テキストマイニングによる判例文書の解析とその結果を用いたＩＴ知財訴訟判決の要因およびＩＴ知財訴訟判決が経営に与える影響の分析

代表研究者	増　山　　繁	豊橋技術科学大学　大学院　工学研究科情報・知能工学専攻 教授
共同研究者	酒　井　浩　之	豊橋技術科学大学　大学院　工学研究科情報・知能工学専攻 助教
共同研究者	野　中　尋　史	豊橋技術科学大学　大学院　工学研究科情報・知能工学専攻 産学官連携研究員

1 はじめに

松下電器(現パナソニック)社とジャストシステム社のいわゆるアイコン特許訴訟など情報通信分野における知財紛争（以下，ＩＴ知財紛争）が注目を集めはじめている．この背景には，９０年代からのコンピュータ関連特許，特にソフトウェア特許の出願増加がある．このようにＩＴ知財紛争が頻発する中で，各企業においては，知財訴訟が経営へ及ぼす影響を見定め，過去の判決要因の分析を行い，また，原告勝訴・敗訴の確率を踏まえながら訴訟戦略の意思決定を行うことが求められている．また，学術的な立場からも，判決を左右する要因分析や判決が経営に与える影響分析を行うことは，知財法運用のメカニズムを解明し，さらにその判決が経済に及ぼす効果を明らかにすることにつながり非常に重要となる．

そのためには，過去の判例を用い，ＩＴ知財紛争，特に民事訴訟において，原告がどの程度勝訴しているか，その勝率を見積もることや，経営への影響分析（知財訴訟判決が経営指標，たとえば，株価に対し，有意に影響を与えるかなどの分析を行う必要がある．しかしながら，現状では，原告勝率のような司法統計・司法データベースの整備はなされていない.一方で,膨大な判例から人手により統計をとることもできるが,多大なコストがかかるため実施は容易ではない．さらに，これに派生した分析もほとんど行われていない.

よって，このような判例の解析を人手に依らず，計算機を用いて自動的に行うことが求められている．しかしながら，現状では，上記のような分析を自動的に行う手法は存在しない.

そこで，本研究では，テキストマイニングを用いて大規模な判例データを対象とした判例文書からの自動的な情報抽出手法の開発と IT 知財訴訟の分析[野中　2009-a]，[野中 2009-b]，[野中 2010]，さらに，計量ファイナンス論における手法を併用した経済的影響の分析を行う[野中 2011].

2 テキストマイニングを用いた判例文書からの自動的な情報抽出手法の開発と IT 知財訴訟の分析

本研究では，判例文書からの情報抽出手法として野中らの手法を用いる．特に問題となるのは，民事行政訴訟別，権利種別などの分類や，訴訟における勝訴敗訴の判定である．これらの情報は判例中において，明示的にタグ付けされているわけではなく，判例文書中の主文などを解析することによりはじめて得られる情報である．そのため，本研究では，機械学習アルゴリズムを用いた分類・判定手法を開発した．具体的には，判例文書を形態素解析（各文書を単語単位に分割すること）した上で，判例文書の構造中，適切な部分を切り出し，その部分の特徴語のみを素性（機械学習において学習する際の特徴量）としてサポートベクターマシン(以降 SVM と表記)[Boser 1992]による学習を行い,学習により作成した分類器により,民事行政訴訟別,権利種別などの分類や，訴訟における勝訴敗訴の判定を機械的に行う．ここで，機械学習アルゴリズムとして SVM を使用したのは，汎化能力が高い，すなわち，教師データに学習が過度に依存しない優秀な学習モデルの１つであることによる．機械学習を行う上で，その素性は特徴を判別するものでなくてはならない．このため，全文を対象にするよりも，分類を行うのに，よりふさわしい部分のみを選択する方が望ましい．そこで，分類の特徴に併せて適切な部分の選択を行う．民事行政訴訟別の分類を行う際は，主文，請求，事案の概要部の要旨を用いる．これは，民事訴訟の場合，主文や請求，事案の概要部の要旨には，「被告は○○を払え」などの文言が多く登場し，行政訴訟の場合は，「特許庁が下した審決を取り消す.」などの文言が出現しやすいなどの統計的偏りがあるためである．また，権利種別の分類においても同様の部分を使用する．こ

れは，上記と同様に権利種の特徴を現す語が多く出現するためである．例えば，商標の場合，「標章」など商標固有の表現が多く出現する．一方，判例において，原告が勝訴（一部認容も含む）したかどうかを判定するには，主文を用いればよい．

本研究では，「事案の概要」部中の特徴語を抽出し，それを素性として分類を行う．具体的には，以下のステップで行う．

Step 2-1:2 値（一方を正例と定義し，他方を負例と定義する.）にわけて教師データを用意する

Step 2-2:重み $Wp(t_i, S_p)$の計算（式（1））を行う．同様に負例の重み $W_n(t_i, S_p)$も計算し，その２倍超になる語を特徴語とする．以下，式の説明を行う．　□

$$W_p(t_i, S_p) = P(t_i, S_p)H(t_i, S_p) \quad (1)$$

$$P(t_i, S_p) = \frac{Tf(t_i, S_p)}{\sum_{t \in T s s_p} Tf(t, S_p)} \quad (2)$$

ただし，

$P(t_i, S_p)$: 正例の文書集合S_p における語t_i の出現確率

S_p: 訓練データにおいて正例に属する文書集合

$Tf(t_i, S_p)$: 正例の文書集合S_p に含まれる語t_i の数

T_{sSp}: 正例の文書集合S_p に含まれる語の集合

$H(t_i, S_p)$: 正例の文書集合S_p に含まれる各文書における語t_i の出現確率に基づくエントロピー

ここで，エントロピーは統計的な偏りを表す量

であり，偏りがなく満遍なく出現する語ほど大きな値となる．このため，Step 3-2 を満たすものは，正例集合に満遍なく出現し，かつ，出現頻度も大きいものであり，逆に，負例集合に出現するのは，少数の文章にしか出現しないものとなり，正例に関する特徴語といえる．上記により，正例に偏って出現する特徴語を素性として，効果的・効率的な分類を行うことができる．

正確度は，0.94 と非常に高い性能を示した．

ここでは，上記で評価した識別器を使用して知的財産民事訴訟における原告勝訴の判例を抜き出し，データベースを構築し，年度ごとの原告の勝率導出に応用した例を用いて，知財司法運用システムの考察を行う．本論文では，特に知財訴訟全体のトレンドと特許民事訴訟に焦点を当てて解説を行う．

知財民事訴訟全体について本手法を適用して得られた結果を表 1 に示す．

なお，IT 知財訴訟の分類は，「プログラム」，「ソフトウェア」の 2 語のどちらかが含まれる判例文書と定義した．

表１．結果

知財訴訟全体原告勝訴の割合	0.343	IT 知財訴訟全体原告勝訴の割合	0.415

表１より，全体の原告の勝率は，3 割程度であることが分かる．これは，訴訟になった場合，権利範囲の認定が厳しく，原告の勝率は 5 割に満たず 3 割程度であることも示している．このことは，特許権をはじめ，知的財産権は,権利取得・維持に多大なコストを要するものの，権利取得により無条件に権利侵害を防ぐものとはならないことを示唆している．よって，知財訴訟を通じて，権利侵害を防ぐ強い知財権を確立するためには，出願時の権利範囲の検討を十分に行うことが非常に重要であることが分かる．

一方，IT 知財訴訟における原告勝訴の割合は全体訴訟の原告の勝訴の割合より高い水準となっている．これは，IT 知財訴訟においては，特許権のみならず著作権訴訟が絡んでいること等が要因として考えられる．すなわち，権利範囲の認定が厳しい特許権の訴訟と比較し，著作物のデッドコピーのケースが多い（権利範囲における侵害）著作権訴訟の場合のほうが原告として勝訴しやすいと考えられるためである．

3.ＩＴ関連企業の訴訟における経済的影響の分析

本研究では，訴訟が与える経済的影響について金融工学的手法の一つであるイベントスタディ法[Mackinlay 1997]を用いることでその分析を行う．イベントスタディ法は，株式を上場している企業において，ある出来事（イベント）が生じたことによって，引き起こされる経済的影響について，株式時価総額が企業価値と等価であるという仮定に基づいて，イベント期間におけるインデックスと比較した当該企業の株価推移の異常収益の統計的検定を行うことにより分析するものである．イベントスタディ法は簡便であり，その意味が明確であることから幅広く使用されており，例えば，鈴木[鈴木 2008]は，イベントスタディ法を利用し，新薬開発が承認された企業について，そのニュースが株価に及ぼす影響を調べている．また，藤村ら[藤村 2009]は業績要因となる文をプレスリリースより抽出した上で,経常的な業績要因文と一時的な業績要因文に分類し，イベントスタディ法に基づく分析により，一時的な業績要因よりも経常的な業績要因の方が株式市場に与える影響が大きいことを明らかにしている．しかしながら，年間 1,000 件にも達する知財訴訟が起きながら，原告勝訴・敗訴の別等，イベントスタディ法を適用することで必要となる情報データベースが存在しない状況である．このため，手作業で大量の訴訟データから情報を抽出する必要があるが，知財訴訟の判決文を読解するには専門知識が求められ，かつ，1 文書あたりの文章量も多いという問題があった．本研究では，自然言語処理技術を利用して，あらかじめイベントスタディ法を適用する上で必要な情報を抽出し，データベースを構築するものである１節で述べた手法を利用した．

イベントスタディ法は，以下の手順で，イベントが株価に影響を及ぼしたかどうかを検証する．

イベントスタディ法

Step 3-1:イベントが起きなかった場合に得られたと推定される正常収益率を計算

Step 3-2:イベントが起きたことによる収益率と正常収益率との差である異常収益率を計算

Step 3-3:異常収益率（の各日における平均）が 0 である仮説を立て，棄却できるかどうかの検定　□

異常収益率の計算手法はいくつかある．この中で一番簡単で，使用頻度が高い手法がマーケットモデルを用いたものである．マーケットモデルとは，市場全体の株価収益率を説明変数とし，個別銘柄（以下では i とする）の株価収益率を被説明変数とした線形回帰モデルのことを指す．なお，回帰係数など各種推定値の推定は，イベント前の適当な期間で行う．これによる異常収益率の計算を以下の式で行う．

ここで，$AR_{i\tau}$：異常収益率，$R_{i\tau}$：イベント発生後の収益率，$R_{m\tau}$:市場全体の収益率である．

$$AR_{i\tau} = R_{i\tau} - \hat{R}_{i\tau} = R_{i\tau} - (\hat{\alpha}_i + \hat{\beta}_i R_{m\tau})$$

各日において導出した平均異常収益率(平均異常収益率)について推定期間の誤差項で標準化し統計量(漸近的に正規分布に従う）について検定を行うことで各日においてイベントの影響があったかどうか検証する．

まず,原告勝訴の判決における原告企業の株価変動を対象にイベントスタディ法を適用し,実験を行った．原告勝訴の判決における原告企業の株価変動を対象としたのは，本ケースが他のケースと比較し，経済的影響（ほぼ間違いなく企業収益拡大方向に動く）が大きいと目されることもあり，本ケースを最初の分析に選択した.結果(イベント日の営業日ベース 10 日前からのイベント日毎の標準正規分布に漸近的に従う統計量)を表２に示す．なお，イベント期間はイベント日の前後 10 日づつを選択し，マーケットモデルの導出期間はイベント日 90 日前から 60 日間で行った．対象企業は 91 社(2002 年～2008 年判決)である．

表２．各日の平均異常収益率 AR_t(いずれも有意水準 10%において統計的有意差無し)

10 日前	0.0021162	1 日後	0.0011669
9 日前	-0.0022367	2 日後	-0.004275
8 日前	-0.0021153	3 日後	0.000547
7 日前	0.0007203	4 日後	0.0001392
6 日前	-0.0001464	5 日後	0.003739
5 日前	-0.0002887	6 日後	0.0007836
4 日前	0.0003133	7 日後	0.0037545
3 日前	0.0034918	8 日後	0.001445
2 日前	-0.0010507	9 日後	-0.002618
1 日前	0.0011417	10 日後	0.000059
当日	-0.0004314		

結果より，イベント期間内それぞれの日において原告勝訴の判例が株価に影響がないことが分かった．これより，権利者は，株主利益最大化の点で，知財権を最適に運用しているかどうか疑義があることがわかる．また，法運用は，侵害者寄りであり，経済的影響を与えるほどの損害賠償額を司法側では認容していないことが分かり，少なくとも特許法 102 条 1 項新設の目的は上場企業においては必ずしも達していないことが分かった．

次に IT 業界における実験を行った．結果を表３に示す．ここでも，統計的有意な変動はイベント期間で見られなかった．これには，上述した全体の結果と一致し，業種による差異もないことを示している．

表３．IT 業界における各日の平均異常収益率 AR_t(いずれも有意水準 10%において統計的有意差無し)

10 日前	0.002312
9 日前	−0.00715
8 日前	−0.00309
7 日前	0.000195
6 日前	−0.00039
5 日前	0.001548
4 日前	−0.00025
3 日前	−0.00333
2 日前	−0.00278
1 日前	0.003025
当日	−0.00124
1 日後	0.001223
2 日後	−0.00619
3 日後	−0.00124
4 日後	0.001621
5 日後	−0.00554
6 日後	0.000453
7 日後	0.006694
8 日後	−0.00014
9 日後	−0.0013
10 日後	−0.00135

４.まとめと今後の予定

本研究では，自然言語処理を利用して判例文書より原告勝訴・敗訴の別などの有効な情報を抽出した上で，

IT 業界の訴訟の分析を行った．また，抽出した情報とイベントスタディ法を利用することにより知財訴訟が経済的な影響をもたらすかどうかの分析を行った．

前者の結果により，IT 業界は全体の訴訟よりも原告が勝訴しやすいことが分かった．一方，イベントスタディ法の結果により，IT 業界においても知財訴訟全体と同じように原告勝訴の判決における原告側企業の株価変動に対して訴訟の影響はないということが分かった．今後は，被告企業等への対象を拡大した分析を行うこと，法律別・企業規模別の分析を行うこと，及び，企業規模と損害賠償額の相対的比較等を行うことにより，より訴訟が与える経済的影響分析を詳細に行っていく．

【参考文献】

[Boser 1992] Boser, B.E., Guyon, I.M., Vapnik, V. , ”A training algorithm for optimum margin classifiers”, Proc of the fifth Annual Workshop on Computational Learning Theory, pp.144-152. ACM Press, (1992)

［鈴木　2008］鈴木公明，新薬関連イベントに対する株価反応に関する実証研究，日本知財学会第六回年次学術研究発表会要旨集 pp334-337，2008.
［藤村 2009］藤村真太郎，酒井 浩之，増山 繁，企業業績要因文の経常的か否かに基づく分類とイベントスタディ法に基づく分析，第 23 回人工知能学会全国大会，2009.

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
［2011 野中］野中尋史，酒井浩之，増山繁，知財訴訟判例文書からの情報抽出とそれを利用した知財訴訟判決が与える経済的影響の分析	第 38 回 OR 学会中部支部研究発表会	2011 年 3 月
［2010 野中］野中尋史，酒井浩之，増山繁，知財訴訟判例文書からの判例統計情報抽出と知財訴訟分析への応用	第 1 回特許情報シンポジウム	2010 年 12 月
［2009-a 野中］野中尋史，酒井浩之，増山繁，自然言語処理技術を用いた知財判例からの情報抽出と知財訴訟トレンド分析への応用	知財学会第 7 回年次学術研究発表会	2009 年 6 月
［2009-b 野中］野中尋史，酒井浩之，増山繁，テキストマイニング技術を用いた判例文書の分類および情報抽出－判例統計作成のために	情報ネットワークローレビュー，vol.8，pp.74-85	2009 年 5 月