

修士論文 2022 年度 (令和 4 年度)

個別文脈依存蓄積差分確認方式による利用規約読解時間の短縮

慶應義塾大学大学院 政策・メディア研究科
渡邊 聡紀

個別文脈依存蓄積差分確認方式による利用規約読解時間の短縮

インターネット利用者は年々増加しており、それに伴いインターネット上で提供されるサービスも増加している。インターネット上で提供されるサービスを利用するためにはほとんどの場合利用開始前に利用規約への同意をする必要がある。しかし、利用規約はあまり読まれていない問題があり、そのためにトラブルが発生する可能性がある。本研究では、利用規約を読むために必要する時間を減らし、かつ、問題のある条項を発見するための手法として、同意した利用規約を記録していき、新たに利用規約を読む時に読んだことのある条文と同じ意味の文を抽出してそれ以外を注目するようにする手法を提案する。これにより、個人個人に合わせた形での利用規約の読解支援を提供でき、より利用しやすいインターネット環境を作り出すことができる。

キーワード:

1. 利用規約, 2. 法的文書, 3. 自然言語処理, 4. 読解支援

慶應義塾大学大学院 政策・メディア研究科
渡邊 聡紀

Reduction of reading time of Terms of Service by individual context-dependent accumulated difference confirmation method
--

The number of Internet users is increasing every year, and the number of services offered on the Internet is also increasing accordingly. In most cases, it is necessary to agree to the Terms of Service before using a service offered on the Internet. However, there is a problem that the Terms of Service are not read very often, which can cause problems. In this study, as a method to reduce the time required to read the Terms of Service and to discover problematic clauses, we propose a method to record the Terms of Service that have been agreed to, so that when reading new Terms of Service, we can extract clauses that have the same meaning as those we have read before and focus on the rest of them. This can provide assistance in reading the Terms of Service in a form that is tailored to each individual and create a more user-friendly Internet environment.

Keywords :

1. Terms of Service, 2. Legal Document, 3. NLP, 4. Reading Support

Keio University Graduate School of Media and Governance
Watanabe Satoki

目次

第1章	序論	1
1.1	はじめに	1
1.2	本研究での定義	1
第2章	背景	2
2.1	利用規約の法的根拠	2
2.1.1	約款	2
2.1.2	利用規約と民法	2
2.1.3	定型約款	3
2.1.4	個人情報保護法	4
2.1.5	消費者契約法	5
2.2	前提技術	6
2.2.1	自然言語処理	6
2.2.2	Transformer	6
2.2.3	BERT	7
2.2.4	Sentence BERT	8
第3章	問題	9
3.1	利用規約の認知	9
3.2	利用規約の読解時間	10
3.3	利用規約の読解	11
3.4	問題の定義	13
第4章	提案手法	14
4.1	概要	14
4.2	利用規約類似文検出	14
4.3	利用規約警告検出	15
第5章	実験	16
5.1	概要	16
5.2	前処理	16
5.3	文ベクトル生成	17
5.4	コサイン類似度	17
5.5	成形	17

5.6	過去に同意した利用規約データベース	18
5.7	利用規約警告検出	19
5.8	表示方法	19
第 6 章	評価	24
6.1	評価内容	24
6.2	評価のための利用規約	24
6.3	評価のためのクイズ	25
第 7 章	関連研究	26
7.1	機械学習を用いた利用規約からの未知条項の抽出に関する研究 (中村ら 2018)	26
7.2	CLAUDETTE: an automated detector of potentially unfair clauses in on-line terms of service(Lippi ら 2019)	26
7.3	利用規約中の不公平文の自動検出 (青山ら 2019)	26
第 8 章	結論	27
8.1	本研究のまとめ	27
8.2	本研究の課題	27
8.2.1	構想	27
付 録 A	実験に利用した利用規約	28
A.1	Alpha 社	28
	謝辞	29

目 次

2.1	Transformer における self-attention の仕組み、文献 [1] より引用	7
2.2	BERT の事前学習とファインチューニングされたタスクごとのモデル、文献 [2] より引用	8
3.1	検索サービスの利用規約をどの程度読んでいるか（回答数：448）、文献 [3] より引用	9
3.2	SNS 等の利用規約をどの程度読んでいるか（回答数：929）、文献 [3] より引用	9
3.3	検索サービスの利用規約の認知（回答数：2,000）、文献 [3] より引用	10
3.4	SNS 等の利用規約の認知（回答数：2,000）、文献 [3] より引用	10
3.5	「利用規約の内容は重要だと思うか」、文献 [4] より引用	11
3.6	「利用規約を読まないことでリスクはあると思うか」、文献 [4] より引用 . .	11
3.7	「サービスを利用する前に利用規約を読むか」、文献 [4] より引用	12
3.8	「利用規約を読まない理由」、文献 [4] より引用	12
4.1	類似条文検出のイメージ	14
4.2	利用規約警告検出	15
5.1	類似条文検出の実装イメージ	16
5.2	Alpha 社の利用規約/どの利用規約も既読ではない	20
5.3	Alpha 社の利用規約/Beta 社が既読	21
5.4	Alpha 社の利用規約/Beta 社、Gamma 社が既読	22
5.5	Alpha 社の利用規約/Beta 社、Gamma 社、Delta 社が既読	23

表 目 次

5.1	過去に同意した利用規約データベースの内容	18
6.1	実験用利用規約の一覧	24

第1章 序論

本章では本研究の背景、課題及び手法を提示し、本研究の概要を示す。

1.1 はじめに

インターネットの利用者は年々増加しており、それに伴い、インターネット上で提供されるサービスも増えている。利用できるサービスが増加すると便利な機能や新しい体験をすることができるが、新しくサービスを利用したり、アカウントを作成するためにはほとんどの場合、利用規約を読み、同意することが求められる。しかし、サービスを利用するのを優先するためにその前に表示される利用規約は読まれないような状況に陥っている。将来的に同意したサービスを利用した時に問題が発生した場合、ほとんど読んでいない利用規約をもとに裁判を行う必要が生じてしまうため、このときに自分が納得することができない条項などが含まれていても手遅れになってしまい、当然に同意した事実の方が優越してしまう。これにより、利用規約を読むことは重要といえるが、それを利用者に求めた場合、新しいサービスを利用するたびに利用規約を読む時間や労力を求めることになってしまう。よって、本研究では、より安心してインターネット上のサービスを利用するために、利用規約の読解支援を行う。

1.2 本研究での定義

本研究における「利用規約」とは、インターネット上で特定のサービスを利用する際に同意を求められる事業者がサービスの利用に関する規則を記載した定型約款である。なお、定型約款については2章で詳しく述べる。本研究では他国の利用規約まで対象にした場合は、他国で制定された法律についても取り扱う必要がありまた、利用規約の法的立ち位置は国によって異なるため取り扱うことが困難である。よって、本研究は日本の法律の下で提供されるサービスの日本語での利用規約を対象とする。

第2章 背景

本章では本研究の背景について述べる。

2.1 利用規約の法的根拠

本節では、利用規約について法的な立場の整理を行い、その重要性について論じる。

2.1.1 約款

約款とは、一般に、大量の同種の取引を迅速、効率的に行うなどのために作成された、定型的な内容の取引を行う場合に示す契約条件のことである。

2.1.2 利用規約と民法

「利用規約」という用語は法令用語ではなく、法的には約款の一種であると考えられる[5]。2020年4月1日以前の民法には利用規約に関する規定は設定されていなかった。当時は当事者間において約款により個別に契約が結ばれていると解釈をなされていたが、これらは事業者が一方的に作成したものであり、利用者が条項の内容を認識していないことが多い。本来、両者の認識のもと合意に基づいて契約がなされるという前提により法的拘束力があるとみなされるべきである。しかし、実態として条項の内容を認識していない利用者に個別の契約交渉をさせるということは困難であるが、一方で定型的な内容が想定される契約類型においては約款の法的拘束力を認めないと、円滑な取引を阻害させることになる。また、約款に含まれる条項が契約内容になることが争われた裁判では、それぞれのケースごとに判断が分かれるなど透明性にも課題があった[6]。

さらに、「この約款は当社の都合で変更することがあります。」のような条項は一般的に含まれていたが、この条項が有効であるかについても議論が分かれていた。このような条項がある取引は一般に長期にわたって継続するため、法令の変更や経済情勢、経営環境の変化に応じて約款の内容を事後的に変更をする必要がある。民法の原則によれば、契約内容を事後的に変更するには、個別に相手方の承諾を得る必要があるが、多数の顧客と個別に変更についての合意をすることは実務上困難である。このような条文は基本的には顧客の利益保護のために行われているが、合理的な場合に限る必要があり、条文が利益を損なうことがないようにする必要がある[7]。

これらの要請をもとに、2020年4月1日に民法の改正が行われた。これにより、利用規約のような不特定多数と契約を執り行うような約款を「定型約款」として定義された。定型約款については2.1.3節で詳しく述べる。

2.1.3 定型約款

前節で述べた改正民法により、定型約款に関する規定がなされている。規定されている部分を以下に示す。

（定型約款の合意）

第五百四十八条の二

定型取引（ある特定の者が不特定多数の者を相手方として行う取引であって、その内容の全部又は一部が画一的であることがその双方にとって合理的なものをいう。以下同じ。）を行うことの合意（次条において「定型取引合意」という。）をした者は、次に掲げる場合には、定型約款（定型取引において、契約の内容とすることを目的としてその特定の者により準備された条項の総体をいう。以下同じ。）の個別の条項についても合意をしたものとみなす。

一 定型約款を契約の内容とする旨の合意をしたとき。

二 定型約款を準備した者（以下「定型約款準備者」という。）があらかじめその定型約款を契約の内容とする旨を相手方に表示していたとき。

2 前項の規定にかかわらず、同項の条項のうち、相手方の権利を制限し、又は相手方の義務を加重する条項であって、その定型取引の態様及びその実情並びに取引上の社会通念に照らして第一条第二項に規定する基本原則に反して相手方の利益を一方的に害すると認められるものについては、合意をしなかったものとみなす。

民法改正時の議論では、従前に存在しなかった、約款全体についての定義について議論がなされていたが、最終的にまともらずに、約款全体についてを民法上で規定することは見送られた。それにより、定型取引以外で用いられる約款のみに関する規定が導入された。定型取引以外で用いられる約款に関する問題については、従前通り裁判所の判断に委ねられることとなった。[8] 条文上で定型約款の要件が述べられているが、非常に抽象的であり、改正民法下での判例や政令が増えない限りは具体的にどのようなものが定型約款に該当するかは不明瞭な状態が続くと見られている。しかし、国会審議などを通して、以下のような約款が当たると考えられている。[9]

- 旅客運送約款¹
- 電気供給約款²
- 保険約款³

¹「東日本旅客鉄道株式会社旅客営業規則」「国内旅客運送約款」（全日本空輸株式会社）など

²「特定小売供給約款」（東京電力エナジーパートナー株式会社）など

³「普通保険約款」（損保ジャパン株式会社）など

- 普通預金規定⁴
- インターネットサービスの利用規約

定型約款には当たらないものとして、事業者間取引の契約書ひな型や就業規則、労働契約書などが挙げられている。

定型約款はいわゆる「みなし合意」が認められる。顧客が定型約款にどのような条項が含まれているのか認識をしていなくても、

- 定型約款を契約の内容とする旨の合意をしたとき。
- 定型約款を準備した者があらかじめその定型約款を契約の内容とする旨を相手方に表示していたとき。

以上の2点のうちどちらかが満たされたとき、約款についての合意をしたとみなされる。⁵ただし、第548条の2第2項に示されているように、信義則に反するような利益を一方的に害する不当な条項はみなし合意が認められない。⁶

インターネットサービスで一般的に利用前に読む必要がある利用規約は以上のことから定型約款として法的に定められており、みなし合意が認められるため、信義則に反しない限りは利用規約の合意が基本的に認められてしまう。よって、利用規約に合意をする際は、その内容が自身にとって問題がないか慎重に読む必要があるといえる。

2.1.4 個人情報保護法

日本では、個人の権利、利益の保護と個人情報の有用性のバランスを図るために、個人情報の保護に関する法律(以下、個人情報保護法)が定められている。個人情報保護法では、事業者が個人情報の適正な取り扱いを行うための方針が定められている。

個人情報

個人情報保護法において、「個人情報」とは、生存する個人に関する情報で、氏名、生年月日、住所、顔写真などにより特定の個人を識別できる情報のことである。これらは、他の情報と容易に照合することができ、それにより、特定の個人を識別することができるようになるものも含まれている。たとえば、生年月日単体では個人を特定することは不可能であるが、これに、氏名などを組み合わせることで、特定の個人を識別できるため、個人情報に該当すると考えられている。また、メールアドレスもユーザー名やドメインにより個人を特定できる場合は、個人情報に該当する。⁷[10]

⁴「普通預金規定」(株式会社三井住友銀行) など

⁵民法第548条の2第1項

⁶民法第1条第2項 権利の行使及び義務の履行は、信義に従い誠実に行わなければならない。

⁷個人情報保護法第2条第1項

個人識別符号

文字、番号、記号その他符号などでその情報単体から特定の個人を識別できる情報のうち政令、規則で定められたものを「個人識別符号」といい、個人識別符号が含まれる情報は個人情報となる。個人識別符号は大きく 2 つに大別することができる。1 つ目は DNA を構成する塩基の配列 (いわゆるゲノムデータ)、本人を認証することを目的とした装置やソフトウェアにより本人を認証することを目的とする顔認証データ、指紋認証データ、虹彩などの身体の一部の特徴を電子処理のために変換した符号である。2 つ目はサービス利用や書類などで割り振られる符号で、パスポート番号やマイナンバー、保険証の番号などがこれに当たる。⁸

要配慮個人情報

「要配慮個人情報」とは、不当な差別や偏見その他の不利益が生じないようにその取扱いに特に配慮を要するものである。内容としては、人種、信条、社会的身分、病歴などが定められている。⁹

本人の同意

これらの「個人情報 (個人識別符号を含む)」、「要配慮個人情報」を取得する場合、また、個人データの第三者提供や個人関連情報の第三者提供に関しては、原則として本人の同意が求められることが、個人情報保護法において定められている。¹⁰ よって、これらの情報を取得するサービスは、利用規約もしくはプライバシーポリシーなどで同意を求める必要がある。

2.1.5 消費者契約法

民法では、契約自由の原則を採用しているが、この原則からすると両者の間で契約内容を自由に決定することができる。しかし、消費者と事業者の間の情報の質及び量及び交渉力の格差があることから、この部分に限定して、民法の特別法として、消費者契約法が定められている。この法の下で、利用規約において契約自由の原則が修正されている。[11]

消費者保護法では事業者の努力義務として、先述の格差を念頭に、消費者の権利義務その他の消費者契約の内容が明確なものでかつ消費者にとって平易なものになるよう配慮することが求められている。また、個々の消費者の知識及び経験を高所した上で、消費者の権利義務その他の消費者契約の内容について必要な情報を提供することが求められ

⁸個人情報保護法 (以下、法) 第 2 条第 2 項、個人情報の保護に関する法律施行令 (平成 15 年政令第 507 号)(以下、政令) 第 1 条、個人情報の保護に関する法律施行規則 (平成 28 年個人情報保護委員会規則第 3 号) (以下、規則)

⁹法第 2 条第 3 項、政令第 2 条、規則第 5 条

¹⁰法第 18 条第 1 項 個人情報取扱事業者は、あらかじめ本人の同意を得ないで、前条の規定により特定された利用目的の達成に必要な範囲を超えて、個人情報を取り扱ってはならない。

る。¹¹[12] このような規定は、消費者が不利な契約条件に同意しないことを目的としている。よって、利用規約も消費者にとって平易でかつ権利義務について明確に記す努力義務がある。

このような背景をもとに、消費者庁では、契約条項の分かりやすい表示について検討がなされている。分かりやすさの 1 つとして、消費者が定型約款にアクセスしやすくするものがある。これに加えて重要な契約条項について、消費者に分かりやすく表示することを事業者にも必要であると指摘されている。[13] その中でも、特定業種については、すでに個別の法令で規定がなされている。例えば、保険業における保険業法¹²、電気通信事業者における電気通信事業法¹³などが挙げられている [14]。これらの個別の法令については、説明をするべき事項が列挙されており、ある程度の型をもとに説明をする必要がある。これらを総合して契約条項の分かりやすい表示の検討が行われているが、具体的な規定には未だに至っておらず [15]、先述したように、現状では分かりやすい契約条項の明記は特定の業種を除いて消費者契約法に基づく努力義務に留まっている。

2.2 前提技術

本節では、本研究において前提となる技術について述べる。

2.2.1 自然言語処理

人工言語とは、コンピュータで用いるコンピュータ語のような、形式言語のような言語である。これに対して、自然言語とは、人間がコミュニケーションを取るために、民族や国家などにより自然に運用されてきた言語のことである。例えば、日本語や英語のことである。自然言語のコンピュータ処理に関する学問分野、研究開発分野を自然言語処理とよぶ [16]。本研究では、自然言語処理を利用して、利用規約の解釈を行う。

2.2.2 Transformer

Transformer[17] は、自然言語処理用の深層学習で使われているモデルである。Transformer 以前の自然言語処理では、回帰型ニューラルネットワークが主に用いられていた。これは、自然言語が順序性を持っていることに着目をし、入力された文章を端から端まで次の単語を読む時間ステップごとに処理した内容を記憶していく神経ユニットを持っているモデルである [18]。しかし、このモデルは離れた位置にある情報を考慮して処理を行うのが難しいという問題がある。これに対して Transformer はテキストの文脈を用いているが回帰型ニューラルネットワークのような時系列的な方法は使用していない。入力されたある単語が与えられると、その周りの全ての単語を見て self-attention(注意機構、自己注

¹¹消費者契約法 第 3 条

¹²保険業法第 294 条、同法施行規則 227 条の 2

¹³電気通信事業法第 26 条、電気通信事業法施行規則第 22 条の 2 の 3

意) と呼ばれるある 1 文の単語だけを使って計算された単語間の関連づけをするための概念を用いて、文章の文脈に関して表現をする。例えば、以下の文を翻訳したい入力文だとする。

“The animal didn’t cross the street because it was too tired”

この文章の it が何を指しているのかは人間には簡単に文脈から理解することができるが、“street”のことを指しているのか、“The animal”を指しているのかはコンピュータが理解することは困難である。モデルが“it”という単語を処理しているとき、self-attention は“it”を“The animal”を関連づけることができる [1]。Transformer はこのように文脈をモデル化することができるため、他の深層学習に比べて表現能力が高いことからよく利用されている。

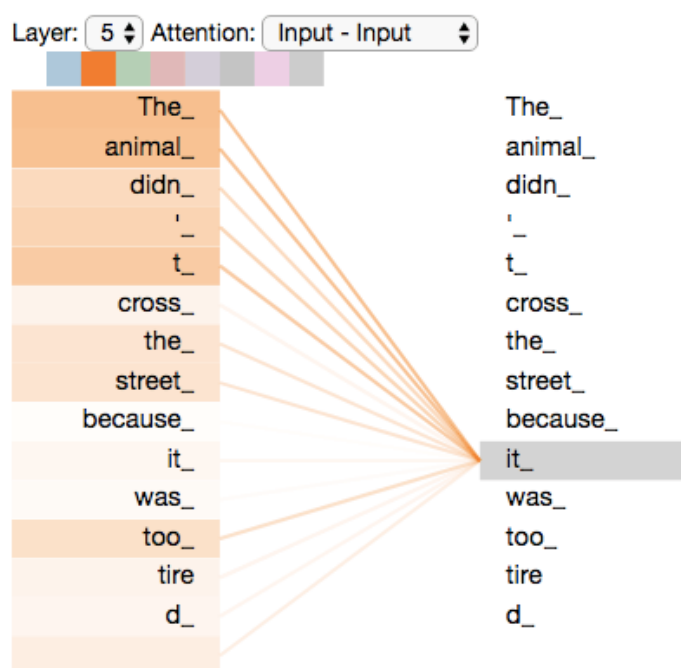


図 2.1: Transformer における self-attention の仕組み、文献 [1] より引用

2.2.3 BERT

Transformer が発表された後に自然言語処理では Transformer を用いた様々な発展モデルが発表された。そのうちのひとつが Bidirectional Encoder Representations from Transformers(BERT)[2] である。BERT は、非常に大規模な Transformer のモデルを、文の一部をそれ以外の部分を使って予測することで、教師なしで学習 (事前学習と呼ばれる) するというものである。これによって、言語の高レベルなニュアンスをエンコードできるようになる [19]。図 2.2 の左側は事前学習されたモデルである。このモデルは、右側に示

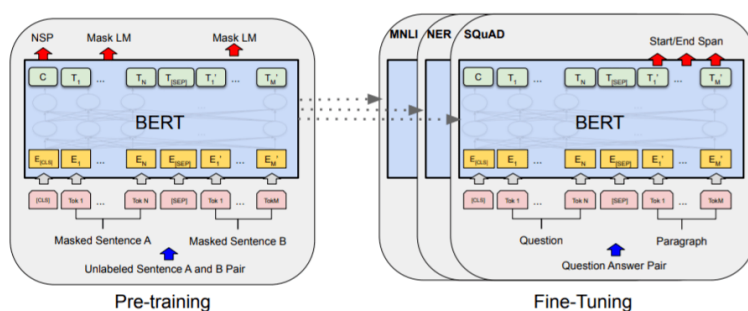


図 2.2: BERT の事前学習とファインチューニングされたタスクごとのモデル、文献 [2] より引用

されているようなテキスト分類 (MNLI)、固有表現認識 (NER)、質問応答 (SQuAD) のような自然言語処理タスクのためにファインチューニングを行う。事前学習された膨大な量の知識により、BERT は先述したようなさまざまなタスクのために知識を効率的に移転をすることができる。

2.2.4 Sentence BERT

Sentence BERT とは、事前学習済み BERT の BERT に Pooling 層を加え、自然言語推論タスクで構築される。

第3章 問題

本章では、サービス利用者がインターネットサービスを利用する際に利用規約を読む場面において、本研究の解決する問題について述べる。また、利用規約が読まれないことに対しての問題点や原因を明確化した後に、問題解決のための要件を述べる。

3.1 利用規約の認知

2020年に消費者庁が「デジタル・プラットフォーマーの取引慣行等に関する実態調査」の一環として、デジタル広告分野についての実態調査を行い、検索サービス及びSNSの利用者向け(消費者向け)アンケート調査を行った。この調査の結果を以下に示す。

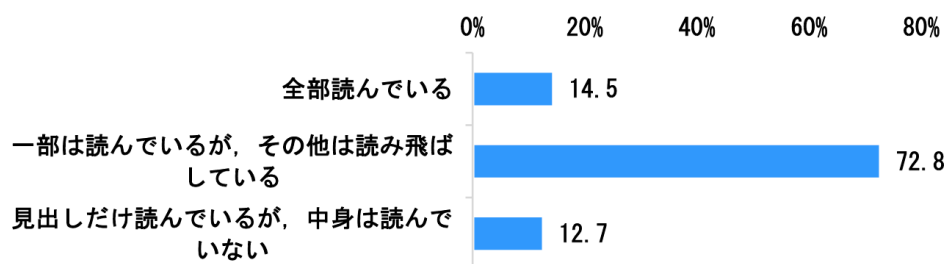


図 3.1: 検索サービスの利用規約をどの程度読んでいるか (回答数: 448)、文献 [3] より引用

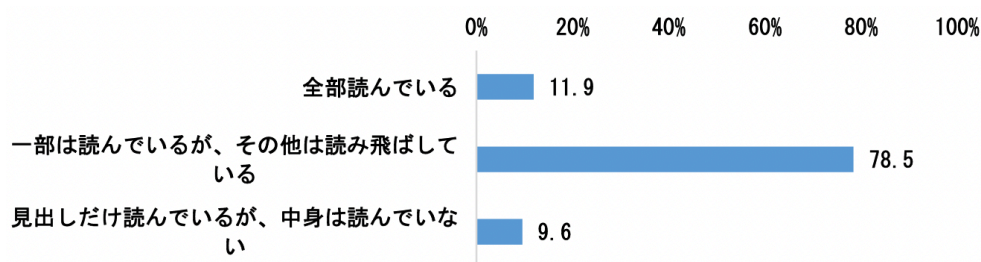


図 3.2: SNS 等の利用規約をどの程度読んでいるか (回答数: 929)、文献 [3] より引用

図 3.1、3.2では、それぞれ検索サービス、SNSについて「どの程度読んでいるか」についての調査が行われている。なお、この質問の前段では、「利用規約を認知しているか」についての質問があり、これについて「知っている」を選択した人がこの質問に回答して

いる。このような調査では一般的に、社会適応バイアスにより高めの数値となってしまうことが多いが、それを指し引かなくとも、ほとんどの人が読み飛ばしているもしくは読んでいないという結果が示されている。

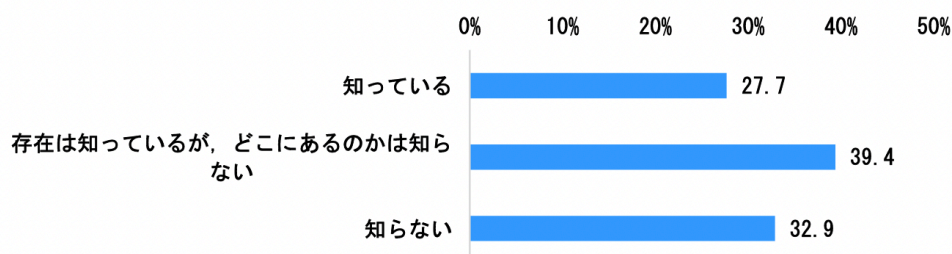


図 3.3: 検索サービスの利用規約の認知 (回答数: 2,000)、文献 [3] より引用

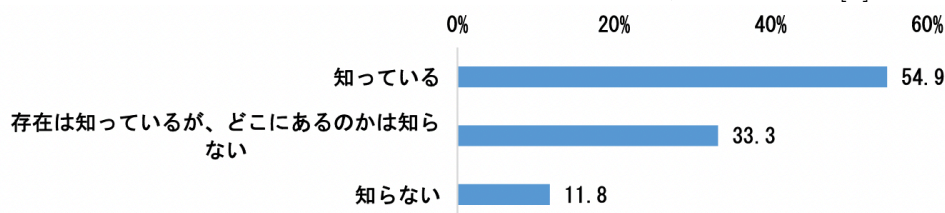


図 3.4: SNS 等の利用規約の認知 (回答数: 2,000)、文献 [3] より引用

本研究では、「利用規約の認知」についてはアプローチが非常に困難であることから、この認知を前提として、「利用規約を読み飛ばしてしまう」事象を問題として定義し、それに対するアプローチについての検討を行う。

3.2 利用規約の読解時間

ネットマイル社が2012年に「ネットサービスの利用規約・プライバシー調査」を行った[20]¹。これによると、「利用規約の内容は重要だと思うか」という質問に対し、94.5%の人が「重要だと思う」もしくは「サービスによっては重要だと思う」と答えており、利用者は利用規約に関する重要性について認識をしている(図3.5)。しかし、「利用規約を読まないことでリスクはあると思うか」という質問に対しては、「把握している」と答えた人は54.6%となり重要性についての認識は重要だと思っている人の割合からは低くなっている(図3.6)。この理由の一つについて、実態として利用規約を読んでいないために具体的なリスクについて述べるできないということが考えられる。本調査においても「サービスを利用する前に利用規約を読むか」という質問がなされており(図3.7)、読むと答えた人が15.0%となっている。また、「利用規約を読まない理由」についても質問がなされており、全体で一番多い回答が「めんどくさい」(87.6%)、次に「時間がない」(26.0%)が

¹詳細なアンケート結果はネットマイル社のサイトから削除されているため、この結果について報じている「TECH+」(マイナビ社)のニュースサイトより引用している。

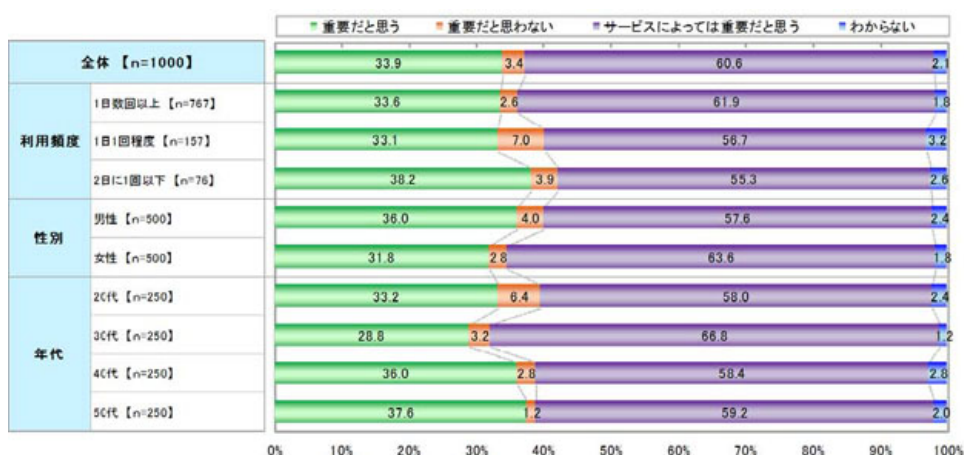


図 3.5: 「利用規約の内容は重要だと思うか」、文献 [4] より引用

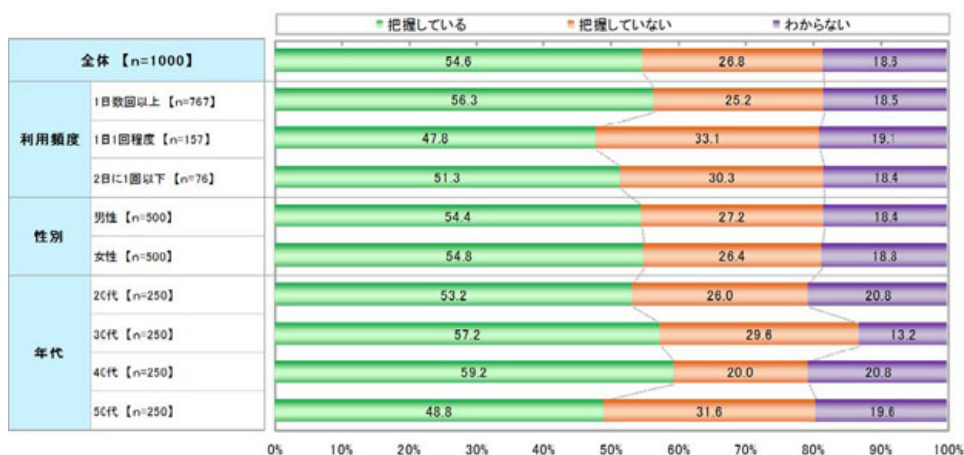


図 3.6: 「利用規約を読まないことでリスクはあると思うか」、文献 [4] より引用

挙げられている。この調査で挙げられている理由をもとにすると、利用規約を読む重要性を感じている人は多いと考えられるが、「めんどくさい」、「時間がない」などの理由により利用規約を読まないことが多いと言える。「めんどくさい」に関しては、利用規約を示す側のリスクや法的問題なども関わっており、即座に解決することは難しい。よって、本研究では、調査では次点であった「時間がない」に関して利用規約を提示するインターネット環境においての問題点として提起する。

3.3 利用規約の読解

オムリ・ベン＝シャハーほか (2022)[21] は、利用規約をはじめとして、食品表示ラベル、著作権表示警告、金融機関の取引前リスク告知、医療行為の場面で行われるインフォームドコンセントなど、法令などで規制されている同意を求める行為をまとめて「開示主義」



図 3.7: 「サービスを利用する前に利用規約を読むか」、文献 [4] より引用



図 3.8: 「利用規約を読まない理由」、文献 [4] より引用

と定義し、これらの規制を「義務的情報開示」と定義している。その中で、開示主義が失敗していると述べている。(要追記)

3.4 問題の定義

利用規約を読む重要性を認識している人は多いが、その実態は利用規約を読まない人が多い。理由の一つとして、利用規約の読解のために多くの時間がかかるという点が挙げられる。このような現象により、インターネットサービスを安心して利用できる環境ではないと考えられる。よって、利用規約を読むために多くの時間がかかるということを問題定義とし、本研究では利用規約の読解時間短縮を目指す。

第4章 提案手法

本章では提案手法について述べる。

4.1 概要

本研究では、利用規約の読解時間短縮のために、本研究では「個別文脈依存蓄積差分確認方式」を提案する。これは、利用規約の個別の文脈についてを蓄積していき、その差分についてを確認することである。具体的には、以前同意した利用規約を記録し、それに類似する条文を読んだときに類似していることを検出し、それ以外に注目することで、利用規約の読解時間の短縮ができるのではないかと考えた。この提案手法は大きく2つの機能を持つ。1つ目は「利用規約類似文検出」、2つ目は「利用規約警告検出」である。詳細については各節で述べる。

4.2 利用規約類似文検出

仮説のうち、類似条文を検出する部分について述べる。

図4.1では、イメージを示している。ここでは、提案するシステムを利用し始めて、初めに、A社の利用規約を読んだときは条文の背景が全て黄色で目立ち、次にB社の利用規約を読むと3がA社の1と類似する条文であるため、それ以外の条文が背景が黄色で目立ち、最後にC社の利用規約を読むと1,3,4,5が類似する条文であるため、2が背景が黄色で目立つイメージとなっている。

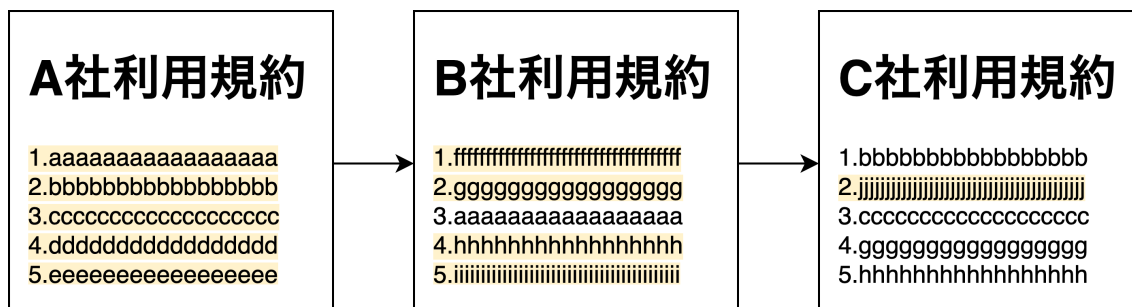


図 4.1: 類似条文検出のイメージ

4.3 利用規約警告検出

4.2 項では、利用規約の類似文を検出していったが、この仕組みの問題点として、問題があるが、サービスを利用したいために仕方なく利用したという状況が考えられる。このような状況に陥ってしまうと、提案するシステムは正しく挙動することができないため、その対策として、利用規約のうち、問題があると利用者が考えた部分について、ユーザーに指定してもらいうことで、利用規約の問題のある部分を目立たせることができる。図 4.2 では、利用規約警告検出のイメージを示している。A 社の 1 が問題ある条文であるとユーザーが考えたため、1 にチェックをすると、次に B 社の利用規約を読んだときにその条文と類似している 3 の条文の背景を水色にして目立たせる。最後に、C 社の利用規約を読むと、再び類似する条文 1 について水色が目立つイメージとなっている。

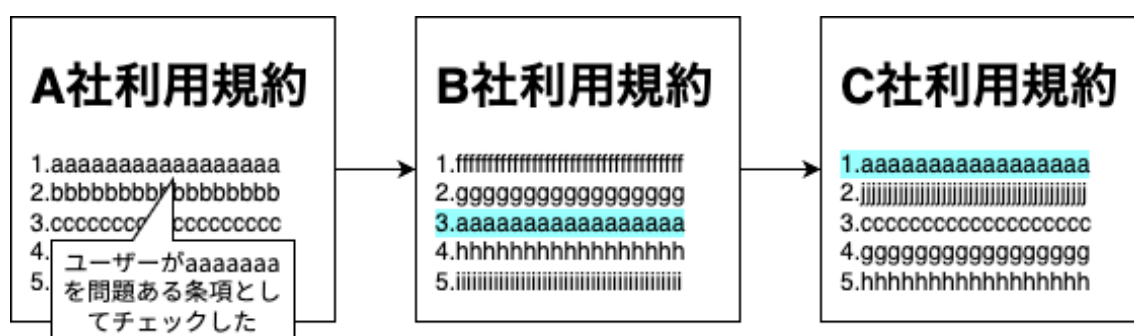


図 4.2: 利用規約警告検出

第5章 実験

本章では提案手法の実験について述べる。

5.1 概要

本研究の仮説をもとに、システムの実装をおこなった。

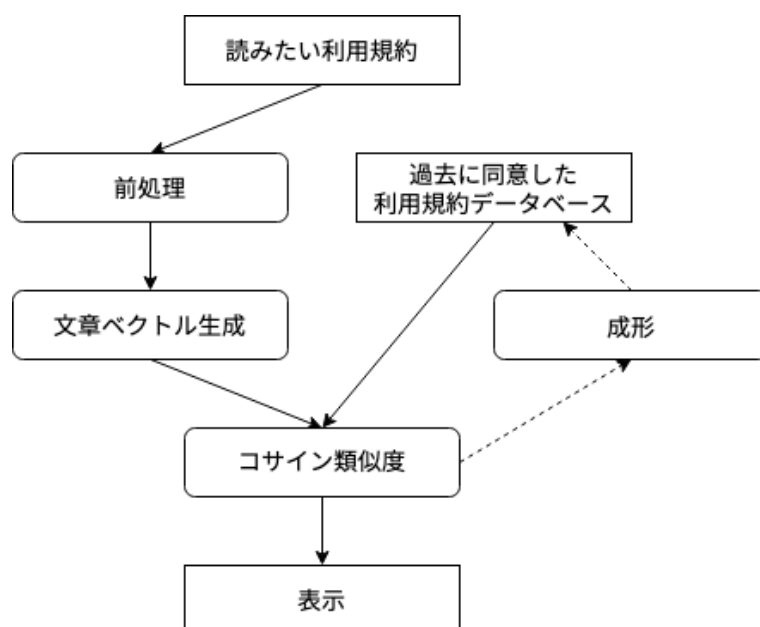


図 5.1: 類似条文検出の実装イメージ

各節にて、図 5.1 に基づいた実装の詳細について説明をする。利用規約を表示した際に出てくる各条項を「読みたい利用規約」として分解を行う。それぞれの条項に対して、図 5.1 のシステムを通し、表示を行う。なお、本章では利用規約の項や号についてを条項と表現する。

5.2 前処理

まず、読みたい利用規約について前処理を行う。これは、条項が自然言語処理を行うために適さない文章である場合があるためである。具体的には、正規表現を利用し、以下の

ような処理を行なった。

- 空白の除去
- URL の除去
- かっこ (「、() など) の除去
- 「ユーザー」「ユーザ」「利用者」などの表現の統一
- 社名の置き換え

以上は処理の例である。このような処理を通して、類似度の精度の向上を行う。

5.3 文ベクトル生成

類似度を求めるために、文ベクトルの生成を行う。文章ベクトルの生成には、質問モデルなどで精度の高かった Sentence-BERT を利用する。また、日本語の事前学習モデルについては、芝山ら [22] の研究において精度が高かった、東北大版の BERT¹ をもとに生成された Sentence BERT モデル² を利用して生成を行う。

5.4 コサイン類似度

生成した文ベクトルは、以下の式でコサイン類似度を求めることにより、与えられた文同士の意味の近さを数値化することができる。

$$sim = \frac{x \cdot y}{|x||y|}$$

本研究では、この数値を Score として最終的に表示を行う。また、この数値を一定の閾値において表示の目立たせる切り替えを行う必要がある。閾値に関しては、実際に利用規約の表示を行い比較を行ったところ、 $score = 0.89$ を閾値としている。この値は、モデルの差異などによっても異なっているため、さらに調整が必要になる可能性がある。

5.5 成形

類似度の表示を行なった後に、今後読む利用規約のために、過去に同意した利用規約データベースに追記を行う。(なお、この処理は、利用規約に「同意する」ボタンをクリックした場合について行い、「同意しない」ボタンをクリックした場合は追記を行わない。)5.2 節での処理などが十分に行われているのかを確認し、データベースに追記する問題がない

¹<https://github.com/cl-tohoku/bert-japanese>

²<https://huggingface.co/sonoisa/sentence-bert-base-ja-mean-tokens-v2>

この確認を行う。おおむねは 5.2 節での処理で十分であることが多いが、様々な利用規約を調査したところ、これらの処理では不十分な場合があり、これ以外にも前処理を手作業で行なった部分がある。以下にその例を示す。

モイ株式会社 サービス利用規約 第 3 条第 1 項

モイは、本サービス運営上の合理的必要性があるとモイが判断した場合、本利用規約を変更できるものとします。

モイ株式会社のサービス利用規約においては、全文にわたってモイ株式会社について「モイ」もしくは「当社」と表現している。「モイ」のままであると類似度を求めるにあたって不都合なため、「当社」と置き換える処理を行う。これらに関して、固有表現抽出を使用して社名を認識することができるのではないかと検討を行ったが、以下のような利用規約では適用できないことが判明した。

株式会社 Z 会ソリューションズ 利用規約

ユーザがクレジットカード決済、コンビニ決済を選択した場合、当社は請求業務を GMO ペイメントゲートウェイ株式会社に委託できるものとします。

このような利用規約の場合、「GMO ペイメントゲートウェイ株式会社」のような表現がある場合、これについても会社名と認識してしまう場合がある。このとき、一般的な利用規約の場合は「当社は請求業務を代行業社に依頼することができる」といった表現になるため、これらについての置き換えは現状困難であると考えられる。よって、これらについて手作業を入れて確認を要する。

5.6 過去に同意した利用規約データベース

5.5 節において問題なく成形された条項に関して、その後に読む利用規約と比較するためにデータベースへの登録を行なう。データベースへは以下の情報を登録する。

表 5.1: 過去に同意した利用規約データベースの内容

形式	項目
数値	条項 ID
数値	条項の形式 ³
文字	条項
ベクトル	条項の文ベクトル

データベースに文ベクトルを保存しておくことにより、効率的に類似度の計算を行うことができる。

³条項であるか、第 n 条などの見出しであるかを区別する。見出しは文ベクトルの生成を行わず、類似度の比較も行わない。

なお、「同意しない」を選択した利用規約の条項に関してはデータベースへの登録は行わない。

5.7 利用規約警告検出

利用規約警告検出についても同様の処理を行う。なお、閾値を超えた場合に文を目立たせる処理を行うため、上記までと逆の判定をおこなっている。

5.8 表示方法

表示方法については、利用規約の類似条文検出に加えて、判断の材料にするために、類似度の高い順に3つの条文とその条文をもつサービス名、コサイン類似度を表示した。3つに設定した根拠としては、他サービスの条文と比較するときに複数サービスの条文と比較することで、より類似度を利用者が比較しやすいと考えたからである。

また、利用規約を目立たせる表示として、一定の閾値のコサイン類似度を超えた背景をオレンジ色とすることにした。

利用規約警告検出は、オレンジ色と区別をできるように、水色とした。また、この区別は利用者が能動的に指定をおこなっているため、類似条文検出よりも優先して表示をおこなっている。

利用規約3つのスコア表示と利用規約警告表示の図を追加する。

Alpha

第1条（適用）

1. 本規約は、ユーザーと当社との間の本サービスの利用に関わる一切の關係に適用されるものとします。
2. 当社は本サービスに関し、本規約のほか、ご利用にあたってのルール等、各種の定め（以下、「個別規定」といいます。）をすることがあります。これら個別規定はその名称のいかんに関わらず、本規約の一部を構成するものとします。
3. 本規約の規定が前条の個別規定の規定と矛盾する場合には、個別規定において特段の定めなき限り、個別規定の規定が優先されるものとします。

第2条（利用登録）

1. 本サービスにおいては、登録希望者が本規約に同意の上、当社の定める方法によって利用登録を申請し、当社がこの承認を登録希望者に通知することによって、利用登録が完了するものとします。
2. 当社は、利用登録の申請者に以下の事由があると判断した場合、利用登録の申請を承認しないことがあり、その理由については一切の開示義務を負わないものとします。
 - i. 利用登録の申請に際して虚偽の事項を届け出た場合
 - ii. 本規約に違反したことがある者からの申請である場合
 - iii. 反社会的勢力に該当したとき、または反社会的勢力と関係を有している者からの申請である場合
 - iv. その他、当社が利用登録を相当でないと判断した場合

第3条（ユーザーIDおよびパスワードの管理）

1. ユーザーは、自己の責任において、本サービスのユーザーIDおよびパスワードを適切に管理するものとします。
2. ユーザーは、いかなる場合にも、ユーザーIDおよびパスワードを第三者に譲渡または貸与し、もしくは第三者と共用することはできません。当社は、ユーザーIDとパスワードの組み合わせが登録情報と一致してログインされた場合には、そのユーザーIDを登録しているユーザー自身による利用とみなします。
3. ユーザーID及びパスワードが第三者によって使用されたことによって生じた損害は、当社に故意又は重大な過失がある場合を除き、当社は一切の責任を負わないものとします。

図 5.2: Alpha 社の利用規約/どの利用規約も既読ではない

Alpha

第1条（適用）

1. 本規約は、ユーザーと当社との間の本サービスの利用に関わる一切の關係に適用されるものとします。
2. 当社は本サービスに関し、本規約のほか、ご利用にあたってのルール等、各種の定め（以下、「個別規定」といいます。）をすることがあります。これら個別規定はその名称のいかんに関わらず、本規約の一部を構成するものとします。
3. 本規約の規定が前条の個別規定の規定と矛盾する場合には、個別規定において特段の定めなき限り、個別規定の規定が優先されるものとします。

第2条（利用登録）

1. 本サービスにおいては、登録希望者が本規約に同意の上、当社の定める方法によって利用登録を申請し、当社がこの承認を登録希望者に通知することによって、利用登録が完了するものとします。
2. 当社は、利用登録の申請者に以下の事由があると判断した場合、利用登録の申請を承認しないことがあり、その理由については一切の開示義務を負わないものとします。
 - i. 利用登録の申請に際して虚偽の事項を届け出た場合
 - ii. 本規約に違反したことがある者からの申請である場合
 - iii. 反社会的勢力に該当したとき、または反社会的勢力と関係を有している者からの申請である場合
 - iv. その他、当社が利用登録を相当でないと判断した場合

第3条（ユーザーIDおよびパスワードの管理）

1. ユーザーは、自己の責任において、本サービスのユーザーIDおよびパスワードを適切に管理するものとします。
2. ユーザーは、いかなる場合にも、ユーザーIDおよびパスワードを第三者に譲渡または貸与し、もしくは第三者と共用することはできません。当社は、ユーザーIDとパスワードの組み合わせが登録情報と一致してログインされた場合には、そのユーザーIDを登録しているユーザー自身による利用とみなします。
3. ユーザーID及びパスワードが第三者によって使用されたことによって生じた損害は、当社に故意又は重大な過失がある場合を除き、当社は一切の責任を負わないものとします。

図 5.3: Alpha 社の利用規約/Beta 社が既読

Alpha

第1条（適用）

1. 本規約は、ユーザーと当社との間の本サービスの利用に関わる一切の關係に適用されるものとします。
2. 当社は本サービスに関し、本規約のほか、ご利用にあたってのルール等、各種の定め（以下、「個別規定」といいます。）をすることがあります。これら個別規定はその名称のいかんに関わらず、本規約の一部を構成するものとします。
3. 本規約の規定が前条の個別規定の規定と矛盾する場合には、個別規定において特段の定めなき限り、個別規定の規定が優先されるものとします。

第2条（利用登録）

1. 本サービスにおいては、登録希望者が本規約に同意の上、当社の定める方法によって利用登録を申請し、当社がこの承認を登録希望者に通知することによって、利用登録が完了するものとします。
2. 当社は、利用登録の申請者に以下の事由があると判断した場合、利用登録の申請を承認しないことがあり、その理由については一切の開示義務を負わないものとします。
 - i. 利用登録の申請に際して虚偽の事項を届け出た場合
 - ii. 本規約に違反したことがある者からの申請である場合
 - iii. 反社会的勢力に該当したとき、または反社会的勢力と関係を有している者からの申請である場合
 - iv. その他、当社が利用登録を相当でないと判断した場合

第3条（ユーザーIDおよびパスワードの管理）

1. ユーザーは、自己の責任において、本サービスのユーザーIDおよびパスワードを適切に管理するものとします。
2. ユーザーは、いかなる場合にも、ユーザーIDおよびパスワードを第三者に譲渡または貸与し、もしくは第三者と共用することはできません。当社は、ユーザーIDとパスワードの組み合わせが登録情報と一致してログインされた場合には、そのユーザーIDを登録しているユーザー自身による利用とみなします。
3. ユーザーID及びパスワードが第三者によって使用されたことによって生じた損害は、当社に故意又は重大な過失がある場合を除き、当社は一切の責任を負わないものとします。

図 5.4: Alpha 社の利用規約/Beta 社、Gamma 社が既読

Alpha

第1条（適用）

1. 本規約は、ユーザーと当社との間の本サービスの利用に関わる一切の関係に適用されるものとします。
2. 当社は本サービスに関し、本規約のほか、ご利用にあたってのルール等、各種の定め（以下、「個別規定」といいます。）をすることがあります。これら個別規定はその名称のいかんに関わらず、本規約の一部を構成するものとします。
3. 本規約の規定が前条の個別規定の規定と矛盾する場合には、個別規定において特段の定めなき限り、個別規定の規定が優先されるものとします。

第2条（利用登録）

1. 本サービスにおいては、登録希望者が本規約に同意の上、当社の定める方法によって利用登録を申請し、当社がこの承認を登録希望者に通知することによって、利用登録が完了するものとします。
2. 当社は、利用登録の申請者に以下の事由があると判断した場合、利用登録の申請を承認しないことがあり、その理由については一切の開示義務を負わないものとします。
 - i. 利用登録の申請に際して虚偽の事項を届け出た場合
 - ii. 本規約に違反したことがある者からの申請である場合
 - iii. 反社会的勢力に該当したとき、または反社会的勢力と関係を有している者からの申請である場合
 - iv. その他、当社が利用登録を相当でないと判断した場合

第3条（ユーザーIDおよびパスワードの管理）

1. ユーザーは、自己の責任において、本サービスのユーザーIDおよびパスワードを適切に管理するものとします。
2. ユーザーは、いかなる場合にも、ユーザーIDおよびパスワードを第三者に譲渡または貸与し、もしくは第三者と共用することはできません。当社は、ユーザーIDとパスワードの組み合わせが登録情報と一致してログインされた場合には、そのユーザーIDを登録しているユーザー自身による利用とみなします。
3. ユーザーID及びパスワードが第三者によって使用されたことによって生じた損害は、当社に故意又は重大な過失がある場合を除き、当社は一切の責任を負わないものとします。

図 5.5: Alpha 社の利用規約/Beta 社、Gamma 社、Delta 社が既読

第6章 評価

本章では、5章で行った実装のの評価について述べる。

6.1 評価内容

評価として、以下の内容を検証する。実装が実際に利用できるという観点から、1つ目は類似度の計算が一般的な利用に問題ない時間であるかを測定する。2つ目は、実際に利用をしてもらい、読解時間を短縮することができるかについて検証を行う。

6.2 評価のための利用規約

提案手法の評価のために、利用規約を Alpha 社、Beta 社、Gamma 社、Delta 社、Epsilon 社、Zeta 社、Eta 社の 8 つ用意した。これらは、実在する会社やサービスの利用規約を実験用に会社名やサービス名を置き換えたり、一部条項の書き換えをおこなったものである。また、評価で利用規約の読解に要した時間を測定するために、条文の数を揃える作業を行なっている。利用規約については、以下のような観点から選んだ。

- 条文数が極端に多くない
- 可能な限り違う種類のサービスを展開している

表 6.1: 実験用利用規約の一覧

実験用社名	主に元にした利用規約
Alpha 社	揭示版・SNS 系向きの利用規約の雛形（ひな型） ¹
Beta 社	Annict ²
Gamma 社	鎌倉新書 ³
Delta 社	クックパッド ⁴
Epsilon 社	第一法規 ⁵
Zeta 社	IRIAM ⁶
Eta 社	三越伊勢丹 WEB 会員規約 ⁷
Theta 社	Z 会ソリューションズ ⁸

6.3 評価のためのクイズ

本提案手法により利用者の理解度が下がっていないかを確認するために、クイズを設定した。クイズに内容は、評価のための利用規約の中から広く選び、また、一部書き換えた条文などをクイズとして出題をした。具体的なクイズの内容は付録に記す。

¹<https://kiyaku.jp/hinagata/sns.html>

²<https://annict.com/terms>

³<https://www.kamakura-net.co.jp/servicepolicy/>

⁴<https://cookpad.com/terms/free>

⁵<https://www.daiichihoki.co.jp/support/rules/>

⁶<https://www.live.iriam.com/terms>

⁷https://www.mistore.jp/shopping/help/guide/terms_h.html

⁸<https://www.zkai.co.jp/assess/terms>

第7章 関連研究

本章では、本研究の関連研究を示す。

- 7.1 機械学習を用いた利用規約からの未知条項の抽出に関する研究 (中村ら 2018)
- 7.2 CLAUDETTE: an automated detector of potentially unfair clauses in online terms of service (Lippi ら 2019)
- 7.3 利用規約中の不公平文の自動検出 (青山ら 2019)

第8章 結論

本章では、本研究のまとめと今後の課題を示す。

8.1 本研究のまとめ

8.2 本研究の課題

本節では、本研究で提案したシステムの課題と展望について述べる。

8.2.1 構想

本研究の提案をより、社会実装しやすくするためのものとして、利用規約図書館のような構想がある。この構想は渡邊 2020 で提案されている利用規約をユーザー同士でとりまとめて保存しておくシステムである。このような仕組みを構築することで、同じ約款を大量の人が同意する点に着目し、以下のようなメリットが考えられる。

1. 利用規約の変更などを利用者同士で確認し、追従することができる。
2. 本研究の実装の自然言語処理などの手順を誰か 1 人が取ることで、その成果を他の利用者也利用することができる。

本研究では特に、文ベクトルの生成に利用者の計算リソースを消費してしまうという課題があり、さらに、文量が非常に多い約款の場合はシステムの利用のための時間がかかってしまうため、この仕組みをもとにシステムを稼働することができれば、より利用しやすいシステムになると考えられる。しかし、この提案にも維持者や管理人が必要になるために、このコストを誰が負担するかという問題がある。論文では、この問題を克服するために、適格消費者団体のような、現在利用規約の監視や裁判を消費者側の立場で行なっている団体が管理することを提案している。

付 録 A 実験に利用した利用規約

実験に利用した利用規約を以下に示す。なお、利用した利用規約は実在する利用規約をベースに社名やサービス名などを置き換え、実験用に条項の一部を変更したものである。

A.1 Alpha社

謝辭

謝辭

参考文献

- [1] Jay Alammar. The illustrated transformer – jay alammar – visualizing machine learning one concept at a time. <https://jalammar.github.io/illustrated-transformer/>, 6 2018. (Accessed on 12/27/2022).
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018.
- [3] (令和3年2月17日) デジタル・プラットフォーム事業者の取引慣行等に関する実態調査 (デジタル広告分野) について (最終報告):公正取引委員会. <https://www.jftc.go.jp/houdou/pressrelease/2021/feb/210217.html>, 2021. (Accessed on 12/17/2022).
- [4] マイナビ. 利用規約をきちんと読んでいる人は15% - google 規約変更は8割が知らない — tech+ (テックプラス) . <https://news.mynavi.jp/techplus/article/20120423-a054/>, 2012. (Accessed on 12/28/2022).
- [5] 板倉陽一郎. 個人情報の取扱いに関する利用規約上の定めに関する考察. **研究報告電子化知的財産・社会基盤 (EIP)** , 2013-EIP-62(4):1-6, 2013.
- [6] 橋本阿友子. 民法改正による新制度 (第1回) - 定型約款 — j-net21[中小企業ビジネス支援サイト], 2021. Accessed: 2022-10-25.
- [7] 民法の一部を改正する法律 (債権法改正) について. https://www.moj.go.jp/MINJI/minji06_001070000.html, 2020. Accessed: 2022-10-26.
- [8] 定型約款の規定の新設, 2019.
- [9] 尼崎西宮総合法律事務所. 改正民法の定型約款について解説! 判断のポイントは? https://www.hanshin-law.com/blog/cat54/post_249.html. (Accessed on 12/21/2022).
- [10] 政府広報オンライン. 「個人情報保護法」をわかりやすく解説 個人情報の取扱いルールとは? <https://www.gov-online.go.jp/useful/article/201703/1.html>, 8 2022. (Accessed on 12/22/2022).
- [11] 弁護士法人ファースト&タンデムスプリント法律事務所. その利用規約は有効? 消費者契約法と有効な利用規約の作り方. <https://ec-lawyer.com/620/#i>, 3 2022. (Accessed on 12/22/2022).

- [12] 消費者庁. 消費者契約法 逐条解説. https://www.caa.go.jp/policies/policy/consumer_system/consumer_contract_act/annotations/, 2 2019. (Accessed on 12/22/2022).
- [13] 消費者契約法改正に向けた専門技術的側面の研究会. 消費者契約法改正に向けた専門技術的側面の研究会報告書. https://www.caa.go.jp/policies/policy/consumer_system/consumer_contract_act/review_meeting_002/, 9 2019. (Accessed on 12/22/2022).
- [14] 消費者庁. 契約条項の表示・不当条項について. https://www.caa.go.jp/policies/policy/consumer_system/consumer_contract_act/review_meeting_002/, 5 2020. (Accessed on 12/22/2022).
- [15] 消費者委員会 オンラインプラットフォームにおける取引の在り方に関する専門調査会. オンラインプラットフォームにおける取引の在り方に関する専門調査会報告書. https://www.cao.go.jp/consumer/iinkaikouhyou/2019/houkoku/201904_online_pf_houkoku.html, 4 2019. (Accessed on 12/22/2022).
- [16] 黒橋禎夫. **自然言語処理（改訂版）（放送大学教材）**. 放送大学教育振興会, 3 2019.
- [17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.
- [18] Andrej Karpathy. The unreasonable effectiveness of recurrent neural networks. <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>, 2015. (Accessed on 12/27/2022).
- [19] Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, and Harshit Surana. **実践 自然言語処理 ー実世界 NLP アプリケーション開発のベストプラクティス**. オライリージャパン, 2022.
- [20] 株式会社ネットマイル. ネットマイル | ネットサービスの利用規約・プライバシー調査～個人情報漏れ警戒するも、面倒で利用規約を読まない利用者像、明らかに～. https://biz.netmile.co.jp/news/press_2012/press_release120420.html, 2012. (Accessed on 12/28/2022).
- [21] オムリ・ベン＝シャハー and カール・E・シュナイダー. **その規約、読みますか?: 義務的情報開示の失敗**. 勁草書房, 2022.
- [22] 新納浩幸 芝山直希. 日本語 sentencebert の構築とその評価. Technical Report 7, 茨城大学大学院理工学研究科情報工学専攻, 茨城大学大学院理工学研究科情報科学領域, jul 2021.