

報道関係各位

株 式 会 社 U B I C  
(東証マザーズ 証券コード: 2158)  
(NASDAQ ティッカーシンボル: UBIC)

## 「行動情報科学」でビッグデータ解析を変革する 「バーチャルデータサイエンティスト」を発表

「行動情報科学」的アプローチで、「事実の発見」から「未来の予測」へ

米ナスダック、東証マザーズ上場で国際訴訟支援サービスを手掛ける株式会社 UBIC（本社・東京都港区、代表取締役社長・守本正宏）は、昨今産業界が注目する「ビッグデータ解析」ビジネスにおいてデータ分析を行う「データサイエンティスト」の人材不足が顕在化する中、人工知能（AI）搭載のコンピュータが分析作業を代行するソフトウェアの開発プロジェクトを始動させました。

同ツールは、弊社が国際訴訟の現場で培ってきたノウハウに自社開発のデータマイニング技術「Predictive Coding®」（プレディクティブ・コーディング）を融合することで発展させ、『行動情報科学』と命名した世界でも例のないアプローチにより、訴訟のほかにも幅広い分野でビッグデータの解析と活用を図るものです。

現在弊社がサービス展開している訴訟支援事業は、ビッグデータ解析事業の一例です。なぜなら、訴訟支援で扱うデータは、一般的な「ビッグデータ」の定義に当てはまっており、弊社はこのビッグデータを解析しているからです。つまり弊社は、会社設立以来ずっとビッグデータ解析に携わり、独自の技術とノウハウを蓄積してきているのです。

こうした長年の知見から、将来的には M&A や医療、インテリジェンス（安全保障支援）などの分野でも弊社の活躍が期待されます。現在、弊社 R&D センター内で開発を進めており、2014 年度初頭に新製品を投入する予定です。

### ビッグデータ時代の到来と「データサイエンティスト」の出現

インターネットの普及と IT 技術の進化により「ビッグデータの時代」が到来しています。いまやデータ（情報）は、企業の経営三大資源である 3M「人（Man）・物（Material）・金（Money）」と同等の価値を持つとされています。

総務省が今年 7 月に発表した「平成 25 年度 情報通信白書」によると、ビッグデータがフルに活用された場合、関連ビジネスの経済効果は現状でも年間 7 兆 7700 億円が見込めると試算されています。ビッグデータをどのようにビジネスに活用するかが将来の日本企業の競争力を左右するといっても過言ではないのです。

ところが、ビッグデータを取り扱う現場では今、大きな問題点が指摘されています。



それは、大量のデータを分析・処理する「データサイエンティスト」と呼ばれる専門家の人材不足です。

「ビッグデータの活用」とは、単に高速なハードウェアや高度なソフトウェアを利用することではありません。データを「資産」として実際の経営戦略に取り入れることを意味しており、その任務を担うのがデータサイエンティストです。マーケティングや商品開発以外でもビッグデータ解析へのニーズは急速に高まっており、今後、世界的にデータサイエンティストへの需要も高まると予想されています。しかし、米国調査会社のガートナー社によると、「2015年までに、ビッグ・データの需要は、世界的に440万人の求人をもたらすが、こうした求人はわずか3分の1しか満たされない」とされています。

こうした現状に産業界も危機感を抱き始め、今年7月には、産学連携の組織「一般社団法人データサイエンティスト協会」が発足、人材不足解消のための取り組みに乗り出しました。こうした中、弊社では従来とは異なる発想・アプローチで日本企業の「ビッグデータ」ビジネスを支援すべく、『バーチャルデータサイエンティスト』プロジェクトを立ち上げました。

### **米訴訟社会で培ったノウハウ・技術をビッグデータ解析に応用**

弊社は2010年3月、eディスカバリ（電子証拠開示）支援システム「Lit i View®」（リット・アイ・ビュー）をリリースし、国際訴訟における電子データの証拠保全・調査・分析サービスを展開してまいりました。2013年5月には、米ナスダック市場に日本企業として14年ぶりに上場しました。同年9月には、犯罪捜査で証拠となるメールやドキュメントファイルを収集・解析する次世代型フォレンジックソフトウェア「Lit i View XAMINER」（リット・アイ・ビュー・エグザミネー）を開発し、警察など法執行機関へ提供しています。

このように、弊社はeディスカバリやフォレンジック分野におけるデータ収集・解析で世界屈指の実績を持ちますが、米国の厳しい訴訟社会で蓄えたノウハウと技術は、「ビッグデータ」ビジネスにおいても十分応用できると確信しています。なぜなら、国際訴訟において膨大な電子データを解析するドキュメントレビューの作業自体がデータサイエンティストの仕事と非常に似ているからです。

これまでeディスカバリにおけるドキュメントレビューは、多数の弁護士が多額のコストをかけて目視で行ってきました。しかし、紙の時代とは異なり、電子データに含まれる情報量は桁違いです。1台のパソコンに保存された情報を紙に換算すると2トントラック4台分にも相当します。そこで弊社は、AI応用技術「Predictive Coding®」（プレディクティブ・コーディング）を自社開発し、レビュー作業に投入しました。

これは、ベテラン弁護士が電子データの中から証拠を見つけ出す際の判断や検査のパターンをAIに教え込ませることで、膨大な量の電子データに対する解析作業の大半をベテラン弁護士に代わってコンピュータに肩代わりさせるというものです。その処理速度と精度は、人によるレビュードキュメント数の4000倍以上の効率化、90%以上の精

度を実現すると同時に、従来のキーワード検索や、人によるレビューでは見落とされてしまうことのあった、重要な証拠データの確実な抽出をサポートします。現在、こうしたデータ解析に利用可能なプレディクティブ・コーディング技術を自社開発した企業は、世界でも弊社を含めて数社である、と弊社では認識しております。さらにアジア言語に完全対応しているのは弊社のみです。

### 「行動情報科学」が生み出した『バーチャルデータサイエンティスト』とは

「ビッグデータ」の定義に関して我々は、単なる「二進数」で表わされた数学的世界ではなく、人間の思考や行動がもたらす結果の集合体ととらえています。それらを分析するためのアプローチとして、新たに「行動情報科学」という概念を考案しました。

「行動情報科学」とは、「情報科学」（統計学・数学・データマイニング・パターン認識技術など）と「行動科学」（心理学・犯罪学・社会学など）を融合したものです。

従来型のアプローチが、過去のイベントを分析して「事実を抽出する」ことに留まるのに対し、我々が提案する行動情報科学では、人間の行動やコミュニティの生成パターンを通じて、より人間の思考に即して「未来を予測する」ことを可能にします。

こうした新しいアプローチによって、これまでデータサイエンティストが行ってきた作業の大部分をコンピュータで代行できると見ております。

データサイエンティストに求められる知識は①IT・情報通信、②統計学、③ビジネスの3点です。『バーチャルデータサイエンティスト』はこれらの知識を活かし、能力を発揮します。まず①に関して、さまざまな場所に散在するデータを非構造データも含めて容易に収集することができます。②に関しては、すでに十分な分析機能を持っており、操作側の人間に統計の専門知識がなくても自動的に分析方法を選択できます。③については、機械学習にプレディクティブ・コーディングを活用し、コンピュータにデータを教え込みながら解決策を探ります。

以上のように、これまで人間が行っていたデータ準備や解析等を『バーチャルデータサイエンティスト』に任せることで、データサイエンティストは、人間にしかできないコンサルティング業務、すなわち、分析結果に基づくビジネス戦略の立案等に注力できます。

このように、ビッグデータ解析における「AI化」は、将来、データサイエンティストの業務の置き換えを意味するものではなく、コンピュータが人を「アシスト」することを意味しております。よって、データサイエンティストが優秀であればあるほど、弊社の技術はより活用されると考えております。

それでは、こうした任務を遂行するために、『バーチャルデータサイエンティスト』にはどのような機能が必要なのでしょう。

まずは、ビッグデータ解析が現在直面している、以下のような技術的課題の解消を目指します。

- ① データが統合されていない（テキスト・画像・音声等メディアの違いや、データ保存先の分散）
- ② データ分析が柔軟に行えない（分類の重複や複数にまたがる分類、中間的な分類といった問題への対応が困難）
- ③ 事例横断的な分析が行えない（データの傾向の定量分析を重視する従来のデータマイニングでは、測定データの中にある程度の偏りや傾向がないと分析が困難）
- ④ データ分析に膨大な時間・人的コストがかかる

弊社はこれらの課題をクリアするため、プレディクティブ・コーディング技術を投入してまいります。

### 「ユニーク」なチーム構成が生み出す「フレキシブル」な発想

弊社の開発プロセスには、他社にはない「強み」があります。それはプロジェクトチームのメンバー構成の「ユニークさ」です。

まず、研究開発を指揮する最高責任者は「哲学」「心理学」「犯罪学」「社会学」の専門家です。また、彼の下では、大学・専門機関等の研究員として素粒子物理学の研究に従事した経験を持つ理学博士が、データ分析機能のメインロジックを構築しております。さらに、日本語・韓国語・英語・デンマーク語と日欧4カ国語に精通した、計算言語学博士のデンマーク人が言語解析を担当しています。こうした多種多様な分野におけるエキスパートが集結して初めて「行動情報科学」が確立できると考えています。

### 行動情報科学が変えるデータマイニングの世界

従来のデータマイニングでは「定量的」な関係性に基づいており、たとえばeショッピングなどでは、同一プロファイルの顧客に対して、同一商品をリコメンドするといった戦略を立てていました。これに対して、行動情報科学に基づくデータ分析では、さらに一歩踏み込むことで、顧客の購買行動に関連する背景や、顧客が属するコミュニティとの関連性、といったことまで把握することが可能になります。

さらに企業のM&Aにおいては、企業合併の前後に発生する「派閥」を把握し、将来起こりうる問題を探るとともに、コラボレーションが円滑に進むよう対策を講じるといったことが可能となります。

また、医療分野では、カルテや投薬情報、検査データを統合的に管理することで、医療費削減や医療事故の防止に寄与できると考えています。

### 「ディスカバリ専門企業」から「行動情報データ解析企業」へ

弊社では今後、様々な産業分野におけるソリューションを開発・提案し、ビッグデータの蓄積とビジネスへの活用を考えている企業・機関を支援していきます。同時に、ビッグデータ解析の専門会社等とのコラボレーションやシンクタンクと共同で各業界の市場分析等を実施することも考えています。

さらには、各種ツールの活用方法を教える講座を開催するなど、データサイエンティストの人材育成にも積極的に取り組んでまいります。

これらの活動を通して、「ディスカバリ専門企業」から「行動情報データ解析企業」へと進化し、「フューチャー・ディスカバリ（未来発見）」企業への飛躍を図ります。



ビッグデータ解析ビジネスの市場規模は、2015年には全世界で1.7兆円に達すると予測されています。こうした市場の中で弊社は、将来的に1000億円以上の売上げを目指しております。

#### 【UBICについて】

代表取締役社長：守本 正宏 東京都港区港南 2-12-23 明産高浜ビル  
URL: <http://www.ubic.co.jp/>

株式会社UBICは、国際的カルテル調査や連邦海外腐敗行為防止法（FCPA）に関連する調査、知財訴訟、PL 訴訟などで要求される電子データの証拠保全及び調査・分析を行うeディスカバリ事業（電子証拠開示支援事業）のほか、電子データ中心の調査を行なうコンピュータフォレンジック調査サービスを提供する、リーガルテクノロジー総合企業。アジア言語対応能力では世界最高水準の技術と、アジア圏最大の処理能力を有するラボを保有。2007年12月米国子会社を設立。アジア・米国双方からアジア企業関連の訴訟支援を実施。2009年末には企業内でも国際訴訟における電子証拠開示が可能な電子証拠開示支援システム「Lit i View®」（リット・アイ・ビュー）を自社開発し、2011年10月からはクラウドサービスとして「UBICリーガルクラウドサービス」の提供を開始。また、2012年3月に、アジア言語に対応した「Predictive Coding®」（プレディクティブ・コーディング）技術を世界で初めて独自開発し、実用化に成功。

2003年8月8日設立。2007年6月26日東証マザーズ上場。2013年5月16日NASDAQ上場。  
資本金1,095,358,941円（2013年06月30日現在）。

〈本件に関するお問合せ先〉

株式会社UBIC 広報部 勝部 TEL: 03-5463-6344 FAX: 03-5463-6345