



フェアな合意を実現するための 契約データセット構築とその課題点

稲村和樹

MNTSQ株式会社

2022.03.18

MNTSQ株式会社 - 会社説明

トップローファーム × トップテクノロジー企業のタッグで
法務のベスト・プラクティスをアップデートする

提供
プロダクト

大企業の法務業務を一気通貫でデジタル化する SaaSプロダクト “MNTSQ”

資本業務
提携

NAGASHIMA OHNO & TSUNEMATSU

長島・大野・常松 法律事務所

リーガルリソースの提供

PKSHA

TECHNOLOGY

自然言語処理技術の包括ライセンスの提供

経営陣

代表取締役 板谷 隆平 (長島・大野・常松法律事務所 / 弁護士)
取締役 安野 貴博 堅山 耀太郎 (PKSHAの自然言語処理部門Founder / ex-CEOら)
社外役員 杉本 文秀 (長島・大野・常松法律事務所マネージング・パートナー)
社外役員 藤原 総一郎 (長島・大野・常松法律事務所パートナー)

実導入実績
(一部)

TOYOTA

KOMATSU

SMBC
三井住友銀行

MITSUBISHI
ELECTRIC

大阪ガス

三菱商事

ENEOS ENEOS

稲葉総合法律事務所
Inaba & Partners

その他東証一部上場企業 (社名非公開)

フェアな合意を実現するための契約データセット構築

日本語で契約書のデータセットを作成・公開する上では、多くのハードルがある

- 本発表ではこの点を主に概説・議論したい
- 以下では前置きとして、MNTSQ社が契約書を自動解析/計量化したいと考える背景を説明する

課題感:

世の中の契約の多くは十分フェアでなく、契約交渉や紛争解決に余計なコストがかかっている

- 自社の「スタンダードな契約交渉」を十分検討できていないことに由来するのでは？
 - 共通して参照されるスタンダード指標が不在なために契約が平準化されていない
 - 契約書に対して計算される各種契約条件の統計量(**条項の意味解析** に由来)
 - 事業制約力の強い条項有無に関する集計
 - 権利義務が定める金額、期限、持株比率等の数量条件の平均値、等
 - また具体の契約事例にあたりにいけるだけの **意味検索** インフラも未整備
 - 高品質の交渉を成立させる契約作成(ベストプラクティス)は、現実の取引を反映したスタンダード指標をふまえた上で吟味・更新されるべき

MNTSQのビジョン:

契約のスタンダード提示とベストプラクティス更新を循環させることで、契約をフェアにする

- 契約書を分析可能な形で自動解析して(スタンダードの算出)
- 各組織の法務業務従事者にベストプラクティスを届ける(よりフェアな契約の吟味)
- 契約品質の向上を通じて契約交渉・紛争解決を円滑化し、社会全体の生産性向上を企図

契約書に対するアノテーション概要

契約類型の分類

商標及び著作物使用許諾契約書

Aテクノロジー株式会社（以下、「甲」という。）と、B商事株式会社（以下、「乙」という。）とは、甲が権利を有する「〇〇〇〇」に関する商標と「〇〇〇〇」に関する著作物とに関する使用許諾契約（以下「本契約」という）を、次のとおり締結する。

第1条（定義）

本契約における用語を次の各号のとおり定義する。

- （1）「本件商標」とは、甲が権利を有する「〇〇〇〇」に関するすべての商標をいう。
- （2）「本件著作物」とは、甲が権利を有する「〇〇〇〇」に関するすべての著作物をいう。

第2条（使用許諾）

1. 甲は、乙に対し、本件商標及び本件著作物を甲が承認した商品に使用するために本契約に従って使用することを許諾するものとする。
2. 甲は、乙以外の第三者に本件商標及び本件著作物の使用を許諾してはならない。ただし、乙の承諾がある場合には、この限りでない。

第3条（競業避止義務）

1. 乙は、本契約の有効期間中、事前に甲の書面による承諾を得た場合を除き、甲と競業する第三者の保有する商標及び著作物を利用した商品を取り扱ってはならない。
2. 乙は、第三者の保有する商標及び著作物を利用した商品を取り扱おうとしている場合において、当該第三者が甲の事業と競業するか否かにつき疑義があるときは、甲に対し事前に通知し、甲の見解を求めなければならない。

第4条（許諾された使用の対価）

甲が乙に許諾することに対して、乙は次の各号にしたがってその対価を支払う。

- （1）本件商標及び本件著作物を使用する商品の税抜き希望小売価格の〇〇%を製造数量に乗じた金額を対価とする。
- （2）前号の対価は、各年度の総額を各年度末から90日以内に、乙から甲に対して支払うものとする。

第5条（販売）

1. 本件商標及び本件著作物を使用した商品の販売において、乙は甲の定める指針に従わなければならない。

契約の基本情報抽出:

契約当事者
契約締結日
契約期間
契約自動更新
取引金額
準拠法の国・地域,
etc.

条項構造解析 条項の意味解析

英文契約の公開データセット事例 CUAD

Contract Understanding Atticus Dataset (CUAD*1):

- CUAD: 有志の弁護士がアノテーションした英文契約データセット
- データソースとしては米国のEDGARをデータソースとしている
- データセットおよびアノテーションマニュアル*2は公開されている
 - transformerによるモデルベンチマークも付随

EDGAR*3について:

- 米国証券取引委員会(SEC)により管理される開示システム
- 日本の有価証券報告書開示システムであるEDINETの元となった
- 日本と違い米国では契約書提出義務があり、契約書の公開データソースが存在
 - 一方の日本は「経営上の重要な契約」の概要を記す程度の指示にとどまる
 - 金融庁の報告資料*3を読む限り、日本企業へのグローバル投資活動の促進という面で契約書データ公開が進む線は望み薄なように見える

*1 [Dan.H+, “CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review”, NeurIPS2021.]

*2 CUAD Labeling Handbook: <https://www.atticusprojectai.org/labeling-handbook>

*3 EDGAR: <https://www.sec.gov/edgar.shtm>

*4 https://www.fsa.go.jp/singi/singi_kinyu/disclose_wq/siryoku/20220119/01.pdf

日本語契約書データセット収集・アノテーション上の課題点

日本語契約書の集約・公開を阻む壁: 契約書が各企業組織内に分散・閉塞している

- データ集約ハードルに加えて、組織ガバナンス上のハードルを乗り越える必要
 - プライバシー、組織の契約ノウハウ保護、等が主論点

契約データの品質: 以下の各観点に対してカバレッジを担保することが実用上は重要

- 「契約類型」
 - 例: M&A、業務提携、ライセンス契約、研究開発、...等
 - 課題: 複雑度の高い契約ほど解析できる価値が高い一方、稀少性も高い
 - 超大企業であっても、1社内にはM&A契約は数件しかないなどざら
 - 米国EDGARには、上記の「重要な契約」が集約されている差は大きい
- 「産業上の業種ドメイン」
 - 例: 金融、保険、不動産、製造業、商社、...等
 - 課題: 主要な業種を網羅するために、多数の企業の契約書をカバーする必要がある

アノテーション作成コスト: 仕様決定時だけでなく作成時にも専門家の知見が必要

- アノテーション仕様の吟味には専門家の意見が必須(境界事例の検討等)
- 契約内容に関するアノテーション作成は、専門家以外にとって解読難易度が高い

日本語契約書データセット構築上の課題まとめ

契約解析の最終目的:

NLP/検索技術を駆使して、フェアな合意形成を支援し交渉・紛争コストを下げる

- 契約交渉のスタンダード指標の提供⇒ 指標を踏まえた契約書品質の向上
- 契約条件の指標化のために、契約書の自動解析が必須

実用的なデータセットを構築する上で乗り越える必要がある壁

- 専門家の雇用による、高コストなアノテーションデータの仕様吟味・作成
- 各組織に散逸したデータの集約、および組織ガバナンス上の妥協点探し
- カバレッジ品質を担保するデータ収集方針の策定
 - 「契約類型」や「産業上の業種ドメイン」というカバレッジ観点
 - データの希少性を克服するために多数の企業との連携が必須
 - 米国EDGARには「重要な契約」がパブリックに集約されている差は大きい

いずれの点においても、主要企業の契約をカバーするプラットフォームの存在が必須