大学共同利用機関法人 情報・システム研究機構
統計数理研究所

統計数理研究所
NOE（Network Of Excellence）形成事業
顧問会議 報告集

平成25年3月

N O E

統計数理研究所
NOE形成事業

統計数理研究所
NOE (Network Of Excellence) 形成事業
顧問会議 報告集

■ まえがき …… 1

■ プログラム …… 2

■ 出席名簿 …… 3

■ 会議風景 …… 4

■ 開会 …… 7

■ 概要説明

NOE 形成事業 …… 9

リスク科学 NOE …… 21

次世代シミュレーション NOE …… 37

調査科学 NOE …… 55

統計的機械学習 NOE …… 73

サービス科学 NOE …… 87

Network Of Excellence

まえがき

2012年9月24日(月)、東京會舘11階ゴールドルームにおいて「NOE(Network Of Excellence)形成事業顧問会議」が開催されました。

統計数理研究所では、2010(平成22)年度から、リスク科学、次世代シミュレーション、調査科学、統計的機械学習、サービス科学の五つの分野におけるNOE(Network Of Excellence)の形成、これらの研究分野における新しい方法論の確立及び異分野交流のハブの役割を果たすことを目的とした「NOE形成事業」に係る活動を第二期中期目標・中期計画の中で行っています。この事業推進におきましては、大所高所の視点からの様々な指導・助言を受けるべく、産学官の有識者の方々からなる顧問会議を設置しております。

2012(平成24)年1月1日の改組において、研究所内の研究施設の戦略的研究センター五つすべてをNOE型として分野間知識移転の仕組みが整ったところで、これまでの各NOE活動に係る概要説明、意見照会及び討論の機会として、このたび本会議を開催いたしました。

顧問9名中、理化学研究所脳科学総合研究センター特別顧問の甘利俊一氏、神戸大学大学院システム情報工学研究科計算科学専攻特命教授の小柳義夫氏、滋賀大学長の佐和隆光氏、東京大学名誉教授の鈴木基之氏、国際公共政策研究センター理事長の田中直毅氏、青山学院大学総合文化政策学部教授の真鍋一史氏、科学技術振興機構研究開発戦略センター長の吉川弘之氏、そして日本銀行金融研究所長の吉田知生氏の8名に御出席いただき、研究所から所長及びNOE形成事業に係る中心メンバー等11名が出席しました。統計数理研究所からは、研究所及びNOE形成事業全体、また、リスク科学、次世代シミュレーション、調査科学、統計的機械学習及びサービス科学の各NOEに関する概要説明を行い、顧問からの質疑及び意見への応答、それに係る討論が、自由な形式で展開されました。

各顧問からは、専門分野の立場から、また異分野の立場から、非常に本質的で核心をつく指摘をいただき、多岐にわたる闊達な意見交換が行われました。今後の当研究所NOE形成事業推進及び活動に係る指針が示され、会議が非常に有意義かつ貴重な機会となりましたことを、心からありがたく思っております。このことを記録にとどめるために、当日のNOE概要説明のスライドに加え、その説明内容、また各顧問からの質疑と意見、それに係る応答、意見交換をまとめ、「統計数理研究所NOE(Network Of Excellence)顧問会議報告集」を発行しました。今後、ますます重要性を増す分野間知識移転のネットワーク型研究、新しい共同研究スタイルの確立及びそれに伴う新学術領域の創成を目指し、統計数理研究所は、このNOEに係る活動をさらに活発化し、事業を推進していく所存です。

いっそうの御理解と御支援を賜りたく、どうぞよろしくお願い申し上げます。

2013年3月

統計数理研究所長

樋口　知之

NOE 形成事業 顧問会議 プログラム

日 時：2012 年 9 月 24 日（月）15：00 ～
場 所：東京會舘 11 階「ゴールドルーム」

15：00	開会	司会：椿　　広計 （副所長 /NOE 推進室長）
15：00 ～ 15：05	所長挨拶	所長　樋口　知之
15：05 ～ 15：15	出席者紹介	
15：15 ～ 15：25	NOE 形成事業 概要説明	所長　樋口　知之
15：25 ～ 15：40	リスク科学 NOE 概要説明	リスク解析戦略研究センター センター長　椿　　広計
15：40 ～ 15：55	次世代シミュレーション NOE 概要説明	データ同化研究開発センター センター長　樋口　知之
15：55 ～ 16：10	調査科学 NOE 概要説明	調査科学研究センター センター長　吉野　諒三
16：10 ～ 16：25	休　憩（15 分）	
16：25 ～ 16：40	統計的機械学習 NOE 概要説明	統計的機械学習研究センター センター長　福水　健次
16：40 ～ 16：55	サービス科学 NOE 概要説明	サービス科学研究センター センター長　丸山　　宏
16：55 ～ 17：45	自由討論	
17：45 ～ 17：50	御礼挨拶	所長　樋口　知之
18：00 ～ 20：00	意見交換	

出席名簿

NOE 形成事業 顧問

甘利　俊一	理化学研究所 脳科学総合研究センター 特別顧問	出 席
小柳　義夫	神戸大学大学院 システム情報学研究科計算科学専攻 特命教授	出 席
佐和　隆光	滋賀大学 学長	出 席
鈴木　基之	東京大学 名誉教授	出 席
田中　直毅	国際公共政策研究センター 理事長	出 席
手代木　功	日本製薬工業協会 会長	–
真鍋　一史	青山学院大学 総合文化政策学部 教授	出 席
吉川　弘之	科学技術振興機構 研究開発戦略センター センター長	出 席
吉田　知生	日本銀行金融研究所 所長	出 席

敬称略五十音順

NOE 形成事業 運営委員会委員

樋口　知之（委員長）	所長 / データ同化研究開発センター センター長 / 教授	出 席
椿　　広計（副委員長）	副所長 / NOE 推進室長 / リスク解析戦略研究センター センター長 / 教授	出 席
吉野　諒三	調査科学研究センター センター長 / 教授	出 席
福水　健次	統計的機械学習研究センター センター長 / 教授	出 席
丸山　　宏	副所長 / サービス科学研究センター センター長 / 教授	出 席
田村　義保	副所長 / データ同化研究開発センター 副センター長 / 教授	出 席
山下　智志	リスク解析戦略研究センター 副センター長 / 教授	出 席
中村　　隆	データ科学研究系 研究主幹 / 調査科学研究センター 教授	出 席
松井　知子	モデリング研究系 研究主幹 / 統計的機械学習研究センター 副センター長 / 教授	–

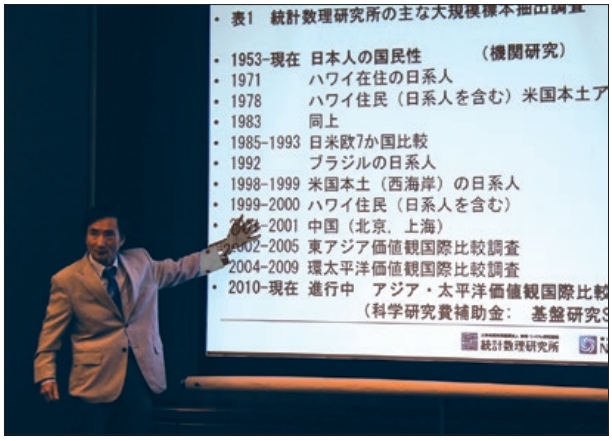
陪 席

栗木　　哲	数理・推論研究系 研究主幹 / 教授
中野　純司	統計科学技術センター センター長 / 統計思考院 院長 / 教授
内山　　亮	共通事務センター センター長

会議風景

樋口 知之 統計数理研究所長

椿　広計 リスク解析戦略研究センター長



吉野　諒三 調査科学研究センター長

福水　健次 統計的機械学習研究センター長

丸山　宏 サービス科学研究センター長

会場風景

出席した８名の顧問

甘利　俊一 顧問

小柳　義夫 顧問

佐和　隆光 顧問

鈴木　基之 顧問

田中　直毅 顧問

真鍋　一史 顧問

吉川　弘之 顧問

吉田　知生 顧問

NOE 形成事業運営委員会委員他

【開会】

○ **司会**

それでは、定刻となりましたので、これより、統計数理研究所 NOE 形成事業顧問会議を開催したいと思います。本日は御多用のところ、また、遠路はるばる本会議に御出席を賜り、誠にありがとうございます。司会進行を務めます統計数理研究所副所長、NOE 推進室長の椿広計でございます。本日はどうぞよろしくお願いいたします。

開会に先立ちまして、統計数理研究所長の樋口知之より御挨拶申し上げます。

○ **樋口所長**

統計数理研究所の樋口でございます。私、昨年の４月から統計数理研究所の所長を務めております。現在、情報・システム研究機構、機構長の北川先生が前所長でした。北川先生の後を受けまして、昨年の４月からは所長を務めております。初めて御挨拶させていただきます先生もいらっしゃいますので、今後とも御指導・御鞭撻をよろしくお願いいたします。

今日はこの後色々詳しく御説明させていただきますが、NOE（Network Of Excellence）事業はまだ始まったばかりです。現在どのような形で進められているのか、また、どのようなことを考えて行っているのかということを御説明させていただきたいと思っております。

非常に重々しい雰囲気ですけれども、評価のための会議ではありませんので、先生方、アドバイス、あるいは期待感等を、御自由に御発言いただければと思っております。本来ならば、このような会議は、統計数理研究所が研究活動を行っております立川で開催するべきだと思いますけれども、ここから立川まで、片道１時間ちょっと、往復では２時間ちょっとかかるということで、時間の有効利用ということからこの会議場を借りての開催となりました。

短い時間でも、我々の今考えていることを先生方に紹介させていただき、色々御意見を頂戴したいと思っております。

これから、各センターから紹介させていただきますが、それらに対しまして、先生方、どうぞ御自由に、アドバイス、期待感等々御発言いただければと思います。どうぞ今日はよろしくお願いいたします。

○ **司会**

それでは、ここで御出席の顧問の先生方を、五十音順でございますが、順に紹介させていただきます。先生方におかれましては、御着席のままでけっこうでございます。

理化学研究所　脳科学総合研究センター　特別顧問　甘利俊一先生
神戸大学大学院　システム情報学研究科計算科学専攻　特命教授　小柳義夫先生
滋賀大学　学長　佐和隆光先生
東京大学　名誉教授　鈴木基之先生
国際公共政策研究センター　理事長　田中直毅先生
青山学院大学　総合文化政策学部　教授　真鍋一史先生

科学技術振興機構　研究開発戦略センター　センター長　吉川弘之先生
日本銀行金融研究所　所長　吉田知生先生

本日は、顧問９名中８名の先生方に御出席賜りましたこと、統計数理研究所一同、心からありがたく存じております。

続きまして統計数理研究所からの出席者、陪席者からの自己紹介をさせていただきます。それでは、吉野調査科学研究センター長より順番にお願いいたします。

（吉野調査科学研究センター長ほか、出席者・陪席者が自己紹介）

○　司会

本日の議事は、お手元の議事次第にございますとおり、NOE 形成事業全体及び NOE 概要説明の後、討論という形で進めてまいりたいと思っておりますので、よろしくお願いいたします。

この今回の顧問会議に関しましては、顧問の先生方から御意見をいただける大変貴重な機会と考えております。今後の NOE 形成事業推進の指針とすべく、報告書を作成させていただく予定でございます。併せて御承知おきいただければ幸いです。また、この記録のために、本日写真撮影などをさせていただきますので、何卒御了解いただければ幸いです。

Network Of Excellence

統計数理研究所　NOE形成事業 顧問会議

NOE形成事業

概要説明

統計数理研究所
所長　樋口　知之

統 計 数 理 研 究 所
NOE形成事業概要説明

統 計 数 理 研 究 所
NOE形成事業

統計数理研究所
所長　樋口　知之

大学共同利用機関法人

- 人間文化研究機構
- 自然科学研究機構
- 高エネルギー加速器研究機構
- 情報・システム研究機構
  - 統計数理研究所
  - 国立情報学研究所
  - 国立遺伝学研究所
  - 国立極地研究所
  - 新領域融合センター
  - ライフサイエンス統合DBセンター

設置目的・沿革・活動

設置目的

○ 統計数理に関する総合研究

沿 革

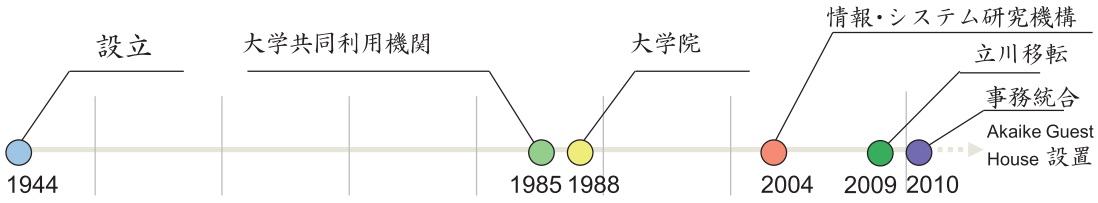
- ●1944年：文部省直轄の研究所として設立
- ●1985年：大学共同利用機関に改組転換
- ●1988年：総合研究大学院大学創設
- ●2004年：法人化，機構化
- ●2009年：立川移転
- ●2010年　極地研と事務統合，Akaike Guest House設置

主要活動

○ 研究活動
- ・我が国における統計数理の中核拠点
- ・先端的な研究を推進

○ 共同利用
- ・多様な分野との共同研究
- ・スパコン，ソフトウェア，乱数

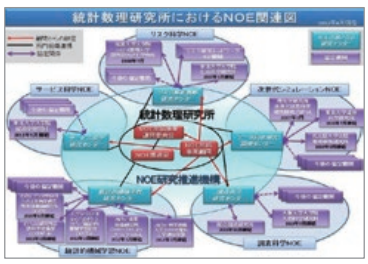
○ 人材養成
- ・総研大における大学院教育
- ・公開講座



中期目標・中期計画

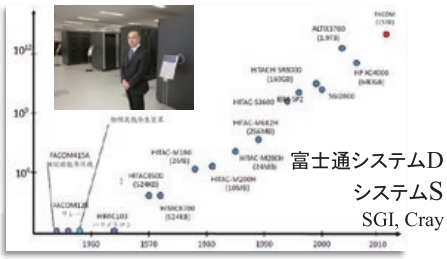
1. 大規模・大量データ（ビッグデータ）を活用する研究方法（**データ中心科学**）の確立と実践
2. 統計数理を中核とするNOE型共同研究システム（Network Of Excellence）の確立と統合リスク科学，次世代シミュレーション科学，調査科学，統計的機械学習，サービス科学などの新分野推進
3. 統計思考力を備えたT型人材育成による融合研究の推進
4. 世界最先端の統計計算基盤（クラウド）の構築と提供

事業① NOE形成事業



事業② 統計思考力育成事業

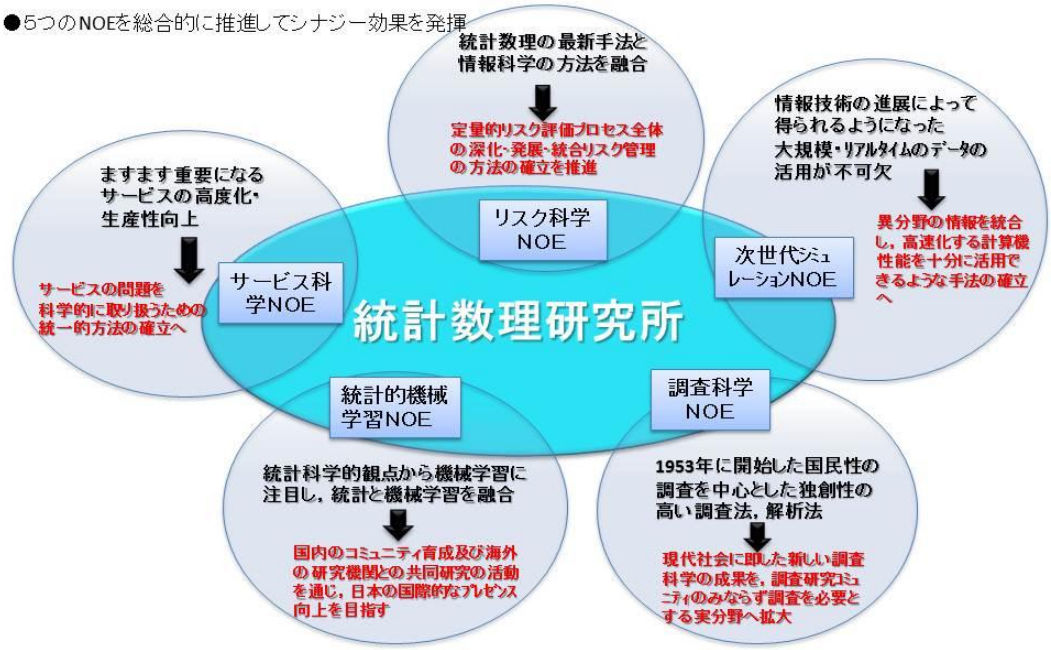
計算リソースの提供



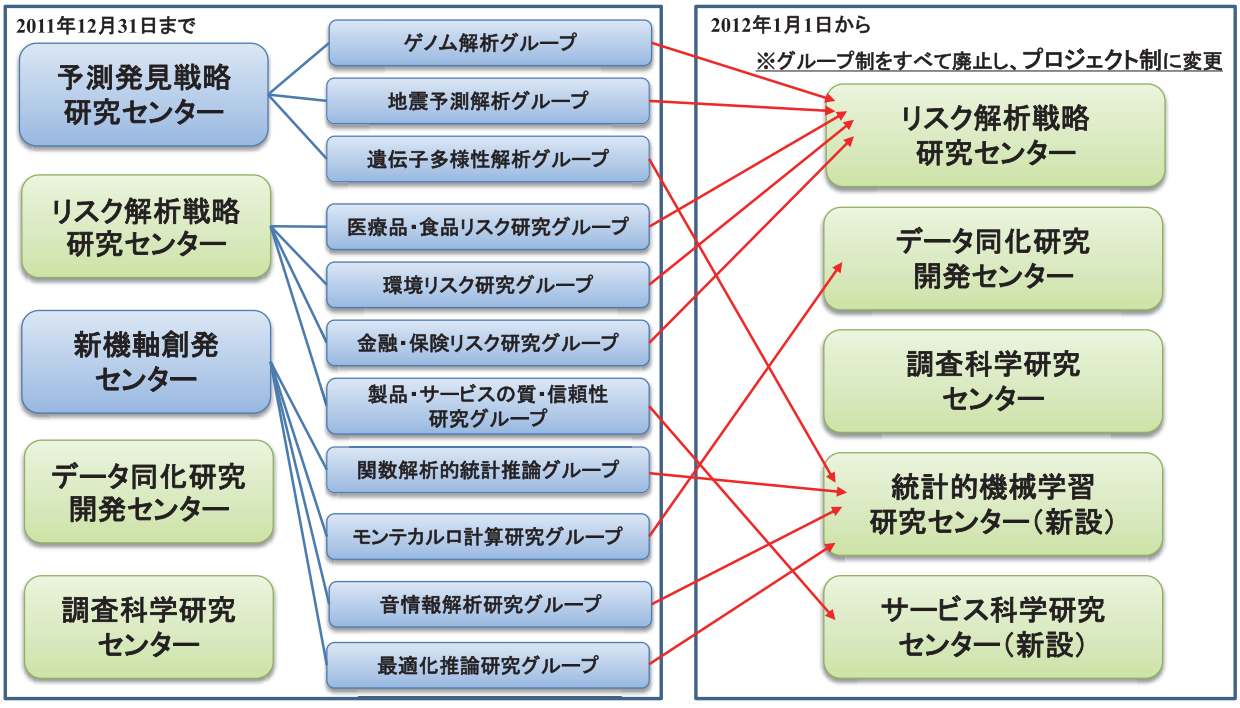
統計数理ＮＯＥ形成事業

コミュニティから支持。共同利用機関として大学が求める要望を反映した拠点。

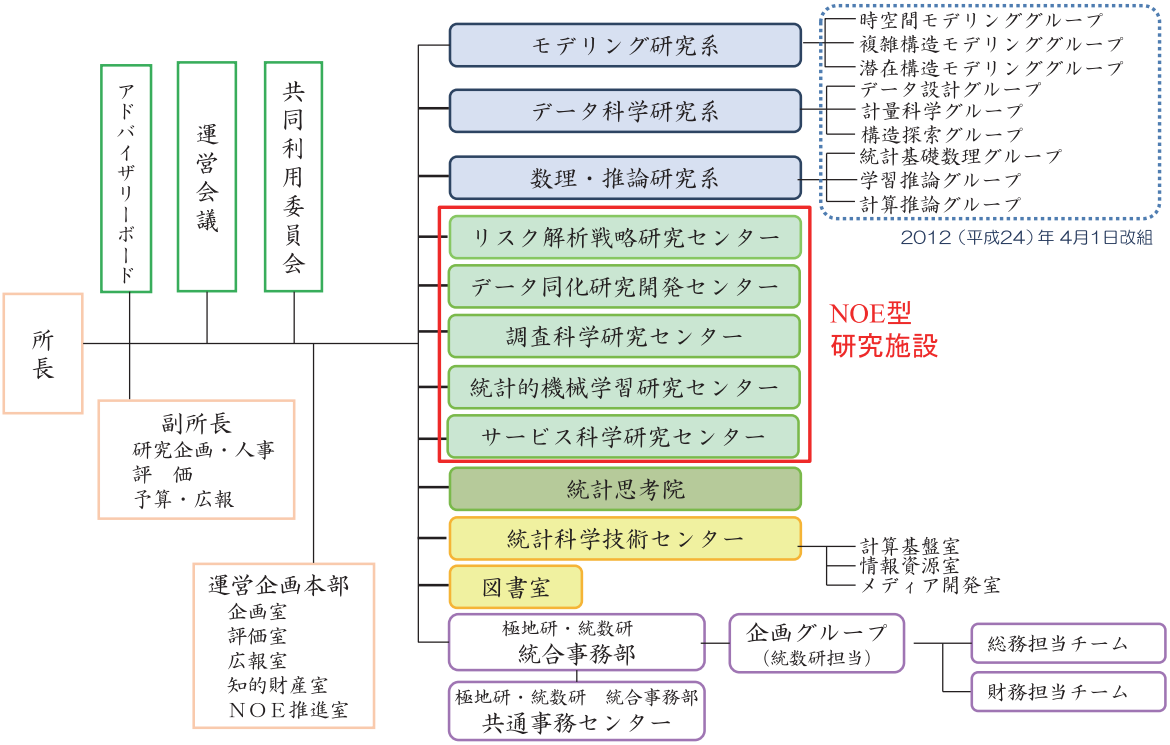
●現代社会の複雑な問題の解決には，大規模データの活用と融合研究の実現が不可欠
●大規模データの利用技術を確立しつつある統計数理が中心となる分野横断型のNOE構築が必要
●5つのNOEを総合的に推進してシナジー効果を発揮



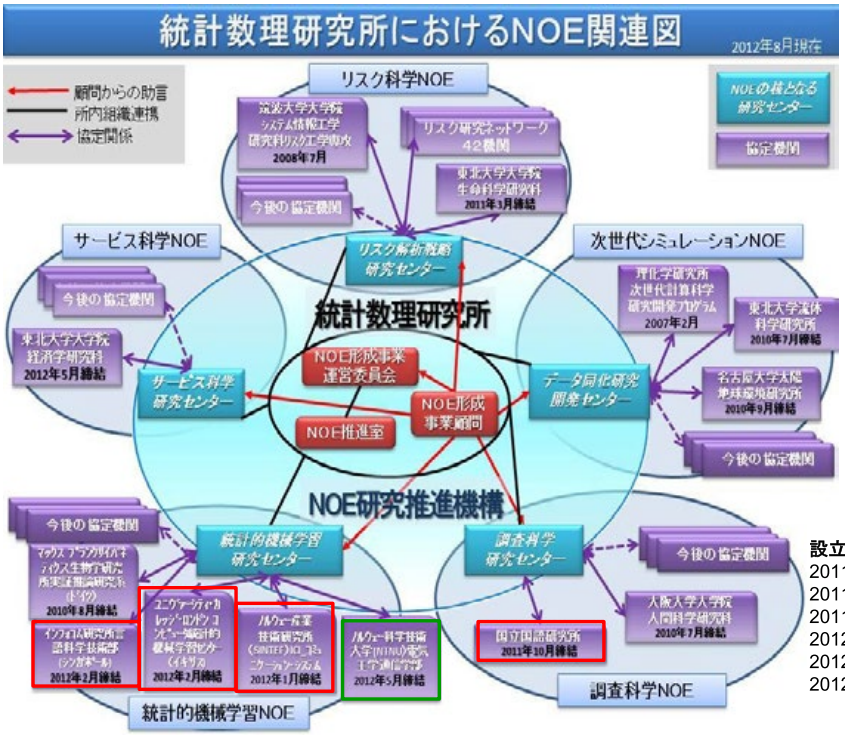
戦略的研究組織の改組



組織構成



統計数理ＮＯＥ形成事業　運営体制



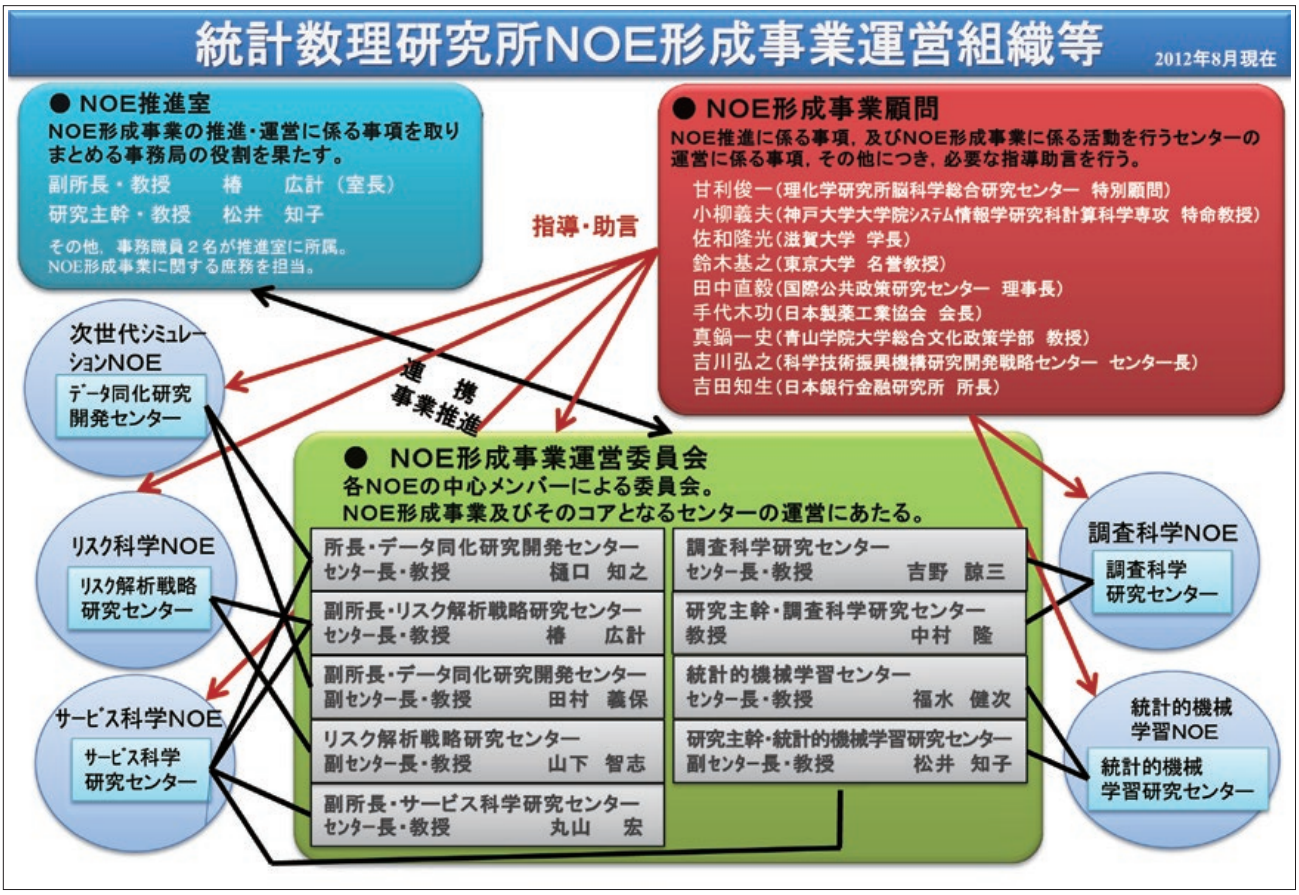
2011年度MOA締結
国内機関：1機関　海外機関：3機関

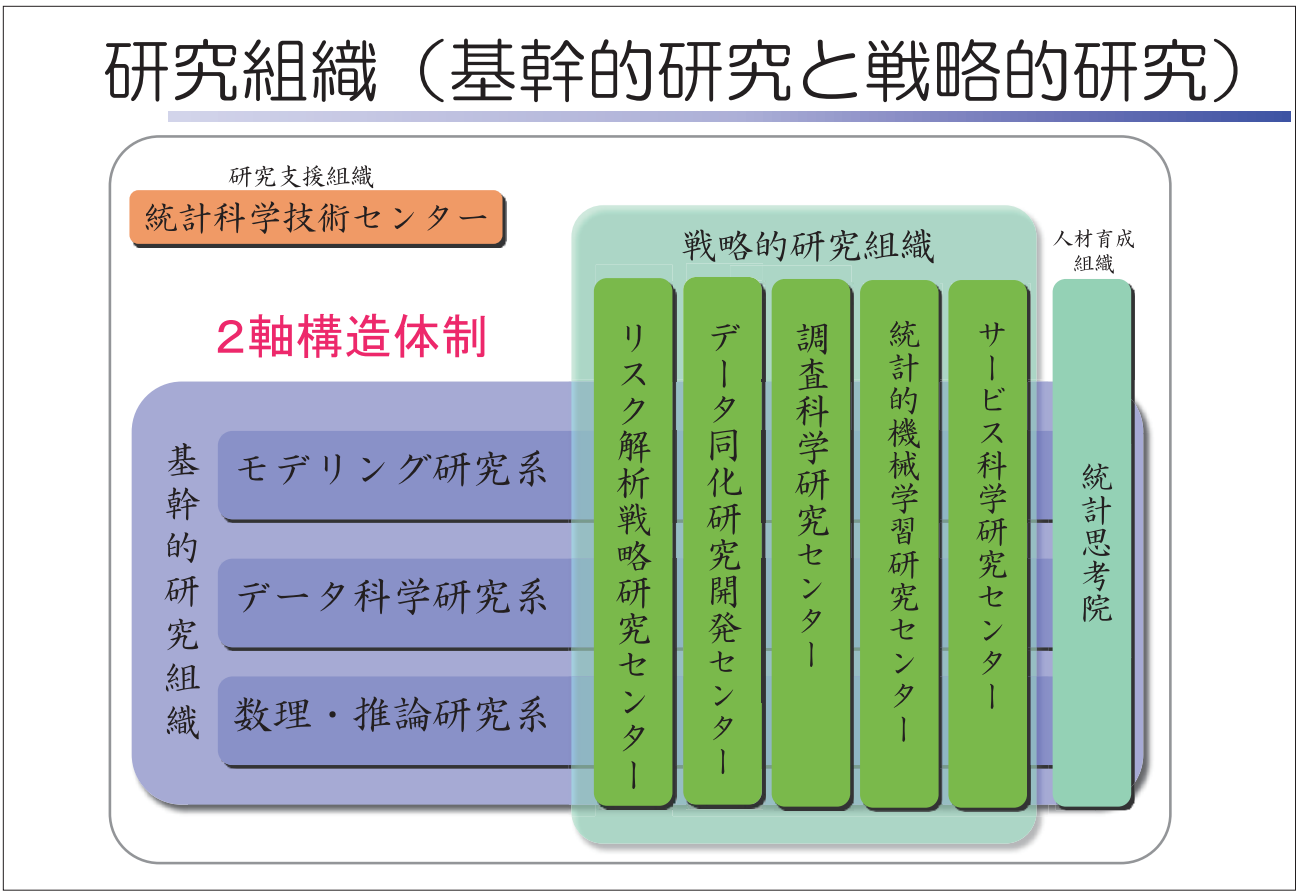
2012年度MOA締結
海外機関：1機関
（秋以降，海外2機関との締結内定済み）

NOE研究協力協定締結機関
国内機関：8機関
リスク研究ネットワーク：42機関
海外機関：5機関　　（2012年8月現在）

http://noe.ism.ac.jp/

設立記念シンポジウム等
2011.03.11 データ同化研究開発センター設立記念シンポ
2011.06.13 リスク研究ネットワーク総会
2011.06.27 調査科学研究センター設立記念シンポ
2012.01.30 サービス科学研究センター設立記念シンポ
2012.02.23 統計的機械学習研究センター設立記念WS
2012.03.18 融合研究ワークショップ2012
（流体科学研究所 流体融合研究センター
統計数理研究所 データ同化研究開発センター共催）


統計数理研究所NOE形成事業運営組織等
2012年8月現在
● NOE推進室
NOE形成事業の推進・運営に係る事項を取りまとめる事務局の役割を果たす。
副所長・教授　椿　広計（室長）
研究主幹・教授　松井　知子
その他，事務職員２名が推進室に所属。
NOE形成事業に関する庶務を担当。
指導・助言
● NOE形成事業顧問
NOE推進に係る事項，及びNOE形成事業に係る活動を行うセンターの運営に係る事項，その他につき，必要な指導助言を行う。
甘利俊一（理化学研究所脳科学総合研究センター　特別顧問）
小柳義夫（神戸大学大学院システム情報学研究科計算科学専攻　特命教授）
佐和隆光（滋賀大学　学長）
鈴木基之（東京大学　名誉教授）
田中直毅（国際公共政策研究センター　理事長）
手代木功（日本製薬工業協会　会長）
真鍋一史（青山学院大学総合文化政策学部　教授）
吉川弘之（科学技術振興機構研究開発戦略センター　センター長）
吉田知生（日本銀行金融研究所　所長）
次世代シミュレーションNOE
データ同化研究開発センター
連携事業推進
リスク科学NOE
リスク解析戦略研究センター
サービス科学NOE
サービス科学研究センター
● NOE形成事業運営委員会
各NOEの中心メンバーによる委員会。
NOE形成事業及びそのコアとなるセンターの運営にあたる。
所長・データ同化研究開発センター　センター長・教授　樋口　知之
副所長・リスク解析戦略研究センター　センター長・教授　椿　広計
副所長・データ同化研究開発センター　副センター長・教授　田村　義保
リスク解析戦略研究センター　副センター長・教授　山下　智志
副所長・サービス科学研究センター　センター長・教授　丸山　宏
調査科学研究センター　センター長・教授　吉野　諒三
研究主幹・調査科学研究センター　教授　中村　隆
統計的機械学習センター　センター長・教授　福水　健次
研究主幹・統計的機械学習研究センター　副センター長・教授　松井　知子
調査科学NOE
調査科学研究センター
統計的機械学習NOE
統計的機械学習研究センター


研究組織（基幹的研究と戦略的研究）
研究支援組織
統計科学技術センター
2軸構造体制
基幹的研究組織
モデリング研究系
データ科学研究系
数理・推論研究系
戦略的研究組織
リスク解析戦略研究センター
データ同化研究開発センター
調査科学研究センター
統計的機械学習研究センター
サービス科学研究センター
人材育成組織
統計思考院

【NOE 形成事業概要説明】

○ **司会**

それでは、これより NOE 形成事業及び各 NOE の概要説明に入らせていただきます。

はじめに、所長の樋口より、当統計数理研究所及び NOE 形成事業についての概要説明をいたします。これに関するスライドを映写いたしますので、お手元の資料と併せて御覧いただければ幸いです。

なお、これは貴重な意見照会の場ですので、顧問の先生方におかれましては、もし説明の途中でも、質疑・御意見等ありましたら、御発言いただいてけっこうでございますので、よろしくお願いいたします。

○ **樋口所長**

私から、NOE 形成事業の全体につきまして説明させていただきます。

１ページ目のここにありますのは、作成しました NOE 形成事業のロゴです。五つの分野が融合して、周りの研究者コミュニティの方々を巻き込んで行っていくイメージを表しているものです。

まず、NOE 形成事業の御説明に入る前に、統計数理研究所について、少しだけ御説明させていただきたいと思います。

情報・システム研究機構は、統計数理研究所、国立情報学研究所、国立遺伝学研究所、国立極地研究所の四つの国立研究所で構成され、統計数理研究所は情報・システム研究機構の中にあって活動しております。

国の大学共同利用機関法人、これは、ステークホルダーとしましては、大学の先生方、大学の学生の方々、あるいは社会の方々、そういった方々になります。

国には、この機関法人が 17、附属施設も含めますと 19 ありますが、それが四つの機構に分かれております。情報・システム研究機構は、この統計数理研究所、あるいは国立情報学研究所という分野横断的な研究を中心とする２研究所と、生命、ライフサイエンスを専門分野とする国立遺伝学研究所及び環境、地球環境等々を専門とする国立極地研究所の応用分野が明確な２研究所といった、縦と横が組み合わさった、他の機構等にはない、非常におもしろい構成になっております。

統計数理研究所は、非常に歴史は古く、1944 年、終戦ちょっと前に設立されました。ここに歴史が書いてありますが、大きく分けて、三つに分けることができると思います。

まず設立の 1944 年から 1985 年まで。この間には、統計数理研究所は、国の、すなわち文部省の直轄の研究所として活動して参りました。1985 年に先ほど申し上げました大学共同利用機関に改組されました。ここまでが第一期と言えると思います。

2004 年の大学の法人化とともに研究所は情報・システム研究機構に参画しましたが、その 2004 年までを第二期ということができると思います。この間の非常に大きなイベントとしましては、黄色く書いてありますが、1988 年に総合研究大学院大学に統計科学専攻として、当初から参画したことが挙げられます。今日はあまり触れませんけれども、日本には、統計科学を専門とする学科は統計数理研究所以外にないという、日本は先進諸国において

は稀有な存在ですが、1988 年に初めて統計科学専攻が日本にできました。

2004 年以降の法人化以降は、大きなイベントとしましては、2009 年の立川移転があります。だいたい移転して３年になります。先ほども申しました通り都心から片道約１時間ちょっとかかりますけれども、郊外型の研究所として、ゲストハウスを作る等色々行いまして、多くの方々にこれまで同様に研究活動に参画してもらっております。

統計数理研究所の中期計画・中期目標は、情報・システム研究機構に含まれる形になるわけですが、これが統計数理研究所の部分の主たるものになります。大きく四つにまとめることができます。

まず、大きな研究活動の目標は、昨今の言葉でいうと「ビッグデータ」、大規模・大量データを活用する研究方法の推進を掲げています。この研究方法は、データ中心科学とも最近呼ばれますが、その確立と実践を研究所の基礎研究の大きな目標としております。

事業におきましては、今日御紹介いたします、Network Of Excellence 形成事業、もう一つ、統計思考力育成事業を主に行っています。先ほど私、申し上げました通り、日本におきましては、統計科学の専攻がこの統計数理研究所しかありませんので、統計数理に係る人材育成をこの統計思考力育成事業で行っております。この二つが大きな事業です。

あともう一つ、施設としましてスーパーコンピュータを持っております。

研究所の研究活動は、大きくこれらの四つに分けてまとめることができます。

さて、Network Of Excellence 形成事業ですが、統計数理というのは、分野横断的な学際性の高い学問という特性があります。ということは、その応用分野を特定するわけではなく、例えばリスク科学を取り上げますと、色々な分野でリスク科学研究がされておりますが、その共通するような部分というのを取り出し、まとめて研究することによって、そこで得られたものを各分野に還元していくという活動が重要であることは論を待ちません。前所長の北川先生が、まずはリスク科学 NOE というのを構成するのが重要だということで、NOE 形成事業に関する活動を始めました。だいたい６～７年前だと思います。

本研究所は、常勤職員が 50 人弱、つまり教授、准教授、助教の総数が 50 人弱の小さい研究所ですので、色々な NOE が考えられますが、この五つの分野の NOE 形成を行うことにいたしました。リスク以外は、次世代シミュレーション、調査科学で、これらの NOE 形成のために中心たる研究センターを昨年（2011 年）の１月１日に設立しました。また残りの二つの分野は、統計的機械学習とサービス科学で、これらの NOE 形成促進のために今年（2012 年）の１月１日に研究センターを立ち上げました。この五つの分野を中核として NOE 形成事業を進めております。

さて、運営をどのような組織体制で行っているかを次に説明します。先ほど申し上げましたように、リスク科学 NOE を進めるために、リスク解析戦略研究センターを６～７年前に設立いたしました。次世代シミュレーション及び調査科学 NOE 形成事業を進めるために、昨年の１月１日にデータ同化研究開発センターと調査科学研究センターを設立いたしました。それまでの戦略センターとして、この青色で書いております予測発見戦略研究センターと新機軸創発センターがありましたが、今年の１月にこの２センターを廃止しまして、新しく統計的機械学習研究センターとサービス科学研究センターを設立いたしました。センター内におきましては、これまで色々グループがあったのですが、それを組み替えることによって新しい体制を整えました。

また、このセンターを作るに当たりまして、これまであったグループ制というのを廃止し、プロジェクト制にしました。それは二つ目的があります。一つは、例えば外部資金を獲得したときにそのプロジェクトをやりやすいようにするということと、また、プロジェクト制にすることによって、萌芽的なボトムアップのプロジェクトもすくい上げていくことができます。この二つを目的としまして、プロジェクト制といたしました。

これが研究所の組織構造になっております。常勤研究職員はこの青で書いております三つの研究系のどれかに必ず入っております。NOE 型研究センターは 5 センターありますが、これは、常勤職員が兼務で入っております。また、今日は説明いたしませんが、統計思考力人材育成事業を中核的に進める組織として、統計思考院が昨年の 11 月に発足いたしました。詳しくはお手元の茶色いパンフレットにこの思考院の解説があります。

これが NOE 形成事業を行っている運営体制です。真ん中に色々な事務的なこと等を詰める運営委員会、また、今日いらしている先生方で構成される NOE 形成事業顧問、あと教員、事務がサポートする NOE 推進室があります。この真ん中の部分がこの事業を担っている中核となっています。その周りに五つの NOE が示してあり、その NOE の真ん中に各センターがあります。右下は各センターが設立された日付と、その設立に合わせまして開催しましたワークショップの開催日が示してあります。

これが、さらに真ん中の部分を詳しく描いたものです。真ん中は、予算を決めたり、年度計画を立てたり等を行います NOE 形成事業運営委員会になります。右側は、今日の会議に集まっていただいていますメンバーが大所高所の立場からアドバイス等々を行います NOE 形成事業顧問に、また左上が、色々なたたき台等の案を作る推進室になります。これが事務作業準備を中心となって進めております。

研究所の研究活動の再掲になりますけれども、教員はこの横軸となります基幹的研究組織の三つの研究系のどれかに必ず入っております。縦の NOE 型組織の五つのセンターには兼務として常勤職員は参加します。ですから、教員によっては、例えば、二つの NOE 型センターに入っていることもあり得ます。各々自分の興味の持つところ、あるいは得意とするところで、NOE 型研究組織に入るようになっています。

以上が、NOE 形成事業のだいたいの説明になります。

一つ一つのセンターの活動につきましては、各センター長から説明していくことになります。以上です。

○　司会

ただいまの総括的な説明に関しまして、もし御質問、御意見等ございましたら、顧問の先生方に御発言いただきたいと思います。いかがでしょうか。

佐和先生。

○　佐和顧問

ちょっと関心があるので、お尋ねしたいのですが、研究所が 1944 年に創設されたということですが、終戦の前の年に、どういう目的で統計数理研究所が設立されたのでしょうか。

○ **樋口所長**

どうしてそのタイミングなのか正確には私もわかりませんけれども、それより前に、日本において統計の研究所を作るべきであるということで、年数は失念しましたけれども、色々な学術の方で動きがあり、建議もあって、それで、タイミング的にその時期となったようですが、なぜ本当に終戦直前なのかはわかりません。

○ **佐和顧問**

戦中のアメリカで「科学動員」という言葉がありました。要するに、数学のできる人には、兵隊にするよりも、オペレーションズ・リサーチを研究させた方がよいということで、戦後、数理統計科学者や経済学者として大成された方々が動員されたのです。経済学者のミルトン・フリードマンの名をご存じかと思いますが、そのフリードマンが Sequential Analysis のパイオニアリングな仕事を、科学動員されている最中にやっていたそうです。日本でも、科学的戦略論というか、オペレーションズ・リサーチを研究する場所を作ったらどうかという意図があったのかなと思ったのですが、いかがでしょうか。

○ **樋口所長**

断定するまではできないのですが、そういう流れが水面下であり、学術的にも 10 年来の建議等々があってというふうに記録には残っています。

○ **司会**

いかがでしょうか。

甘利先生。

○ **甘利顧問**

大学院が 1988 年にできたということになっていますが、今、言われた統計科学の基礎をやるのと同時に NOE の事業をやるという動きがあるとすると、もっと基本的な学生を教育するというものとの整合性を知りたい。現状は本当にうまくいっているのかどうか、もっと平たく言うと、例えば生物系の大学院、たぶん他もそうでしょうけれど、手助けをしてくれる学生がほしくて仕方がないところが多い。しかし、統数研は大学ではないし、そういうことはないと思うのです。相当の意欲を持って教育しないと新しい人材は育たない。むしろ、それは次世代のためにやっているということになるのだと思います。そういう整合性、大学院の位置付けをどうするかというのをちゃんと考えておかないと、だんだんちぐはぐになっていくような気がするのです。現状は大学院学生はどれくらいいるのですか。

○ **樋口所長**

現状、５年の博士課程も始めました。５年の博士課程の定員が２人で、後期３年の博士課程の定員が３人、合計５人になります。それは定員であって、実際は定員の２倍弱ぐらい、入学を認めています。全５学年、５年を超えた学生も入れると 30 人くらいが在籍しています。先生の御指摘はもちろんですが、ステークホルダーとしてはまずは大学の先生方、次は大学院生、というところで、大学共同利用機関法人の第一義的なミッションとしての

研究に集中したいという気持ちがあります。これと、教育の基礎的な部分をどの程度しっかり負担していくかというところで、この二つをどうすり合わせていくのかがなかなか難しいです。これは私感ですが、この 10 年間でだいぶ雰囲気が変わってきたのではないかと思っています。どういうことかというと、社会全体の価値を生むものが、目に見える「もの」というより、もう少し抽象化された、システムとか数理的なものに価値を生む、というふうに社会全体が変わってきているような気がします。そうしますと、企業等々、中央研究所等々も比較的、統計数理研究所の目指す方向に共感し、統計数理を理解したという学生が、社会人学生ですけれども、増えてきているような気がします。現在、統計科学専攻には社会人学生がけっこう多いのですが、そういう期待に答える意味で、社会人学生を受け入れ、教育し、送り出す観点は、だいぶうまくいっているのではないかと思います。

○ 司会

いかがでしょうか。鈴木先生お願いいたします。

○ 鈴木顧問

たぶん、あとで個別の五つのセンターについては、色々お話が出ると思うのですが、最初に伺っておきたいのは、このスライドの３枚目のあたり、横軸が研究所の基幹的な研究組織で、そして、それに乗って、五つのセンターが存在する。横軸は基幹的ということですが、これまでずっと継続して活動してきている研究室ということでしょうが、この戦略的研究組織の五つはどうやって現れたのか、社会的なニーズとどのように関わっているのか、あるいは、予算の配分等々による何か動機づけがあるのか。御自分たちがおやりになりたいというだけのセンターではたぶんないのだろうと思うのですが、この五つをどう戦略的に選ぶ必要があったのか、それはどういうふうに体系づけられているのか、個別のセンターが何をしておられるかのお話はこの後伺うとして、全体としてこの５本立てはどういう哲学で、テーマが選ばれているのか、その辺のお話をまず伺っておきたいと思います。

○ 樋口所長

御質問、どうもありがとうございます。このリスク解析戦略研究センターは、６～７年前に、色々な分野でリスク解析という視点で研究するのが重要であるという北川先生のイニシアティブで設立されました。

その後、データ同化研究開発センターと調査科学研究センターが設立されました。

この調査科学研究センターは、所属するメンバーを中心に５年に１回、日本人の国民性調査をやっておりまして、以前は事業的にきっちりと文部科学省からもサポートしていただいていたという経緯もありますが、事業的にこれからもしっかりと継続していくことが必要であるということもあり、事業主体を研究系から切り離してセンターに位置付けることによって、よりビジビリティが高まることも狙い、センターにいたしました。このセンター活動に関する補足的説明はまたあとで吉野センター長等からしていただければと思います。

統計的推論が逆問題的にある中で、シミュレーションというのは前向き推論、つまりフォワードタイプですので、その両者を組み合わせるというのが色々な場面で重要になって

きているということ、また、統計数理クラウドの開発を概算要求として文部科学省にしまして、残念ながら限定的な予算しかいただけませんでしたが、予算をいただきました。この2点により、データ同化研究開発センターを設立いたしました。

先ほどビッグデータというのがありましたが、大規模データ解析に必須の統計数理の研究分野をより強化し、さらには従前のオペレーションズ・リサーチ領域との連携を機械学習の観点から強化するのは統計数理研究所としても非常に重要であると判断し、統計的機械学習センターを設立いたしました。

サービス科学研究センター、これは、文部科学省、あるいは経済産業省等々国のほうで、サービス科学やサービス工学が今後の科学技術として重要であると研究推進されていることもあり、また統計数理がその分野に多面的に貢献できるのではないかということで設立いたしました。これは研究諸環境を意識しながら、現在、当該分野を専門とする教員スタッフがたくさんいるわけではなかったのですが、将来を見越して設立いたしました。以上です。

○ 司会

よろしいでしょうか。まだ若干質問があるかと思うのですが、この後、五つのセンターが控えておりまして、その後でまた討論の時間がございますので、御容赦いただければ幸いです。

それでは、次の説明に移らせていただきます。

Risk Research NOE

統計数理研究所　NOE形成事業 顧問会議

リスク科学NOE

概要説明

リスク解析戦略研究センター
センター長　椿　広計

統計数理研究所
リスク科学NOE概要説明

統計数理研究所

NOE形成事業

リスク解析戦略研究センター
センター長　椿　広計

リスク科学NOEの概要

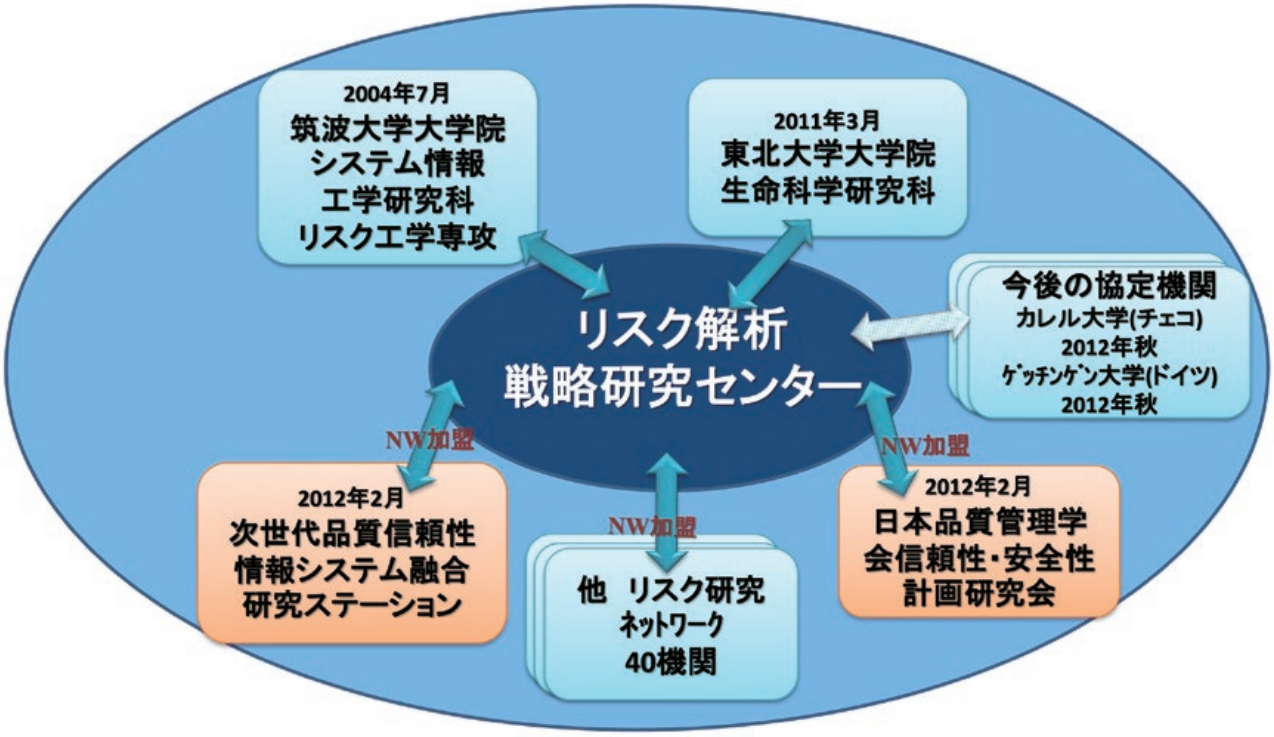
■リスク科学NOEの目的は，**リスク解析戦略研究センター（2005年4月発足）**をハブとして様々な分野の研究者・実務家と共同して，統計的・数理的側面を有するリスク研究プロジェクトを推進し，リスクに関する公共的活動を企画・実施することです。

■具体的活動は，リスクに関わる科学的研究並びにその支援を受けた社会活動を促進するための事業（科研費共同提出など）・プロジェクト協力事業・広報活動の推進などです。

■リスク科学NOE形成活動の大半は，2005年(平成17年)11月に発足した**リスク研究ネットワーク（現在42機関）**を基盤とした，緩やかな連携を行ってきました。近年，国内学の大学などとMOAを結ぶ，より強い連携も開始しているところです。

リスク研究ネットワーク

リスク科学NOE協定図



10プロジェクトの研究活動

1. データ中心リスク科学基盤整備

（PL：椿広計センター長　他8名）

リスク科学共通の理念とデータ基盤生成のあり方を明らかにします。

■公的統計匿名化事業等への協力とオンサイト分析拠点形成

高度なセキュリティ環境を実装し，機密性の高いデータを分析できるオンサイト分析室を設置し，そこでの公的統計データ・レセプトデータなどの分析を可能にしています。

■リスク情報・システム科学の基本理念形成

諸リスク科学を横断する概念，情報学的方法論をリスクNOEのメンバーと議論・整備しています。

2. リスク基盤数理

（PL：栗木哲教授　他4名）

リスク科学を横断する数理と計算手法の研究を推進します。

■ホットスポット検出問題

ホットスポットの統計的有意性を正確に計算するためのアルゴリズムの開発しています

■極値統計学

リスク管理のためには，稀に起こる極端な事象の研究が不可欠です

■共同研究集会「極値理論の工学への応用」

極値理論に関わる研究者やその応用に関わる研究者の交流の場を毎年提供しています。

3. 食品・医薬品リスク
（PL：逸見昌之准教授他7名）

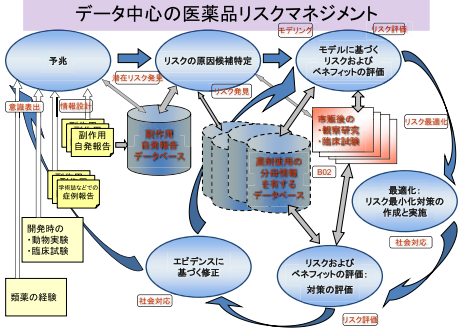
食品・医薬品の健康影響研究の基本的枠組みの提供を目指します。

■医薬品・医療のリスクに関わる実証研究支援
レセプトデータベースなどに基づく医薬品のリスク評価などの共同研究を行っています。

■食品の安全性評価に関する研究
食中毒リスクの低減を確実にし，経済合理性をもつ計数抜取検査技術の検討などの共同研究を行っています。

■医薬品臨床評価データベース構築とそれに関わる統計的方法の開発
医薬品のリスク統合評価に資するデータベースを構築し，それを用いた統計的問題検討を行います。

データ中心の医薬品リスクマネジメント

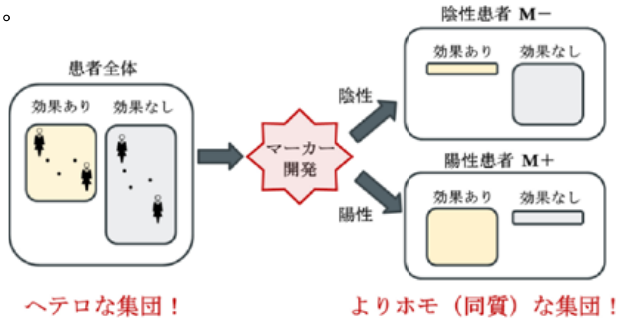


4. 予測医療の実現に向けた新しい生物統計学領域の創生
（PL：松井茂之教授他10名）

治療効果・副作用発現の診断法開発と診断法を用いた新しい臨床試験の枠組みを提案し，仮説検定ベースの従来の方法論に予測解析の要素を取り入れ，融合した新しい生物統計学領域を創成します。

■予測医療開発のための統計的枠組みの構築
治療法の臨床試験における効果予測マーカーの開発とその妥当性評価の方法論の研究に取り組んでいます。

治療法臨床試験における効果予測マーカーの開発の意義



5. 自殺とメンタルリスク
（PL：椿広計センター長，
SPL：久保田貴文特任助教　他5名）

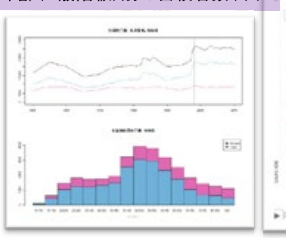
自殺やその背後にあるメンタルヘルス上の問題について現状と問題点を統計的に明らかにします。

■集積性の検出
（時）空間集積性，特定の地域，時間的差分，希少地域の集積性を明らかにします。

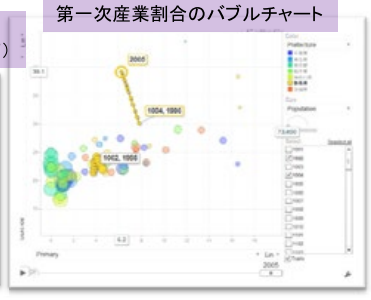
■要因分析
婚姻関係，職業，気象，地理，経済などのデータとリンケージし，因果分析，共分散構造分析を行い自殺の原因に関する共同研究を推進します。

■見える化

自殺者数の年次推移（上）と年齢（5 歳階級）別の自殺者数（下）



関東地方における自殺率と第一次産業割合のバブルチャート

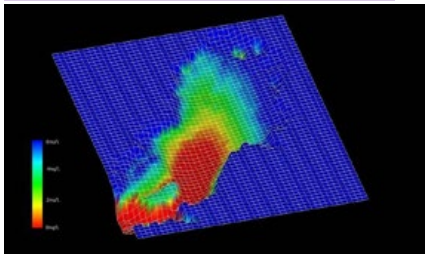


6. 環境情報に対する統計解析手法開発
（PL：金藤浩司教授　他12名）

環境科学分野との横断的協調により，環境課題に対して計量的な解析・評価手法の提供を目指します。

■閉鎖性海域に於ける底層溶存酸素量基準に対する達成率評価
評価対象となる海域においては，高々2，3地点での連続的測定データと広範囲での離散的測定データを統合し評価する方法論の確立が望まれています。
対象となる海域において最適な離散的測定のサンプリング回数の決定問題や環境基準判定の時空間的評価手法の開発を行っています。

東京湾の溶存酸素量分布



7. 資源管理リスク分析
（PL：吉本敦教授　他7名）

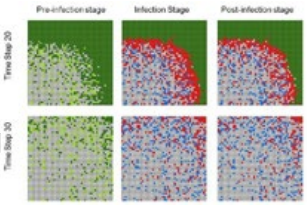
最適化による制御モデルの構築を中心にフィールドワークを通して循環型社会経済システムにおける資源管理リスク分析，評価に関わる研究を推進します。

■資源管理リスク評価

森林リスクの外的要因の時間的・地理的変化を組み込んだリスク評価モデルの構築と妥当性検証を行っています。

■撹乱現象発生シュミレーションモデルの構築

病虫害の拡散予測



■3D技術による樹木構造型モデルの構築

根曲や新芽枯死メカニズムの考慮

8. 金融・保険リスクの計量化と戦略的制御
（PL：山下智志副センター長　他13名）

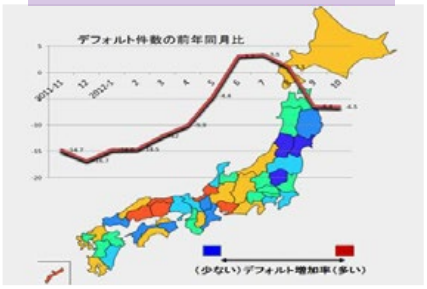
金融リスク計量化モデルのユーザーの目的に合ったモデルを選択するためのモデルの評価方法や評価基準を実務的な視点から整理・開発し，金融機関などに提供します。

■信用リスクデータベースの構築とモデル化

金融機関や保証協会が保有するデータをもとに，信用リスクの推計を行います。

バーゼル規制や国際会計基準など社会制度に準拠したモデリングにより，実務的にも利用可能なモデル開発を行っています。

2011年のデフォルトの前年同月比と都道府県別増加率



9. 地震予測解析
(PL：庄建倉准教授　他7名)

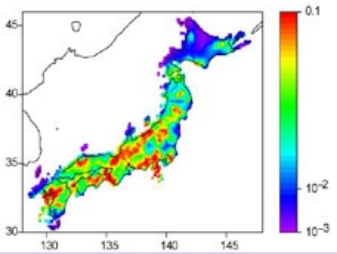
■地震の確率予測と統計モデル

地殻内部の断層やストレス状況が直接的に見えないうえ，それらが複雑で地域的に多様であるため，地震予知は難しさが増しています。

しかし，地震の発生は全く不秩序ではなく，確率的な予測は可能です。

時空間ETASモデルは過去のデータを使って将来の地震発生率を予測する標準的地震活動モデルです。

防災上要請に見合うように，リアルタイムの確率予報を実用化します。



M4以上の内陸直下型地震
今後1年間起きる単位面積当りの確率予測

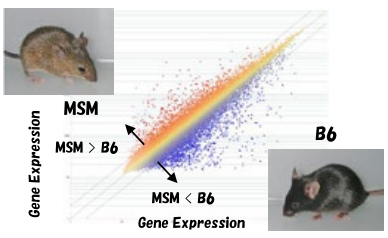
10. ゲノム解析
（PL：間野修平准教授　他12名）

■遺伝子発現データ解析

遺伝子発現データは，大規模であるがために，データ解析をする際に多くの問題を同時に内在しています。そのような問題に効率的にうまく対処するための統計的手法を確立することを目指しています。

■ゲノム情報解析による生物多様性の理解

生物多様性を理解するためには，進化的な視点からゲノム情報を解析する必要があり，そのための統計的な方法の開発を行っています。



マウスのB6とMSMの遺伝子発現データ
（国立遺伝学研究所提供）

リスク研究ネットワーク

■ 加盟組織（総会承認）：42組織

- 大学：19組織
- 研究機関：4組織
- 学協会：14組織
- 団体：5組織

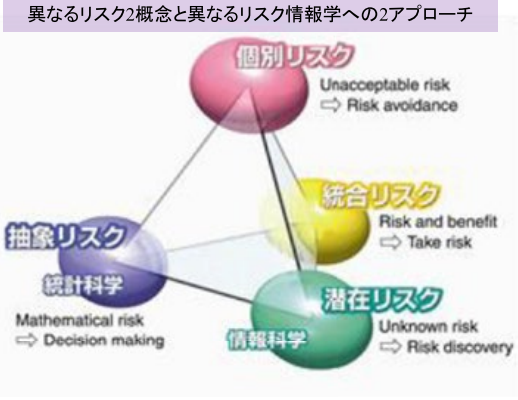
■ 運営体制

- 運営委員長
  - 宮本 定明 氏：筑波大学システム情報工学研究科リスク工学専攻
- 運営委員
  - 椿 広計 氏：応用統計学会
  - 蒲生 昌志 氏：産業技術総合研究所 安全科学研究部門
  - 小宮山 靖 氏：日本製薬工業協会 医薬品評価委員会 統計・DM部会
  - 鈴木 和幸 氏：電気通信大学次世代品質信頼性情報システム融合研究ステーション
  - 津本 周作 氏：島根大学医学部付属病院医療情報部
- 事務局
  - リスク解析戦略研究センター

■ NOEを成立させている仮説

- 概念も学術としてのスタイルも NOE組織内で多様だが 統一的接近の形成は可能かつ有益
- リスク情報システム科学創生活動

異なるリスク2概念と異なるリスク情報学への2アプローチ



【リスク科学 NOE 概要説明】

○ 司会（樋口所長）

それでは、リスク科学 NOE の概要説明を行いますが、センター長の椿が今から説明させていただきますので、司会は私が務めさせていただきます。お手元のリスク科学 NOE の資料と、前のスライドと併せて御覧いただければと思います。それではよろしくお願いいたします。

○ 椿リスク解析戦略研究センター長

よろしくお願いいたします。リスク科学 NOE に関して紹介させていただきます。先ほどからありますように、リスク科学 NOE、リスク解析戦略研究センターは 2005 年 4 月に発足しました。このセンターはもともと発足当初から、NOE 活動、各研究機関とのネットワーク形成を目的としておりまして、2005 年 11 月には、リスク研究ネットワーク設立総会を開いた次第でございます。

もともとこのリスクという分野に関しましては、後ほど述べるかもしれませんが、リスクをテイクする立場、リスクをアヴォイドする立場、それぞれが諸学術分野、産業界に散らばっている中で、統計数理の立場から、リスクに関する統一的な枠組みのようなものを議論することができないか、という形で、ゆるい NOE 連携を目指して活動してきたところです。

基本的にはリスクに関係する理念の整理、リスクに関わるプロジェクト活動の推進ということを行って、NOE に属する組織間の方法、連携を主眼として行っているところです。現在リスク解析戦略研究センター周辺に関しましては、先ほど言いました非常にゆるい連携としてリスク研究ネットワークを作っております。これに 42 組織が入っております。ここには 40 と書いてありますが、2012 年 2 月に電気通信大学の次世代品質信頼性情報システム融合ステーション、それから 2012 年 2 月に日本品質管理学会信頼性・安全性計画研究会が加盟して 42 になったところです。

これを見ていただいてわかりますように、実は所属組織は、大学研究科レベルから、場合によっては科研費の研究班というようなもの、それから、団体、学会という形で、どういうものが入るということに関しては、一切制約がありませんで、リスクに関わる研究をやっているもののゆるやかな組織体です。2004 年 7 月に筑波大学大学院システム情報工学研究科リスク工学専攻との研究協定というもう少し強い連携、2011 年 3 月に東北大学大学院生命科学研究科との協定、そして現在、チェコのカレル大学とドイツのゲッティンゲン大学との国際連携協定を進めているという形です。上にありますのは、今、統数研全般で行っている、わりと強い組織連携、この両者を使い分けて、もともとは弱い連携から開始したのですが、強い連携の方も最近リスク解析戦略研究センターもやっているという形になっているわけであります。

センターの中には、現在 10 のプロジェクトが動いております。もともとは四つのグループだったのですが、先ほど所長から説明がありましたように、それを 10 のプロジェクトという形に再編いたしました。その中に二つの横断的なプロジェクト、八つの分野型のプロジェクトが配置されています。二つの横断型のプロジェクトというのはデータ中心のリス

ク科学の基盤整備にあたるプロジェクトで、これは、データのようなものをきっと集めていって、使えるようにするということです。現在行っておりますのは、特に、いわゆるセキュリティの厳しい個人情報、法人情報に係る公的統計のデータの収集、それからレセプトのナショナルデータベース、こういう大型のデータを収集して研究に利用できる可能性を探っているところです。もう一つの基盤整備は、リスクの理念、リスク科学の在り方というような概念的な議論も、リスク科学 NOE の中で行っているところです。

もう一つの基盤的横断的なプロジェクトがリスク基盤数理と呼ばれているもので、リスク科学に横断する数理の計算手法に関する共同研究等を現在行っています。いわゆるホットスポット、非常にリスクの大きい地点、空間のようなものを検出する問題、それから、いわゆるエクストリームバリュー、きわめて巨大であるけれど希にしか起きないようなものに関する確率数理、そのための研究会などを行っているところです。これが２つの横断的なプロジェクトでございます。

それから残りの八つのプロジェクトですけれども、これは分野型のプロジェクトです。これはおおむねリスク科学 NOE の中の共同研究プロジェクトとして推進しているものでありまして、最初は食品・医薬品に関するリスクというものです。この医薬品に関するリスクというものは、先ほど申し上げましたレセプトのデータベースから、安全に関わる情報の抽出を行うとか、それから食中毒に関わる安全性評価の位置に対応する研究などを推進しているところです。

それから予測医療の実現に向けた新しい生物統計科学領域の創生ということがございます。これは予測医療の開発のためのマーカーというものをどのように作っているかということの共同研究を推進するとともに、国内の大学病院関係の生物統計学の研究者のネットワークの形成のような事業も行っているところです。

次に自殺とメンタルリスクという分野。これはある意味で医療保健関係になるわけですが、内閣府の自殺予防プロジェクトと共同して自殺に関する時空間集積性とか、その対応というようなものに関して、これもデータを作る事業と、要因分析を行い、対策につなげていくということをやはり共同研究という形で推進しているところです。時空間集積性というのは自殺にはホットスポットあるいはクールスポットがあるわけです。一方、徳島県においては、一つの境界、村を隔てただけで、自殺のワースト地域と、非常に少ない地域というふうに分かれていくようなことがありまして、こういったものが要因分析の必要性ともなっているところです。

それからもう一つ、環境に関わるプロジェクトというのも二つ配置しておりまして、環境情報に対する統計解析手法の開発ということで、これは大気とか水質に関して、折に触れた共同研究が実施されているのです。現在は水質に関する研究というものが非常に進んでいる、国立環境学研究所と一緒にやっているという形になっております。

もう一つの環境に関するものが、森林を中心とした資源管理のリスク分析ということで、これは東南アジア諸国を中心とした共同研究が進んでおります。資源管理、森林リスクの森林に関する、森林自体の伐採を含めた外的要因、時間的な地理的変化を組み込んだ評価モデルというのを開発しているところです。今のが、環境に関する二つのプロジェクトです。

それから、もう一つ大きな分野として、金融・保険リスクがあります。ここは、他のリ

スク分野がどちらかというと、リスクをアヴォイドする分野なのに対して、リスクをテイクするという側面を持っている分野です。この分野のリスクの計量化とか、戦略的制御という形で信用リスクのデータベースの構築、並びにモデル化というところに注力しているところです。

それから、もう一つの分野、実はこの９番、10 番のプロジェクトは、昨年 12 月までは私どもリスク解析戦略研究センターではなくて、予測発見戦略研究センター等にあったプロジェクトが移行してきたものであります。近年の関心事であります、地震予測の解析に関わるプロジェクト。これは地震の確率予測並びに統計モデル等を開発しているところでございます。それからゲノム解析のプロジェクト。遺伝子の発現データ解析及びゲノム情報解析により、いわゆる生物多様性問題の理解ということに関する学術的研究をやはり共同研究という形で推進しているところです。

以上、10 のプロジェクトを紹介しましたけれども、お手元に、そのプロジェクトを担当する人間が、この私どものパンフレットの最後のページにあるのですが、見ていただくとわかりますように、我々の研究所自体は小さなところなのですけれども、客員教員や外来研究員を中心とした共同研究という形で、これらのプロジェクトを支えているということが理解できるのではないかと考える次第です。この種の外来研究員、あるいは共同研究員をある意味で供給してくださるところというのが、リスク研究ネットワークという先ほど申し上げました非常にゆるい組織、共同体なのですが、現在、大学が 19 組織、研究機関が４組織、学協会が 14 組織、団体が５組織という形で運営しているところです。

運営委員長に関しましては、筑波大学のリスク工学専攻が出していただいているという形になっております。私どもリスク解析戦略研究センターは事務局を務めるということが内規で決まっているという形になっております。

運営委員に関しましても、学会、私自身はここでは学会として出ておりますけれども、産総研、製薬協、産業界、島根大学、電通大の科研費の特殊なチームというような形で、そういう色々な分野の方に運営委員を担当していただいているということです。リスク科学 NOE では、リスク概念も学術としてのスタイルも非常に多用なのですけれども、おおむね統一的な接近の形成が、少なくとも有用であるということを信じて活動している、コミュニケーションが出来ているのではないかと自負しているところです。

異なるリスクの概念、先ほど言いましたようにアヴォイドするリスクとテイクするリスク、それを潜在化させてマイニングするという情報学の方もいれば、それから我々のようにマセマティカルリスクという形で数学的に提示する、そういう中で活動を進捗させているところであります。

以上、大変簡単ではありますが、私どもリスク科学 NOE の活動の紹介とさせていただきます。御清聴ありがとうございました。

○　司会

それでは、顧問の先生方から御意見、御質問等ありましたらお願いいたします。

小柳先生。

○ **小柳顧問**

大変膨大な分野をカヴァーしていらっしゃるということに感銘を受けました。色々な専門家の前でこんなことを言うのは口幅ったいのですが、ざっと拝見したときの私の印象を申し上げます。リスクの発見とか分析ということがこれは大変重要なことなのですが、昨今の例えば去年の 3. 11 のことを考えますと、その次の意思決定とか決断とか、そこに結びつかないとリスク研究が完結しないのではないかという気がするのですが。特に、統数研では開設以来制御ということを研究所任務の一つとしていらっしゃったので、そういう視点がもしあったらな、というのが素人としての感想でございます。

○ **椿リスク解析戦略研究センター長**

どうもありがとうございます。もちろん私どものセンターの中だけですべてのことができるわけではないのです。けれども、リスク科学 NOE で議論している横断的な情報システム科学の中では、先生のおっしゃった最適化、特にリスク対応に、リスクに対して事前に最適な対応を行う、それから、レディジュアルリスクという形で、ひとたび外に出てしまったことに対して事後的に最適な対応を行う、といった二側面に分けて研究を行うのがよいということになっています。そういう形でのディスカッションないしは研究のコミュニティの形成を考えているところです。

もちろん我々の研究所内の多くが、モデリングというものを専門にしているのですけれども、最適化、特にパレート最適化という、色々なステークホルダーがいる中での最適化をどう考えていくかということについて、コミュニティの中で議論しているところです。

○ **司会**

田中先生。

○ **田中顧問**

リスク解析とインフォームドディシジョンとの関係が私は重要だと思っています。我々の社会が的確な意思決定に至るには、少なくとも関心ある人に対して、充分な情報がいきわたって、どういう選択肢を選べば、どういうリスクがある、それについての諸説はあったとしても、いくつかの説が述べられていることが重要であると思います。

一例を挙げますと、医師不足と言われていますが、とりわけ小児科と産科にお医者さんがいかないと言われています。まったく素人ですが、その分野の方に聞くと、子どもは大人と違っていろんなリスクというのですか、診断が難しい、事態が急変する。産科について言うと、出産に関わって周辺のファクターを誘発することがあって、そもそも難しい、要するにリスクが多い。それに裁判事例が重なって、誤診判断、今、最高裁事務総局で、ある判決が結果として何に至ったか、多少論理的に調べているのですが、彼らの解析によっても、産科あるいは小児科において、医師に対して過失に係る厳しい判断を下したがゆえに、お医者さんがこの分野を避ける。だから、一般的に医師が不足するというのではなくて、小児科と産科を医師が嫌がる。要するにリスクについて充分社会にインフォームドされず、医師の責任というふうに持ち込む。この点については専門家の間でははっきりしているらしいのですが、それが多くの人にシェアされていない。リスクの所在について、

研究発表がなされていないことが、我々のディシジョンをゆがめている可能性があると思うのです。

ですから、リスク解析とインフォームドディシジョンとの関係を問いながら、我々の社会の持っている欠陥というか、ついつい陥りがちな隘路について指摘するというのは、私は非常に重要だというふうに思うのですが、例えばそういうことをお願いしたいと言ったら、どうですか。どういう問題意識で、どういう研究チームで、リスク解析戦略研究センターで扱ってもらえるのですか。

○ 椿リスク解析戦略研究センター長

リスク科学 NOE の中では、今先生がおっしゃることは非常に重要なことだと思っているのです。先ほど、色々なステークホルダーがいる中でのパレート最適化ないしは多目的最適問題ということに関して触れたところです。例えば医薬品の場合、ベネフィットを受ける人間と副作用が出てしまう人間は同じなのです。一方、リスクを取る人間、それを享受してベネフィットを得ようという人間は、環境科学物質の場合ではそうではない可能性があります。使う人間はベネフィットを得るけれども、被害者というのは別に出るという可能性があったりする。その辺のステークホルダーが多様にある時に、従来ですと、ある種の価値関数、リスク関数を設定して、それを最適化問題にしてしまっているという部分があるわけです。むしろ色々なステークホルダーがあって、そのウエイト付けが色々変わった場合、その結果として起きる行動も、別のものになる、色々なフロンティアがありうるということを可能な限り可視化して、提供するというのが、システム科学側の役目ではないかということに関する議論はしているところです。

一方で、そういうフロンティア、あるいはパレート最適なフロンティアを提供したときに、何を選択するかということの社会側の役割というものに対して、我々システム科学や理工学の人間は必ずしもよいアイディアを持っていなくて、その点に関して、厚生経済学とかいろいろな文科系の知、あるいは社会学の知というのが必要なのではないかということを強く最近感じています。我々も、限界自体を感じているところなのです。ですから、ぜひ、そういうテーマの御指導を賜れば、と強く思うところです。

○ 司会

佐和先生。

○ 佐和顧問

８番の金融保険リスクについてですが、先ほど御説明の中で、その他のプロジェクトはおおむねリスクを回避することをねらいにしているけれども、これはリスクに挑戦することをねらいとしているというふうにおっしゃいましたね。それではこのプロジェクトは、投資家のための研究プロジェクト、すなわち投資家にモデルを提供するためのプロジェクトなのでしょうか。そうであっては困ると私は思います。国際金融危機が 2008 年に起こりました。ああいう危機を回避するためには、金融システムをどのようにしてロバストにするのか、金融システムをどのように改編すべきかについての研究をぜひやっていただきたいと思います。

○ **椿リスク解析戦略研究センター長**

ありがとうございます。私どももいわゆる数理ファイナンスのコミュニティの方々と共にそういうことを考えていければと思いますので、その方向で動かせていただければと思います。

○ **司会**

せっかくですので、副センター長に解説をお願いいたしましょうか。

○ **椿リスク解析戦略研究センター長**

そうですね。このプロジェクトのプロジェクトリーダーの山下副センター長がおりますので、山下先生お願いします。

○ **山下リスク解析戦略研究センター副センター長**

当初ファイナンスについて、リスクをテイクするというような解説がありましたけれども、考え方によっては、他のリスクについても、（費用削減など）何らかのベネフィットがあるからあえてリスクを取るという行動というのはあると思うのです。

ただ、ファイナンスについて見れば、リスクを取るというのはかなり昔から伝統的に計量化されておりまして、リターンが欲しい場合は、追加リスクを取るというのが世の中でセンサスになっている。単にリスクが好きだからそちらに行っているわけではなくて、リターンを得るための費用としてリスクをテイクするという文脈で、今のリスク選好ということを理解していただければと思います。

○ **司会**

真鍋先生。

○ **真鍋顧問**

リスク解析戦略研究センターの研究に関しまして、すでに色々御質問や御意見が出ています。私は、少し別の視点から質問させていただきます。実は、かつて昔林知己夫先生の御指導のもとに、「中国における不安感・科学観・環境意識」というテーマの共同研究（1993年～1998年）を実施したことがあります。そこで、我々が採用した方法論的な視座の一つが、「確率論的リスク評価――ある事件・事故・危険が実際に起こる確率――」と「知覚されるリスク評価――人々がそれらの事件・事故・危険がどれくらいの確率で起きていると思っているか――」を区別するというものでした。そして、「病気」「交通事故」「戦争」「原発事故」などについて考えるならば、「病気」「交通事故」に比べて、「戦争」「原発事故」では、「確率的リスク評価」よりも「知覚されるリスク評価」の方が高くなるが、それは後者の場合は①人々の日常生活における注意や制御をはるかに超えた問題であるということと、②その影響の及ぶ範囲が極めて広いということ、が大きく関わっているという議論を展開しました。そのような点からするならば、「リスク解析戦略研究センター」の活動は、様々な研究領域との研究交流が必要になってくる分野だと思いますが、いかがでしょう。

○ 椿リスク解析戦略研究センター長

私どものNOEに関しては、特にリスク研究学会、リスクコミュニケーションを研究している分野がその種の研究はかなり進んでいると思います。ただ、伝統的な理工学の確率の中でも、いわゆる客観的確率以外に主観確率というものがあって、事象の確率以外に、行動をもとに決める確率があるわけです。今、リスクコミュニケーションのようなものを考える場合には、客観的な確率以外の、その種の行動に関する確率というものの位置付けをもう一度見直さなくてはならない、これも実はリスクのネットワークの中で議論のあるところです。

○ 真鍋顧問

社会科学の領域では、オブジェクティヴとサブジェクティヴと言いますけれど。

○ 椿リスク解析戦略研究センター長

先生のおっしゃるとおりです。

○ 吉川顧問

簡単になのですが、御存知、低放射線の健康被害というものがけっこうありますよね。それに対して、統計学というのはどういう立場に立っているのかというのをリスク問題ということで。

○ 椿リスク解析戦略研究センター長

これは、今年の９月に札幌で行われました統計関連学会連合大会の特別セッション自体の最後のテーマが低線量放射線の影響に対して、統計科学がどれくらい寄与しているのか寄与していないのか、ということでした。これはNOEの活動とは少し違う側面であることを御容赦いただければと思いますけれども、現時点において、広島、長崎のデータに基づいて、低線量の放射線に関する健康影響のモデルというものは構築されているところなのですけれども、このモデリングに関してですら、放射線影響研究所のモデルと、広島大学のモデルにおいては若干の解釈の差といいますか、影響に関する差が出てきているというのが現状でございます。

つまり統計モデルといっても、もともとのデータの持っている情報量というのがそれほど大きくない状況においては、かなり不確実性の高い結論となり、解釈も分かれてしまうという状況が起きている、これが、少なくとも私の知りうる現状でございます。そういうものに対して、統計科学、データの科学と同時に、システムの科学、つまり健康影響に関する数理的モデリングの知を補っていくことによって、もう少しそこに迫れるのではないか、という考え方があります。それには私どものすぐ後になりますけれども、データ同化のようなものが非常に有効ではないかと期待されているところです。現時点で、疫学的方法論で、この種の低線量放射線の影響を調べようとすると膨大な調査費用が掛かるであろうということは間違いないことかと思います。

○ **司会**

吉田先生。

○ **吉田顧問**

日本銀行の吉田です。ここで発言するのがよいのかどうかわからないのですけれども、金融に関してのいくつかの議論になっているので、金融研究所として一言だけ申し上げると、今議論になっている中に、リスクを取るというのはリターンの裏返しであるということがあって、それはもちろんその通りなのですけれども、リーマンショック以降起きたことを振り返ってみると、一体何が問題だったのだろうかという研究が今、ファイナンス分野とか経済の分野で皆さんやっていらっしゃるのですけれども、私どもでも多少やっておりまして、その中でどういう研究をやっているか１個だけ紹介すると、個別の金融機関で最適なリスク量を取っているというふうに判断して取っていたとしても、マクロでみたときに負の外部性みたいなものが働いて、全体としてみるとリスクを取りすぎるというそういうようなモデルを作って、ソーシャルプランナーみたいなのがいて、最適リスク量を計算したときの場合と、個々の金融機関が取ったリスク量の差みたいなものが生じるようなモデルを作るというのをやっております。私どものスタッフも、京都大学の先生との共同研究等でそういう論文を書いたりして、それであるからこそ、マクロプルーデンス政策と我々は呼んでいるのですが、何らかの規制をかけて金融業をコントロールしていくということが必要であり、実際に国際的にも各国で共同作業として進められているということです。

○ **椿リスク解析戦略研究センター長**

どうもありがとうございます。勉強させていただいて我々のグループも一緒にやらせていただきたいと思います。

○ **司会**

鈴木先生。

○ **鈴木顧問**

リスク科学とは何かということなのですが、統計学、統計的な手法というものが、一体リスクを、例えば軽減させる、あるいはリスクを評価する、あるいはリスクとベネフィットに関してのある考え方を打ち出せる、そういうものになり得るのか。要するに、膨大なデータがあるときに、ナイフを研いで、きれいにこう切って見せる、これは統計学としては非常に得意なのでしょうが、統計学そのものがリスクの本質に迫るということはあり得るのか、というあたりに私は違和感を覚えるのです。環境情報のように、まさに地球環境関連のデータなど、温暖化もそうなのですが、膨大なばらつきのあるデータの中から、何か優位なものを読み取って、それを我々の知恵にしていく、そこまではよいのでしょうが、結局それは、統計学の範囲でしかなくて、要するに「リスク科学」というようなものに迫るものとして本質的なものがあるのかどうか。

○ 椿リスク解析戦略研究センター長

最初に申し上げましたように、リスク科学自体は、各分野でそれなりにきちんと発展しているものと思うところです。統計科学の人間の特徴というのは色々な分野に対してインターフェースを持っているということなのですよね。環境の方に対しても食品の方に対しても、医薬品の方に対しても、医療の方に対しても、金融の方に対しても持っている。我々としては、もちろん個別のリスク科学の個別性というものが非常に重要なものであるということを認識したうえで、その中を横断するような方法論の知とかそういうものがあるのではないかと考えて活動しているところなのです。先生のおっしゃる、本当にそういうものが具体的に誕生するのかどうかということに関しては、我々自身も日夜悩みながら進んでいるということは事実でございます。

○ 司会

甘利先生。

○ 甘利顧問

今の話とも関係するのですけれども、例えば地震予知にしても、ゲノム解析、それから金融系、それぞれの分野で個別の研究は今までもやってきたわけですよね。統計的な手法に主眼をおいて、それをここでやるというだけではなくて、特にリスク管理という視点から、それにない何か新しい研究の方針なり、仕方をぜひ作ってくれないと、ただの普通のサイエンスになってしまっては、残念だと思うのです。たぶんそれは新しい科学を作るという意気込みがあってやっておられるのだろうと思うのですけれども、その辺の意気込みを説明してください。

○ 椿リスク解析戦略研究センター長

はい。普通は、統計科学はシステムのモデルかと思うのですけれども、モデリングと推論、あるいはデータの収集という要素をつなぐ一連のプロセスというもの、リスクに対してアプローチするプロセスのモデルというようなものが、ある程度の範囲の中で有効ではないかとういことを期待して動いているつもりではあるのです。もちろん先ほど言いましたように、個別の領域には個別の領域の中でチューニングすべきことが、たくさんあると思うのです。その中での第一歩として、ファンダメンタルな部分として、標準的なプロセスのモデルというものを当面作りたいという意欲は持っているところです。

○ 司会

それでは、次のセンターの説明に移らせていただきます。

Next-Generation Simulation NOE

統計数理研究所　NOE形成事業 顧問会議

次世代シミュレーションNOE

概要説明

データ同化研究開発センター
センター長　樋口　知之

統 計 数 理 研 究 所
次世代シミュレーションNOE概要説明

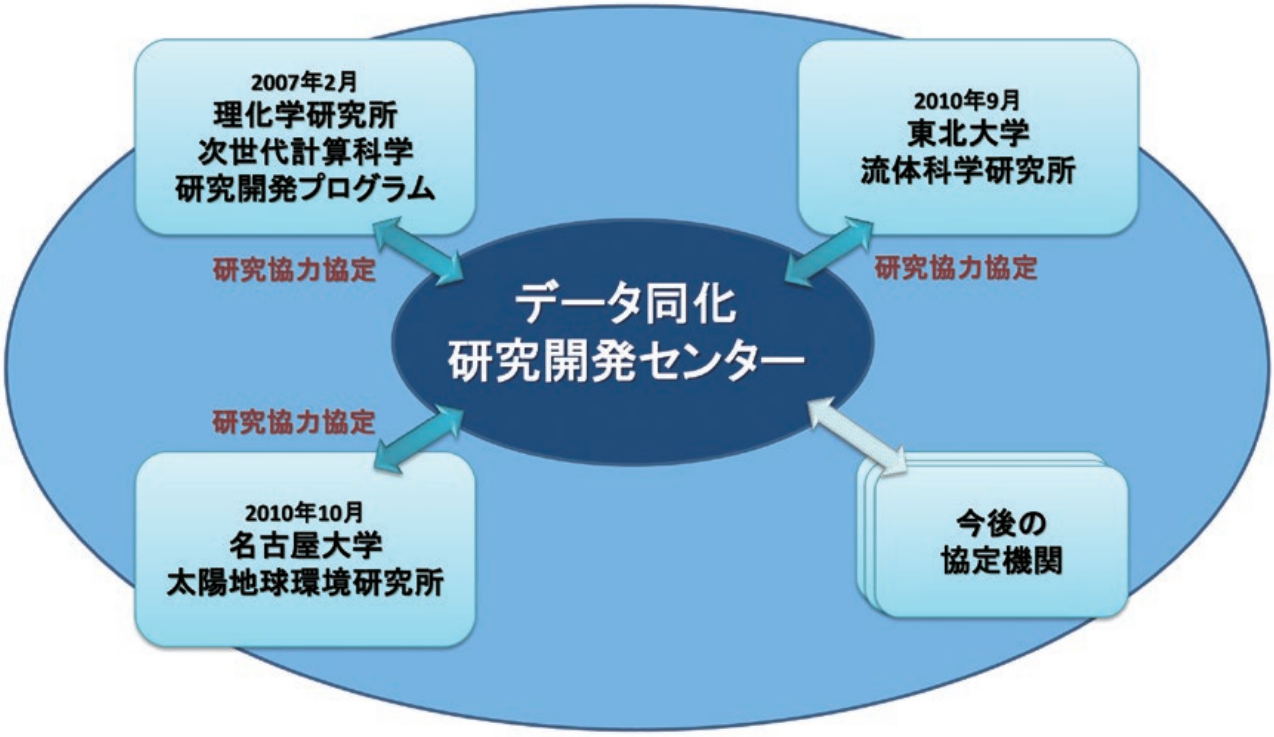
統 計 数 理 研 究 所
NOE形成事業

データ同化研究開発センター
センター長　樋口　知之

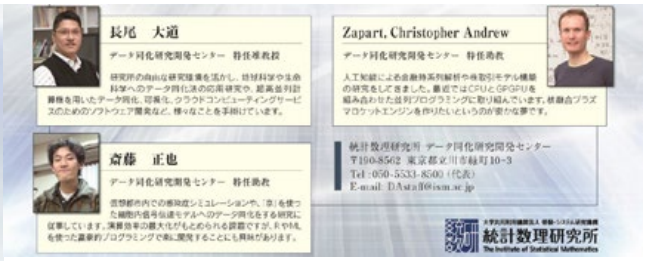
次世代シミュレーションNOEの概要

次世代シミュレーションNOEは，数値シミュレーションと観測データを融合するための基盤技術である「データ同化」をキーワードとし，大規模並列計算による数値シミュレーションを実施している国内外の研究者との研究協力体制を確立することにより，データ同化の応用分野を拡大していくことを目的としています。具体的な活動として，国内外の大学・研究機関との定期的なワークショップの共催や，学会や各種メディアを通じた広報活動を実施しており，データ同化研究で得られた成果を，専門家のみならず，一般社会へ還元することにも力を注いでいます。本事業は，2011年（平成23年）1月に発足したデータ同化研究開発センターに所属する教員ならびに研究員が中心となって推進しています。

次世代シミュレーションNOE協定図



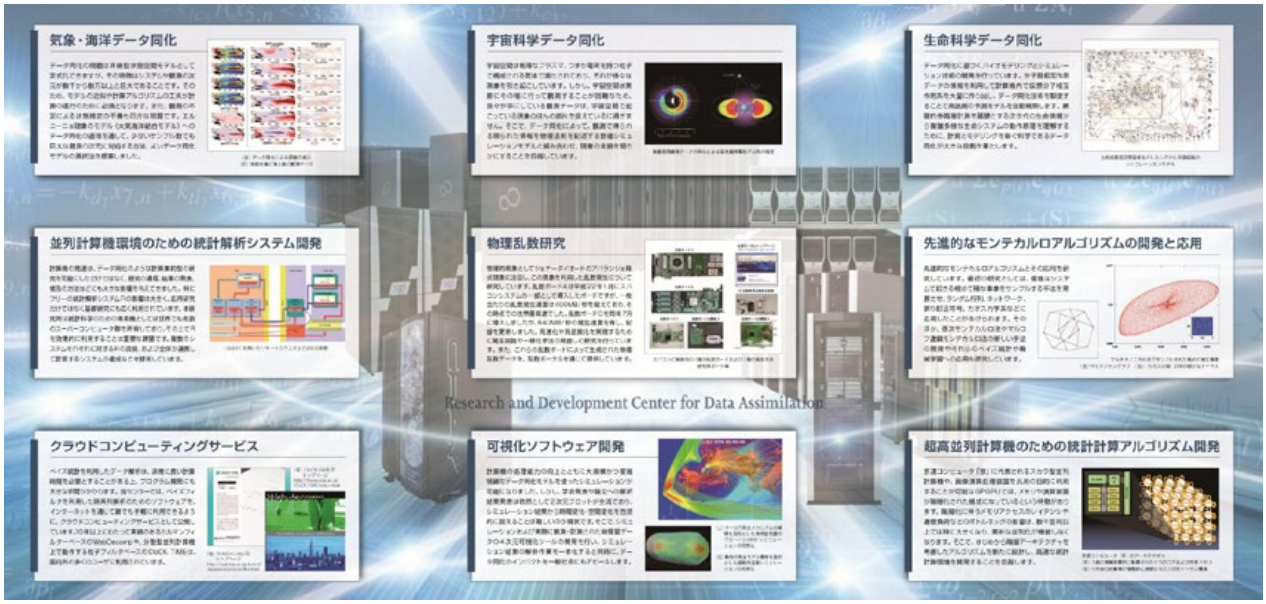
次世代シミュレーションNOE 構成メンバー



特任研究員：才田聡子　鈴木香寿恵

研究支援員：大石まどか

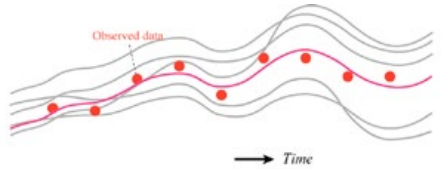
データ同化研究開発センター　プロジェクト一覧



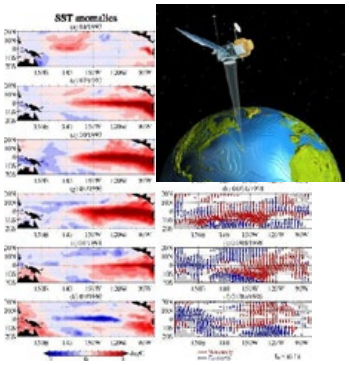
データ同化研究開発センター広報パンフレットより

データ同化技術の基盤構築

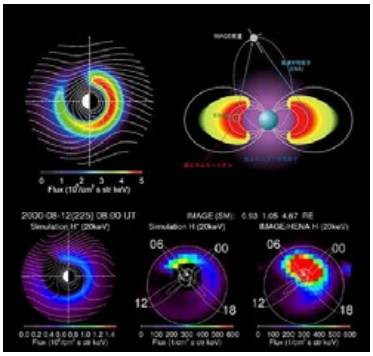
文部科学省研究開発局 新世紀重点研究創生プラン「人・自然・地球共生プロジェクト」（平成14～18年度），および科学技術振興機構 戦略的創造研究推進事業「シミュレーション技術の革新と実用化基盤の構築」（平成16 ～21年度）を通じて，データ同化技術の基盤構築および応用研究を実施しました。



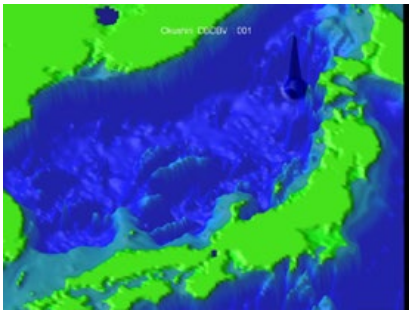
$$x_t = f_t(x_{t-1}, v_t), \quad v_t \sim p(v \mid \theta_{sys})$$
$$y_t = h(x_t, w_t), \quad w_t \sim p(w \mid \theta_{obs})$$



気象・海洋データ同化



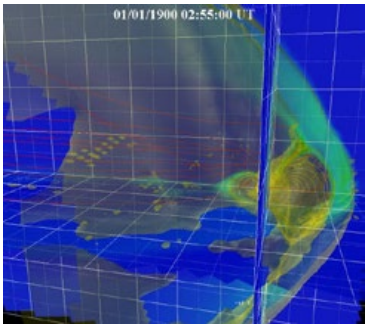
宇宙科学データ同化



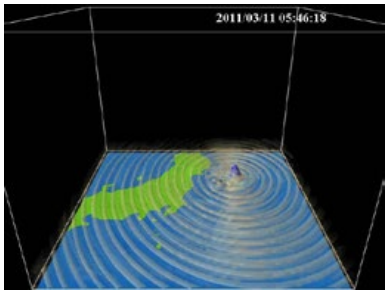
津波データ同化

融合研究プロジェクト

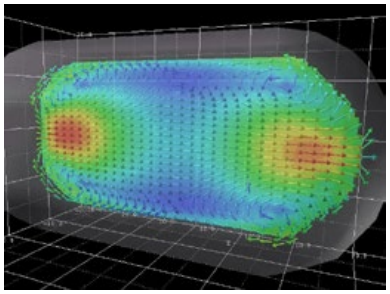
情報・システム研究機構の**新領域融合研究センター**の新領域融合プロジェクト「データ同化による複雑システムの定量的理解と計測デザイン」（プロジェクトディレクター：中野純司教授）において，情報・システム研究機構内の各研究所に所属する研究者と連携し，データ同化に関する共同研究を実施しています。



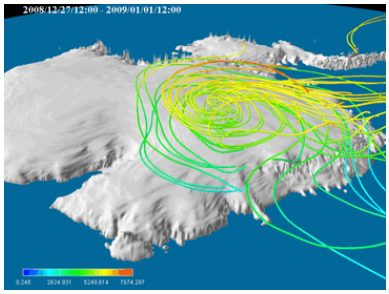
オーロラ発生メカニズム解明を目的とした磁気圏・電離圏データ同化



東日本大震災に伴う大気変動データ同化



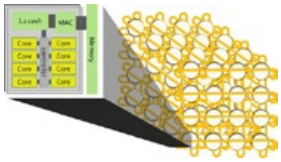
細胞質流動データ同化

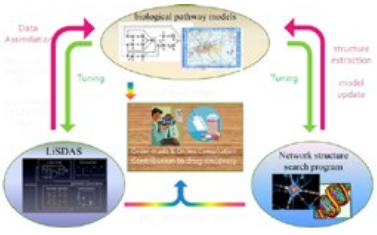


南極氷床上空の大気移流データ同化

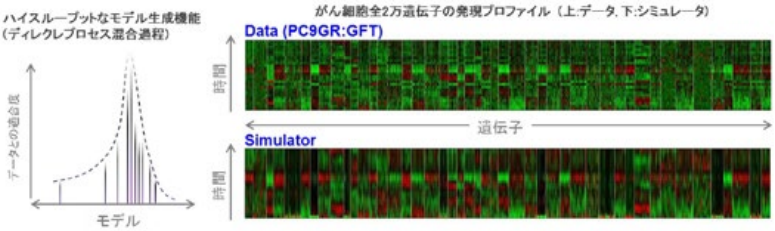
コンピュータ「京」プロジェクト

「次世代生命体統合シミュレーションソフトウェアの研究開発」（平成18年度～）において，約2年間にわたって世界最速を誇ったスーパーコンピュータ「京」のためのソフトウェア「**LiSDAS**」および「データ同化ライブラリ」を開発しています。「京」のアーキテクチャに最適なデータ同化アルゴリズムを開発し，遺伝子ネットワークモデルの構築や，データ同化アルゴリズムの普及を目指しています。





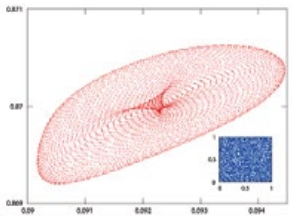
LiSDASの概念図



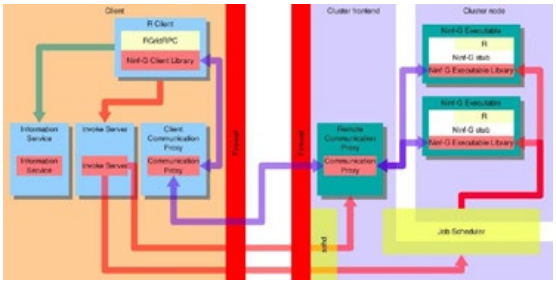
データ同化を支える基礎研究

データ同化研究に必要不可欠な様々な基礎研究を実施しています。具体的には，高速に高品質な物理乱数を生成するための研究，並列計算機を使って効率的な統計計算を実施するための研究，先進的なモンテカルロアルゴリズムの研究，そしてクラウドコンピューティングによる統計計算サービス構築の開発が挙げられます。

物理乱数ポータルと乱数ボード



先進的モンテカルロアルゴリズムの研究



並列計算機を利用するための統計計算言語Rの開発

統計計算のためのクラウドコンピューティングサービスの構築

出版活動

（2011年4月刊行）

（3刷　計3,600冊）

（2011年9月刊行）

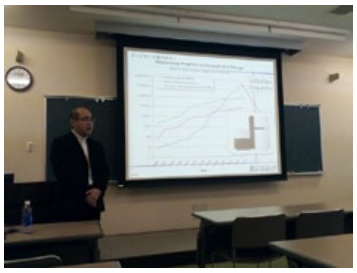
（3刷　計1,500冊）

（2011年11月刊行）

学会・ワークショップの共催

研究協力協定を締結している大学・研究機関と，国際学会特別セッションの設立や，定期的なワークショップの共催を実施しています。2011年3月に開催したデータ同化研究開発センター設立記念シンポジウムでは，樋口所長の基調講演の他，協定締結機関である東北大学流体科学研究所の早瀬所長，および名古屋大学太陽地球環境研究所の荻野教授から，招待講演を賜りました。

データ同化研究開発センター設立記念シンポジウム（**2011年3月11日**）



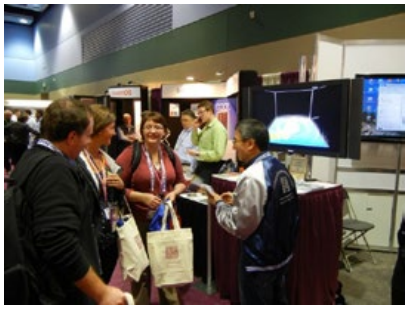
流体研・統数研共催「融合研究ワークショップ」（2012年3月）

国際会議***FUSION*2011**（シカゴ）の特別セッション「データ同化」における樋口所長と加藤博司氏（東北大流体研）の講演。***FUSION*2012**（シンガポール）でも企画，開催。

SC（Supercomputing）ブース展示

毎年11月に米国で開催されるSC(Supercomputing)は，スーパーコンピュータに関係する大学・研究機関や企業が一同に会する大規模な国際学会です。SCでは，次世代シミュレーションNOEのメンバーが中心となって本研究所のブースを出展しており，研究成果の国際的な広報活動を実施しています。

樋口所長による基調講演



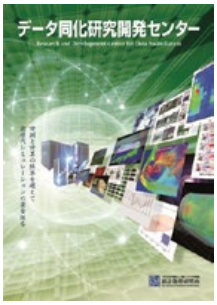
外国人来訪者への広報活動

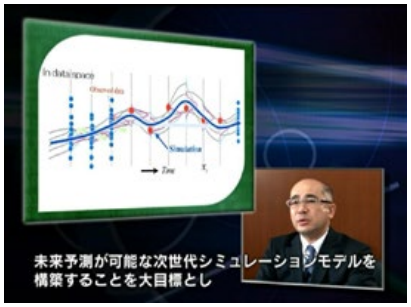
研究成果発表「Research Digest」

その他の広報活動

データ同化研究開発センター広報用のパンフレットとビデオを制作し，研究集会やメディア等で，研究成果を積極的に公開しています。また，専門家のみならず，一般の方々にもご参加頂けるシンポジウムや講義を開催しています。

地元住民への広報活動（サイエンスカフェ＠国分寺の茶店）





データ同化研究開発センター広報パンフレットと広報ビデオ

広報ビデオを動画サイト上（You Tube&ニコ動）で公開

【次世代シミュレーション NOE 概要説明】

○ 司会（椿副所長／NOE 推進室長）

それでは次に、次世代シミュレーション NOE の概要説明を行います。これは所長であり、データ同化研究開発センター長を務めております樋口知之より御説明いたします。

○ 樋口データ同化研究開発センター長

すでに最初のセンターの研究活動に対しまして、先生方から大変鋭い指摘や、たくさんの宿題をいただきまして、私も今から説明させていただきますが、色々な宿題等をいただくのではないかと思っております。

NOE としましては次世代シミュレーション、その中核機関としてデータ同化研究開発センターで活動を行っております。

データ同化では、シミュレーションを扱います。色々な形のシミュレーションがありますが、たいていシミュレーションというのは、初期条件、境界条件、パラメータ等々与えますと、前向きに計算する、フォワードタイプの推論です。一方、先ほどから色々話題になっていますが、統計科学等々はデータから逆読みしていくという、後ろ向きの推論という特性があります。データ同化は、この前向きの推論と後ろ向き推論を融合するということによって、各々の苦手なところを各々の強みで克服するということを狙った手法です。

データ同化について、もう少し時間を使って説明させていただきます。シミュレーションモデル自体の様々な不確実性や、シミュレーションでもって何か表現したときの現実からの乖離をどのように検討しモデルを修正すればよいのか等々の研究というのは、モチベーションは 1960 年くらいから、古くからあったわけです。1990 年代に入りまして、人工衛星等により地球を外から、グローバルに色々な観測ができ、その観測データが利用できるようになった。1990 年代、特にエルニーニョが非常に多く起こりまして、気象・海洋の分野では、シミュレーションもやはり限界があるし、また、データの方も時間、あるいは空間解像等で限界があるので、その両者を合わせることによって、よりよい予測、あるいはモデルの持つ不確実性を少なくすることができるのではないかということで、1990 年代からデータ同化の研究が活発になってきました。

現在におきましては、アメリカ、ヨーロッパ及び日本も含めて、気象予測の現業官庁がすでにデータ同化の手法を採用しております。時々刻々入ってくるデータとシミュレーションを併せることによってよりよい予測をすることにデータ同化手法が使われております。

データ同化研究開発センターは、同化計算の基本となる計算アルゴリズムの研究を中心的に行なっています。シミュレーションとデータを組み合わせるということは、気象の分野だけではなく色々な分野に適応できますので、例えば、ものづくりにも適応できるわけですので、基礎的なアルゴリズムの研究開発はもちろん色々な分野への応用の二つを大きな目標として研究活動を行っております。

先ほどリスク解析戦略研究センターのリスク科学 NOE は椿から説明がありましたけれども、この次世代シミュレーションの NOE は特色として、東北大学流体科学研究所、あるいは名古屋大学の太陽地球環境研究所という、大学の附置研、特に文部科学省の分類では拠点研究所との学術協定を結ぶことにより NOE 形成を推進する点が特色となっておりま

す。2007 年に理化学研究所と結んでおりますが、これはスーパーコンピュータの「京」の上で動く応用ソフトの開発をするということで、締結したものです。

これが構成メンバーで、左側が常勤職員、右側が特任教員の構成メンバーとなっております。詳しくはパンフレット等を御覧いただければと思います。

これがデータ同化研究開発センターで行っているプロジェクトです。上の方に並んでおりますのが、データ同化の応用分野です。左上が気象・海洋、真ん中が宇宙科学、また我々は世界に先駆けて生命科学に対してもデータ同化の手法が有効であるということを示し、こちらの方も研究を行っております。また、左に並列計算機上で動く統計解析システムの解説を示しています。統計数理研究所は昔から物理乱数の装置も持っておりますし、それに関連する技術開発も行っておりますので、これもセンターで継続して技術開発を行っています。また、今年（2012 年）1 月に改組しました結果、新しいメンバーが入ってきましたが、この中にモンテカルロ計算アルゴリズムを非常に得意としている者がいますので、一緒にモンテカルロアルゴリズム研究も行っております。

下の方に示したものは、開発した計算アルゴリズム等々をクラウドサービスとして提供している様子や、開発した可視化ソフトウェアや、スーパーコンピュータ「京」で動く特別な計算アルゴリズム等々になります。

先ほどどうしてこのセンターを立ち上げたということに関し御質問がありましたけれども、比較的大きい研究資金をいただいてデータ同化に関する研究開発を行ってきましたので、継続的に研究を推進する意味もありセンターを設置いたしました。我々は平成 14 年から 18 年、文科省の研究開発局で、地球共生プロジェクトというのがありましたが、これは JAMSTEC の「地球シミュレータ」を使って、いわゆる、我々とは違う形のデータ同化のアルゴリズム研究を行っているプロジェクト研究チームに参加しました。また、平成 16 年から 21 年度は JST の CREST を、私が代表として、研究資金を得ることができましたので、そこでもってデータ同化技術の開発を行ってきました。また、後でまた説明しますが、我々は現在、スーパーコンピュータの「京」上のアプリケーションソフト応用開発に関する国プロにも参加しています。

このように比較的大きい外部資金をいただきながら研究を行ってきましたが、また、新たに統計数理クラウドという概算要求で特別経費の研究予算をいただくことができましたので、これがこの研究センターを立ち上げた理由の一つです。

この新領域融合研究センターは、情報・システム研究機構の方で、四つの研究所が色々関連、協力してプロジェクトをやるバーチャルなセンターですが、この中に融合研究プロジェクトというのがあります。繰り返しになりますが、情報・システム研究機構は四つの機構の中でも非常にユニークな研究機構で、横型の統計数理研究所と情報学研究所、また縦型の極地研究所と遺伝学研究所の組み合わせ方によって、多様な共同研究、融合研究がやりやすい素地が整っております。データ同化も、その素地を活かしまして、色々な融合研究をこのセンター内で行っております。右側がオーロラのデータ同化研究、これは極地研の先生と行っているもの、左下は、東日本大震災で津波の後に非常に音波が大気中に伝わって行きましたが、その音波の影響が及んだ現象に関するデータ同化の結果の様子です。大気中が電磁層派に伝わって行きますと、電磁層に揺らぎが生じます。GPS のデータは電子密度の揺らぎも観測できますが、音波による電子密度の揺らぎが生じるとリング状の電

子密度の濃淡が見えます。その揺らぎと音波を発生するシミュレーションモデルをデータ同化させる研究を行っています。これは極地研究所及び電通大の先生と一緒にやっています。真ん中は、これは遺伝学研究所の先生とやっている共同研究です。現在、一つの細胞の中の動きを、いわゆる PVI という、蛍光物質で細かい分子の動きを測定することができます。この分野は、近年測定テクノロジーが著しく発達していますが、それと一つの細胞の中の動態を記述するシミュレータを同化させることで、測定できないものを間接的に推定するということを可能にする研究をやっております。右がこれも極地研と一緒にやっているものです。

あと我々は「京」のソフト開発の国プロにも参画しておりまして、ライフサイエンスの研究者の方々にデータ同化のアルゴリズムを提供するために、我々は長年にわたってプロジェクトに参画してきました。“仕分け”など、研究プロジェクトの推進上、つらいときも二度ほどありましたけれども、ポスドクと一緒に、我々チーム一丸となって頑張ってきました。このライフサイエンスにおいては、遺伝子、あるいはタンパク質の色々なつながり、つまりネットワークというのは常微分方程式でモデル化されていますが、その常微分方程式中の係数パラメータというのはアンノウンなものばかりです。それらを、遺伝子発現データやタンパク発現データを使いまして推定していく挑戦を行っております。

あと、基礎研究の方は、繰り返しになりますが、モンテカルロ計算のアルゴリズム、あるいは乱数等の発生アルゴリズムの研究開発を行っています。できたアルゴリズムをクラウドにもっていく等も実施しています。そして左にありますように、R という統計が得意とする言語をより多くのコミュニティの方に使ってもらうために、並列計算で動くシステムの開発と言った、非常に実質的な研究も行っています。

データ同化というのは、色々な分野に適応できるということで、産業界等々からのニーズが非常に高く、アウトリーチ活動内の出版活動としましても、私、あるいは我々のチームで教科書を出しました。まだ 1 年ちょっとですけれども、販売は順調です。

また、色々なワークショップのオーガナイズや、国際研究集会内でデータ同化のセッション企画を行ったり、スーパーコンピュータの SC という見本市と研究発表をかねた会議がありますが、そこにブースを出して、毎年成果発表を行っております。

広報活動としましては、左下にありますが、成果公開用ビデオを作りました。10 分間のビデオですけれども、You Tube とニコニコ動画にアップロードしておりますので、もしお時間がおありのときに、御確認いただければと思います。以上です。

○ 司会

以上、次世代シミュレーション NOE に関しまして、御質問、御意見等おありでしたらお願いいたします。小柳先生。

○ 小柳顧問

次世代シミュレーション NOE ということなのですが、シミュレーションというのはもちろん、計算科学、コンピュータサイエンスの色々な分野で研究されているわけですが、統数研の NOE がどういう点で、特徴を出すのでしょうか。何かお考えがあるのだろうと思いますが、どういう点を軸に、こういう分野に参画していくかというその辺をもう少し御説

明いただけるとありがたいのですが。

○ 樋口データ同化研究開発センター長

データ同化というのはフォワード推論とバックワード推論をつなぐ技術ですが、フォワード推論、つまり通常のシミュレーション研究を重点的にやる方策もあり得ると思うのですが、我々は、データとシミュレーショナルをつなぐアルゴリズム研究開発に注力していきたいと思っております。

もう一つは色々な応用分野への技術展開をしたいということで、統計数理研究所はスーパーコンピュータのインフラムを持っておりますが、もちろん理化学研究所、あるいはもっと大きいスーパーコンピュータを持っているところはたくさんありますけれども、データ同化に適したやりやすいアーキテクチャというのがあるのではないかと思っています。そこらが他のいわゆる前向き計算に重点を置いた研究機関との違いであると思います。

また、MOA を結んでおります、東北大学流体科学研究所の、今、所長をされている早瀬先生という方は、融合計測シミュレーションというものを 10 年くらい前からやっておられますが、それは、実体を模倣する形式できっちりとシミュレーション計算するのではなくて、計算と観測値のずれをある意味外部ドリブンの力として、整合性のある形でシミュレーションに入れ込む研究になります。これは、データ同化の特殊なアルゴリズムと考えられますが、このような外部の研究所とも具体的に密な情報交換をやっていきたいと思っています。

○ 司会

吉川先生。

○ 吉川顧問

私もよくわからなくて質問なのですけれども、今のお話は、要するに、人間が政策的に介入できないような対象になるのですか。

○ 樋口データ同化研究開発センター長

いえ、そんなことは。

○ 吉川顧問

それも含むわけですね。要するにデータというのは、例えば気象データなどは向上できないわけですよね。

○ 樋口データ同化研究開発センター長

それはもうデータしかないわけなので。

○ 吉川顧問

そのデータからモデルを自動的に作る、これをバックワードと言っていらっしゃるのですね。例えばモデルというのは論理的にできる面もあるわけですから、それだけでなくて、

わからないことがいっぱいあるので、それをデータから逆に読み取るということになるわけですね。

○ 樋口データ同化研究開発センター長

学習するということになります。

○ 吉川顧問

そういう問題ですよね。そうすると、現在、本当に知りたい例えば経済の不安定ですとか、あるいは生物多様性とか、そういった緊急の課題があるのだけれども、そういったものも皆人間のコントロールできない部門で、ディズアスターが起こっているわけで、逆に言うと、人間の知の範囲を超えた判断というのをデータからとるという、非常に大事な話ではないか思うのですが、そういうものが取り上げられていないというのが、これもどういうテーマを選ぶかというその基準を何か伺いたい。

○ 樋口データ同化研究開発センター長

我々は、そういう問題を避けているわけではないのですけれども、今までは、比較的シミュレータの素性がよくわかっているものを取扱いながらアルゴリズムの研究をやってきたというところがあります。

○ 吉川顧問

例えば経済学というのは非常に立派な理屈があって、その経済学のセオリーができれば、経済は制御できるのだという、そういう非常に勇敢な分野ですよね。だけどそうでない場合は、いわゆる人間の論理的なものを作る能力だけでは取れずに、もちろんそれをする必要はないけれど、別にどこか人間の知らないところで動くのが残っている、むしろそれはデータとして把握するしかできないのだということになると、例えば経済学の同化シミュレーションという話が出てきてもいいような気がするのですが、これは経済学者に遠慮しているということですか。

○ 樋口データ同化研究開発センター長

実際そのような研究の動きがあり、ファイナンスで使われているいろいろなモデルの不完全性というのがあったならば、それをデータからある程度読み取るというような試みは少しずつされつつあると思います。

○ 司会

真鍋先生。

○ 真鍋顧問

本来、我々の役割というのは、それぞれのセンターの研究活動にコメントすることを通して、そこから新しい研究の方向が出てくるとか、これまでの研究がさらに発展するということにあるのだろうと思いますが、私の場合は、むしろよくわからないことをここで改

めて教えていただけるとてもありがたい機会になっていると思います。さて、私の質問というのは次の点です。今、シミュレーションというのは「前向きの推論」、統計科学はデータから「逆読みしていく後ろ向きの推論」で、データ同化では、この「前向きの推論」と「後ろ向きの推論」を融合することだという御説明をいただきました。私のイメージだと、これまでの科学の基本的な考え方が「要素主義」といいますか、研究対象を様々な要素・次元・側面に分けた上で、それぞれの要素・次元・側面の相互の関係について分析をするという行き方をとってきたのに対して、シミュレーションは方法論的には「全体主義」といいますか、対象の要素・次元・側面の細分化を踏まえない形での対象の法則化の方向を模索するという行き方というように思われます。そうるすと、「データ同化」というのは、このような「要素主義」と「全体主義」との方法論的な統合の試みというようにも理解されると思うのですがいかがでしょうか。

○ 樋口データ同化研究開発センター長

ええ。データ同化がやるものとして、そういう機能もあると思います。

○ 司会

鈴木先生。

○ 鈴木顧問

あるいは今の真鍋先生の御質問とも関連するかもしれませんが、現実の世界というのは、なかなか方程式、微分方程式であれ、何であれ、書き下ろして、それをひも解いて再計算するといっても、かなりの近似が色々なところに入ってくるわけですね。人類が膨大な計算能力を手にしたものだから、ますますモデルを複雑化し、複雑化すればするだけ高級な計算をしている気になっていく、しかし、そこにはまたわからないパラメータがいっぱい入ってくるわけですね。

結局のところ、現実と、計算結果を適当に合わせていくというのは、アジャスタブルパラメータをカリブレーションというかベリフィケーションしていきながら計算するという、そういうようなことをなさることがデータ同化ということの中心なのでしょうか。つまり私が伺いたいのは、巨大コンピュータに対して、別にどういう意図があるつもりはないのですが、我々としてはともかく勘であったり、電卓であったり、それくらいで計算できて現実が目に浮かぶように、ある程度の精度で判断できればそれで充分だし、そっちの方が望ましいのですよね。ともかくデータに合わせてシミュレーションの方の精度を高めて行き、パラメータを合わせていったところで、何が面白いのかというあたりが、ぴんとこないところがあります。

○ 樋口データ同化研究開発センター長

合わせることが主目的ではないです。

○ 鈴木顧問

合わせることによって、モデルを完成に近づけるということではないのですか。

○ **樋口データ同化研究開発センター長**
合わせることによって、例えばデータとの乖離の視点で、遊びの多い部分のパラメータですとか、あまりいじっても結果に影響のないパラメータというのをシステマティックに読み解くようなことができます。

○ **鈴木顧問**
そういうセンシティビティ・アナリシスをやっていってモデルを簡略化していくことができるということは理解できます。

○ **樋口データ同化研究開発センター長**
簡略化できればよいかもしれませんが、実は私は、簡略化されたモデルがよいとは思いません。簡略化された上で充分予測能力が高いモデルも作ることができます。

○ **鈴木顧問**
ただ、それは、特定の場合にのみ有効となるということでしょうか。要するに一般化ができないですよね。そのデータの範囲の中では簡略化されたけれども、ちょっと条件が変わればまたせっかくのものが違ってきますよね。

○ **樋口データ同化研究開発センター長**
それは状況に依存したものになります。医学への応用の文脈ですと、患者さんごとによって全部変わってくることになります。ですが、私は変わってくることこそが本質的ではないかと思います。その変わってくる中で、どれが一番トリガーになっているのかというところを、シミュレーションを通して逆読みするということが重要ではないかと思います。

○ **鈴木顧問**
そうすると、データ同化という言葉の意味は、何を指すのでしょうか。

○ **樋口データ同化研究開発センター長**
モデルを通してデータをモデルに同化させる。違った言い方をすれば、データをモデルの中にいれる、つまりモデルがデータから学習するということを指します。

○ **真鍋顧問**
データ同化という場合、「同化」は英語ではどうなりますか。

○ **樋口データ同化研究開発センター長**
「Assimilation」になります。

○ **田中顧問**
量的だけではないのでしょう。関数系そのものが、人間が、気がついていないものも教えてくれるという。

○ **樋口データ同化研究開発センター長**

そこまでの機能はないです。やはり手持ちの有限の集合の中から、自分たちが思いもつかなかったような組み合わせとか、そういうことは教えてくれますけれども、思いもつかないものを手持ちの集合の外から探し出すまでの能力はありません。

○ **司会**

吉田先生。

○ **吉田顧問**

コメントと質問なのですが、吉川先生が、経済分野への応用ということで質問されたので、それに関連して知りたいと思うのですが、私どもも一応金融政策の当局なのですが、モデルも持っていますし、モデルで予想もして、あたらないことも多いのですが、毎日出てくる経済統計を見て一喜一憂しながらやっているということなのですが、データの同化という作業をなさる中で、今私が伺っているところ、刻々入ってくるデータというものは、それはもうその時点で数字として確実なデータが入ってきているようなものではないですか。

○ **樋口データ同化研究開発センター長**

そんなことはありません。もちろん計算上は、どの程度不確実性があるかということをある程度分かっていた方が計算は早いですが、その不確実性までもアンノウンにして、同化させることによって、どこに不確実性があるかという、ある意味センシティビティ・アナリシスみたいなことは可能です。

○ **吉田顧問**

そうすると、すでに私の質問にすでに答えられているような気がするのですが、実は私どもが見ているデータの中で、一回発表されると、あとで絶対変わらないもの、例えば短観 DI とか、二度と調査しない、変わらないのですけれども、鉱工業生産とかはあとで変わりますし、一番緩いのが GDP ですよね。プラス成長だったのがマイナス成長に後になったりするように、後になって変わってしまったりするデータがあるのですが、そういう不確実なデータが入ってきても、この同化というような作業は基本的に問題なくできるということでしょうか。

○ **樋口データ同化研究開発センター長**

合わせることによってシステムの中で、どういう癖があるのかというものを逆読みしようという、それがデータ同化の目的の一つです。

○ **吉田顧問**

そうすると、むしろ我々は勘というか、このデータあまり信用できないから、話半分くらいにしておこうとやっているのをシステマティックにできるようになると。

○ **樋口データ同化研究開発センター長**
そういうのを目指していく研究領域です。

○ **吉田顧問**
ありがとうございました。

○ **司会**
甘利先生。

○ **甘利顧問**
大変面白くて、今の説明にあったように、同化モデルは順モデルである、観測されるのはその結果出てきたデータであって、データは順モデルから生成されるから、逆モデルと順モデルをうまく融合させることが必要である、それは非常にすっきりした説明だと思いますね。

しかし、今の例えば金融データなどを見ると、本当にうまい順モデルがあるのかというようなものがいっぱいあるのですね。金融データ、地球環境データ、それから例えばゲノムの発現データ、細胞内の流動ならよいのかもしれないけれど、そういうふうに順モデルがあまりはっきりしなかったり、いくらでも作れるというような状況の中で、ただ一方で、現実にデータは非常にたくさんある、そういうときにデータ同化アプローチというのはどういうふうに発展させていくべきなのか、その展望をぜひお聞きしたい。

○ **樋口データ同化研究開発センター長**
非常に重要な御質問だと思います。特にライフサイエンスのモデルというのは、他の分野での色々なシミュレーションモデルと比較しますと、表現能力の観点から、なかなか、プアであるといわざるを得ない、先生がおっしゃったような状況です。そうしたときには、単なる逆だけで、データからだけでもモデルを作ることはできますので、そういうモデルとどちらが優越なのか、データ同化のようにフォワードモデルを入れるのが有益かということはわかりません。

ただし、ある意味、入れることによって、そのモデルの有効性というのは量的に判断することができるので、モデル自体はデータ説明能力といった全体的視点からはプアでさほど使い物にならないかもしれないけれど、どれくらい有効なのかというのを量的にはかることができるのではないかと思います。ただ、先生のおっしゃったところの分野では現時点では、あまりデータ同化がうまく機能しない研究領域だと感じています。

○ **田中顧問**
赤池弘次先生に私は色々教えていただいたのですが、今、データというのが色々あるのだというお話なのですが、もし、それまでそれなりに説明力のあるモデルがあっても、今まで考えてもいないような状況になったら、過去のデータを落として、直近のデータで推測、例えば、海を航行している船長さんが、突風が突然来た、そのときにベテランの船長は、普通は起きないことが起きたのだから、突風が吹いてきた方向に行先をシフトさせる、

次また来たときに、そのときに、船の安全を維持するという経験値を持っている。

モデリングからいくと、今まで起きないことが起きた場合は、過去のデータを落としてしまって、直近のもののいくつかでパラメータの再推計を行ってそれに備えると。データはもちろんいっぱいあるのだけれども、モデルは全部を予測しきれていないわけだから、何かが起きた場合は、それに向き合うために、データ処理を、先ほど言われたフォワードとバックワードはそういうことと理解してよいですか。

○　樋口データ同化研究開発センター長

そういうことも部分的にはあると思います。イナーシャとしては、過去の知見等もフォワード推論の種になります。データが入ってきたときに、ある程度の過去のものを捨てて、データから学ぶというのがある意味逆向きの推論と言え、それをイナーシャとどう帳尻を合わせて按配をうまく決めるのか、それがデータ同化であると、そのように私は解釈しています。

○　司会

まだ議論を続けたいところなのですが、だいぶ時間の方も押してまいりました。この後、調査科学 NOE の説明を行う予定だったのですが、ここでもう時間の半分くらいですので、一度休憩を入れさせていただければと思います。

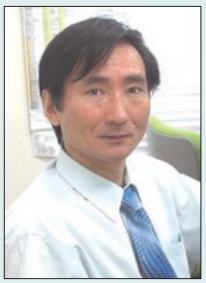
（１５分間の休憩）

Survey Science NOE

統計数理研究所　NOE形成事業 顧問会議

調査科学NOE

概要説明



調査科学研究センター
センター長　吉野　諒三

統 計 数 理 研 究 所
調査科学NOE概要説明

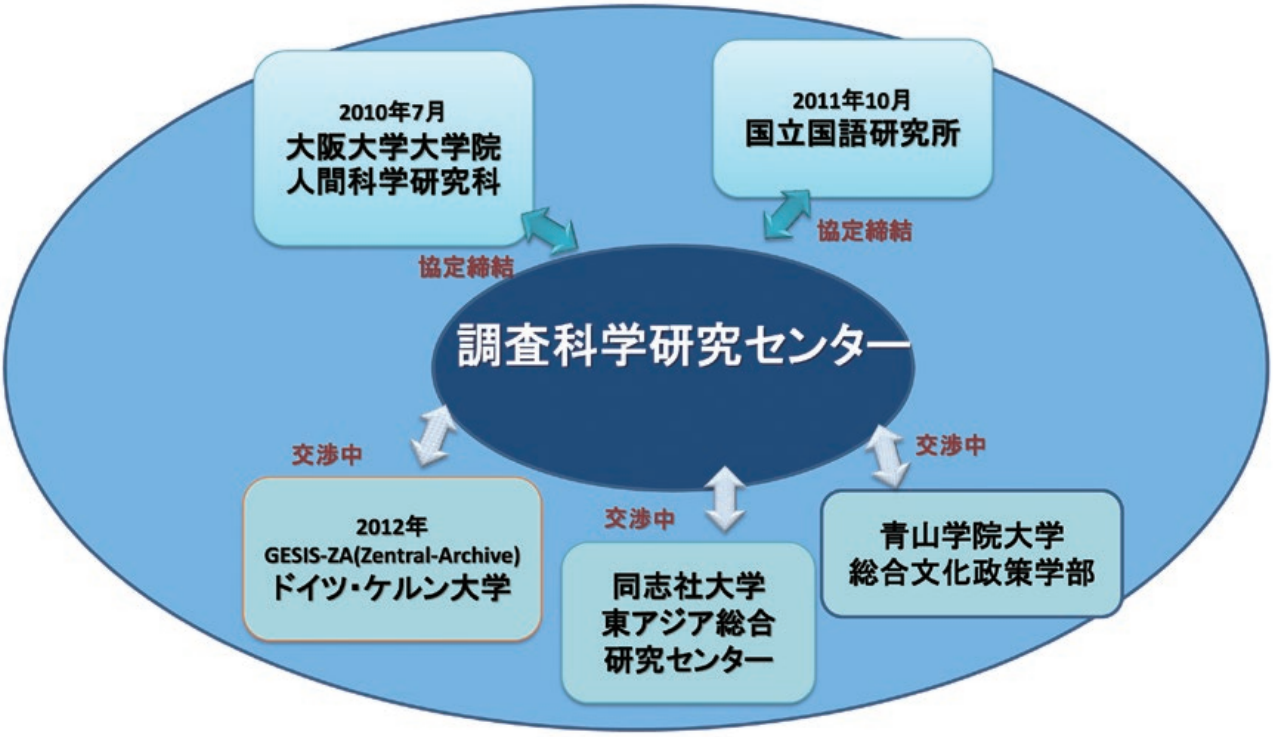
統 計 数 理 研 究 所
NOE形成事業

調査科学研究センター
センター長　吉野　諒三

調査科学NOEの概要

統計数理研究所の半世紀以上にわたる社会調査研究の成果を基盤として，そのさらなる発展と関連諸機関との連携，及び社会的貢献の促進のために，調査科学NOE (Network Of Excellence)の構想が持ち上がり，その効率的な実現を目的として，2011年(平成23年)1月に「調査科学研究センター」を設置しました。
国内外の大学や調査機関との連携を展開し，
(1)調査科学NOEの構築
(2)調査科学の学術的基盤の充実
(3) 連携研修調査を通じた若手人材の育成
(4)調査リテラシーの向上
(5) 調査データ資料の蓄積
(6)国際比較調査を通じた国際的相互理解の促進
等を念頭に各種の調査プロジェクトを推進しています。

調査科学NOE連携協定



- **表1　統計数理研究所の主な大規模標本抽出調査**

- **1953-現在　日本人の国民性　　（機関研究）**
- 1971　　ハワイ在住の日系人
- 1978　　ハワイ住民（日系人を含む）米国本土アメリカ人
- 1983　　同上
- 1985-1993　日米欧7か国比較
- 1992　　ブラジルの日系人
- 1998-1999　米国本土（西海岸）の日系人
- 1999-2000　ハワイ住民（日系人を含む）
- 2001-2001　中国（北京，上海）
- 2002-2005　東アジア価値観国際比較調査
- 2004-2009　環太平洋価値観国際比較調査
- **2010-現在　進行中　アジア・太平洋価値観国際比較調査**
  （科学研究費補助金：　基盤研究S）

◎ 調査科学研究センター／活動目的

- □ 調査データ資源の有効活用
  - ● 統計数理研究所において，50年以上にわたり集積されてきた日本人の国民性調査および国際比較調査，関連する諸調査による調査データ資源を順次，一般社会で広く有効活用できる体制に整える。
- □ 共同利用研究制度の活用
  - ● 共同利用研究制度を通じて，高度な解析研究が可能な調査データ資源をできる体制を整える．特に国内外の大学・研究所・諸機関との全国レベルの調査遂行のための連携を進め，実践的社会調査の研究者及び学生の教育・啓発を図る。
- □ 調査方法論の研究基盤の有効活用
  - ● 調査データ資源の集積過程で，研究・開発してきた標本調査法・社会調査法の研究を深化させ，その展開と普及を図る。特に，各地方の拠点となる国内大学と連携し，社会調査に携わる研究者及び学生に実践的調査研究の機会を提供する。
- □ 調査データ解析方法論の有効活用
  - ● 研究・開発してきた社会調査データの統計的分析方法を基盤に一層の展開と普及推進を図る。
- □ 公開講座を通じた教育・啓発
  - ● 一般社団法人社会調査協会と提携し，学生・一般社会人が同協会の「社会調査士」及び「専門社会調査士」を取得するための幾つかの科目となる講座を担当し，社会調査教育の充実および社会人の啓発を図る。

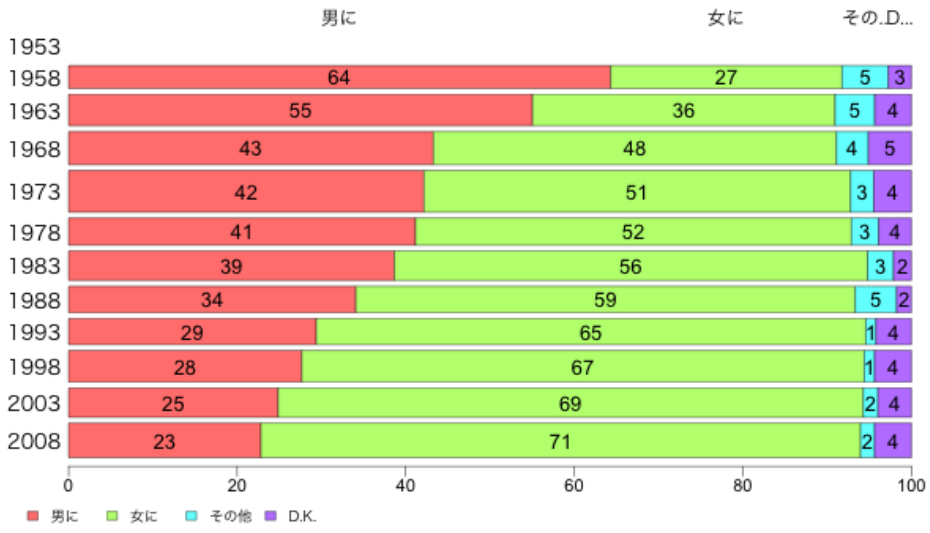
The Institute of Statistical Mathematics

◎ 調査科学研究センター／プロジェクト

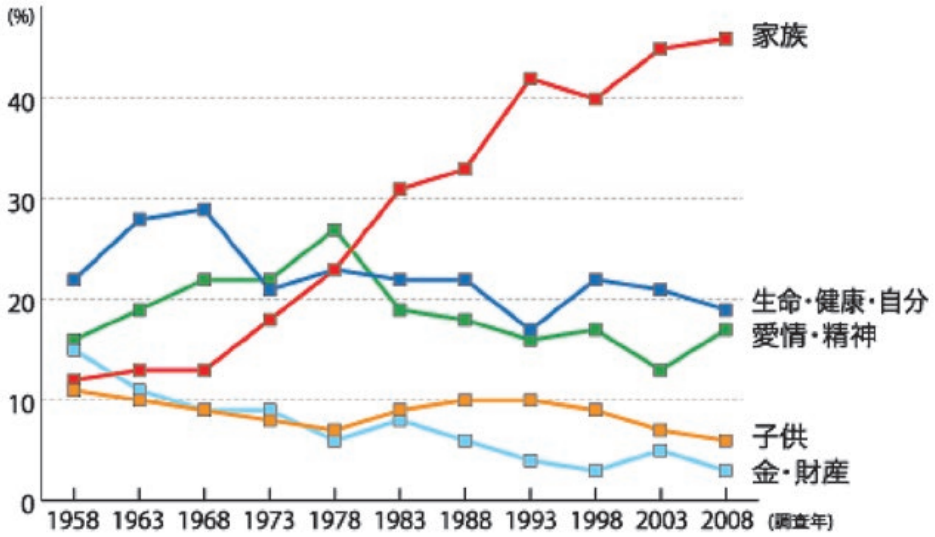
- □ 日本人の国民性調査プロジェクト　（リーダー：　中村　隆）
  - ● 「日本人の国民性調査」を5年ごとに実施し，日本人の意識動向を明らかにする。
- □ 国民性の国際比較プロジェクト　（リーダー：　吉野諒三）
  - ● アジアを中心とした諸外国において国際比較調査を実施し，異文化理解の促進に役立てる。
- □ 社会調査情報集積プロジェクト　（リーダー：　土屋隆裕）
  - ● 社会調査の結果データ等を集積・データベース化し，二次分析等のために提供していく。
- □ 連携研修調査実践プロジェクト　（リーダー：　前田忠彦）
  - ● 他機関と連携しながら社会調査を実施し，各地の若手研究者を実践的に育てる。
- □ 社会調査情報活用プロジェクト　（リーダー：　尾崎幸謙）
  - ● 社会調査データを活用して新たな統計分析の手法を開発し，統計科学の発展に寄与する。

Research Organization of Information and Systems
The Institute of Statistical Mathematics

もういちど生まれかわるとしたら，あなたは男と女の，どちらに，生れてきたいと思いますか？



あなたにとって「一番大切なもの」は何ですか?



■■■■国際比較調査データアーカイブ■■■■

Dataarchive of
Cross-National Comparative Survey of National Character

・
・
はしがき
Preface
・
・
環太平洋価値観国際比較調査
The Asia & PacificValue Survey (In English)
(2005-2008年)
報告書正誤表(Corrigenda)
環太平洋価値観国際比較調査総合報告書(Summary Report)
・
・
東アジア価値観国際比較調査：2004年日本A・B調査
East Asia Value Survey: Japan 2004 A/B Survey (In English)
(2004年)
東アジア価値観国際比較調査総合報告書(Summary Report)
・
・
東アジア価値観国際比較調査
East Asia Value Survey (In English)
(2002-2005年)
・

◎ 調査科学研究センターの 連携調査

□NOEの構築：研究分野における新しい方法論の確立と，異分野交流のハブの役割を果たすことを目指す。

□NOE活動の一環としての「連携調査」「連携研修調査」

統計数理研究所だけの調査研究ではなく

- より開かれた調査研究の実践
- 調査科学の研究の裾野を広げる（人材育成）

Tadahiko Maeda (June 27th, 2011)

Research Organization of Information and Systems
The Institute of Statistical Mathematics

◎ 現在の連携調査の関係メンバー

□ 大阪大学との連携（2010年度調査）

吉川 徹 （客員准教授，大阪大学人間科学研究科）

□ 国立国語研究所との連携（2011年度言語調査）

米田正人（国立国語研究所名誉所員）

●阿部貴人（客員准教授，国語研究所）

□ 東アジア等「国際比較調査」での連携

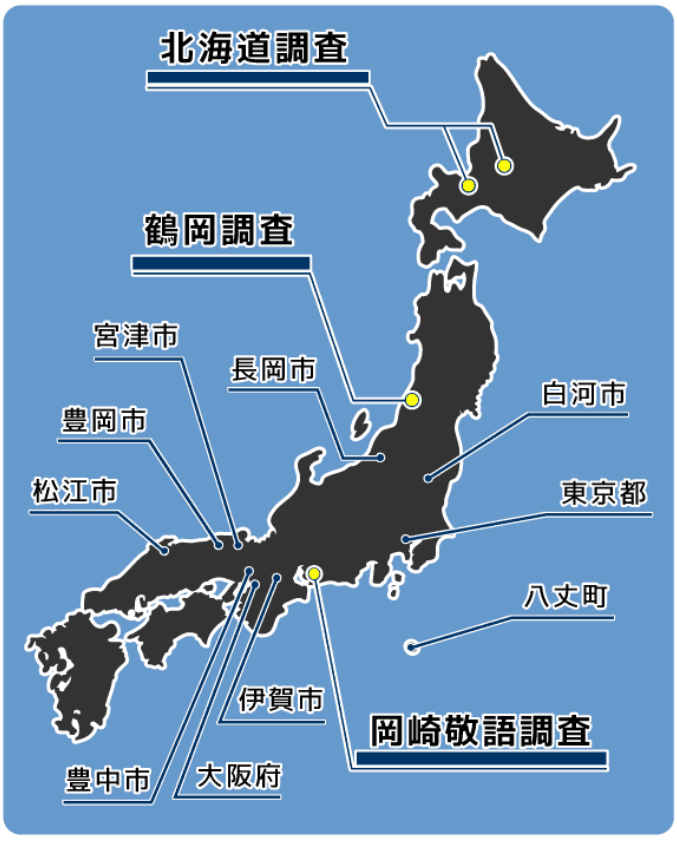
●鄭　躍軍（同志社大学東アジア総合研究センター長）

Research Organization of Information and Systems
The Institute of Statistical Mathematics

科学研究費補助金研究　**基盤研究(S)**　**2011年度より**

SSPプロジェクト

総格差社会日本を読み解く調査科学



鶴岡2011年度調査

国立国語研究所は昭和23（1948）年に設立されました。設立の目的の1つは，日本の言語生活の実態を明らかにすることでした。設立の翌年（1949年）には，福島県白河市をモデル地域に選んで言語生活の実態調査を行い，多くの知見を得ることができました。

そこで得られた知見は，日本全国の他の地域にもあてはまるものなのでしょうか。それを確かめるためには，他の地点でも同様の実態調査を行う必要があります。国立国語研究所は，地域社会における言語生活を明らかにする実態調査を全国に展開することを考えました。福島県白河市に次いで実施されたのが，**昭和25（1950）年の山形県鶴岡市における調査**です。鶴岡市の言語生活の実態を知るために，サンプリング（無作為抽出）という方法で鶴岡市民の代表を選びました。

日本の地域社会には**方言**があります。一方で，首都圏を中心とする地域では，あるいは地域社会でも相手や状況によっては，**共通語**が使われることがあります。日本における言語生活を捉えるうえで，この2つの使用実態を知ることが不可欠です。そのため，方言と共通語に関する調査項目を設け，鶴岡市の言語生活の実態を明らかにしました。

日本学術会議　大型研究計画申請がマスタープランに採択

社会科学統合データベース・ソリューション網の形成
（Network Building of the Integrated Social Science Database Solution)

今田高俊　　東京工業大学大学院社会理工学研究科

5か年計画　予算　90億円

社会科学の研究拠点を結んだ「データベース・ソリューション網」を整備し，人材育成や技術開発にともなって生じる諸課題の解決および制度づくりの提言をおこなうことで，持続可能な社会づくりの先端研究を推進する。

学術会議の各学会の連合で，いわば「社会科学のNOE」を展開し，それを支える基盤として統計数理研究所が期待されている。

【調査科学 NOE 概要説明】

○ 司会

それでは、再開させていただければと思います。次は調査科学 NOE の説明をさせていただきます。資料はスライドと併せて御覧ください。

○ 吉野調査科学研究センター長

吉野です。よろしくお願いいたします。調査科学 NOE の概要を御覧下さい。統計数理研究所の半世紀以上にわたる社会調査研究の成果を基盤として、そのさらなる発展と関連諸機関との連携、及び社会的貢献の促進のために、調査科学 NOE の構想が持ち上がり、その効率的な実現を目的として、2011 年（平成 23 年）1 月に正式に発足しました。2010 年度に文科省の大型の科学研究費の基盤研究 S を五か年計画でいただきまして、実際には活動としては 2010 年から始まっておりましたが、翌年の 1 月にセンターが発足という形になりました。

調査科学研究センターの大義名分が、いくつか書いてございます。調査科学 NOE の構築、調査科学の学術的基盤の充実、連携研修調査を通じた若手人材の育成、調査リテラシーの向上、調査データ資料の蓄積、国際比較調査を通じた国際的相互理解の促進です。

統計数理研究所は、冒頭で樋口所長からお話があったように、戦時中にできました。戦争が始まる前に、「世の中の課題に直面して、それにふさわしい統計を活用したり、それにふさわしい統計を開発したりするという、『統計数理』という新たな概念が必要である」という統計科学の議論がありました。この流れが、戦時体制に結びつき設立されたのでした。

しかし、戦後は、日本の民主主義を発展させるために、科学的な世論調査の方法論を確立させないといけない、日本人の調査関係者の皆で、科学的な方法論に基づいた世論調査のシステムを確立せよ、という指令があったそうで、統数研は戦後の新たな使命を帯び、その後、御存知のように、本日、皆が説明しているように、統計の非常に幅広い分野で活躍しています。我々の調査の部門は、一応そのような世論調査のシステム確立の流れの中での仕事が続いています。

昨年（2011 年）の 1 月にセンターが発足して、まず 2010 年に大阪大学大学院と連携協定を結びまして、調査研究を行っています。これは後にまた説明いたします。さらに翌年度に国立国語研究所とも連携協定を結びまして、色々な調査を進めています。国語研ももともと各種の立派な調査をしていたのですが、ここのところそのような調査研究が低調になっていたようですが、統数研との連携に刺激を受けて、新たに色々な重要な調査が継続され、また新たなプロジェクトが始まっています。さらに海外では、ドイツ、ケルン大学、ここはデータ・アーカイヴの主要な拠点があるところですが、ここととの連携を模索しているところです。

また、同志社大学東アジア総合研究センターは、私の部署にいた中国人研究者が移動し、向こうの大学で新たなセンターを作り、中国との関係もありまして、現在そことも連携を交渉しています。青山学院大学総合文化政策学部は、本日御出席の真鍋先生との関係でこれまで国際比較研究を協調してやってきましたが、これをもっと正式な形での連携を模索しているところでございます。

我々の調査研究をお話ししますと、1953（昭和 28）年に「日本人の国民性調査」が開始され、これは機関研究という形で、統数研の資金で継続しています。これは 1948（昭和 23）年に「日本人の読み書き能力調査」というのがございました。「日本人の読み書き能力調査」というのは、日本の民主主義を発展させるために、人々が自分の意見を自由に言える必要条件としての能力があるかという調査をしたわけです。そのときに、国勢調査のような全数調査をやっていたら時間がかかりすぎるので政策立案に間に合わないから、サンプリング調査をしなくてはいけない。そのために、実践的なサンプリング方法を統数研が中心になって研究して、その当時、国立国語研究所はまだできていなかったのですが、後の国語研究所の有力な先生になる若い言語学者たち、社会学者、心理学者、教育学者等、広い分野の人たちが絡まって調査を行いました。そのときに確立した方法論を利用して、「日本人の国民性調査」が昭和 28 年に開始されたのでした。

その後、ずっと 5 年ごとに継続されてきて、来年が 60 年目になり、現在、次回の調査の準備をしているところでございます。そして、40 年ほど前から、日本人だけ調べていて、日本人の国民性について本当にわかるのか、いかに他の国と違っているかを比べてみる必要があるのではないかと、ハワイ在住の日系人との比較調査というものが始められました。

いきなり全然違う国々と比べても、ただ違うという話で終わるだけです。ある程度似ているところもあるだろうけれど、違うところもあるだろうという人々を比べてみて、初めて計量的比較の意味があるという発想で、始められました。そういう比較の連鎖を拡張していけば、やがては世界中の人々も比較できるのではないかという発想です。

「日本人の国民性調査」から、一つ見本の質問をいたします。「もし、生まれ変われるとしたら今度は男がいいですか？今度は女がいいですか？」

不思議なことに、50 年以上にわたって男性の回答は、全然変わらなくて、9 割方は「また男がいい」といいます。数パーセントだけが女がいい、あと 2～3 パーセントは「わからない」とか「今度生まれ変わったら神様がいい」とか、そういう回答です。

ところが、女性の回答は、1958（昭和 33）年には、「今度は男がよい」というのが 64%、「また女がよい」というのが 27%。それが時代とともに、どんどんどんどん変わってきて、90 年代初めには、今言った数字が全く逆転しているわけです。その後も、さらには、「また女がよい」というのが、変化はわずかになりましたが、増えている傾向です。

考えてみますと戦後、日本は色々な面で変わってきました。良いことも悪いこともありました。しかし、いずれにせよ、この回答分布データは、男女の問題だけではなくて、日本の戦後の社会の大きな変化を象徴しているのではないかと思います。

表 1 は「日本人の国民性調査」から始まる統計数理研究所の主要な調査のリストですが、1970 年ごろに国際比較調査が海外で開始されました（注：国内での国際比較は先行してやってあった）。その開始時には、私もちろん統数研に入る前で、まだ学生だったのですが、平成元年に私が統数研に入りまして、この国際比較の部門に入り、林知己夫先生と一緒に調査を続けてきました。この過去 10 年間ほどは東アジアに着目しまして、まず中国調査に入って行き、中国での協力者を求めて展開してきました。

先ほど申しましたが、「日本人の国民性調査」は機関研究として、統計数理研究所の資金でやっていますけれども、国際比較調査は、科学研究費をその都度申請して、採択されたときに実施しているのでこのように不定期な形になります。「東アジア価値観国際比較」、

「環太平洋価値観国際比較」、そして現在、「アジア・太平洋価値観国際比較」を実施していますが、主として、日本、韓国、中国は全土では調査できませんが、大陸では北京、上海、そして香港、台湾、さらにシンガポール、オーストラリア、アメリカ、それからインドをカヴァーしています。統計的なサンプリング調査ができ得るところを選んでいます。

調査科学研究センターの他の任務としまして、「調査データ資源の有効活用」は、過去に統数研に蓄積してきたデータを有効に活用しましょうということです。「共同利用研究制度の利用」は、統数研の共同利用機関としてのシステムがございまして、一つ一つは小規模ですが、統数研のメンバーと各大学や研究機関の人たちとの「共同研究制度」をさらに活用していこうという意図です。「調査方法論の研究基盤の有効活用」というのは、今まで調査してくる中で開発してきた標本調査法や社会調査法の研究を深めていって、その展開と普及を図り、それに基づいて国内外の大学と連携して、人材を育成しましょうというのが趣旨です。次の「調査データ解析方法論の有効活用」というのは、そういった調査の中で開発した統計、分析方法をさらに深め、また普及することを図りましょうということです。「公開講座を通じた教育・啓発」というのは、我々のセンターに限りませんけれども、統数研では、長い間、公開講座を通じて、大学の先生も含めて、一般の方々に、講演して、色々な統計の分野のことを啓蒙してきました。今は各分野で大先生になっている方も、若い頃これに出たことがある方もいらっしゃると思いますけれども、調査の部門でもこれをさらに進めて行きましょうというところです。

我々の分野は、センターといいながらも、予算も、メンバーも限られておりますが、プロジェクト制にして、それぞれが一つのプロジェクトのリーダーになって、進めるようにしています。「日本人の国民性調査」、「国際比較」、「社会調査情報収集」、「データベースとその公開」ですね。「連携研修調査実践プロジェクト」というのは、他の大学や機関と連携しながら、社会調査を実施して、その中で若手の研究者を育てていきましょうというものです。最後の「社会調査情報活用プロジェクト」は、社会調査データを活用して新たな統計分析の手法を開発して統計科学の発展に寄与しましょうというものです。

御覧の印刷資料は、統数研の HP から持ってきたものですが、一番初めは 90 年代半ばだと思いますが、我々が先輩たちと一緒にやった「日米欧 7 か国国際比較調査」というのがございまして、それを、初めて統数研の研究紹介として HP に載せました。そのときは私たちのものしかなかったので、当時の所長もそれを所外の皆さんに説明したり、啓蒙したりしていたのですが、その後、システムも技術も発達して、どんどんお金や手間をかけてきれいなものが出来ていますので、もしお時間があったらいくつか見ていただければありがたいと思っています。

資料の次のページは、先ほど話に出ました「もう一度生まれ変わるとしたら、あなたは男と女のどちらに生まれてきたいと思いますか？」の調査のデータでありますが、各質問に関して、このようなグラフがでています。もう一つ「あなたにとって『一番大切なもの』は何ですか？」と聞くと、「家族」が浮かび上がってきました。どうもデータから見ますと、「家族」の大切さというのが、世界に普遍的な価値観であるようです。

時間が限られて来ましたので、調査科学 NOE について今後を占う上で重要だと思う成果を二つだけ述べておきます。もともと資金が限られていたのですが、他大学とあるいは他研究機関との連携の中で、まず大阪大学との連携が始まった時、大阪大学は調査資金がな

く、我々の資金で小規模な全国調査をやったのですが、その後、それをもとに、大阪大学の先生が、大型の科学研究費を得ることに成功しました。そういう意味では、小さな種をまいて、大きな成果を得たといえると思います。それからもう一つ、国立国語研究所の話です。国語研究所と統数研の関係は長く、昭和 20 年代から重要な調査をずっと進めていたわけなのですが、残念ながら、重要な調査を遂行すべき時期が来た年だったのですが、内部のメンバーの人たちが変わってきたせいなのか、関心を持ってくれなかったため、国語研 OB の人が私のところに相談に来ました。そこで我々のセンターとの連携で、調査をやろうということになりました。その OB の人が、その話を国語研究所に持って帰ったら、統数研とやるようなそんなに重要な調査だったのかと気が付いて、「今年は統数研の資金による調査になるが、来年度は自分たちでやろう」、と、また火が点いたというところです。そのように小規模ではありますけれども、各方面でまいた種が少しずつ花が開き始めたかというところでございます。

資料の最後のページのところは、現時点では、詳細なことは全部言えませんけれども、日本学術会議の方でも社会調査データ関係の話でデータベース機関を作ることで、少し大きな話が進んでいます。まだ確定ではありませんけれども、当然我々も関与しているところです。

以上です。

○ 司会

それでは調査科学 NOE に関して御質問、コメント等あればよろしくお願いいたします。吉川先生。

○ 吉川顧問

質問なのですが、最近よく、社会調査ということなのですが、医学的な調査というものがあるわけですね。コホートというのがあって、これを伺っているといわゆる医学といっても病気のデータを取ればよいのではなくて、どんな生活をしているとか、イギリスでは確か、100 年以上、ものすごく長くやっているわけですが、子どものときからある人を特定したら、それをずっと一生追跡してやっていて、そのためには、現在、どういう生活をしているか、連絡できないといけないわけですね。それをどうやっていやっているのか聞いたら、毎年必ずクリスマスカードを送っている、そうしないとすぐ連絡できなくなってしまうわけです。そういったような調査が今、医学的な分野に進んでいるのですが、そういった方面はどういうことをお考えですか。

○ 吉野調査科学研究センター長

我々の分野では、先ほど申しましたように、中身は学術調査なのですが、方法論としては世論調査としての方法論を研究するところなので、そちらに専念しています。ただ、共同研究として他の大学の先生たちと関わっている場合で、吉川先生の御指摘のような調査はよくあると思います。本研究所の中村先生はパネル調査（コホート研究のための追跡調査）のときに関わっています。我々のような社会調査でも、数年か 5 年か 10 年のタイムで生活調査をする場合はパネル調査になるわけです。そういう場合はおっしゃったようなク

リスマスカードを用いたり、あるいは違う方法で、登録サンプルをやはり更新していきます。そのときは個人情報の保護を含め、名簿の管理が大変になります。普通の世論調査とは違ってきます。

○ **司会**

調査科学研究センターでは、コホートという観点からの研究というと中村先生が専門かと思いますので、中村先生。

○ **中村教授**

日本人の国民性調査は 1953 年来 5 年ごとに行われておりまして、その回ごとに、異なる対象者を調査しているのですが、こういう継続的な調査データの中から年齢によって意見が変わる部分とか、世代（コホート、コウホート）によって異なる意見を持っているとか、そういったことで世代効果とか年齢効果とか、そしてその時々に皆の意見が変わる時代効果とかを分ける方法とについて私自身が研究してまいりました。

それから、もう一つの研究、コホート研究というのは、同じ対象者を追跡していくのですが、ちょうど国立国語研究所との、最近復活した共同調査なのですが、1950 年からほぼ 20 年ごとに調査がされておりまして、その調査回ごとにサンプリングによって新たな対象者を選ぶ以外に、1 度サンプリングされた同じ対象者を、1950 年からほぼ 20 年経った 1971 年、また 20 年たった 1991 年と追跡して、鶴岡方言がどのように共通語化していくかというプロセスを捉えるパネル調査というのもあり、ちょうど 60 年後にあたる昨年（2011 年）に第 4 回調査をやりました。第 1 回の調査対象者は 500 名くらいだったのですが、60 年後の第 4 回の調査では 4 人くらいになりました。そういう調査も行っておりまして、方言というのは言語生活なので、比較的変化が緩やかで、しかも、世代によって方言を話す割合というのが違うのですが、そういう中で、個人の変化も抑えられるという研究方法といいますか、調査方法をとっています。

医学や疫学の分野ではだいぶコホート研究が日本でも行われるようになっております。普通の世論調査でも、今回の震災のようなときに、個人の意見がどのように変わって、またそれが落ち着いてくるのかということがあり、我々も来年（2013 年）実施する第 13 次日本人の国民性調査の周辺に、パネル調査タイプの研究も入れまして、個人の意見の変化というのをとらえてみようと、本年度（2012 年度）獲得した科学研究費 A でやるようにしております。

○ **司会**

佐和先生。

○ **佐和顧問**

日本人の国民性調査のデータを、私自身、著書の中で引用させていただいたことがあります。ネットのページを見ればわかることかもしれませんが、例えば、調査の方法、回収率、回収率が低下しているのか上昇しているのか、調査方法の抜本的な見直しがあったのかについて教えてください。

○ **吉野調査科学研究センター長**

先ほども申しましたように、我々の調査は、中身は学術調査なのですが、世論調査の方法論の研究として遂行しています。世論調査の場合は、成人は一人一票の原則です。そうすると、サンプルとして抽出確率が個人レベルで等しい限り、各個人の回答者に調査を拒否されたら、その拒否した個人の分は補わない、代替サンプルを用いないわけです。国民性調査も確かに回収率だけ見ると、時代とともにどんどん下がって、今では 50%をかろうじてまだきってはいない程度のところまで落ちてきました。そこだけ見ると、日本人の半分しか答えてないのに、日本人の意見、世論といえるのかというコメントが出るわけです。

しかし、世論調査に限っては「拒否も回答のうち」と私は考えています（注：選挙で投票率が低くとも、選挙自体は無効にならない）。ただし、拒否だけではなく、回答者として選ばれた個人に、面接員がなかなか接触できない人が増えているところは大きな問題であるところです。普通の学術調査であれば、性別、年齢の偏りというのが、ある問題に関して決定的な要因と分かっているのであれば、その知見を利用して補正したり、サンプリングのウエイトを工夫したりすることがあり得ます。けれども、世論調査に限っては今のところ、やはり誰でも一人一票、等しいウエイトのサンプリングをするという理念を尊重して、ウエイト調整などの加工は行わないということでずっと続いています。それは政府の調査もそうです。

○ **佐和顧問**

民間のシンクタンクなどが世論調査をするときに、昨今では、インターネット調査が多くなっていますよね。私自身は、インターネット調査には有意なバイアスがあると、かねがね思っています。

もう一点、新聞社や放送局が例えば政党支持率、内閣支持率の調査をやりますが、乱数を発生させて、固定電話にかけているわけです。固定電話に出るのはどういう人かを考えてみると、相当程度のバイアスがありそうです。その点を弁えた上で、各社とも、属性を尋ねていますから、しかるべき調整をします。生のデータをそのまま載せるのではなく、だいぶいじっているみたいです。その辺の調査の難しさについて何かコメントをいただけないでしょうか。

○ **吉野調査科学研究センター長**

はい。戦後の調査が始まったときの説明で、先ほど「科学的な世論調査」を強調しました。「科学的」というのは、標本抽出誤差（データの精度）が計算できるということです。回収率の問題は別としておいて、サンプリング計画の時点で標本抽出誤差が計算できるかということです。その場合、ユニバース（調査対象）、ポピュレーションは日本人の有権者全体です。ところが、ウェブ調査も、RDD 電話調査も標本抽出誤差の計算ができていないわけです。ウェブにアクセスできない人はそもそも省かれるし、RDD もそうで、日本人の有権者全体をカヴァーしていない。その問題は、再々議論してきました。

世論調査に電話が入ってきたのは 90 年代初めくらいかと思いますが、初め導入されてきたときは、あくまで試行的に、本当に面接調査をやったときと、電話調査のデータを比べるとどれだけ違うか違わないかと慎重に検討を重ねていました。しかし、55 年体制が崩壊

し、政党がバラバラになって、小選挙区制が入ってきました。そうすると、統計的には各小選挙区が無限母集団となるので、選挙予測では、調査すべき無限母集団（調査区）が激増してしまいました。

ある新聞社の RDD 調査ですが、国政選挙での予測調査を行うと５億円かかるといいます。もし、それを面接調査で行うとしたら、それよりもはるかに巨額な資金がかかります。従って、現実的な問題として、マスコミでは、RDD 電話調査以外できなくなってしまっているのです。データバイアスなど、RDD 電話調査の方は世論調査協会のメンバーたちが戦後お互い監視してチェックしてきた体制があるのですが、ウェブ調査の方は、実験比較調査をしたところ、先ほど先生がおっしゃったように、バイアスやウエイト調整等にも、色々な方法が出てきます。

○ 司会

真鍋先生。

○ 真鍋顧問

「調査科学研究センター」の活動に関して、私は、３点に分けて、コメントさせていただきます。

まず第１点は、吉野先生のお話にありましたように 1953 年に「日本人の国民性調査」が始まったということで、この時点でこれだけ綿密で内容の深い大規模な質問紙調査が企画され、実施されることになり、それが現在まで継続されていることは、世界的に見ても例のない、まさに日本が誇るべきこの研究領域における学術成果であると思います。

さらに、この調査研究活動は、日本の調査から、「国際比較調査」へと発展していきました。このような視座の導入を通して、林知己夫先生の Cultural Link Analysis、そして吉野諒三先生の Cultural Manifold Analysis という理論的そして方法論的な枠組みが確立されてきます。この研究領域における世界の現状に照らして、このようなデータの蓄積と理論的・方法論的枠組みの構築は、まさに社会調査の科学の領域における日本の財産と言えるものであると考えます。この意味で、中村隆先生を中心とする「日本人の国民性調査」をはじめ様々な調査活動の継続と発展にさらに期待していきたいと思います。

次に第２点は、社会調査のインフラストラクチャーということです。ここでインフラストラクチャーというのは大規模な社会調査の実施と、そのデータ・セット、そしてデータ・ドキュメンテーションの作成・発表、データ・アーカイヴィングの諸活動を支援する機関ということです。吉野先生は、ドイツの GESIS――German Social Science Infrastructure――の例を出されましたが、GESIS は ALLBUS と呼ばれる「総合社会調査」を継続的に実施するとともに、世界最大のデータ・アーカイヴの活動とともに、データ分析の活動を幅広く行っている機関として世界にその名を馳せています。我々は、日本においては統計数理研究所こそが、このような「包括的なインフラストラクチャー」として発展していくことを心から願っています。

最後に、第３点として、「社会調査データ」といわゆる官庁統計をはじめとする「統計データ」との統合ということです。ドイツの GESIS では連邦・州の諸官庁とのネットワークの上に、このような官庁統計と社会調査データとの統合が行われています。それによって、

社会のオブジェクティヴな側面とサブジェクティヴな側面も含めて、いわゆる「マルチ・レベル・アナリシス」が可能になっています。統計数理研究所においても、このような「マルチ・レベル・アナリシス」が試みられるような方向でNOEが発展していくと素晴らしいと思います。

○　吉野調査科学研究センター長

ありがとうございます。重要な点で、今説明しているセンターそのものが直接するかどうかは別として、官庁統計の方は、例えば椿先生やそのグループの方々が、経済統計の先生とともに、学術会議の方も頑張っています。私も関与していますが、その方面ではこの数年でずいぶん進んでいまして、システムもどんどんできています。

○　椿副所長

はい。基本的に国の官庁統計のデータを研究利用する制度の改善が今進んでいますし、研究所の方でもそれを分析に利用できる環境整備というのを進めているところです。

○　吉野調査科学研究センター長

真鍋先生がおっしゃったポイントは、オブジェクティヴデータとサブジェクティヴデータを同時に解析できる体制のことだと思います。例えば、現在、人々の「格差」が問題になっていますが、経済的な客観的データも、人々の主観的な生活満足度等も同時に見て、多面的に、現実を浮かび上がらせなくてはいけないのだけれども、そこに至っていないのではないかという問題点ですね。ぜひ、そちらの方向へ進めなくてはならないのではないかと強く思っています。

○　司会

どうもありがとうございます。だいぶ時間の方が押して、心配になってまいりましたが、意見交換のときに顧問の先生方からコメントをいただきたいと思っている次第です。

Statistical Machine Learning NOE

統計数理研究所　NOE形成事業 顧問会議

統計的機械学習NOE

概要説明

統計的機械学習研究センター
センター長　福水　健次

統 計 数 理 研 究 所
統計的機械学習NOE概要説明

統 計 数 理 研 究 所

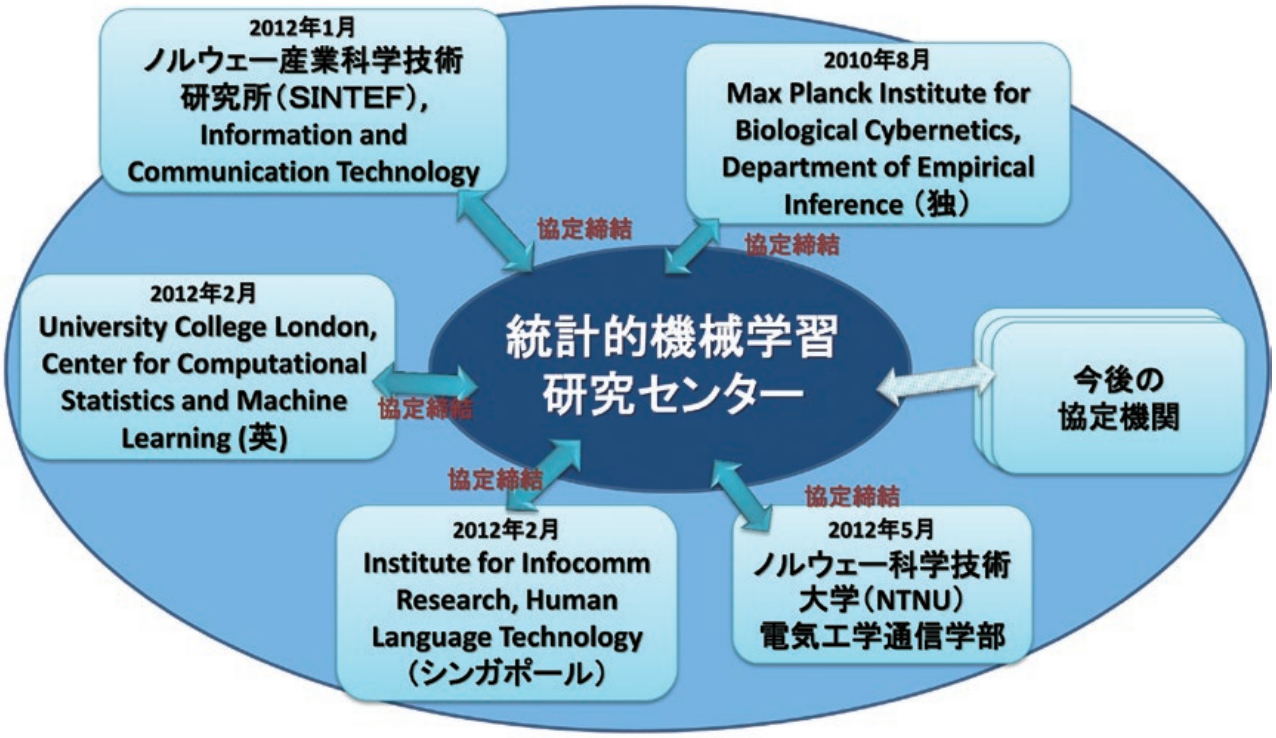
NOE形成事業

統計的機械学習研究センター
センター長　福水　健次

統計的機械学習NOEの概要

統計的機械学習NOEでは，統計科学的観点から機械学習に注目し，国内の研究ミュニティーの活性化や海外の研究機関との共同研究の推進を行っています。それらの活動を通して，統計的機械学習に関する中核的なネットワークを形成し，日本の国際的なプレゼンスを向上させることを目指しています。また，統計科学 ／最適化／データ解析／情報処理／自然科学などの分野を横断して，統計的機械学習の理論から応用，人材育成にわたる研究の進展に資することを目指していま す。統計数理研究所では2012年（平成24年）1月に発足した統計的機械学習研究センター所属の教員・研究員が，その事業の推進を担っています。

統計的機械学習NOE協定図



センターの体制

教員10，PD 1，客員 8，補佐員 2

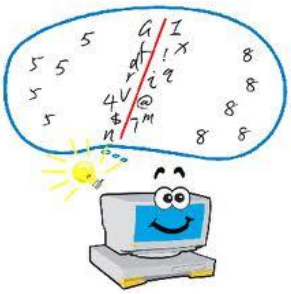
所内

福水健次（センター長，教授）
松井知子（副センター長，教授）
伊藤聡（教授）
江口真透（教授）
宮里義彦（教授）
池田思朗（准教授）
持橋大地（准教授）
小林景（助教）
小山慎介（助教）
伏木忠義（助教）
西山悠（特任研究員）

客員

池上敦子（客員教授，成蹊大）
後藤真孝（客員教授，産総研）
土谷隆（客員教授，政策研大）
津田宏治（客員教授，産総研）
和田山正（客員教授，名工大）
金森敬文（客員准教授，名大）
品野勇治（客員准教授，Konrad-Zuse-Zentrum）
Arthur Gretton（客員准教授，Univ. College London）

機械学習とは？



機械学習 Machine Learning

- 経験・データに基づいて自動的に学習を行うシステムに関する研究分野
- 基盤：「統計科学（データからの推論）」
  ＋ 「計算機科学（アルゴリズム）」
- 「いかに計算するか？」 を念頭においたデータ解析へのアプローチ
  歴史：1960年代から，人工知能の一分野として発生
  (Intern. Conf. ML 2012 = 29th)
  → データと計算機環境の変化：「統計的」機械学習が主流に.
- 広範な応用： ロボット行動制御，情報処理（音声・画像・文書），脳科学，インターネット上のサービス（検索，推薦etc）...
  Google, Yahoo, Microsoft, …

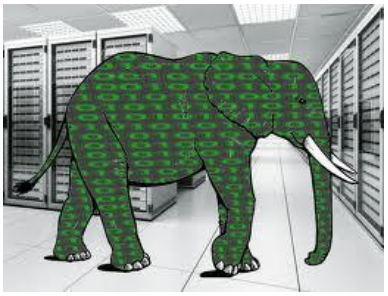
一つの例

Netflix Prize（推薦システム）

Netflix社（米大手レンタルDVD会社）による推薦システムコンペティション

- データ： ユーザ数48万人，映画数18000，
- 見た映画の5段階評価，評価数 1億
- 目標： Netflix社の推薦システムの性能を10%改良する
- 賞金： $1million

2006年10月開始.
2009年6月10%達成



➢ Big data 時代のデータ解析

「いかに計算するか？」が重要

NOE活動

➢ 国内研究コミュニティの活性化

- 電子情報通信学会・情報論的学習理論と機械学習研究会との連携
  情報論的学習理論ワークショップ（年1回）共催など
- 文部科学省「数学・数理科学と諸科学・産業との連携研究WS」開催
- 統計的機械学習セミナーの開催
  2010年11月から8回開催 （講演者17名，うち外国人12名）
- 国際会議，研究会の主催，共催

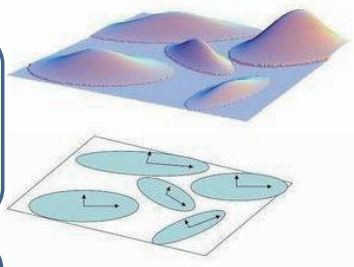
➢ 国際的NOEとして

- 海外研究機関との研究協定 （現在5機関）
- PDを含めた国際共同研究 （Max-Planck Institute, Univ. College London)
- Machine Leaning Summer School 2012共催
  国際的サマースクール（初の日本開催）

研究プロジェクト

- ➢ 数理基盤 - 方法 - 応用までを含めた研究推進
- ➢ 統数研の強みを活かした7つの研究プロジェクト

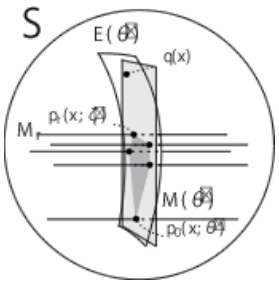
(1) 最適化推論 （伊藤，宮里，池田，伏木，池上，土谷，和田山，品野）
統計的機械学習の各領域を横断的に支えるための，大規模数値計算による新たな推論技術の開発。



Local Learning

(2) 情報幾何と機械学習 （江口，池田，小林，福水，金森）
確率モデルを幾何的対象として扱う「情報幾何」を用いて，機械学習で用いられる高度な学習アルゴリズムの統計的な性質を解明。

(3) グラフ上の確率推論 （福水，池田）
ベイズ推論や復号アルゴリズムに用いられる，グラフ構造を用いた計算アルゴリズムに対する，情報幾何，位相幾何，グラフ理論などの数理的アプローチによる研究。



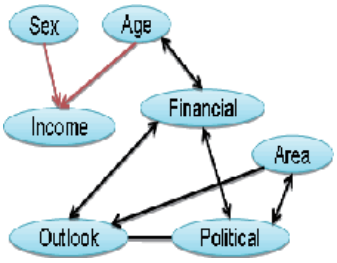
研究プロジェクト（2）

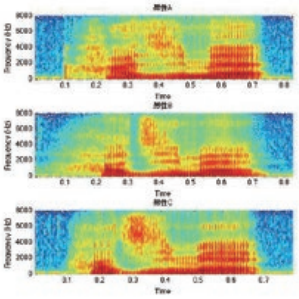
(4) カーネル法の理論と応用 （福水，持橋，小林，西山）
高次元データに対する計算効率の高い非線形データ解析手法である「カーネル法」による新しい統計的推論の方法の展開。

(5) 統計的自然言語処理 （持橋，松井）
ノンパラメトリックベイズ法など深い統計モデルに基づく，言語の潜在構造の推定法と，高度なベイズ推論に必要な計算手法の開発。

(6) マルチメディアデータの判別予測と解析 （松井，持橋）
音声・音楽，映像，テキストなどのマルチメディアデータから，判別予測の目的に応じて，有用な情報を発見するための研究開発。

(7) 機械学習の脳神経データ解析への応用 （小山，松井）
多細胞同時計測技術により獲得可能となった大規模神経活動データに対する機械学習的アプローチの適用による，脳神経情報処理機構の解明。





内外の動向

- 国内では，大学・国立研究所等で「機械学習」を冠した組織は他にないと思われる。

- 海外でも，まだ少数である。
  - CMU, Machine Learning Department
    (School of Computer Science の下)
  - University College London, Centre for Computational Statistics and Machine Learning
    (Computer Science, Statistical Science, Gatsby Computational Neuroscience Unitにまたがる)
  - Alberta Innovates Centre for Machine Learning
    (Computer Science Dept., Univ. Albertaの連携組織)

【統計的機械学習 NOE 概要説明】

○ 司会

それでは、統計的機械学習 NOE の方に入らせていただきます。概要説明に関しましては福水統計的機械学習研究センター長がいたします。

○ 福水統計的機械学習研究センター長

統計的機械学習研究センター　センター長を務めております福水と申します。それでは、統計的機械学習 NOE の活動の概要に関して御説明をしたいと思います。

ここに概要が１枚でまとめられていますけれども、統計的機械学習 NOE では、その活動として、大きな二つの柱があります。一つは、国内の研究コミュニティの活性化、サポートをしていこうというもので、これは共同利用機関としての立場もありますし、そういった活動の一つとして、特に日本の機械学習関係の研究コミュニティをサポートしていきたいというのが一つ。もう一つは、NOE、ネットワークとしては、海外の研究機関とネットワークを組んでいこうという活動を通して、日本の国際的なこの分野におけるプレゼンスを向上させていこうという２本の柱を軸にして活動を行っております。

今日お越しの顧問の先生は、それぞれの分野で第一人者でいらっしゃるのですが、機械学習というのは必ずしも普段聞きなじみがあるとはいえないかもしれません。それほどポピュラーな用語ではありません。日本においてはそれほどコミュニティも強くありません。学会等もその分野に特化した学会等も現状ではないような状況です。従って、活動を通じて、日本のこういった分野の研究を盛り立てて行きたいという、そういう活動をしております。そもそも機械学習とは何ぞやというのはあとで御説明をしたいと思います。

機械学習というのは比較的広い分野でして、それに応じて、この NOE の活動の進展母体になっているのは、今年（2012 年）の１月に発足しました統数研内のセンターであります統計的機械学習研究センターなのですが、ここの研究プロジェクトというのも、数理的な基盤から応用に至るまで幅広いものを扱っています。それもあとで御紹介したいと思います。

まず、体制的な話をさせていただきたいと思いますけれども、NOE の活動としては、現在まで、ここに挙げた５機関と研究協定を結んでおります。御覧いただくとわかりますように、すべて海外の研究機関と研究協定を結んでおります。国内で機械学習としてまとまった大きな組織があまりないというのが一つの理由なのですが、主に海外の研究機関と研究協定を結んで、共同研究を推進しているというのが現状になっております。

これがセンターの体制になります。現在、兼務として、統数研の教員が 10 名、ポスドクが１名、客員として外国人客員１名を含む８名という体制になっております。

では、そもそもの機械学習というのは何かという話を少しだけさせていただこうと思います。機械学習というのは、字面だけ見ただけでは、よくわからない用語ではあるのですが、我々はこういうふうに機械学習というのを考えています。経験やデータなどに基づいて自動的に学習を行うようなシステムに関する研究分野であり、そこに数理、方法論、応用まで幅広い軸が含まれているというふうに考えております。機械学習の基盤になっておりますのは、一つはデータからの推論を扱っている統計科学、それから、もう一つはアル

ゴリズムを扱っている計算機科学、この二つが融合したものが機械学習であると我々は捉えて活動しています。

歴史的なお話を申しますと、もともと Machine Learning、機械学習という用語が興ったのは、1960 年代ごろ、人工知能の分野だったのです。そのころの人工知能というのは、if-then ルールと呼ばれる規則を、データからあるいは人間の知識から抽出して、それに基づいて知的なシステムを作っていくというアプローチが主流でした。しかし、そういったアプローチが行き詰まりを見せまして、現在では、データから統計的な観点に基づいてシステムの学習を行っていくという統計的なアプローチが主流となっております。こういった流れを受けまして、我々は統計数理研究所でこのような機械学習の NOE を立ち上げたということがあります。

機械学習というのは、人工知能ないしは計算機科学が起源になっていますので、当初からいかに計算するか、データを使って解析するというのを単に理論だけで済ませるのではなくて、いかに計算するかという計算の観点を常に念頭においたようなデータ解析のアプローチというふうに考えることもできます。

先ほど統数研の起源に関して色々お話がありましたけれども、どのような理由かは色々議論があるかもしれませんが、統数研では当初から、OR ないしは最適化、あるいは最適制御という分野に力を入れていました。それは古典的な統計科学とは違った観点かもしれないのですが、統数研はそういった研究の歴史を持っている研究所ですので、計算的な観点を念頭においたデータ解析の手法を研究していく場としては最適であろうと我々は考えております。

機械学習ではどういった問題を扱っているかというと、計算機科学を起源にしているという歴史的な背景もあると思いますが、例えば、ロボットの行動制御ですとか、音声・画像・文書といったような情報処理に関わるような分野、それから脳科学のような分野への応用もなされています。最近では非常に大きな応用分野としまして、インターネット上のサービスが挙げられます。Google の検索ですとか、オンラインで買い物をしたときに似た商品を推薦するような推薦システムといったような検索・推薦といった技術の基盤になっているものは、機械学習の分野から多くのアイディアが用いられておりまして、こういった分野が重要な応用分野と考えられます。

一つの例をお話しします。これは別に我々の研究ではないのですが、機械学習を考えるときに一つの典型的な例になると思われるので、御紹介したいと思います。アメリカのネットフリックスという最大手のレンタル DVD の会社が、2006 年にこういうコンペティションを発表しています。彼らは、ユーザーがどういう DVD をレンタルしているか、またユーザーが映画を評価したデータというのも持っているわけです。ざっとユーザー数が 48 万人ありまして、映画のタイトルの総数が１万８千タイトルあります。もちろん全員が全部の映画を見ているわけではありませんので、見た映画の中でも、評価をしたデータしかないわけで、そのテーブルは比較的スカスカのデータなのですが、それでもエントリーとして１億個くらいの数字が入っています。それに基づいて、これから見たくなるような映画を各個人に推薦して、レンタルしてほしい。ネットフリック社は当然、独自の推薦システム、アルゴリズムを持っていたわけですが、それを 10 パーセント改良した人には、賞金 100 万ドルを差し上げますと懸賞をかけました。分野にもよるかもしれませんが、我々の分野

で100万ドルというと大金ですので、非常に有名なコンペティションになったわけです。

これは2006年の問題ですから、今から見ると、ユーザー数48万、映画1万8千というのは、それほど巨大なデータとは言えないかもしれないのですが、実は比較的難しい問題でして、まともに計算をやっていたのではとても間に合わない。計算的な側面、それからもちろん性能を上げなくてはならないという両面から非常に困難な問題で、最終的には2年8か月経ったときにあるグループが10パーセントを達成して100万ドルを手にしました。

今はやりの言葉で言えば、ある種ビッグデータを扱わなければいけないわけです。ビッグデータのデータ解析をやるためには、いかに計算するかというのを非常によく考えてアルゴリズムを作らなければいけない。その意味で、機械学習というのは、このようなビッグデータ時代のニーズにマッチした研究分野であるというふうに考えております。

それでは、具体的にどのような活動を行っているかお話しします。先ほど2本の柱があるといいましたが、もう少し具体的にお話をいたしますと、一つは国内研究コミュニティの活性化ということを考えていまして、国内の研究コミュニティとして、機械学習分野で一番大きなものとして、電子情報通信学会の下部組織であります「情報学学習理論と機械学習研究会」がありまして、ここに多くの機械学習の研究者が集まって研究発表等を行っている。ここと連携を行いまして、年1回のワークショップの共催などを通じてサポートを行っております。もう少し数理的な側面からのコミュニティ形成のための活動としまして、文部科学省の数学・数理科学と諸科学・産業との連携研究ワークショップというものを開催して、数理科学の新たな方向性という側面からもこういった活動を推進しております。

セミナーの開催、国際会議、研究会の共催、主催等は色々行って、国内のコミュニティのサポートを行っています。

それから、第2の柱として、先ほど協定図をお見せしましたけれど、国際的なNOEとして発展させたいということで、現在まで海外の研究機関5機関と協定を結んでおります。単に協定を結んで握手して終わりというのではよくないわけで、具体的に、ポスドクを雇用する際に、海外の共同機関と一緒に共同研究を行う前提でプロジェクトを立ち上げて、ポスドクを選考して共同研究をするという活動も行っております。

機械学習というのは、分野的には新しいこともあって、大学・大学院等でもまとまった教育というのがなかなか難しいというのが現状になっています。それは海外でも同様な状態になっておりまして、そのため、十数年前からなのですが、その分野の主導的な研究者たちを中心として、国際的なサマースクールを年に1回ないしは2回連続して開催しております。今年は日本で初めて京都で開催することになりまして、8月から9月にかけて開催しまして、その共催等を通じてサポートを行っています。

次に我々が行っている研究プロジェクトについて少しだけお話ししたいと思います。今までお話ししたように、機械学習というのは、比較的広い範囲を持っていまして、非常に数理的な基盤から、方法論、それから応用まで含んだようなものになっておりますので、それに応じて統数研内あるいは共同研究者も含めて、様々な軸からの研究プロジェクトを立ち上げております。

一つ目は、機械学習の一つの軸であります、計算、特に最適化に関する計算です。これは伊藤先生を中心にして、機械学習を横断的に支えるための、最適化の技術の開発という

ことをやっております。単に数理的な側面からではなく、そこのテーブルにパンフレットが置いてありますけれども、伊藤先生はこれをスポーツの問題、野球のマジックナンバーなどの問題にも応用しておりまして、今度公開講演会でお話をされるそうですが、そういった応用までを含んだ横断的な基盤になっております。

それから、２番目、３番目は、理論的・数理的な側面でして、２番目は、機械学習の問題に情報幾何と呼ばれている手法を適応しようというものです。情報幾何というのは、今日顧問でおいでいただいております甘利先生が創始者ですが、確率モデルに対する数学的なアプローチでして、甘利先生の影響を強く受けて日本に非常に強いコミュニティがあります。そういった長所を活かして、理論的な、情報幾何関連のプロジェクトを行っております。

３番目は、グラフに関連したアルゴリズムの推論問題、これは理論的なプロジェクトなのですが、一方で、符号の復号問題にこういったアルゴリズムが標準的な方法として使われており、しかしアルゴリズムの解析が難しい、そういう現状にあるので、そういった問題に対する理論的なプロジェクトというものを行っております。

４番目は、理論と方法論を含むようなプロジェクトですけれども、高次元のデータを扱うときに計算というものが非常に問題になりますが、計算効率を意識した新しいデータ解析の方法論としてカーネル法というものがありまして、そのカーネル法の研究は統数研が世界的にも活発な活動を行っておりますので、その分野の研究に関するプロジェクトを行っております。

それから５、６、７番、緑で囲っておりますところが、比較的応用分野なのですが、一つは自然言語処理です。自然言語処理というのは、昔は言語学に基づいたアプローチが主流だったのですが、現在では、大量にある文書そのものをデータと思って統計的なアプローチによって自然言語処理を考えて行こうというのが非常に大きな方法論となっております。そういった分野で非常に高い業績を上げている持橋准教授が昨年の４月に統数研に赴任しまして、一つのプロジェクトをやっているところです。

６番目は、音声・画像・映像・テキストといったマルチメディアデータから判別予測と解析をやるといったプロジェクトです。

７番目は、多少変わった観点のプロジェクトですが、脳のデータというものも、最近では多細胞を同時に計測する技術が発展してきまして、多次元時系列がデータとして与えられるわけです。その観点から、脳神経のデータに関する機械学習的なアプローチというのを、小山先生を中心に進めていただいております。

内外の動向を最後にお話ししておきます。冒頭にも少しお話ししましたけれども、国内では、研究者レベルでは機械学習を研究している先生方というのはいらっしゃいますけれども、まとまって組織として機械学習を推進しているところは、私の知る限りではありません。海外でも、機械学習を大きな組織として研究しているところはまだ少数です。例えばCMUには、Computer Science Schoolの下にDepartmentレベルでMachine Learning Departmentというものが存在しています。それからUniversity College London、ここは協定を結んでおりますけれども、Center for Computational Statistics and Machine Learningというセンターがあります。これは、Computer Science、 Statistical Science、Gatsby、Computational Neuroscience Unitという三つの組織にまたがった、統数研のセ

ンターに近いような全員が兼務で活動しているセンターです。

それからもう一つ、カナダのアルバータ大学に、Alberta Innovates Centre for Machine Learningという組織がありまして、これはアルバータ大のComputer Science Departmentの連携組織としてまとまったセンターがあるということを認識しております。

このように、まだそれほど、Machine Learning の組織として存在しているところは少なく、統数研でこういった組織、センターを作り、かつ NOE 活動をしているというのは非常にユニークな活動だと考えておりまして、そういった特徴を活かした活動というのを今後も推進していきたいと考えております。

以上で私からの説明を終わらせていただきます。

○ 司会

それでは、統計的機械学習 NOE に関して御質問、あるいはコメントを頂戴できればと思います。甘利先生。

○ 甘利顧問

大変わかりやすい説明でよかったと思います。機械学習はまだまだ少数派であると言いましたけれども、今非常にホットなトピックスになっているのではないかと思うのですね。要するに、人間がやっている知的なことを機械が人間を超えてやるにはどうしたらよいのか。人間はそんなに大量のデータは処理できないけれども、機械ならできるかもしれない。それはただ計算するのではなくて、もっとその構造をしっかりと捕まえるような展望がほしいというような願望ですよね。

振り返ってみると人工知能というのがあって、1970 年くらいから 10 年くらい、人工知能ブームがあったけれども、結局限界があって、そのままではだめだ、何か新しいものがいるということになる。その後 1980 年代の終わりから 10 年くらいかけてニューラルネットワークブームというのがあったが、結局これを使った脳に学ぶ技術には限界があった。その後、これらの遺産を継承して、機械学習が出てきたのだと思うのですね。過去にできなかったことを、その基礎の上にもっと発展させたい。機械学習もある意味ではやりなので、10 年すればもっと先に新しいものが出ているかもしれないというわけですよね。目標は非常に大きいのですけれども、今はこれをしっかりやっていこうという話は非常によくわかります。

その中で、特にこの NOE は Statistical Machine Learning と言ってますが、その Statistical という形容詞をくっつけた理由は Statistical に限るのか、Statistical が一番の基礎になってさらに発展するという意味なのか、むしろただ Machine Learning にした方がもっと広い大きな構想になるのではないかという気がしたのです。

一つの質問は Statistical という名前を付けたその意図はどこにあるのかということなのですが。

○ 福水統計的機械学習研究センター長

どうもありがとうございます。一つは、現在、そういう認識をする方が多いかどうかはわからないのですが、機械学習というと、伝統的には人工知能的なアプローチというのが

主流で、今でもそういう研究者の方はいらっしゃるわけです。私の意図としては、それとは違うという意味を出すために、統計的機械学習という名前にしたというのがあります。おっしゃるとおり、統計で閉じてしまっては面白くない、といっては語弊があるかもしれませんけれども、ここに挙げた応用というのは、ある意味で古典的な統計では扱いにくい問題なわけです。精緻なモデルというのはやはり立てづらいわけです。ロボットなんかは比較的物理モデルがあるといいますけれども、実際に走らせてみたら滑ったり等色々とあるわけで、なかなかモデルが立てづらい。情報処理の問題、あるいはインターネットのサービスというのも、きっちりした物理に基づいたモデルというのは立てづらい。そういった問題にもなにがしかの数理的なアプローチをしたい。我々の基盤としては、データというのは統計的である、言い換えれば確率的である、それを一応の仮説としてそこから出発して、こういった扱いづらい問題にも柔軟に対応できるような、新しいパラダイムというものを作りたいという、大きく言えばそのような目的があります。

○　司会

田中先生。

○　田中顧問

Machine Learning についてどうもありがとうございました。大変興味深いお話でした。私が思うのは、これをいくら続けても、データ解析、常に反応があったものを材料にしておられるわけですよね。私どもが一番よく知っている領域で言うと、雑誌の発行部数がどんどん落ちて、編集長になっている人が、データ解析をやって、色々な読者調査をします。どういうテーマが知りたい、どういう筆者に登場してほしい、これをやるとどんどん発行部数が落ちるのです。なぜかというと、読者が反応するというのは自分がすでに知っているテーマなのです。知らないテーマも、知らない筆者も読者の方から出せない。

この分野で、ものすごく展望のある分野がいっぱいあることは知っています。わかりますが、これだけではやはりぶつかるのです。人はクオリティの高いものというのは、出されなければ反応しない、まだ目にしないものには反応しないのです。本当に重要なことはデータがないから、すでに反応したものをいくら解析しても、新しい軸がたぶん立たないのです。

この領域で、たとえばおっしゃったようにノイズを排除して本来のものを引き出す情報処理、でもそれはすでに出たデータですよね。人工知能等はたぶんどんどん実務に近づいていくのだと思うのですが、これだけでは、ブレイクスルーはできないのではないかと、私は直感的に説明を聞いて思ったのですがいかがでしょう。

○　福水統計的機械学習研究センター長

それは非常に厳しい御質問でして、恐らくデータ解析あるいはデータを扱っている人間が常に抱えているジレンマの一つだと思います。我々はデータに基づいて推論をやろうとするのですが、そうすると、おっしゃるように、画期的に新しい過去にないものというのは当然見えないわけです。例えばビジネスの分野で売り上げを伸ばすといったときに、過去のデータを見ていただけではだめかもしれない。ただし、過去にも、画期的な新しいも

のというものはいくつかあったわけで、そういったものにも、ひょっとしたら何かの傾向ないしはデータから抽出できるものがあるかもしれないと、我々、統計的なデータを扱っている人間としてはそういうことを信じてやっていきたいと思っています。

○ **司会**

小柳先生。

○ **小柳顧問**

一言だけコメントですが、先ほどリスク管理の話もありましたが、リスクの発見などもある意味で機械学習の一種だと思います。先ほどのデータ同化シミュレーションの説明がありましたけれども、そういう他の分野との関連なども今後進めていったら色々新しい発展があるのではないかと思います。

○ **福水統計的機械学習研究センター長**

ありがとうございます。確かにおっしゃる通り、色々なところに実際に適応されており、経済の問題等で機械学習を使うという人たちも存在しておりますし、我々としては、現状限られた分野領域のプロジェクトしかやっていませんが、今後、色々なところに拡張していきたいと思っています。

○ **小柳顧問**

こういう分野に機械学習を使うというのは一つのポイントで、それも大事なのですが、私が言いたいのは、リスク発見などというような手法を機械学習の方で取り入れる等、相互交流が必要ではないかということです。

○ **福水統計的機械学習研究センター長**

ありがとうございます。

○ **司会**

真鍋先生、鈴木先生の順番でお願いいたします

○ **真鍋顧問**

Machine Learning というお話を伺いましたが、機械学習 NOE という点で、一つお願いがあります。私の勤務しております青山学院大学では Human Innovation Research Center という研究センターを設置し、国際社会や産官学との積極的な連携を通じて、情報通信技術（ICT）を効果的に活用した研究・教育の向上を目指しています。青山学院の場合は、機械学習の教育の現場への応用ということが中心となると思いますが、統計的機械学習研究センターと何らかの形の交流が持てるといいのではないかと考えています。

○ **福水統計的機械学習研究センター長**

我々としてもぜひ交流をさせていただきたいと思います。

○ 田中顧問

ヒューマンアンラーニングというのがものすごく重要なので、日本の学生さんは、とにかく先生方の業績をフォローしなくてはというラーニングになっているけれども、それだけではブレイクスルーは起きないのです。そこのところをどう考えたらよいかということがあります。そもそも我々の知能とか認識とかに関わるアンラーニングの部分はたぶんマシンじゃないと思うのです。

○ 司会

鈴木先生。

○ 鈴木顧問

こういう ICT の分野で、新しい考え方を開発して、それを使っていく。これはアップルの iPad の開発例みたいに、人が群れれば作れるというものではないのですね。むしろ一人の英才がリードして開発に到る。こういう世界において、NOE というようなことで色々な大学が集まったところで、そこで何か新しいものが生まれるのか、NOE という仕組みはこの分野でどういう意味を持つのかというあたりを、もしお考えがあれば。

○ 福水統計的機械学習研究センター長

この統計的機械学習 NOE に関しましては、必ずしも大きなグループを作って集まりましょうというタイプではなく、どちらかというと自主的な共同研究の輪を広げて行きましょうという形で活動しております。もちろん、群れても仕方がないというのはおっしゃる通りなのですが、 Excellence な人たちと交流を持つと、それなりに新しいアイディアも出てくるのではないかと考えていまして、実際共同研究等ではきちんとした成果というのが上がってきていますので、やっている意味はあるのではないかと私は思っています。

○ 吉川顧問

今の話に関係するのでしょうが、そもそも学習とは何かというと、学習すると先ほどからの新しい着想が得られる、それは過去になかったものだというふうにしないといけないですよね。話が急に古くなって申し訳ないのですが、チャールス・サンダース・パースという人のアブダクションという話があって、それは着想ですね。彼は統計的な機能とそのアブダクション、仮説形成というのは類似性があるということを常に指摘しているのです。

人間の思考、非常に本質的な形式ということを考えていくと、この話題、Machine Learning というのは、非常に大きなチャレンジングな話題に入ってきていると思うのです。ただし、その場合、パースの時代には考えつかなかったビッグデータというのが非常に大きな問題になっていて、ぜひこの話は、ビッグデータと結びつけることによって、田中先生の疑問を払しょくするという話になるのではないかと思います。それをぜひ計画したらどうかと思います。

○ 福水統計的機械学習研究センター長

はい。どうもありがとうございます。

Service Science NOE

統計数理研究所　NOE形成事業 顧問会議

サービス科学NOE

概要説明

サービス科学研究センター
センター長　丸山　宏

統 計 数 理 研 究 所
サービス科学NOE概要説明

統 計 数 理 研 究 所
NOE形成事業

サービス科学研究センター
センター長　丸山　宏

サービス科学NOEの概要

データ中心サービス科学の学術的基盤の充実

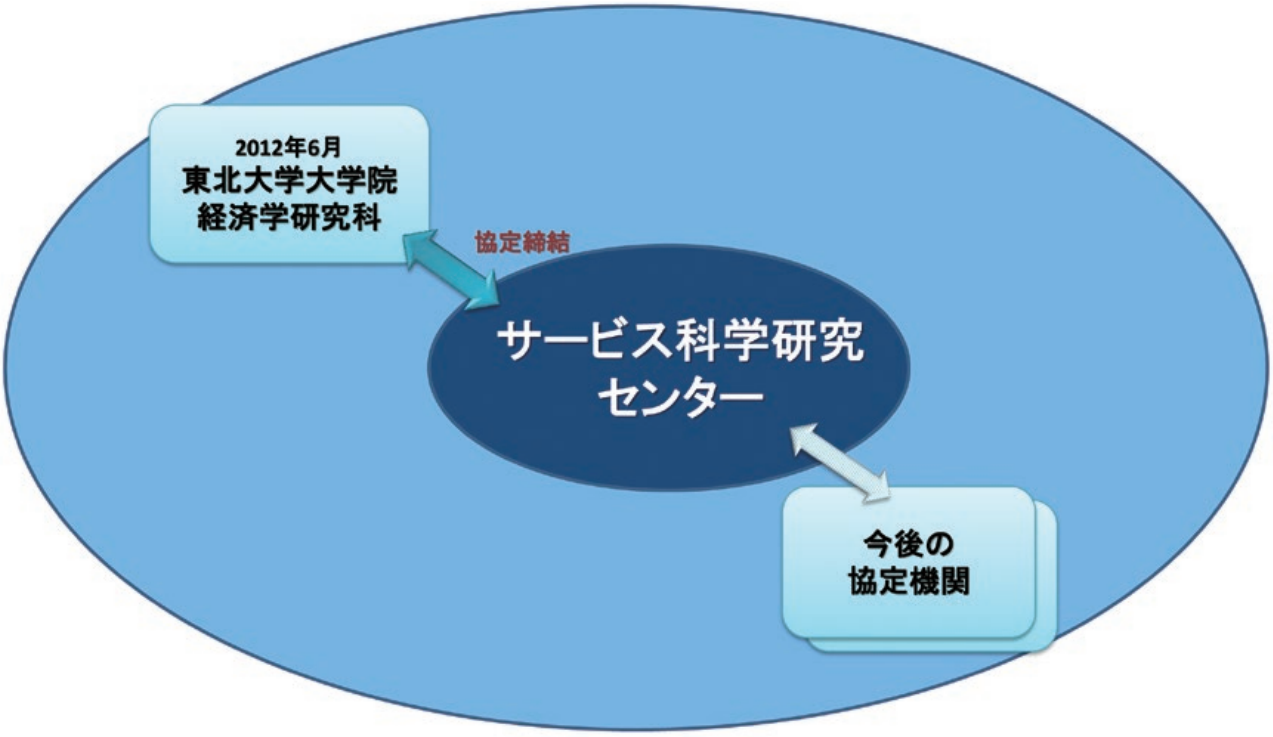
科学の世界では，実験科学，理論科学，計算科学に続く第4の科学のパラダイムとして，データ中心科学の必要性が叫ばれています。サービス科学においても，最新のデータ中心的手法による新しい手法を構築する必要があります。

サービス科学NOE の構築

サービス科学NOEは，統計数理研究所内にとどまることなく，国内の各大学，国内外の研究機関，企業・官庁・NGO などと広く提携し，スケールする支援体制を構築しています。

サービス科学的手法の実証

サービス科学NOE協定図

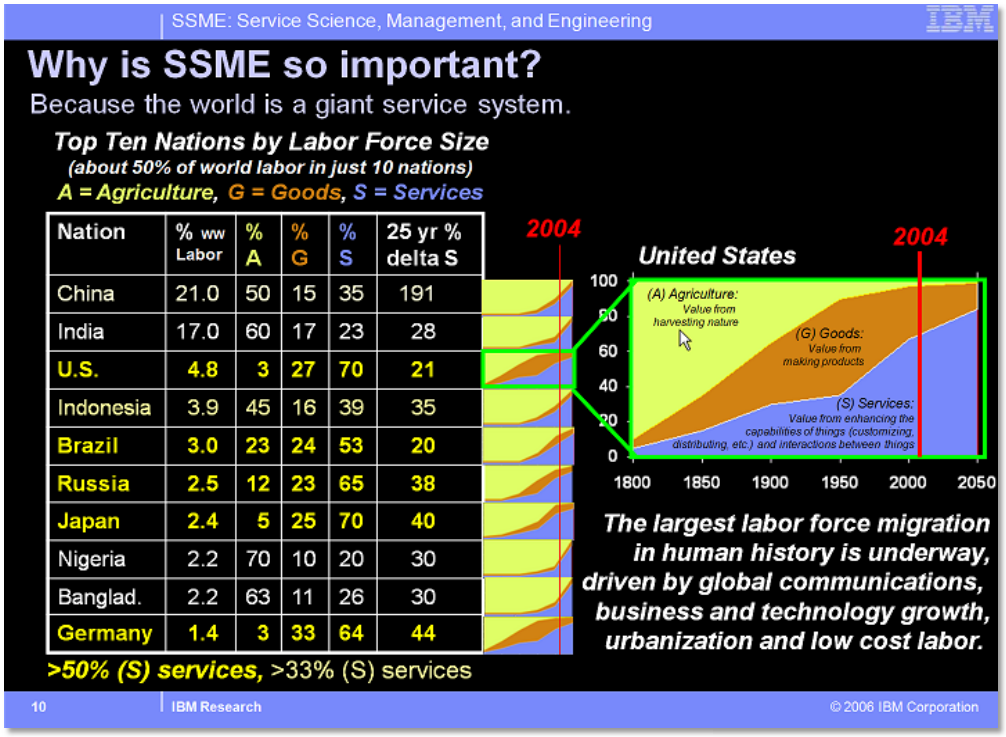


メンバー (as of Jan 2012)

- 教授(兼)　丸山 宏　(センター長(兼)モデリング研究系)
- 教授(兼)　樋口 知之　(モデリング研究系)
- 教授(兼)　椿　広計　(データ科学研究系)
- 教授(兼)　松井 知子　(モデリング研究系)
- 教授(兼)　中野 純司　(データ科学研究系)
- 准教授(兼)　黒木 学　(データ科学研究系)
- 助教(兼)　清水 信夫　(データ科学研究系)
- 助教(兼)　河村 敏彦　(データ科学研究系)
- 客員教授　本村 陽一　産業技術総合研究所
- 客員教授　津本 周作　島根大学 医学部医学科
- 客員教授　照井 伸彦　東北大学 大学院経済学研究科
- 客員教授　山形 与志樹　独立行政法人 国立環境研究所
- 客員准教授　石垣 司　東北大学 大学院経済学研究科
- 客員准教授　佐藤 忠彦　筑波大学 大学院ビジネス科学研究科
- 客員准教授　岡田 幸彦　筑波大学 大学院システム情報工学研究科

プロジェクト・研究活動等

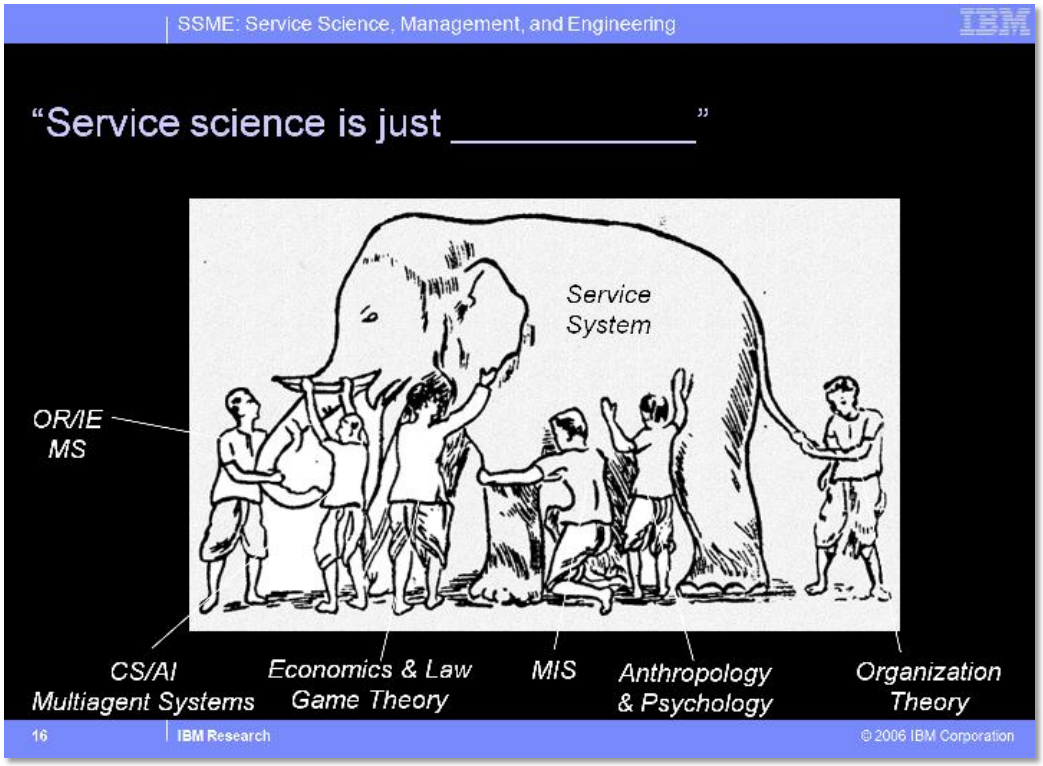
- 製品・サービスの質保証・信頼性研究プロジェクト
- ビックデータ対応型ベイズモデル開発・研究プロジェクト
- レジリエンス社会研究プロジェクト
- 人間社会のコミュニティモデルに関する研究プロジェクト
- サービス産業のためのシンボリックデータ解析手法開発プロジェクト
- データ・キュレーションプロジェクト



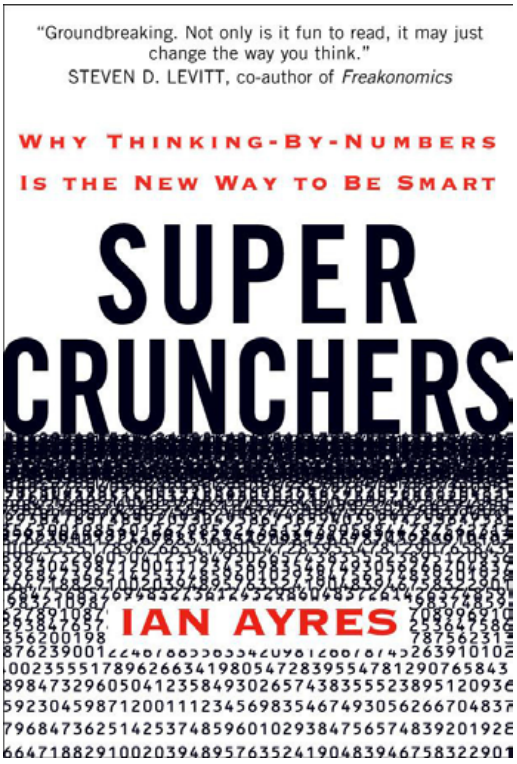
IBM, Jim Spohrer, Dec. 7, 2006

サービス科学研究センター設立記念シンポジウム (2012/1/30)

1. 「人起点のサービス工学 －サービス現場での持続的データ収集と行動モデリング－」持丸 正明 （産総研）
2. 「サービス改革に活用できる実践的なサービス品質とサービスモデル」諏訪 良武（ワクコンサルティング）
3. 「サービス科学の現状と課題」 日高 一義（東工大）
4. 「サービス科学への期待と新しい科学技術イノベーション政策」
5. 有本 建男 (JST)
6. 「ものづくり経営学とサービス・プロセス」 藤本 隆宏 (東大)



IBM, Jim Spohrer, Dec. 7, 2006



「我々は今，馬と蒸気機関の競争のような歴史的瞬間にいる。直感や経験に基づく専門技能がデータ分析に次々に負けているのだ」

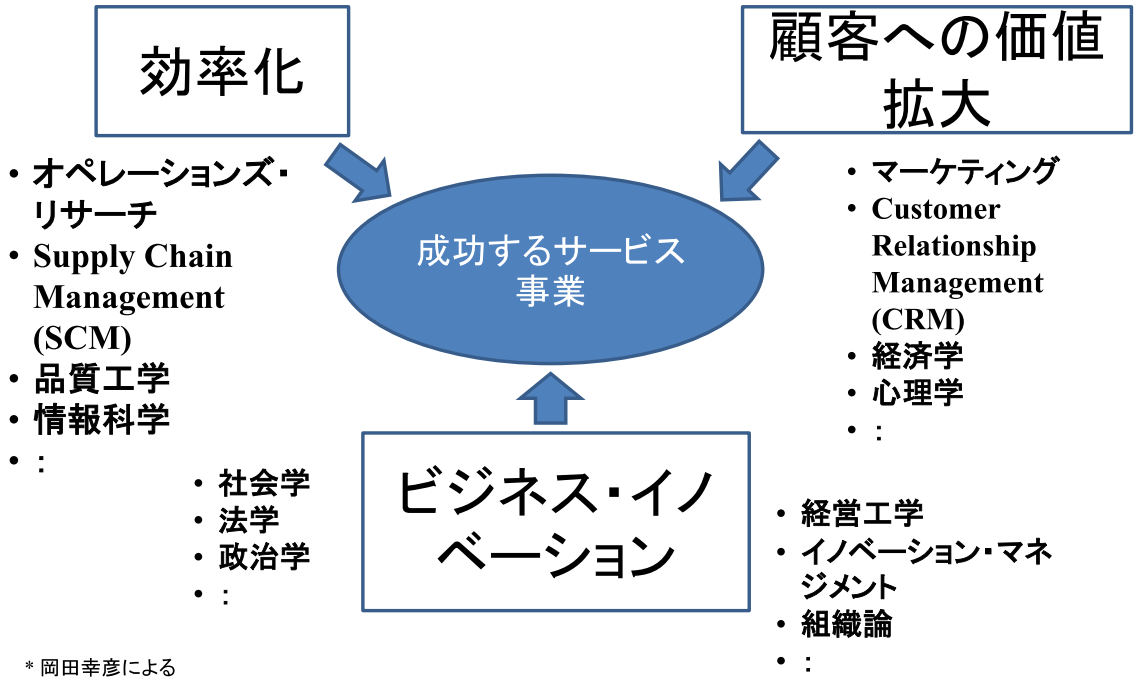
オーリー・アッシェンフェルターによる「モデル」

ワインの質＝12.145+0.00117×冬の降雨+0.0614×育成期平均気温-0.00386×収穫期降雨

Mission

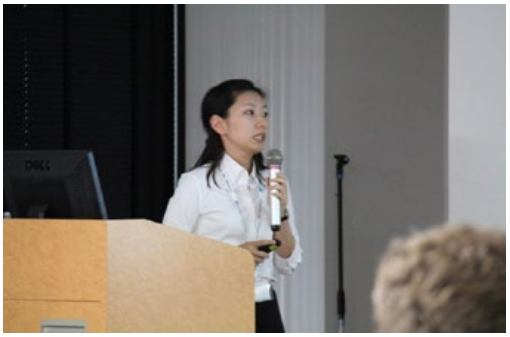
「データ中心」のサービス科学を推進

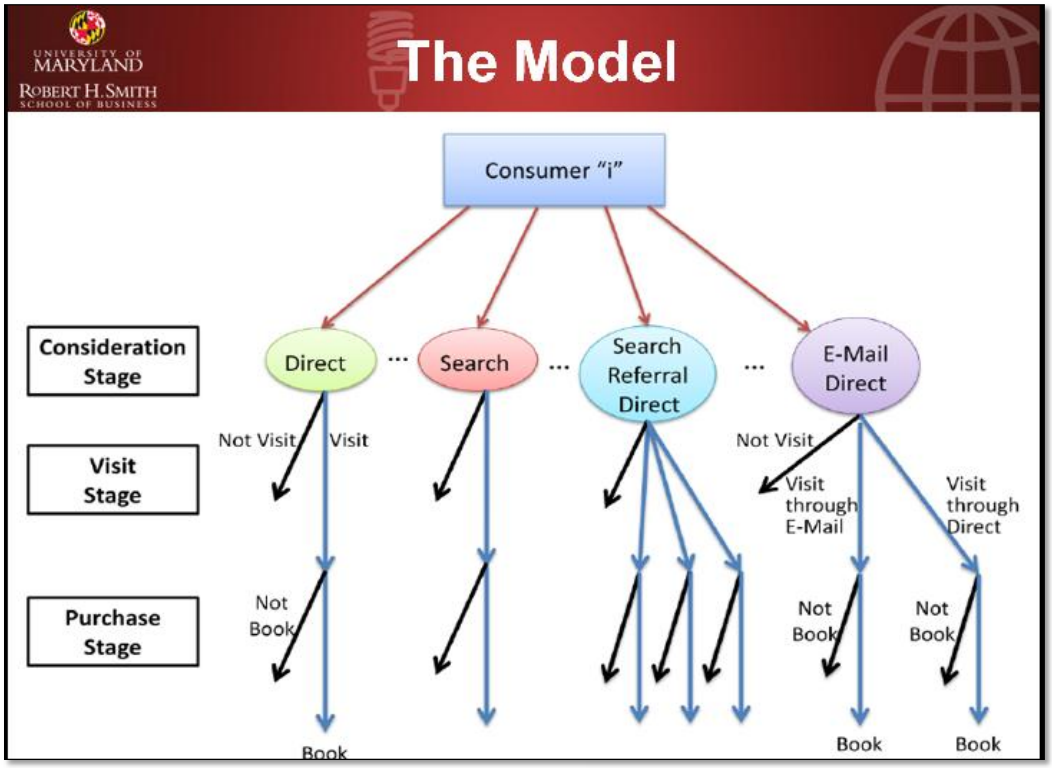
成功するサービスの３要素*



* 岡田幸彦による

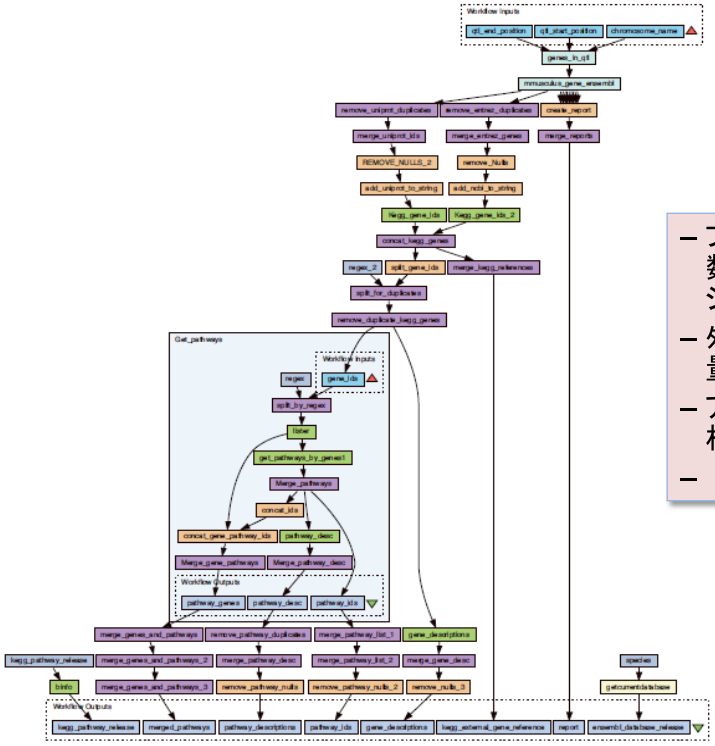
International Workshop on Market Science and Service Research, 2012/7/2-3





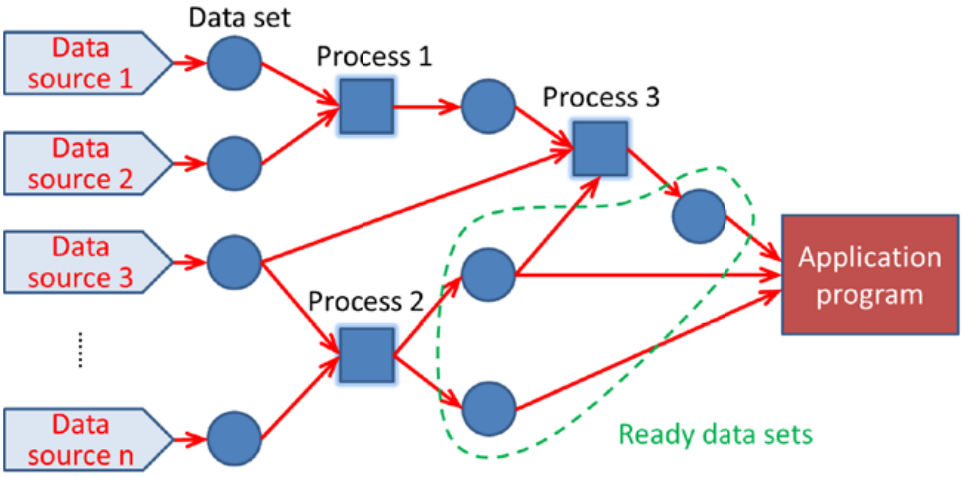
P.K. Kannan, "The Long and Winding Road: Modeling the Conversion Path of Online Customers"

P.K. Kannan, "The Long and Winding Road: Modeling the Conversion Path of Online Customers"



- フォーマット変換，単位変換，数値化，フィルタリング，マージ，…
- 外れ値・欠損値の処理，統計量の計算，…
- アノテーション，より複雑な分析など
- ：

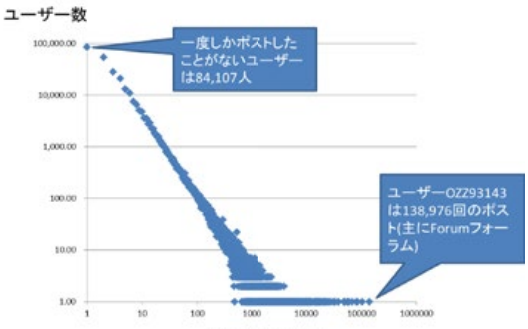
Data Curation = データのワークフローを設計すること

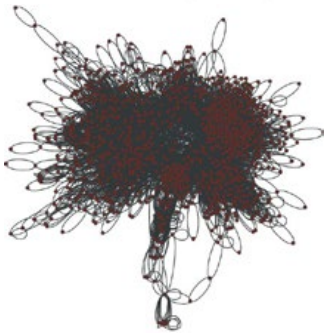


- データ・プロセスの記述
- フロー言語の設計
- データの来歴管理
- ：

実データの入手について

- @Nifty
  - **Niftyフォーラムのデータ**
  - **13年間，1,100万ポスト，**
  - **18万7千ユーザ**
  - **ユーザのサイバー空間上での行動の分析**
- 博報堂
  - **パネルデータを含む各種マーケティングデータ**
- So-net メディア・ネットワークス
  - Real time bidding





いずれも秘密保持契約締結済み

Questions, Comments?

【サービス科学 NOE 概要説明】

○ **司会**

続きまして、最後の概要説明でございますけれども、サービス科学 NOE の説明に入りたいと思います。丸山サービス科学研究センター長が御説明いたします。

○ **丸山サービス科学研究センター長**

丸山でございます。私は 26 年間、IBM という会社におりまして、昨年（2011 年）の４月に統計数理研究所に来たところでございます。まだ色々わからないところがありますけれども、よろしくお願いいたします。

サービス科学ということでございますが、私たちがやりたいことは２点ございまして、一つは、今、第４の科学、データ中心科学といわれているようなものを、サービス科学の分野に持ち込みたいということが一点でございます。もう一つはサービス科学、サービス産業というものは、非常に現場に密着した研究分野だと思います。私ども大学共同利用機関法人として、全国 700 を超える大学があって、その大学の先生方が地場の産業と一緒にサービス科学をやっていらっしゃる、こういうものを何らかの形で御支援申し上げたいというふうに気持ちの上では思っております。まだできたばかりで、これからなのですが、そういうふうに考えていきたいと思います。

サービス科学 NOE の協定ということでございますが、まだ私ども、始まったばかりで、今のところ、東北大学大学院の経済学研究科と一つの協定を結んでおりますが、これからいろいろなところと協定を結ぶなり、あるいは協定を結ばないでも何らかの形で色々なネットワークを広げていきたいというふうに考えております。

今いらっしゃるメンバーの先生方というのは、ある程度統数研と今までつながりがあった先生方が中心でございまして、統数研のメンバー、東北大学の先生方、筑波大学の先生方、国立環境研究所の方という形でメンバーを構成しております。このメンバーにこだわることなく、活動を広げていきたいというふうに思っております。

プロジェクトでございますが、そういう意味では、まだストラトジックにプロジェクトを決めているというよりは、機会主義的に今ある先生方のプロジェクトをまとめているというのが現状でございます。

一番上にあるのが、品質関係のプロジェクト、２番目が主にマーケティング関係のプロジェクトになります。それから３番目、レリジエンスですが、私たちの親機構でございます、情報・システム研究機構の中で、システムズ・レリジエンスという領域横断型のプロジェクトを立ち上げておりまして、先ほどの、色々議論があります、リスクに対してどういうふうに軽減するか、あるいはシステムが生き残るというのはどういうことかということを議論しているところです。

色々なシステムの中で、どうしても人間系というのが非常にビッグなアンノウンになります。その人間を、行動をいかにモデル化するかということが４番目のプロジェクトでございます。５番目のサービス産業のためのシンボリックデータの解析ですけれども、シンボリックデータ、値の範囲を持つようなデータを使う思考がございますが、こういうものを用いて、サービス産業とはどういうものかを解き明かそうというものでございます。

それから、最後のものは、どちらかというと基盤的な考え方でございまして、データをどういうふうに扱うか、どういうふうに前処理するかということに対する考え方です。

実は、サービス科学というものは、私が前におりましたIBMが2006年に提唱した考え方だと思うのですが、考え方としては、世の中の産業、ちょっとこの絵が見にくいかもしれませんが、これはグラフです。1800年から2000年2050年という形で一番上がアグリカルチャーという形で、２番目がいわゆる第２次産業です。ここにサービス産業というのがあります。ずっと年とともにサービス産業の割合が大きくなっているということを示しています。各国がどういう動きをしているかということを見ますと、今の段階では、アメリカ、中国、日本をはじめとして、ほとんどがサービス産業ということになっております。そういうことを考えると、工学におけるいわゆるエンジニアリングに相当するディシプリンがサービスの世界には必要だということを言ったのが、IBMのサービスサイエンス、あるいはサービス・サイエンス・アンド・マネージメント・エンジニアリングというものだと思います。

私ども、サービス科学のNOEを始めて、最初に、2012年１月にシンポジウムを行いました。ここでは、今、日本におけるいわゆるサービス科学の先端を走っていらっしゃる方にそれぞれ来ていただいて、サービス科学の現状ということをお話しいただいたと思います。持丸さんはどちらかというと現場に根差したサービス工学。諏訪さんは、コンサルティング的な経営のノウハウ、あるいはプラクティスについてお話しくださいました。日高さんは、私の前の同僚ですけれども、今サービス科学がどういうところにあるか。有本さんはJSTの方ですが、サービス科学への期待。東大の藤本先生は、ものづくりの先生ですが、ものづくりがいかにサービスに入ってくるかということをお話しくださいました。

全体としてやはり、これらの方々は日本のサービス科学をリードされる方なのですが、どうやら2006年にIBMが言ったことと、まだ実状としてはあまり変わらないのかなと思われます。つまり、サービス科学というのは、私はORのことだと思う、私はマネージメントサイエンスのことだと思う、私は経済のことだと思う、というような色々な人たちが、サービスというものをこんなものだということを考えている、ということは、まだディシプリンとしては固まっていないのかなと思うわけであります。

一方、データ中心の話はビジネスの方にはどんどん入ってきておりまして、例えばイアン・エアーズという人がこんな本を書いておりますが、彼は、この本の中で、今、勘と経験に基づく経営が、データに基づく経営に変わってきている、意思決定がそういうふうに変わってきているということを主張されています。一つの例をお見せしますと、ワインのヴィンテージイヤーかどうかという試飲するエキスパートがいるわけですが、実は、そういうエキスパートよりも、データに基づいて回帰した、これはオーリー・アッシェンフェルターという人は、ワインの専門家でなく経済学者だそうですが、彼が回帰した数式の方が、はるかに精度が高く、ワインのヴィンテージイヤーを予測できるというようなことがこの本に書かれています。

そういうわけで、データをどういうふうに使うかということが、大事なことで、私たちはサービス科学の中で、特にデータ中心的にできるものにフォーカスしようというふうに考えております。

これはメンバーの一人の筑波大学の岡田先生が、サービス科学のランドスケープをこう

いうふうに整理していったらどうかというふうにおっしゃっているわけですけれども、ビジネスですから当然もうけを増やさないといけないのですが、そのためにはコストを下げる、かつ売上を上げる、つまり価値の向上。今のビジネスモデルだけではなくて、先ほどイノベーションの話も出ましたけれども、イノベーションもやらなくてはいけないということで、そういう三つの分野がサービス科学の中に考えられます。が、今までのサービス科学、今までの ICT の使い方は、どちらかというとコスト削減の方に向いていたと思いますが、私たちはどちらかというと、価値向上の方に力を入れたいと思っています。

今年の７月に、東北大学の照井先生が中心になって、マーケットサイエンスのワークショップというのを統数研で開きました。海外からを含め、多くのマーケティングサイエンスの方が来られて、色々な議論をいたしました。この中で、一つだけ理論を御紹介しますと、これはメリーランド大学の P.K. Kannan という先生がやられたものですが、人々がどのように、ホテルの予約をするかというようなことをこういうようなモデルで表現する、それからあるユーザーがダイレクトにウェブサイトに来るその場合に予約する、しない、あるいは、ユーザーがサーチをしてその結果としてあるホテルの予約をする、しない、あるいは、サーチをして出た広告をクリックすることによって来た、来ない、あるいは、電子メールによる案内が来てその中をクリックして来た、来ない、そういうようなモデルを立てて、そのデータを使って学習して、その結果何がおきるかということをやったわけです。

このマリオットホテルは、Google に広告を出すために、年間２千万ドルを払っているらしいのですが、これを払わなくても、実はお客様はマリオットホテルに来てくださる、つまり２千万ドルの価値がないということを彼は示しているわけです。これはこれで大変面白い結果なのですが、残念ながらマリオットホテルは未だに Google にお金を払っているようで、その意味では、結果が出ても、意思決定につながらないところ、というのはこれから経営的なところで考えないといけないと思います。

もう一点、全体の共通基盤についての考え方で、データキュレーションについて、ちょっと話をさせていただきたいと思います。この図はマイクロソフトの第４の科学という本の中に書かれている一つの図ですけれども、どの遺伝子がどの癌に関わるかということに関する論文で、それのためのデータがどのように整備されてきたかというデータのフローを表すグラフです。

色々な測定器からデータが入ってきますけれども、データのフォーマットを合わせたり、外れ値を外したり、データをきれいにしたり、色々なデータをマージしたり、こういうプロセスを経て、最後にデータが出てきて、その結果としてどの遺伝子がどの癌に寄与するかということの結論がこの論文に書かれているわけです。実は、データがたくさんあっても、データの整理、ある意味、データのキュレーションと呼んでもよいと思うのですが、そういうことをしないといけないということです。ここの部分というのはあまり論文にならないところでございますけれども、もう少しアテンションをかけてもよいのかなというふうに思っております。キュレーションという言葉は、いわゆる学芸員、博物館のバックヤードに色々転がっていて、この壺はこういうふうに補修してこういうふうに展示しよう、こういうストーリーで並べようというのがキュレーターの仕事だと思いますが、データに関しても全く同じで、色々なデータがたくさんあるものをいかに整理してそれを価値のあ

るものに変えるか、それがこのデータのキュレーション、キュレーターという職業になるだろうというふうに考えております。

我々はプロジェクトを進める際に、データがもちろん必要なわけで、それに関しては、今、いくつかの会社とお話をさせていただいているところです。例えばニフティが、ニフティフォーラムのデータをお持ちですが、13 年間にわたる、ID 付きの誰がポストしたかわかるという極めて希なデータです。今 Twitter とか、データを処理したという方はたくさんいますが、あれはたかだか数年で、10 年を超えるデータというのはなかなかないわけで、今、この時系列の解析をしているところです。右側に少しあるのがそのデータです。それから博報堂や、その他、ソネットメディアネットワークとの間でも、人間の行動、消費者の購買行動というもののモデルに使えるようなデータについて今、お話を進めさせていただいているところです。

ちょっと駆け足でございましたが、以上でサービス科学 NOE の説明を終わります。

○ **司会**

それでは、サービス科学 NOE に関しまして、御質問あるいはコメントをお願いいたします。吉川先生。

○ **吉川顧問**

ここでお話しすることではないかもしれないのですが、私はそもそもサービス科学の研究のマイノリティの部分にいます。スポーラーとも色々議論しているのですが、少しこれは本編の話とは違ってしまうのですが、ぜひ気を付けていただきたいのは、サービスというものは人と人との関係という、人間が社会をつくる根源的な行為ですよね。そういう立場から、非常に古い還元論的な話を持ち込んで、A さんが、B さんに何かをするという基本行為をサービスだと仮定して、それをプリミティブサービス、原始サービスということでそれが世界をどう構成していくかという話が説明できるという仮説なのです。実は、これは、ロバート・マートンという社会学の人が論じてきた、いわゆる機能論、ファンクショナリティという問題だ、サービスとはいったい何なのか、ということが、何十年というか 100 年以上議論されてきているわけです。私どもはそれを設計という範囲でずっとやってきたわけです。ところが、最近のサービス論というのは、データを扱えるのでサービス現象ばかり集めて、何か問題を作って最適化していくというこういう話でしょう。

現在の瞬間的な状況を是として、それからしか問題が出てこないということになっているわけです。人間というのは社会を作って、猛獣から、外国から身を守るためにそうやって集団を結集してくためにサービスというのが不可欠だったわけでしょう。そういったサービスといったものをもって、非常に根源的な問題をスキップして、サービスに関するデータだけで問題を扱うということに対して、私は、本能的にというと大げさだけれども、どうも耐えられない感じがするのです。ですからスポーラーも、原始サービスというところまでは認めてくれたのですが、もちろんアイデンティファイを持たせる立場としてはそれでよいのだけれど、もし学問としてやるのであれば少し問題ではないかというのが私の問題提起で、特に統計数理、科学という分野でやるのであれば、もう少し扱うデータというのを、人間が根源的に持っている、人間の環境を作っていくファンクショナリティが動

いていくような、そういう話で持っていったらどうかなという気がするのですが、この辺はどうでしょうか。

○ 丸山サービス科学研究センター長

まず、私の個人的な立場としては、サービスとは何かという、哲学的なディベートはさしあたり横に置いておきたいかなという気がいたします。そういうことがたくさん議論されているのは承知しておりますが。その上で、統数研的な立場としてはむしろ、人々の、外部から見える方法というようなものを予測することによって、価値を高めるというようなことにフォーカスしていきたいというふうに考えております。

○ 吉川顧問

価値を高めるということだけれども、先ほどもマートンのことを少し言いましたが、マートンのように、二人しか人間がいないときに、相手のことについてある種の情緒を持つとそれが根源になって、信じるとか信じないというのが社会的に発展していって、それが一つの社会をコントロールする宗教に発展していくというモデルをもっています。そうするといったいどういうサービス産業が生じるのか、という独創的なことを考えようとするならば、今あるサービス産業を一つの普遍物として説いたのでは、また同じような話になります。それは、効率はよくなりますが、本当に必要なのはまだ発展していないサービス産業を発見することですよね。そういう意味では、現在あるサービス産業のデータからだけでは何も出てこないのではないかという気がするのです。

○ 丸山サービス科学研究センター長

それは、その通りだと思います。先ほど、コスト削減、価値向上、イノベーションのお話をしましたが、イノベーションのところが今のところは手がついていない状況でありますけれども、そこは考えていかなくてはいけないと思います。

○ 樋口所長

いずれゆっくり考えたいと思います。

○ 司会

いかがでしょうか。この後また意見交換の際に、顧問の先生方から議論を頂戴するという予定もございます。

自由討論を途中ですべてやってしまい、さらに時間も大幅に超えるという状況になりました。大変貴重な意見を頂戴でき、ありがとうございました。

統計数理研究所

NOE形成事業

Network Of Excellence