では、コミュニケーションにおける重要度を考慮した文化差検出手法を提案した。文化差検出手法の実験では、提案手法は、従来手法にわずかに劣ったものの、今回の手法は、重要度の計算のために、具体的にどういった文化差が存在しているのかというデータを獲得しつつ、従来手法と同程度の精度が得られたため、メリットは大きいと考えられる。

重要度計算の実験では、検索語句に対して、明確な文化差を示している語句においては、良い結果が得られた。しかし、語句レベルに分解してしまっていることで、文章の意味が失われてしまい、適切な検索が行えていないことがわかった。

今後は、文化差検出精度を向上させるとともに、文章レベルで検索する手法を考えていく必要がある。

参考文献

- [1] 藤井薫和, 重信智宏, 吉野 孝: 機械翻訳を用いた異文化空間チャットコミュニケーションにおけるアノテーションの評価, 情報処理学会論文誌, Vol.48, No.1, pp.63-71 (2007) .