
三菱UFJリサーチ&コンサルティング 「化学分野におけるデジタル人材」を 育成する講座の開発」

背景と事業の狙い

背景

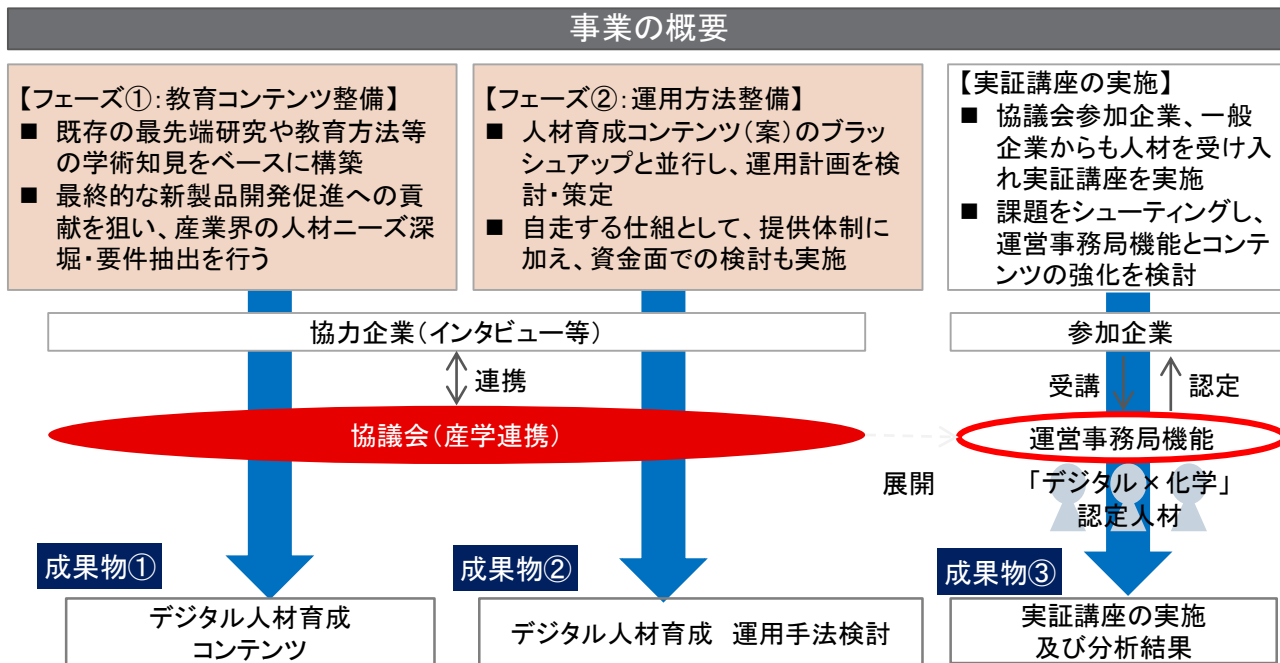
- 化学業界については機能性化学分野での素材開発に取り組む企業が特に欧米で盛んである。新しい素材開発を見つけるためのツールとしてマテリアルズインフォマティクス、及びケミインフォマティクスの取り組みを重要視する傾向にある
- また経済産業省では“Connected Industries”の概念を提唱しており、素材産業のデジタルトランスフォーメーションに取り組む姿勢を示した。こうした方針の基、開発段階での「デジタル×化学」の人材育成が急務とされている

事業の狙い

- 本講座は将来、化学系企業において、データサイエンスを活用した新素材・材料開発や新事業開発の中心メンバーとなることが想定される30代を中心とした若手～中堅の研究者。また、将来、上記研究者のマネジメントを行う層、関連する部署（企画、財務等）の管理層を対象とした
- 上記の初心者及び初級レベルの方に対してマテリアルズインフォマティクス、及びケミインフォマティクスの基礎から応用まで幅広い知識を付けることを目指した講座内容を想定することとした

実施内容

■本事業では、①産業界とアカデミアが密に連携しての産業ニーズの具体化、②将来の自走的運用方法まで見据えた検討、の2点を主眼に置き、①デジタル人材育成コンテンツ、②デジタル人材育成 運用手法検討、③実証講座の実施及び分析結果の検討を行った



成果：概要

達成したい状態

研究開発現場で役立つケモインフォマティクス/マテリアルズインフォマティクス初級講座の開発

実際の達成度

- 達成できたこと
 - 研究開発現場で役立つケモインフォマティクス/マテリアルズインフォマティクス初級講座開催に必要な講座のシラバスの構築
 - 上記検討に必要なと想定した「実証講座」の開催
 - 同講座の来年度以降の運用検討
 - 同講座の来年度以降の運用主体の検討

理由・改善/発展の方向性

- 達成できた理由
 - ケモインフォマティクス/マテリアルズインフォマティクスを必要としている企業をメンバー協議会メンバーとできたこと
 - 協議会において受講者となる上記企業のニーズを把握できたこと
 - 化学に携わる複数の協会からアドバイスを頂けたこと
 - ケモインフォマティクス/マテリアルズインフォマティクスを専門とする有識者から講座開設に必要な知識を提供いただけたこと
 - 実証講座開催により講座運用に必要な有益な情報を得ることができ、その内容を検討結果に反映できたこと

成果：構築できた講座のテーマと講義内容

凡例

講座
テーマ

1	「化学分野における」データサイエンスとは
2	統計学入門
3	多変量データ解析

4	機械学習
5	総合演習 「データサイエンス× 化学」の最前線
6	

各講座の
特徴

座学のみ
座学+実習
実習
グループ討議

初級～中級

1

1 データサイエンスとは(1.5h)
・概念の講義

2 pre機械学習(モチベーションup)(1.5h)
・既存データ利用で大枠をイメージ

3 コンピューター上の化学構造の表現(1.5h)

4 データベース入門(1.5h)
・データベースの活用 世界に散在しているDB

5 環境構築(1h)

2

6 統計学入門(4h)
・統計学
・単変量解析

3

7a 統計結果の表現手法(可視化)(2h)
・分かりやすい表現:
ヒストグラム・ボックスプロット等

7b 教師なし学習を活用した多変量データの把握法(1h)
・多変量解析、多変量データの作り方
・線形代数基礎

7c 教師なし学習を活用した多変量データの把握法(1h)
・正規化テーブル

7d 教師なし学習を活用した多変量データの把握法(1h)
・階層化クラスタリング

7e 教師なし学習を活用した多変量データの把握法(1h)
・PCAなど

7f 教師なし学習を活用した多変量データの把握法(1h)
・解釈及び総合討議(ここまでのラップアップ)

課題創出(グループ討議&発表)(1.5h)

4

8a 機械学習(3h)
・定性:分類問題(判別分析、SVM、ランダムフォレスト等)

8b 機械学習(3h)
・定量:回帰分析(PLS, MRA等)

8c 機械学習(変数選択のオーバービュー)(3h)
・モデル評価(Bayes法、Gaussian Process)
・変数評価(ジニ係数、エントロピーなど)

課題創出(グループ討議&発表)(1.5h)

5

9 マテリアルズインフォマティクス演習(3h+3h)
・データサンプル使用(受講者が自社データの置き換え等で、確認可能)
・グループでの演習・発表

6

10 世界の先進情報の紹介(1.5h)
・マテリアルディープラーニング
・CNN, GCNN, Multilayer NN, 転移学習など

課題創出(グループ討議&発表)(1.5h)

- 1回の講座の受講者数は30人、約6名の班を編成することを想定
- テーマ5における総合演習のほか、テーマ3・4・6の最終講座で課題認識及び定着を意図し、グループ討議&発表を行う

成果：詳細 実証講座の開催内容

- 3頁に記した講座よりエッセンスを抽出し実証講座を開催した

■ 開催概要

- 開催日時：2019年1月25日(金) 13:00~17:50 (12:30~受付)
- 開催場所：三菱UFJリサーチ&コンサルティング(株)2406中会議室
- 定員：28名 (関係者・オブザーバー聴講者除く)
- 受講料：無料

講座講師

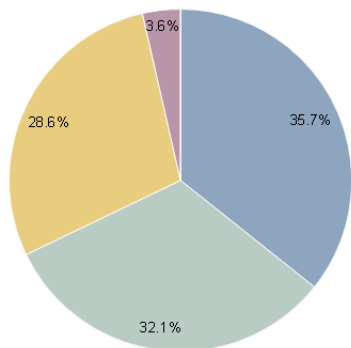
- 講義**： 船津 公人 様 (ふなつ きみと)
東京大学大学院 工学系研究科 化学システム工学専攻教授
- 実習**： 金谷 重彦 様 (かなや しげひこ)
奈良先端科学技術大学院大学 (NAIST) 情報科学研究科
計算システムズ生物学研究室教授

時間	内容	講師	
12:30	開場 (受付開始)		
13:00	データサイエンスとは ~現在置かれている状況~ (20min) 機械学習を使ってできること ~始める前に概観しておこう~ (40min) 変数選択について (20min)	講義	船津 公人
14:20	休憩 (10min)		
14:30	Rプログラミングとは (30min) 統計解析をしながらRプログラミングを理解しよう! (30min)	実習	金谷 重彦
15:30	休憩(10min)		
15:40	化学構造をパソコンに理解させよう! (30min) 階層的クラスタリングでデータを見てみよう! (30min) マシーンラーニング: これであなたもAI研究者? (回帰分析・分類問題) (30min)	実習	金谷 重彦
17:10	実証講座についてのアンケート記入・ご意見交換 (30min)		
17:40	閉会 (10min)		

成果：成果の測定 講座への評価

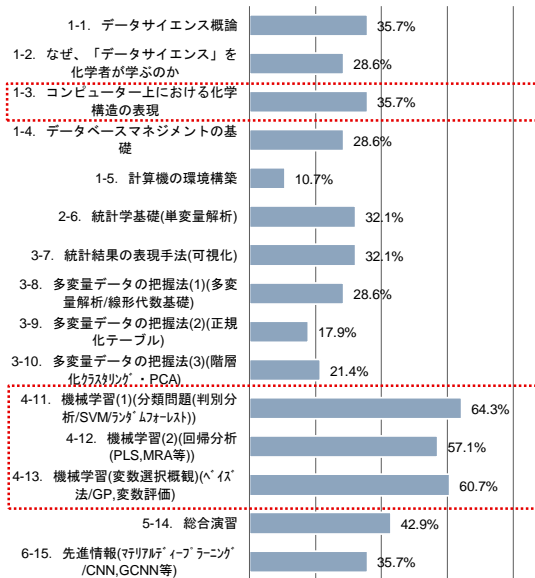
- 頁3の講座内容について実証講座でアンケートを実施したところ、「役立てることが可能」と回答したのは96.4%であった（下記左円グラフ）
- 特に機械学習における分類問題や回帰分析、モデル評価や変数評価を実務へ応用可能と考えている受講者が多く、関心が高かったコンピューター上における化学構造の表現を取り上げることも重視されていた（横棒グラフ）

「講義」の内容について、どの程度実務に役立てることができるか



- 1. とても役立てることが可能と考える
- 2. 役立てることが可能と考える
- 3. ある程度役立てることが可能と考える
- 4. あまり役立てることができない
- 5. 全く役立てることができない

特に関心を持ったテーマ



貴社の現在の業務への活用可能性のあるテーマ

