

科学研究費助成事業 研究成果報告書

令和 5 年 6 月 12 日現在

機関番号：34315

研究種目：基盤研究(C) (一般)

研究期間：2020～2022

課題番号：20K02642

研究課題名(和文) ネットいじめの防止を目的とした子どもの情報モラルの獲得を支援するシステム

研究課題名(英文) A system to support information morality acquisition by children to prevent cyber bullying

研究代表者

西原 陽子 (Nishihara, Yoko)

立命館大学・情報理工学部・教授

研究者番号：70512101

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本研究では、ネットいじめの防止を目的として、子どもの情報モラルの獲得を支援する情報システムに関する研究を行った。ネットいじめの防止では、情報科学分野において、情報フィルタリング技術の研究が進められている。フィルタリングをかけることにより、ネットいじめに関わる情報の投稿を防ぐことができる。しかし、フィルタリングの設定はユーザやサービス提供者に委ねられており、いつでもフィルタリングを外すことが可能である。本研究では、フィルタリング技術がなくとも、ネットいじめに関わる情報の投稿をしないで済むように、子どもに投稿しようとする情報の是非を判断することを支援する情報システムを提案した。

研究成果の学術的意義や社会的意義

本研究の意義は、子どもが情報発信において自ら是非を判断できるようになることを支援する情報システムを提案した点にある。情報技術の多くは人の労力を助けるものであるが(例えば、スマートフォン)、情報技術がない状況ではできないことが増えてしまえば、真に役立つ情報技術といえない可能性がある。本研究で提案した情報システムは、人の能力を鍛えるために使うものである。一度、投稿の是非を判断できるようになれば、この情報システムがない状況であっても、ユーザは適切に投稿の是非を判断し、投稿ができるようになる。

研究成果の概要(英文)：In this study, we conducted research on information systems that support children's acquisition of information morality with the aim of preventing cyber bullying. In the field of information science, information filtering and information recommendation technologies are studied to prevent cyber bullying. Those technologies can prevent the posting related to cyber bullying. However, the usage of the technologies can be decided by users and service providers. The filtering technologies can be removed at any time. In this study, we propose an information system that supports children in judging the pros and cons of information they are about to post, so that they can avoid posting messages related to cyber bullying even without the filtering technologies.

研究分野：ヒューマンコンピュータインタラクション

キーワード：ネットいじめの防止 投稿の可否の判断

1. 研究開始当初の背景

子どもたちがデジタルデバイスを用いてインターネットに頻繁にアクセスするようになった。それに伴い、現実世界でのいじめがネットの世界にも持ち越され、「ネットいじめ」が起こるようになった。ネットいじめとは、インターネット上におけるいじめ、および嫌がらせを指す。ネットいじめに対する有効な対策の検討が早急に求められている。

ネットいじめは、被害者にとって有害な情報の投稿により行われることが多く、情報の有害・無害を評価するフィルタリング技術が研究されてきた。技術の精度は高くなってきたが、フィルタリングの技術だけでネットいじめを防止することは難しい。フィルタリング技術の網の目をかいくぐり、有害情報を投稿することは依然として可能であり、異なる観点からの防止方法が必要となっている。

ネットいじめを真に防止するためには、子どもがネットいじめに関わる投稿を自ら抑止することが不可欠と考えられる。抑止するためには、子どもが情報に対するモラルを持ち、投稿の可否を判断する力を獲得する必要がある。投稿可否判断力の獲得を支援するために、子どもの投稿に対する新しいインタラクションが求められている。

2. 研究の目的

本研究の目的は、子どもに投稿可否判断力を獲得させるために、子どもの投稿に対し評価と指導を行う新しいインタラクションシステムを提案し、その有効性を確認することである。

3. 研究の方法

はじめに、投稿の有害と無害を自動的に判定するコンピュータプログラムを実装する。ネットいじめに関わる有害な投稿の形式は大きく2つに分けられ、直接表現と間接表現がある。直接表現、間接表現のいずれもを判定できるコンピュータプログラムを実装する。プログラムの実装にあたっては、研究者らが以前の研究で作成した辞書式の判定手法[1]や深層学習を用いた判定手法[2]を用いることができる。また、大規模な言語モデルを学習して作成されたAPI (Perspective API[3]など) も利用できる。

続いて、投稿の指導に効果的なメッセージ文を調査する。投稿の可否を判断させ、その判断結果に対する指導をメッセージ文で行う。判断が正しい場合と誤っている場合と、それぞれに対して指導を行うメッセージ文を考案する。判断が誤っている場合のメッセージ文としては、研究者らが提案するセルフトーク型のメッセージ文[4]や、コーチングで使用されている褒めと叱りを利用する事ができる。

最後に、子どもたちに投稿の可否を判断させ、判断結果に応じて指導メッセージを提示し、投稿可否判断に与える影響を評価する実験を行う。実験は、大規模なWebアンケート調査により行うことができる。

4. 研究成果

子どもたちに投稿の可否を判断させ、判断結果に応じて指導メッセージを提示するシステムを構築した。システムは、架空の友人同士の会話と、その後が続いて投稿されようとしているメッセージを提示する。子どもたちはシステムのユーザとして、投稿の可否を判断する。判断結果に応じて指導メッセージが提示され、判断が正しければ褒めるメッセージ、判断が誤っていれば叱るメッセージを提示する。投稿されようとしているメッセージは、単なる悪口と事実に基づく批判の2種類を用意し、悪口は投稿してはならないが、批判は投稿しても良いとした。悪口の投稿文を作成するときは、Web上のフィルタリングシステムで採用されている悪口の辞書を参考にした。

システムを評価する実験を行った。実験の手順は以下の通りであった。

- (1) 被験者は共感度測定アンケートに回答する。
- (2) 被験者は、架空の友人同士の会話とそれに続き投稿されようとしている投稿文を読む。
- (3) 被験者は、投稿文が悪口か批判かを判断する。
- (4) システムは、被験者の回答の正誤に応じて指導メッセージを提示する。
- (5) 被験者は、(2)から(4)を複数回繰り返す。

- (6) 被験者は、指導メッセージの受け入れやすさについて回答する。
 (7) 数日後に被験者は、2回目の実験を実施する。2回目は(2)と(3)を複数回繰り返す。

共感度測定アンケートは、日本語版 IRI [5]を参考に用意した。日本語版 IRI は複数の項目から構成されているが、視点取得に関する7つの項目をアンケートに用いた。7つの回答はそれぞれ5点満点で回答され、合計値を被験者の共感度の数値とした。

被験者に提示された会話は1回目が18種類、2回目が18種類で合計36種類あった。1回目と2回目のそれぞれで9種類の悪口と9種類の批判の投稿文が用意された。会話が提示される順番はランダムになるように設定された。

被験者に提示された指導メッセージのうち、褒めは「この調子で判断していきましょう。これは批判（または悪口）です。」であった。叱りは「投稿の可否をよく考えましょう。これは批判（または悪口）です。」であった。

被験者は3つのグループに分けられ、各グループで指導メッセージの提示の仕方を変えた。グループ1は褒めるのみ、グループ2は叱るのみ、グループ3は併用であった。被験者は15歳から18歳の男女で、有効回答数は267名分であった。267名のうち男性が125名、女性が142名であった。

実験の結果として得られた各グループの判断の正答率のデータを表1に示す。共感度ごとの提示メッセージの受け入れやすさの評価結果を表2に示す。

表1の結果から2回目の正答率が最も高くなるのがグループ3であることが示された。グループ3は指導メッセージを併用するグループである。正しい回答、誤った回答のいずれに対しても指導が行われることで、投稿可否判断の基準をより明確に持たため、2回目の正答率が最も高くなったと考えられる。

表2の結果から最も受け入れやすい指導メッセージは褒めであり、次が褒めと叱りの併用であることが示された。叱りのメッセージは最も受け入れがたいことが示された。正答率を比較すると褒めよりも叱りだけの方が高いが、提示メッセージは叱りの方が受け入れがたい。したがって、叱りだけを指導として行うインタラクションでは、投稿可否判断の力を身につける効果が低いと予想される。一方で、褒めるだけを指導として行っているのは、投稿可否判断の力がこの場合も身につけづらい。したがって、褒めと叱りを併用することが、投稿可否判断の獲得に最も効果があることが確認された。

表1. 実験で得られた各グループの判断の正答率

	1回目	2回目
グループ1	88.6%	84.7%
グループ2	80.5%	85.5%
グループ3	84.1%	89.6%

表2. 提示メッセージの受け入れやすさの評価結果

	受け入れがたい	どちらでもない	受け入れやすい
グループ1	20.0%	23.5%	56.5%
グループ2	57.0%	24.7%	18.3%
グループ3	31.5%	27.0%	41.6%

引用文献

- [1] Yoko Nishihara, Kazuki Iwasa, Junichi Fukumoto and Ryosuke Yamanishi, Harmful Comments Extraction from a Bulletin Board System using Word Meaning and Impression on Thread Context, SCIS & ISIS 2014, pp.1398-1402 (2014).

- [2] Ryuichi Omi, Yoko Nishihara, and Ryosuke Yamanishi, Extraction of Paraphrases using Time Series Deep Learning Method, International MultiConference of Engineers and Computer Scientists 2019, pp. 276-278, (2019).
- [3] <https://perspectiveapi.com/>
- [4] 西原陽子, 藤堂悠杜, 山西良典, セルフトークメッセージの呈示による悪口投稿取り下げの効果分析, 電子情報通信学会論文誌 A, Vol. J104-A, No. 2, pp. 19-26, (2021)
- [5] Himichi, T., Osanai, H., Goto, T., Fujita, H., Kawamura, Y., Davis, M. H. Nomura, M. (2017). Development of a Japanese version of the Interpersonal Reactivity Index. The Japanese journal of psychology, 88, 61-71.

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Yoko Nishihara, Seiya Tsuji, Wataru Sunayama, Ryosuke Yamanishi, and Shiho Imashiro	4. 巻 14
2. 論文標題 A Generation Method for the Discussion Process Model during Research Progress Using Transitions of Dialog Acts	5. 発行年 2021年
3. 雑誌名 International Journal On Advances in Systems and Measurements	6. 最初と最後の頁 17-26
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 西原陽子, 藤堂悠杜, 山西良典	4. 巻 104
2. 論文標題 セルフトークメッセージの呈示による悪口投稿取り下げの効果分析	5. 発行年 2021年
3. 雑誌名 電子情報通信学会論文誌A	6. 最初と最後の頁 19-26
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計9件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 伊藤誠基・砂山渡
2. 発表標題 オンラインチャットにおける誹謗中傷コメントの発信を未然に防止する機能
3. 学会等名 第35回人工知能学会全国大会
4. 発表年 2021年

1. 発表者名 工藤裕登・砂山渡
2. 発表標題 ツイート中の感情表現単語の使用割合に基づく性格推定支援
3. 学会等名 第35回人工知能学会全国大会
4. 発表年 2021年

1. 発表者名 山西良典, 竹元亨舟, 西原陽子, 吉田光男
2. 発表標題 複数コーパスへの尤度分布による意見文の性質表現
3. 学会等名 NII IDRユーザフォーラム 2021
4. 発表年 2021年

1. 発表者名 竹元亨舟, 山西良典, 西原陽子, 吉田光男, 大須賀智子, 大山敬三
2. 発表標題 既存データセットへの尤度分布による意見文の性質表現 -擬似コーパスとしての既存データセット集合の利活用の可能性-
3. 学会等名 第14回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2022年

1. 発表者名 Yoko Nishihara, Xinran Lin, and Ryosuke Yamanishi
2. 発表標題 Do the Number of Creators and Their Conversations Affect Re-Evaluation of a Familiar Place in Making a Tourist Map?,
3. 学会等名 The Fourteenth International Conference on Advances in Computer-Human Interactions
4. 発表年 2021年

1. 発表者名 山中大輝, 砂山渡
2. 発表標題 リフレーミングを用いた文章表現の置換による対人印象改善
3. 学会等名 第26回インタラクティブ情報アクセスと可視化マイニング研究会
4. 発表年 2021年

1. 発表者名 Yihong Han, Yoko Nishihara, and Junjie Shan
2. 発表標題 Human Values Estimation on News Articles through BERT-extracted Opinion Expressions
3. 学会等名 The 27th International Conference on Technologies and Applications of Artificial Intelligence (国際学会)
4. 発表年 2022年

1. 発表者名 Ryosuke Yamanishi, Hoshu Takemoto, Yoko Nishihara, Mitsuo Yoshida, Tomoko Ohsuga, and Keizo Oyama
2. 発表標題 Applying Existing Datasets as a Pseudo Corpus for Sentiment Representation on Social Media
3. 学会等名 26th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (国際学会)
4. 発表年 2022年

1. 発表者名 Junjie Shan, Yoko Nishihara, and Yihong Han
2. 発表標題 Identifying Reply-to Relation in Textual Group Chat using Unlabeled Dialogue Scripts and Next Sentence Prediction
3. 学会等名 The 27th International Conference on Technologies and Applications of Artificial Intelligence (国際学会)
4. 発表年 2022年

〔図書〕 計0件

〔出願〕 計1件

産業財産権の名称 推定方法、及び、情報処理装置	発明者 西原陽子, 岡田志 麻, 山浦一保, 単駿 傑, 韓毅弘	権利者 同左
産業財産権の種類、番号 特許、2022-158881	出願年 2022年	国内・外国の別 国内

〔取得〕 計0件

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	砂山 渡 (Sunayama Wataru) (40314398)	滋賀県立大学・工学部・教授 (24201)	
研究分担者	山西 良典 (Yamanishi Ryosuke) (50700522)	関西大学・総合情報学部・准教授 (34416)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関