jh240067

SINET を介したデータベース基盤と HPC 基盤の連携による医療画像解析基盤実現に関する研究

課題代表者　村尾 晃平（国立情報学研究所）

概要

本研究の目的は、SINET を介してデータベース基盤と HPC 基盤の連携による医療画像解析に必要なインフラを構築し、実際の医療画像を用いた AI／機械学習の開発を実現することである。JHPCN において重要とされる、計算基盤、データベース基盤、ネットワーク基盤の 3 つの基盤を連結し、AI/機械学習のための医療画像解析基盤の実現を目指す。

本研究は 2023 年度からの継続であり、医療画像データを収集・格納している国立情報学研究所（NII）のクラウド基盤と、豊富な計算資源を保有する名古屋大学の情報基盤センターの計算基盤「不老」を SINET6 経由で連携し、データセキュリティを確保したシステム設計・構築を行ってきている。この基盤連携システム上で AI/機械学習の処理や医学的見地からの大量データの分析を進めながら、2024 年度はさらに HPC 基盤を追加できるように、安全性を考慮したネットワーク構成を見直し、計算基盤の連携対象を mdx に広げることができた。

1. 共同研究に関する情報

(1) 共同利用・共同研究を実施している拠点名

名古屋大学 情報基盤センター

mdx

(2) 課題分野

データ科学・データ利活用課題分野

(3) 参加研究者一覧と役割分担

村尾 晃平：代表、全体統括

森 健策：副代表、高性能計算環境設定、AI 開発

合田 憲人：インフラ設計・セキュリティ

佐藤 真一：画像系 AI 開発

大江 和一：インフラ設計・性能検証

大竹 義人：医療系データ整備・AI 開発

明石 敏昭：医療系データ整備・AI 評価

黒瀬 優介：医療系 AI 開発

崇風 まあぜん：医療系 AI 開発

二宮 洋一郎：医療系データ整備

2. 研究の目的と意義

画像診断支援 AI を実現する学習データには多様性が求められ、異なった撮影装置・撮影条件でのデータが必要となる。国立情報学研究所（NII）では、医療画像の提供元を単一の医療機関ではなく、その診療科の複数の医療機関をまとめる医療系学会に求め、全国の医療機関から悉皆的に多彩な医療画像を収集し、データベース上に格納する仕組み「医療画像ビッグデータクラウド基盤」を構築した。以下では単に「クラウド基盤」と表記する。2017 年 11 月の運用開始以来、クラウド基盤はデータ収集を順調に進め、2025 年 3 月末時点で 6 つの医療系学会を通じて 7 億枚以上の医療画像を蓄積している。（図 1）

クラウド基盤は「ネットワークシステム」「ストレージシステム」「データ管理データベ

ースシステム」「コンピューティングシステム」から構成される（図 2）。ネットワークシステムは、解析研究者の利用する SINET L2VPN をコンピューティングシステムの CPU サーバおよび GPU サーバに接続し、医学系学会からのデータ提供を受取るための SINET Azure をストレージシステムの一部であるデータ受入サーバに接続するようルーティングしている。SINET は現在 SINET6 であり、回線は 400Gbps までの世界最高水準の容量があるが、クラウド基盤内のネットワークは 10Gbps の回線容量で接続している。

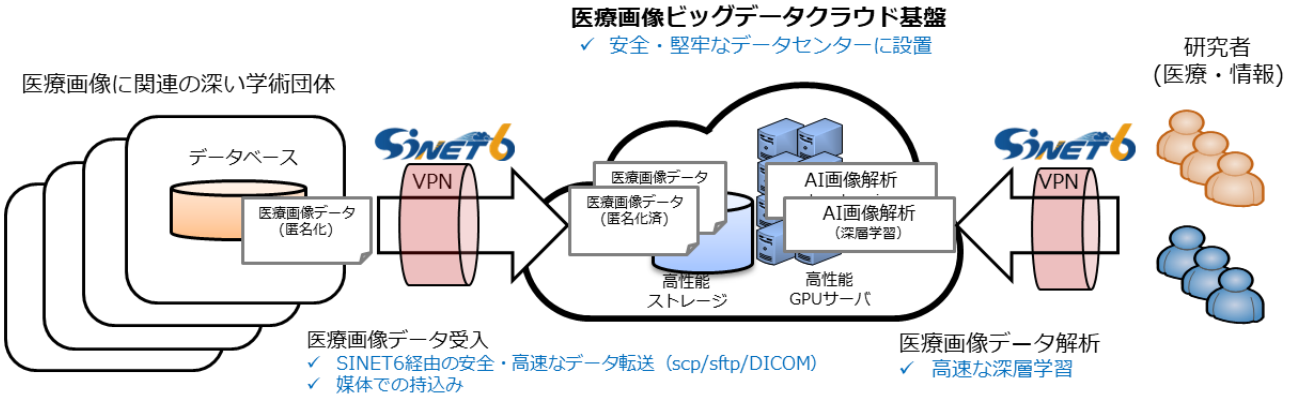

図 1　クラウド基盤の概念図

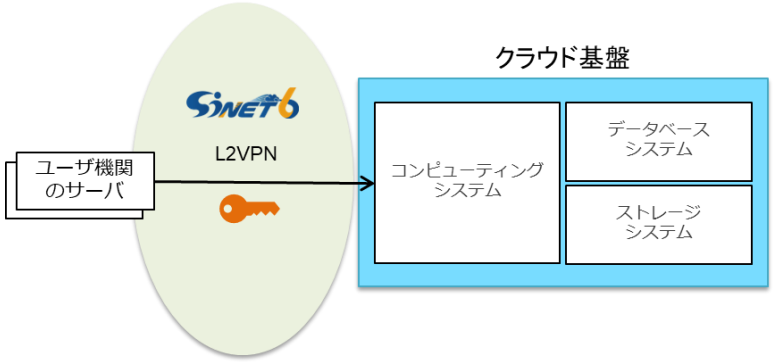

図 2　クラウド基盤の構成システムとログイン経路

データ受入サーバでは，医療画像とそれに付帯する情報（メタデータ）を分離し、画像をストレージに、メタデータをデータベースにそれぞれ格納する。これにより、機械学習に必要な画像データセットはデータベース検索によって揃えることができる。

ストレージはコンピューティングシステムにネットワーク・ファイルシステム（Network File System, NFS）マウントされており、機械学習計算をクラウド基盤内で完遂することで、医療画像やメタデータが外部に持ち出されることを抑止している。

このデータベースには小規模であるが GPU サーバが接続され、NII を含む国内 14 の情報系研究チームが医療 AI 研究開発の課題に取り組んでいる。しかし、データの増加に伴って計算資源に対する需要は増大しており、マルチモダリティ解析や時系列解析といったタスクの高度化に伴う処理能力への要求に対して、単一拠点の計算・データベース資源だけで対応することは困難となりつつある。

そこで、データを蓄積する拠点と HPC 資源を提供する情報基盤センターとを SINET

を介して接続することで分散計算基盤を構成することは、有効な解決策と考えられる。しかし、医療 AI 研究が求めるセキュリティと性能を確保するための分散計算基盤を実現する方式は確立されていない。そこで本研究では、医療 AI 研究開発のための分散計算基盤型の医療画像ビッグデータクラウド基盤を構築することを目的とする。2024 年度は、2023 年度に構築した SINET6 の L2VPN を活用し、拠点間のデータベース基盤－HPC 基盤連携について、安全性を高めると共に、複数の HPC 環境の同時利用へ拡張性を考慮したネットワーク設計を行った。

3. 当拠点の公募型共同研究として実施した意義

本研究では、単に各拠点の計算資源を利用するのではなく、クラウド基盤とオンラインで結び、ネットワークの構成やデータ連携方法を試行しながら構築する必要がある。拠点のメンバーと密に議論しながらシステム構築を進めるために、共同研究の枠組みが有意義であった。

4. 前年度までに得られた研究成果の概要

2023 年度は基盤連携の設計方針を定め、それを基に実際に名古屋大学情報基盤センターのスーパーコンピュータ「不老」と国立情報学研究所のクラウド基盤を連携させたシステムを構築した。

連携の実装にあたって、計算後に医療情報をクラウド基盤の外部に残さないようにするため、次の要件を掲げた。

(1) 利用者は、計算に必要な時のみにクラウド基盤に接続する。
(2) 医療情報をクラウド基盤の外部に残さない。
(3) 外部の計算環境にて、クラウド基盤にアカウントの無いユーザはクラウド基盤にアクセスできない。

具体的には、不老が保有する多数の計算ノードのうち、特定の計算ノードの先にゲートウエーサーバを導入し、そこに SINET6 L2VPN の回線を引き込んでクラウド基盤に接続し、不老でジョブを投入した時のみにオンデマンドでクラウド基盤に sshfs 接続する仕組みを構築した。これにより、上記の要件を全て満たすシステム構築を実現できた（図 3）。なお、ネットワーク遮断などのロバスト性能実験用に NII の柏分館に用意したサーバへも別の SINET6 L2VPN を設定し、ポート切り替えで接続できるようにした。

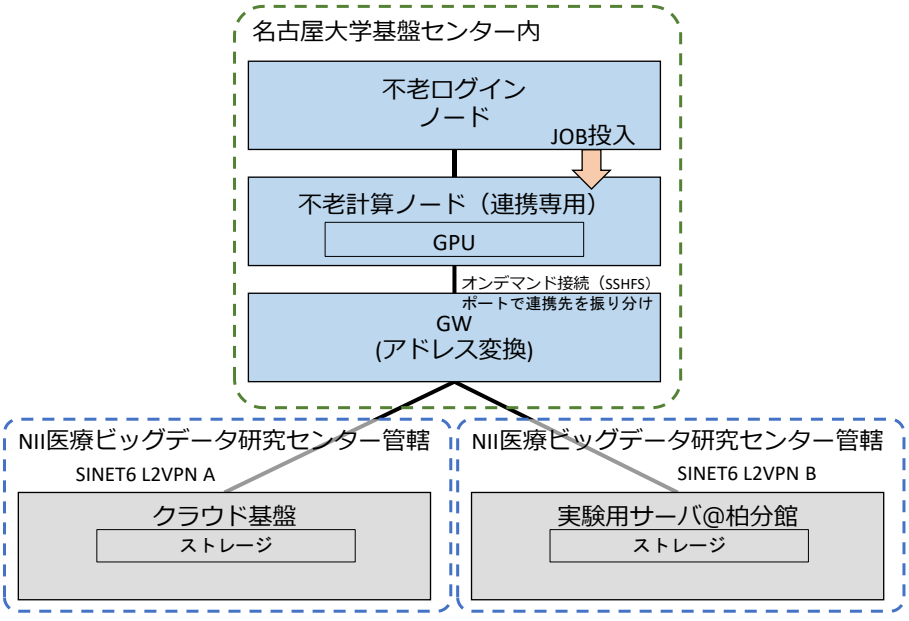

図 3 不老とクラウド基盤、実験用サーバとの接続概念図

このシステムを使って、実際に深層学習の AI 開発のベンチマーク・テストを行ったところ、頻繁なファイル読み書きが発生するような逐次計算では遠隔によるターンアラウンド・タイムの影響が出るものの、深層学習処理のようにバッチ的に読み書きをする場合には、ローカル環境で計算させるのとほとんど変わらない性能を出せることがわかった。また、計算途中でネットワークの断絶が数分程度あっても回線復旧時には問題なく計算を続行できる耐性もあることがわかった。

5. 今年度の研究成果の詳細

昨年度連携した不老は、ログイン・ノードおよびジョブ管理のシステムが備わっており、セキュリティ的に堅固である。ところが、接続先の SINET6 L2VPN は図 2 に示したようにユーザ機関のサーバが多数つながったネットワークである。2025 年 1 月時点でユーザ機関は 20 以上あり、どこかで接続設定の不備があると、ブロードキャストストームが発生する可能性も否定できない。そのような事象が発生すると、クラウド基盤のコンピューティングシステムに負荷をかけると共に、不老での計算の効率も大幅に下がってしまう。そこで、今年度は新たにログインサーバを設置し、ログインサーバとコンピューティングシステムの間に新たな SINET6 L2VPN を導入した。そして、ここに HPC 基盤を接続することとした（図 4）。

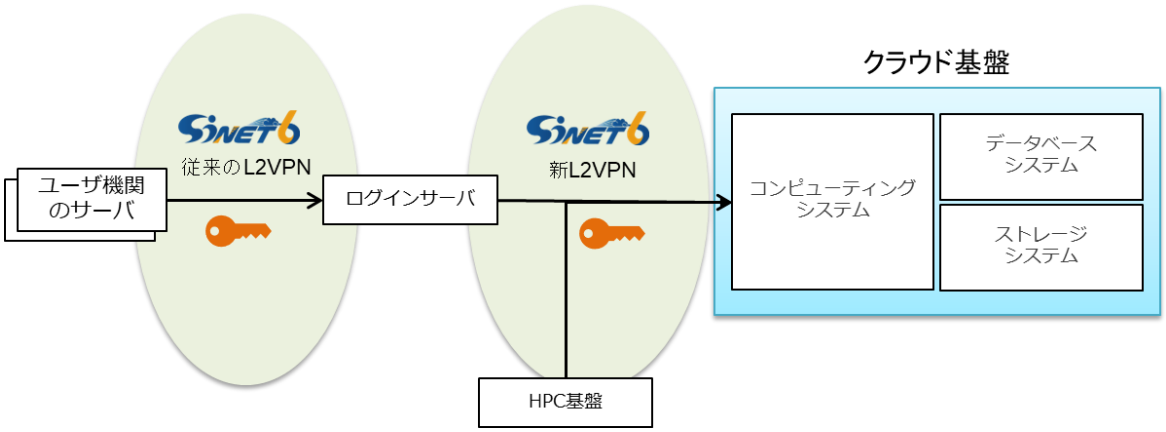

図 4 ログインサーバの設置と新たな L2VPN の設定

一方、東大 mdx のように計算ノードに直にログインし、ジョブ管理なしで使えるような HPC 基盤との連携については、mdx の管理をクラウド基盤の管理者の下で行うことを前提とし、利用者が必ずログインサーバを利用することで、クラウド基盤のコンピューティングシステムの延長のような扱いをすることができると考えた。ただし、ssh_config または PAM (Pluggable Authencication Module) の設定を使い、ログインサーバ以外の経路でアクセスしないような設定が必要である。以上のような考えを基に、全体のネットワーク構築の図をまとめると図 5 のようになる。

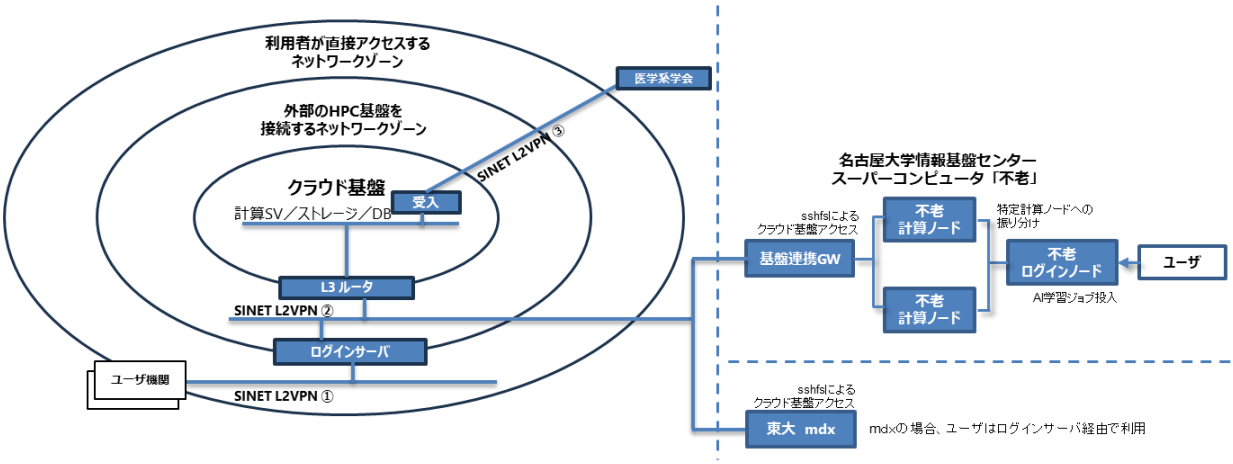

図 5　構築したネットワーク概念図

構築したネットワークに対して、各 HPC サーバからクラウド基盤に向けて RTT (Round Trip Time)を測定した。測定した箇所の模式図を図 6 に示す。また、その結果を表 1 に示す。名大の不老からの RTT が大きくなっているのは、２つの基盤が遠隔地のためである。転送するデータのサイズが小さくても、送受信の回数が多いと RTT の大きさが無視できなくなる。

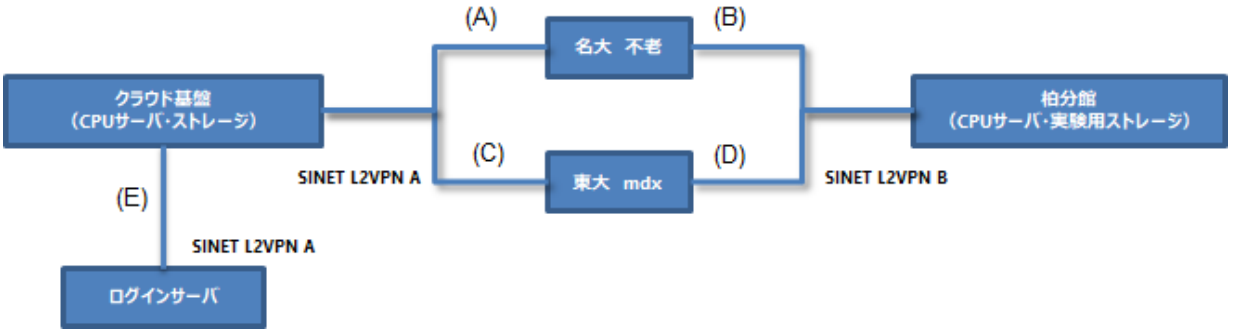

図 6　RTT の実測箇所の模式図

表 1　RTT の実測結果

	クラウド基盤 (ms)			柏分館 (ms)		
	Max	Min	Average	Max	Min	Average
名大 不老	7.256	6.682	(A) 7.043	8.245	8.014	(B) 8.141
東大 mdx	1.480	1.292	(C) 1.392	3.655	3.491	(D) 3.573
ログインサーバ	1.307	1.210	(E) 1.253	−	−	−

次にネットワーク帯域の実測を iperf3 コマンドで調べた。コマンドの設定値はデフォ

ルトの 1 秒間隔 10 回測定とした。ただし、測定時には直接 IP アドレスを指定するのではなく、ssh でポートをトンネリングしているので、暗号化のオーバーヘッドが含まれている。測定した箇所の模式図を図 7 に示す。また、その測定結果を表 2 に示す。回線の経路上は、いずれも 10Gbps で接続しているが、上述の ssh 接続のため、実測結果はおよそ 2Gbps であった。また、利用者の使用制限などは行わずに測定したので、多少の揺らぎは生じている。表中の Cwnd というのは輻輳ウィンドウサイズと呼ばれ、送信側が受信側からの確認応答（ACK）を待たずに、ネットワークに送信できる最大データ量を表している。この値が大きいほどネットワーク負荷が小さいとされている。今回の計測ではこの値は 3～5 Mbytes となった。

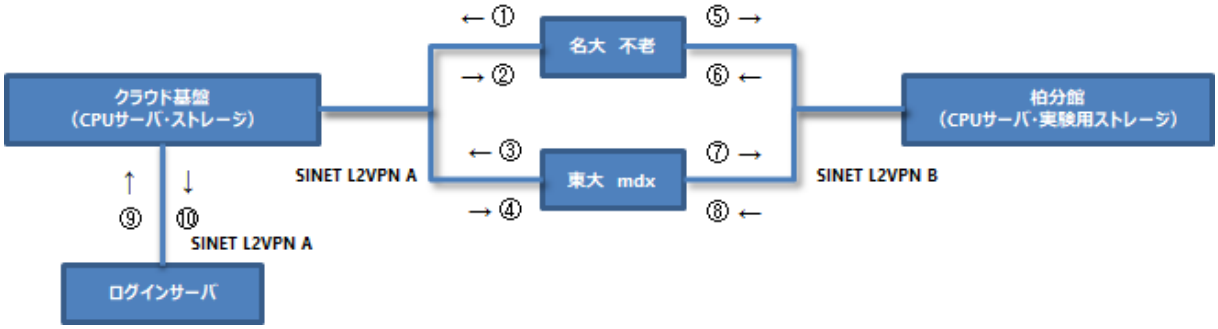

図 7　ネットワーク帯域の調査箇所の模式図

表 2　基盤間の帯域幅の実測結果

	クラウド基盤			柏分館		
	往 (Gbits/sec)	復 (Gbits/sec)	Cwnd (MBytes)	往 (Gbits/sec)	復 (Gbits/sec)	Cwnd (MBytes)
名大 不老	① 1.98	② 1.68	3.19	⑤ 1.62	⑥ 1.66	3.25
東大 mdx	③ 1.99	④ 2.19	3.62	⑦ 1.69	⑧ 1.86	5.00
ログインサーバ	⑨ 2.04	⑩ 2.18	3.25	–	–	–

構築した環境にて計算を実行した。名大の不老にて CT 画像を用いた昨年度末から取り組んでいるファウンデーションモデルの作成を継続して実行した。不老の 2 ノード（GPU ボード 8 枚分）を使って学習させている時の不老側の GW サーバでのトラフィックのスナップショットを図 7 に示す。受信状況（RX）に注目すると、平均的には 2.2 Gbits/sec (272MiB/sec)、瞬間的には 2.6 Gbits/sec (325MiB/sec)の流量が出せていることがわかる。ここで注意すべきは、不老からクラウド基盤への接続は 2 本並列しているということである。これにより、各接続セッションが SINET6 の帯域幅を分割して使用でき、Cwnd のサイズの制約を回避できるため、合計での転送効率を上げることができると考えられる。なお、この計算は、現在、学習のデータサイズを初期の 15 万件から 30 万件、50 万件に順次増やして継続しており、言語モデルと合わせたマルチモーダルな AI 学習への適用も試みている。

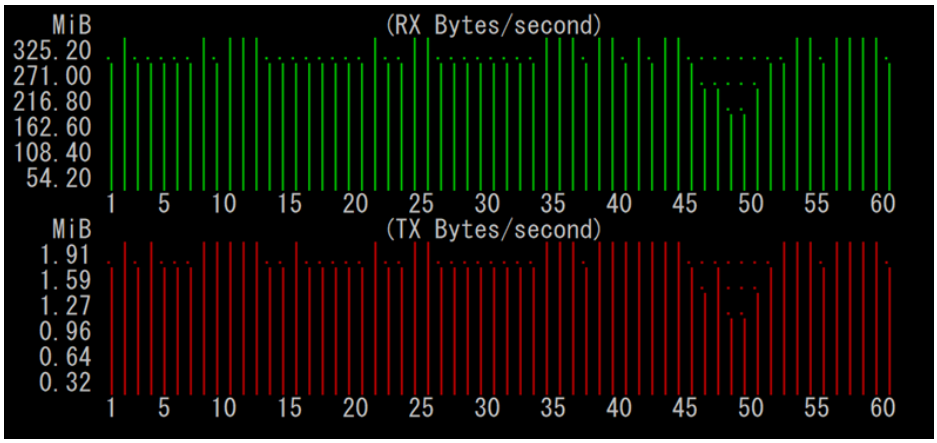

図 8　AI 学習時に不老の GW で観測したトラフィック

このシステムを使い、奈良先端科学技術大学では、日本人の年齢・性別ごとの標準的な筋骨格解析を行った。3 次元的に解析することで体積や平均 CT 値などの解析が正確にできるようになっている。

6.　進捗状況の自己評価と今後の展望

今年度の目標とした基盤連携の HPC 環境の拡張について、ログインサーバの導入、新しい SINET6　L2VPN の導入によって実現が可能となった。具体的には、HPC 基盤としてこれまで名大の不老のみであったが、東大の mdx も利用できるように拡張した。また、回線の帯域測定、実際の AI 学習処理時のトラフィック可視化を通じて、運用上の性能問題が無いことが確認できた。

今後の課題としては、メンテナンスを含めた可用性が挙げられる。今回行ったネットワーク構築作業時にネットワーク機器の保守期限に伴う機器入替などの作業があったが、この時に 5 日間ほどクラウド基盤のサービスを停止せざるを得なかった。その際、AI 学習の処理も計画的に中断しなければいけないのだが、学習用のデータが増大しているので、学習に要する切れ目の良い時期も 2 週間ぐらいに長くなってきている。そこで、メンテナンス時でも AI 学習処理を止めることの無いシステムが求められる。

また、今後の展望として、マルチモーダルのデータを扱うようになって、複数のデータ基盤をまたいで利用する必要性も出てくる。次年度の課題として検討・実験を進めたい。

以上