

平成 2 7 年度実施方針

イノベーション推進部

1. 件 名：学術・産業技術俯瞰システム開発プロジェクト

2. 根拠法

独立行政法人新エネルギー・産業技術総合開発機構法第 1 5 条第 1 項第 8 号、第 9 号

3. 背景及び目的・目標

①政策的な重要性

研究・技術開発のグローバル化、細分化・専門化に伴い、科学技術の進展、変化は加速度的に生じており、研究開発プロジェクトの立案時のみならず、実施時においても常に国内外の技術動向等をモニタリングし、研究開発項目や体制の構築及び柔軟な見直しを行うことが必要となる。上記のグローバルな研究開発動向を把握するに際し、情報量が爆発的に増大していることから、情報技術を用いた論文等の書誌情報の分析を活用することが注目されており、書誌情報の分析手法に関する研究開発が、米国や欧州だけでなく、韓国や台湾等のアジア各国においても実施されている。

国家戦略会議フロンティア分科会においても、知的能力（叡智）を社会や市場につなぐために膨大な情報を効率的に利用できる仕組みの整備についての議論がなされており、情報の効果的活用のための研究プロジェクトの必要性が示唆されている。

②我が国の状況

大学や公的研究機関等において、論文等の情報をもとに、データの体系化や論文ネットワークの可視化といった、現状の整理・分析及び関連技術の開発が行われている。企業や政府からのニーズに応えるためには、現状の整理・分析にとどまらず、将来予測を行う等の新たな分析技術の開発を行い、信頼性・精度・有用性を確保しつつ、実用的なシステムとして構築することが求められる。

③世界の取り組み状況

米国や欧州においても、企業や政府からのニーズを踏まえ、大学や公的研究機関、企業において基礎研究から技術開発、システム化の取り組みが行われている。我が国においても、当該技術の国際競争力強化及び技術経営力強化の観点から、引き続き戦略的・重点的な取り組みが不可欠である。

④本事業のねらい

本事業では、N E D O の技術開発プロジェクトや政府のイノベーション政策、企業の技術経営戦略の立案に貢献することを目的として、計量文献学の手法を活用しつつ、学術論文や特許情報等の様々な情報から、将来的に成長領域となりうる技術領域（萌芽領域）や萌芽領域に関連の深い技術領域、あるいは萌芽領域における有望な研究者及びそのグループを自動特定するシステムの開発を行う。また、専門家等による評価結果を当該システム開発へフィードバックすることで、本格商用化・実用化に要求されるレベルまで自動特定技術の精度を向上させる。

なお、事業全体の目標（中間目標・最終目標）と、実施する研究開発項目及びその目標（中間目標・最終目標）は以下の通り。

〔委託事業〕

○事業全体の目標

【中間目標（平成27年度）】

3つの研究開発項目において開発する自動特定技術につき、それぞれの中間目標を達成する。その上で、それら自動特定技術を統合した中間版「学術・産業技術俯瞰システム」を開発する。また、公的研究機関や民間企業等との間で、当該システムの有効性、有用性、実用性を確認するための共同研究を行う体制を構築する。

【最終目標（平成29年度）】

3つの研究開発項目において開発する自動特定技術につき、それぞれ最終目標を達成する。その上で、最終目標を達成し信頼性や精度が確認された自動特定技術を統合・実装した、「学術・産業技術俯瞰システム」を完成させる。また、当該システムが、萌芽領域等を高い精度で特定し、イノベーションに関する有用な情報を抽出・構造化することで、国内の政策立案機関や公的研究機関、企業等において、政策立案や経営戦略策定に際しての有効性、有用性、実用性が確認され、実用に供されることを目指す。

○研究開発項目及びその目標

研究開発項目①「萌芽領域の自動特定技術の開発」

【中間目標（平成27年度）】

任意の時点・技術領域において実用に足る粒度のクラスタリングを行って得られた部分領域群を母集団として、予測モデル①が判別する「推定萌芽領域」が、「真の萌芽領域」に対し、

- (i) 予測モデル①のパラメータを最適化した場合の適合率（precision）と再現率（recall）の調和平均（F-scores）で0.7以上、
- (ii) パラメータ変化に対して予測モデル①が真陽性率（true positive rate）・偽陽性率（false positive rate）平面上で描く受信者動作特性曲線（ROC）下の面積（AUC）で0.7以上、となることを開発目標とする。

【最終目標（平成29年度）】

中間目標と同様の条件で、予測モデル①が判別する「推定萌芽領域」が、「真の萌芽領域」に対し、

- (i) 予測モデルのパラメータを最適化した場合の適合率（precision）と再現率（recall）の調和平均（F-scores）で0.8以上、
- (ii) パラメータ変化に対して予測モデルが真陽性率（true positive rate）・偽陽性率（false positive rate）平面上で描く受信者動作特性曲線（ROC）下の面積（AUC）で0.8以上、となることを開発目標とする。

研究開発項目②「関連領域の自動特定技術の開発」

【中間目標（平成27年度）】

任意の時点・技術領域において実用に足る粒度のクラスタリングを行って得られた部分領域群を母集団として、事後的に認められる成長領域から遡及的に特定される「真の萌芽領域」に対して予測モデル②が判別する「推定関連領域」が、事後的に認められる成長領域から遡及的に特定される「真の関連領域」に対し、

- (i) 予測モデルのパラメータを最適化した場合の適合率（precision）と再現率（recall）の調和平均（F-scores）で0.7以上、
- (ii) パラメータ変化に対して予測モデルが真陽性率（true positive rate）・偽陽性率（false positive rate）平面上で描く受信者動作特性曲線（ROC）下の面積（AUC）で0.7以上、となることを開発目標とする。

【最終目標（平成29年度）】

中間目標と同様の条件で、予測モデル②が判別する「推定関連領域」が、事後的に認められる成長領域から遡及的に特定される「真の関連領域」に対し、

(i) 予測モデルのパラメータを最適化した場合の適合率 (precision) と再現率 (recall) の調和平均 (F-scores) で 0.8 以上、

(ii) パラメータ変化に対して予測モデルが真陽性率 (true positive rate)・偽陽性率 (false positive rate) 平面上で描く受信者動作特性曲線 (ROC) 下の面積 (AUC) で 0.8 以上、

となることを開発目標とする。

研究開発項目③「有力・有望研究者及びその共同研究体制の自動特定技術の開発」

【中間目標（平成27年度）】

任意の時点・技術領域において実用に足る粒度のクラスタリングを行って得られた部分領域群のうち「真の萌芽領域・関連領域」に対して、予測モデル③が判別する「推定有力・有望研究者及びその共同研究体制」が、事後的に認められる成長領域から遡及的に特定される「真の有力・有望研究者及びその共同研究体制」に対し、

① 予測モデルのパラメータを最適化した場合の適合率 (precision) と再現率 (recall) の調和平均 (F-scores) で 0.7 以上、

② パラメータ変化に対して予測モデルが真陽性率 (true positive rate)・偽陽性率 (false positive rate) 平面上で描く受信者動作特性曲線 (ROC) 下の面積 (AUC) で 0.7 以上、となることを開発目標とする。

【最終目標（平成29年度）】

中間目標と同様の条件で、予測モデル③が判別する「推定有力・有望研究者及びその共同研究体制」が、事後的に認められる成長領域から遡及的に特定される「真の有力・有望研究者及びその共同研究体制」に対し、

① 予測モデルのパラメータを最適化した場合の適合率 (precision) と再現率 (recall) の調和平均 (F-scores) で 0.8 以上、

② パラメータ変化に対して予測モデルが真陽性率 (true positive rate)・偽陽性率 (false positive rate) 平面上で描く受信者動作特性曲線 (ROC) 下の面積 (AUC) で 0.8 以上、となることを開発目標とする。

4. 事業内容

4. 1 平成26年度（委託）事業内容

研究開発項目①「萌芽領域の自動特定技術の開発」

異なる引用の組み合わせによる萌芽領域の引用分析手法の開発を行った。具体的には、直接引用、共引用、書誌結合などの異なる引用関係を組み合わせ、さらにテキストのキーワードの類似度、引用の類似度等により引用関係に重みを付与することで、数万から十万の論文集合からなる萌芽領域の論文引用ネットワークにおける特徴量を用いて、萌芽論文（引用数の伸び率が高い論文）を予測する手法の開発を行った。（実施体制：国立大学法人東京大学大学院工学系研究科）

研究開発項目②「関連領域の自動特定技術の開発」

引用ネットワーク情報とテキスト情報の組み合わせによるコーパス拡張手法の開発を行った。具体的には、引用関係とテキストに基づく潜在意味空間を用いて検索クエリを拡張することにより、元の萌芽領域のコーパスを拡張する手法の開発を行った。本手法を複数の論文データセットに適用し関連領域特定の観点から内容の評価を行

った。(実施体制：国立大学法人東京大学大学院工学系研究科)

研究開発項目③「有力・有望研究者及びその共同研究体制の自動特定技術の開発」

ネットワークの時系列情報から有望研究者・グループを特定する手法の開発の開発を行った。具体的には、著者の同姓同名の問題に対して、著者に識別可能なユニークIDが与えられている Elsevier 社の Scopus データを用いて、大規模な共著ネットワークとその時系列情報を抽出した。ネットワークから抽出された特徴量を用いて著者のパフォーマンスを予測する基本的なモデルの検討及び開発を行った。(実施体制：国立大学法人東京大学大学院工学系研究科)

4. 2 実績推移

	25年度	26年度
	委託	委託
実績額推移 一般勘定（百万円）	28	60
特許出願件数（件）	0	0
論文発表数（報）	14	6
フォーラム等（件）	3	4

5. 事業の実施方式

平成27年度には以下の研究開発を実施する。なお、体制についてはNEDOが国立大学法人東京大学大学院工学系研究科に委託して実施する。

5. 1 平成27年度（委託）事業内容

研究開発項目①「萌芽領域の自動特定技術の開発」

機械学習による萌芽領域のモデル化を行う。任意の時点・技術領域に適用可能な予測枠組みを構築する。具体的には、論文発表年に予測を行うこととし、予測のために用いる論文の特徴（特徴量）を抽出し、過去事例の収集により得た訓練データセットを用いた機械学習により特徴量に重み付けを行って得た予測モデルを予測対象論文に適用して萌芽論文を予測するモデル化を行う（実施体制：国立大学法人東京大学大学院工学系研究科）。

研究開発項目②「関連領域の自動特定技術の開発」

平成26年度に引き続いて、引用ネットワーク情報およびテキスト情報を用いたコーパス拡張手法の開発を行う。具体的には、平成26年度において検討を行った関連領域を特定するためのネットワークおよびテキストの特徴量をそれぞれ用いてコーパスを拡張する手法の開発を行う。

また、引用ネットワーク情報とテキスト情報の組み合わせによるコーパス拡張手法の開発を行う。具体的には、テキスト情報を用いてコーパスを逐次拡張する際に、引用ネットワークの構造情報の変化を指標としてネットワークの性質を保持し、局所最適に陥らないコーパス拡張手法を開発する（実施体制：国立大学法人東京大学大学院工学系研究科）。

研究開発項目③「有力・有望研究者及びその共同研究体制の自動特定技術の開発」

機械学習による有望研究者・グループのモデル化を行う。任意の時点・技術領域に適用可能な予測枠組みを構築する。具体的には、論文発表年に予測を行うこととし、予測のために用いる論文の特徴（特徴量）を抽出し、過去事例の収集により得た訓練

データセットを用いた機械学習により特徴量に重み付けを行って得た予測モデルを予測対象論文に適用して有望研究者・グループを予測するモデル化を行う（実施体制：国立大学法人東京大学大学院工学系研究科）。

研究開発項目④「学術・産業技術俯瞰システムの開発」

システム基盤の開発を行う。萌芽領域の自動特定技術、関連領域の自動特定技術、有力・有望研究者及びその共同研究体制の自動特定技術のそれぞれにおいて所定の性能を満たした要素技術をシステムとして統合する。加えて、論文や特許の効率的な検索、分析のためのデータ抽出を可能とするデータベースの整備を行う。

また、システムの評価及び運用体制整備を行う。共同研究先からの予備評価やユーザー評価を行い、フィードバックを得ることで、ユーザーインターフェース等の改善を行う。

5. 2 平成27年度事業規模

一般勘定 53百万円（継続）

なお、事業規模については、変動があり得る。

6. その他重要事項

（1）評価の方法

NEDOは、「必要性」、「効率性」、「効率性」等の観点から毎年度事業評価を実施するほか、技術的な視点から外部有識者による評価を事業期間中の適切な時期に実施し、本事業の継続可否や今後の方向性等の判断を行うこととする。

（2）複数年度契約の実施

平成25～27年度の複数年度契約を行う。

7. 実施方針の改定履歴

（1）平成27年3月、制定。