

# 金融領域における裁判外紛争解決事例データの構築と和解判断予測

田村 光太郎

株式会社野村総合研究所

k9-tamura@nri.co.jp, k.tamura.phd@gmail.com

## 概要

国内金融 ADR 機関が公開する紛争解決事例を整理・統合し、申立人および被申立人それぞれの主張文と和解の有無から構成されるテキストデータセットを構築した。さらに、双方の主張文を同時に入力とする言語モデルを用い、和解判断を予測する枠組みを検討した。数値実験の結果、複数機関の事例を統合して学習することで、単一機関データを用いた設定と同等またはそれを上回る予測性能が得られることを確認した。本研究は、金融分野の紛争事例を対象とした言語処理研究に向けた基盤的なデータ構成を提示するとともに、和解判断予測タスクの可能性が期待される。

## 1 はじめに

金融分野では、勧誘、営業、取引の各段階において法令遵守や説明責任といったコンプライアンスが重視されている。近年は金融商品の複雑化や取引形態の多様化により、金融機関に対するコンプライアンス要請が一層高まっている。一方で、顧客とのトラブル（以下、紛争）の発生を完全に防ぐことは難しく、金融機関のコンプライアンス向上に向けては、紛争の要因や和解に至る過程を体系的に理解することが求められている。

このような金融分野の紛争を裁判外で調整する制度として、Alternative Dispute Resolution (ADR) が整備されている [1]。日本では、銀行・証券・保険などの分野ごとに業界団体が金融 ADR 機関を運営し、申立人と被申立人双方の主張から和解に向けた交渉が行われる。これらの機関は、事例の概要や当事者の主張、和解の有無を要約した形で公開 [2, 3, 4, 5, 6, 7, 8, 9] している。

言語処理分野では、法的文書や議論を対象とした対立構造や判断結果予測に関する研究

[10, 11, 12, 13, 14, 15] が進展しているが、金融 ADR のように制度や商品構造、説明義務といった産業固有の要素が反映された主張文を対象とした分析は十分に行われていない。

本研究では、国内金融 ADR 機関が公開する紛争解決事例を整理・統合し、申立人文、被申立人文、および和解の有無から構成される統一的なデータセットを構築する。さらに、著者の過去研究 [16] では、単一の ADR 機関で行ったモデルを検討しており、申立人文と被申立人文の対を同時に入力とする言語モデルを用い、和解に至る確率を予測する枠組みを検討する。

本研究の貢献は、金融分野の紛争テキストを対象とした言語処理研究として、(1) 国内金融 ADR 機関が公開する事例を整理し、申立人文・被申立人文と和解可否からなるデータセット構築の方向性を示す点、(2) 二者の主張文を同時に入力とする言語モデルにより、和解判断を推定するというタスクを定式化する点である。

以降、第 2 章で関連研究を整理し、第 3 章でデータセットの概要を述べる。第 4 章でモデルの構成を示し、第 5 章で数値実験、第 6 章で結果を示し、第 7 章でまとめを述べる。

## 2 関連研究

法的文書を対象とした自然言語処理研究は、近年、判断結果や意思決定の予測を中心に発展してきた。裁判文書を入力として判決結果を予測する研究 [10] をはじめ、大規模言語モデルの導入により、訴状や判決理由文の分類・要約などを扱う研究も数多く報告されている [17]。

紛争や対立の言語構造に着目した研究としては、主張・根拠・反論といった構成要素を抽出する Argument Mining が提案されている [12, 13]。これらの研究は、言語的な対立関係や主張構造を捉える点

表1 正規化した紛争事例データ（内容は例となるため、実際のデータと異なる）

正規化済み項目名	内容
事案番号	XXX
事案の概要	入金遅延による差損の補填要求
申立人の主張	入金処理の遅延のため、外国為替差損が（略）
被申立人の主張	照会時に受取人情報が不十分なためと回答し（略）
終了事由	和解
裁定の概要	あっせん委員会において...

で有用である一方、紛争解決という具体的な結果ラベルと結び付けた分析は多くない。

金融分野においても、法令文書、開示資料、顧客対応記録などを対象とした言語処理研究が進められているが、紛争事例そのものを対象とした研究は相対的に少ない。特に金融 ADR のように、産業固有の制度、商品構造、説明義務といった要素が反映された主張文が体系的に公開されているデータを用いた分析は限られている。

金融 ADR 事例を対象とした言語処理研究としては、著者による先行研究 [16] があり、国内金融 ADR 機関である FINMAC が公開する事例を整理し、申立人文および被申立人文の言語的特徴と紛争結果との関係を分析し、入力テキストにおける単語の寄与を評価している。

### 3 データセット

#### 3.1 金融 ADR 機関と公開データ

日本の金融 ADR 制度では、金融庁により指定された複数の ADR 機関が、金融商品や金融取引に関する紛争解決を担っている。金融庁の指定紛争解決機関一覧によれば、2025 年時点で銀行、証券、保険等の分野を中心に 8 種類の金融 ADR 機関が指定されている [1]。各機関は、管轄分野における紛争解決事例をウェブサイト上で公開しているが、公開形式（PDF 等）や記載項目、見出し構成、文章構造は公開時期や機関ごとに異なる。そのため、複数機関のデータを横断的に扱うには、事例本文の抽出と統一的な正規化処理が必要となる。

#### 3.2 データ収集と正規化

本研究では、公開されている金融 ADR 事例のうち、一般社団法人生命保険協会、日本貸金業協会、一般社団法人信託協会、特定非営利活動法人証券・金融商品あっせん相談センター（FINMAC）の 4 機関

が公表する事例を対象として収集・整理を行った。各機関のウェブサイト上でそれぞれ 2001 年度、2010 年度、2014 年度、2011 年度から定期的に公開されている事例を対象としている。手続きが終了した事案が公表されるケースが多く、公表時期は事案の開始時期とは必ずしも一致しない。

各機関が公開する PDF 文書から事例本文を抽出し、正規表現ベースのルールにより見出し表現や表記揺れを補正した。具体的には、「申立内容」「申出内容」「相手方の見解」「事業者の回答」など機関や時期により異なる見出し語を統一的にマッピングし、段落や文頭の書き出しをもとに境界を検出することで、共通フォーマットへ正規化した（表 1）。

申立人と被申立人の主張が明確に区分されていない事例や、概要のみが記載された事例は分析対象から除外した。また、終了事由が独立項目として記載されていない場合は、本文中の表現に基づき正規表現ルールにより分類した。そのため、実際に ADR 機関が発表する和解等の件数とは異なる可能性があることに注意する。特に、終了事由として、「和解」のほか、申立人として不適格であることでの「不受理」や、交渉の不調や訴訟への移行にともなう「打ち切り」などを設定した。

#### 3.3 データセット構造

正規化後の各事例  $i$  において、申立人と被申立人の対になった主張文  $A_i, B_i$  と、それに対応する和解の有無  $y_i$  を利用する。すなわち、モデル化には終了事由のラベルを「和解」とそれ以外という形に 2 値化したうえで、各事例  $i$  を

$$T_i = (A_i, B_i, y_i) \quad (1)$$

として表し、 $A_i$  を申立人の主張、 $B_i$  を被申立人の主張、 $y_i \in \{0, 1\}$  を和解の有無（和解を 1、非和解を 0）と定義する。

また、主張データの有無にかかわらず、手続きが不受理となった事案を除外したうえで、適格性に基づ

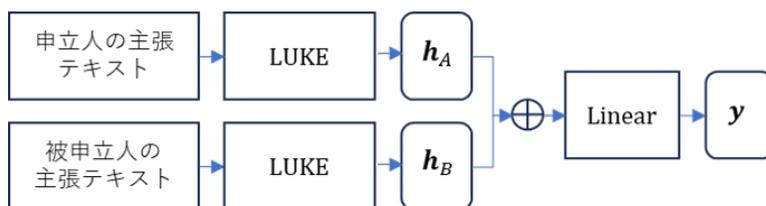


図1 提案モデルのアーキテクチャ

いて受理された事案のみを対象とし、複数機関にまたがる6832件の事例を統合的に扱った。

また、紛争解決委員やADR機関による評価・コメント等は分析対象から除外し、当事者双方の主張文のみを利用した。

### 3.4 データの基礎量

本研究で構築したデータセットは、全体で6832件の金融ADR事例から構成されている。このうち、和解に至った事案は2665件であり、全体の約39%を占める。申立人文および被申立人文の平均文長はそれぞれ228文字および211文字であり、両者の文長は概ね同程度である。

ADR機関別に見ると、生命保険協会が4464件と最も多く、和解率は0.279であった。申立人文および被申立人文の平均文長はそれぞれ270文字、227文字となった。これに対し、FINMACは2350件であるが、和解率は0.600の水準にある。平均文長は148文字および182文字であった。最後に、信託協会、日本貸金業協会はそれぞれ10件、8件と件数が少なくなっている。本研究では当該区分は参考値として扱う。

## 4 提案モデル

著者の過去研究では、FINMACが公開する金融ADR事例データを用い、申立人の主張文および被申立人の文に加えて補償金額情報を入力とし、和解確率および補償割合を同時に推定するモデルを構築した[16]。しかし、複数のADR機関の事例を統合した本データセットでは、補償金額が明示されていない事例も多く含まれる。そのため、本研究では補償金額に関する情報を用いず、申立人と被申立人の主張テキストのみを入力として、和解に至る確率を予測する設定とする。

### 4.1 タスク定義

各事例 $i$ は、申立人の主張文 $A_i$ 、被申立人の主張文 $B_i$ 、および和解の有無 $y_i$ から構成される。本研究の目的は、主張文の対 $(A_i, B_i)$ を入力として、和解確率

$p_i = P(y_i = 1 | A_i, B_i)$ を推定することである。

### 4.2 モデル構成

本研究では、申立人文と被申立人文を同時に処理するモデルを採用する。モデルの全体構成を図1に示す

申立人の主張文 $A_i$ および被申立人の主張文 $B_i$ は、それぞれ日本語事前学習モデルであるLUKE[18]を用いて独立に符号化される。両者のエンコーダは重みを共有し、それぞれ768次元の文埋め込みベクトルを出力する：

$$\mathbf{h}_A = \text{LUKE}(A_i), \quad (2)$$

$$\mathbf{h}_B = \text{LUKE}(B_i). \quad (3)$$

得られた二つの文埋め込みを連結し、

$$\mathbf{z}_i = [\mathbf{h}_A; \mathbf{h}_B] \quad (4)$$

として統合表現を構成する。

### 4.3 出力層と損失関数

統合表現 $\mathbf{z}_i$ は、全結合層を通じて和解確率を出力する：

$$\hat{y}_i = \sigma(\mathbf{w}^\top \mathbf{z}_i + b), \quad (5)$$

ここで $\sigma(\cdot)$ はシグモイド関数である。

学習には二値分類問題としてのバイナリクロスエントロピー損失を用いる：

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)]. \quad (6)$$

二者間の主張テキストのみに基づくモデル構築により、異なるADR機関間の事例を共通の枠組みで分析することが可能となる。

## 5 数値実験

### 5.1 実験設定

構築したデータセットを用い、和解の有無を目的変数とする二値分類問題として数値実験を行った。各事例は申立人文と被申立人文の主張文の対から構

表 2 数値実験の結果（複数回の平均）

学習データ	評価データ	ROC-AUC	Precision	Recall	F1
全体	全体	0.78	0.71	0.69	0.70
全体	生命保険	0.73	0.67	0.65	0.66
全体	FINMAC	0.82	0.80	0.78	0.79
全体	信託	0.61	0.55	0.50	0.52
全体	貸金	0.58	0.49	0.44	0.46
生命保険	生命保険	0.74	0.69	0.67	0.68
FINMAC	FINMAC	0.79	0.76	0.74	0.75

成される。データは ADR 機関ごとの件数割合が維持されるように層化分割し、70%を訓練データ、15%を検証データ、15%をテストデータとして用いることを 5 回行った。本研究では、(1) 複数の ADR 機関の事例を統合した全体データで学習する設定と、(2) 各 ADR 機関のデータのみを用いて学習する設定を比較し、いずれも対応するテストデータに対する性能を評価した。なお、データ件数の少ない 2 機関については個別機関での学習を行わなかった。最適化手法には AdamW を使い、学習率は  $2 \times 10^{-5}$ 、バッチサイズは 4、最大エポック数は 10 とした。検証データにおける損失が最小となるエポックのモデルを最終モデルとして採用した。評価指標としては ROC-AUC、Precision、Recall、F1 を使い、クラス不均衡を考慮して F1 を主要な比較指標とした。

## 5.2 結果

テストデータにおける予測性能の一例を表 2 に示す。ここでは、5 回の試行における精度数値の平均をとっている。(1) 複数の ADR 機関の事例を統合した全体データで学習したモデルと、(2) 各 ADR 機関のデータのみを用いて学習したモデルの性能を比較する。著者の過去研究では FINMAC の補償金額等の数値情報を入力に含め、欠損のある事例を除外していたため、ROC-AUC は 0.86 と高い値が得られていたが、本研究はテキスト情報のみに基づくより制約の厳しい設定であるため、全体データで学習したモデルの全体評価における精度指標はわずかに低下した。一方で、機関別に学習したモデルと同等またはそれを上回る性能を示した。特に FINMAC 分野において、全体データを用いた学習が予測性能の向上に寄与する傾向が確認された。

## 6 考察

本研究の結果から、複数の金融 ADR 機関の事例を統合して学習を行うことで、一部の機関では予測性能が向上する一方、領域によっては性能が横ばいまたは低下する傾向が確認された。説明不足やリスク説明の不十分さに起因する紛争は、商品や業態を越えて共通に存在しており、これらに関する言語的特徴がデータ統合によって補完的に学習された可能性がある。

一方で、金融 ADR の紛争内容は機関固有の制度や商品特性に依存する場合も多く、データ統合による学習が表現の平均化を引き起こすことで、特定領域に最適化された判断規則が弱まり、予測性能が低下する可能性がある。本研究で観測された精度の増減は、共通する紛争構造の学習と機関固有性への適合との間に存在するトレードオフを反映したものと解釈できる。

## 7 まとめ

本研究では、国内金融 ADR 機関が公開する事例を整理し、申立人文、被申立人文、および和解の有無からなるテキストデータセットを構築した。さらに、二者の主張文を同時に入力とする言語モデルにより、和解判断予測が可能であることを示した。単純な構成でも機関横断となる学習の効果が確認された。今後は、他機関も含めた統合的なデータセット構築を行い、補償金額や申立人属性などの構造化情報の付与や、生成 AI を用いた主張内容の整理・分析を通じた応用が期待される。

## 参考文献

- [1] 金融庁. 金融 ADR 制度（金融分野における裁判外紛争解決制度）、（2025-01 閲覧）. <https://www.fsa.go.jp/policy/adr/>.
- [2] 一般社団法人生命保険協会. 裁定審査会が取

- り扱った事案の概要, (2025-01 閲覧). <https://www.seiho.or.jp/contact/adr/item/>.
- [3] 一般社団法人全国銀行協会. 業務の実施状況, (2025-01 閲覧). <https://www.zenginkyo.or.jp/adr/conditions/year/>.
- [4] 一般社団法人信託協会. 苦情処理手続き・紛争解決手続きの取り扱い状況, (2025-01 閲覧). <https://www.shintaku-kyokai.or.jp/consultation/complaint.html>.
- [5] 一般社団法人日本損害保険協会. 苦情・紛争解決等業務に関する実施状況等 (そんぽ ADR センター), (2025-01 閲覧). <https://www.sonpo.or.jp/report/statistics/adr/index.html>.
- [6] 一般社団法人保険オンブズマン. 主な事例 (苦情処理手続・紛争解決手続), (2025-01 閲覧). <https://www.hoken-ombs.or.jp/themaincase>.
- [7] 一般社団法人日本少額短期保険協会. 少額短期ほけん相談室レポート, (2025-01 閲覧). [https://www.shougakutanki.jp/general/consumer/consult\\_report.html](https://www.shougakutanki.jp/general/consumer/consult_report.html).
- [8] 日本貸金業協会. 各種統計資料, (2025-01 閲覧). <https://www.j-fsa.or.jp/personal/statistics/>.
- [9] 金融商品あっせん相談センター. 金融 ADR に関する統計資料, (2025-01 閲覧). <https://www.finmac.or.jp/tokei-siryu/>.
- [10] Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoțiuc-Pietro, and Vasileios Lampos. Predicting judicial decisions of the european court of human rights: A natural language processing perspective. In **Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 1138–1143, 2016.
- [11] Ilias Chalkidis, et al. Lexglue: A benchmark dataset for legal language understanding in english. **arXiv preprint arXiv:2110.00976**, 2021.
- [12] Christian Stab and Iryna Gurevych. Parsing argumentation structures in persuasive essays. **Computational Linguistics**, Vol. 43, No. 3, pp. 619–659, 2017.
- [13] Raquel Mochales and Marie-Francine Moens. Argumentation mining: The detection, classification and structure of arguments in text. In **Proceedings of the 12th International Conference on Artificial Intelligence and Law (ICAIL)**, pp. 98–107, 2009.
- [14] Junyun Cui, Xiaoyu Shen, Feiping Nie, et al. A survey on legal judgment prediction: Datasets, metrics, models and challenges, 2022. <https://arxiv.org/abs/2204.04859>.
- [15] Nasa Zata Dina, Sri Devi Ravana, and Norisma Idris. Legal judgment prediction using natural language processing and machine learning methods: A systematic literature review. **SAGE Open**, Vol. 15, , 2025.
- [16] Koutarou Tamura. Dual-input settlement prediction with word-pair shap: Financial adr cases. **IEEE International Conference on Big Data 2025 (BigData)**, 2025.
- [17] Farid Ariai, Joel Mackenzie, and Gianluca Demartini. Natural language processing for the legal domain: A survey of tasks, datasets, models and challenges, 2025. <https://arxiv.org/pdf/2410.21306>.
- [18] Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki

Takeda, and Yuji Matsumoto. Luke: Deep contextualized entity representations with entity-aware self-attention. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 6442–6454, 2020.