

留学開始までの経緯

高橋秀明¹

2024年 5月

1. はじめに

はじめまして! 高橋秀明と申します。2024年8月から米国のニューヨーク市にあるColumbia Universityの博士課程に進学します。専攻はComputer Scienceで、2024年3月に東京大学教養学部学際科学科B群総合情報学コースにて学士を取得しました。博士課程では人工知能とセキュリティの境界領域について研究をしていこうと考えています。

本報告書では、まず博士課程出願時の過程を簡単にまとめ、その次に私が中学生から大学生にかけてどんなことに取り組み、各学年におけるスキルのレベルがどの程度であったかを整理します。また、コンピューターサイエンス・プログラミングに関するスキルを上げるために有効であると感じた勉強法もおまけとしてまとめておきます。どこまで再現性があるのかは定かでないですし、少しでも読者の皆様のお役に立てることを願っています。

2. 博士課程出願

出願時点(学部四年の冬頃)での主な業績・スコア等は以下の通りでした。なお、出願後でも、論文が新たに採択された時には適宜出願先の研究室の先生方にメールで報告しました。

- 主著一本がCVPR(画像処理・機械学習分野のトップ国際会議)にポスター枠で採択。
- 主著一本(卒論)がAAMAS(マルチエージェント分野のトップ国際会議)で査読中(出願後、合格発表までの間に口頭発表として採択)。
- 共著一本がICLR(機械学習分野のトップ国際会議)で査読中(出願後、合格発表までの間にポスター枠で採択)。
- 主著一本がIJCAI(機械学習分野のトップ国際会議)で査読中。

¹ <https://koukyosyumei.github.io/KoukyosyumeiResume/>

- 主著一本を情報処理学会の研究会で報告。
- 学部時代のGPAは、3.83/4.0（前期・後期通算）および3.91（後期のみ）。
- TOEFLは102 (Reading 29, Listening 26, Speaking 22, Writing 25)

推薦状は以下の三名からいただきました。推薦状に盛り込んでほしい内容は、ある程度具体的に伝えるようにしていました。

- 東大の配属先研究室の指導教授：主に、AAMASに採択された論文のプロジェクト(プライバシーに配慮した経路を策定するためのAIプランニングに関する理論的研究)について言及していただきました。
- 研究インターンをしていた清華大学の准教授：CVPRおよびICLRに採択された論文のプロジェクト(Federated Learningと呼ばれる分散機械学習・深層学習が抱える個人情報漏洩リスクに関する実証的研究)における役割や、普段から英語でコミュニケーションしていたことに言及していただきました。
- リサーチアシスタントをしていたコンサルティング会社の上司(出願時は横浜国立大学の准教授)：実案件におけるシステム構築の経験、他の学生アシスタント達の指導やコードレビューを行っていたことなどに言及していただきました。

出願先のコンタクトの有無および結果は以下の通りでした。面接に呼ばれた3校(Columbia, CMU, UChicago)は全て合格し、出願を考え始めた頃からの第一志望であったColumbiaに進学することにしました。ColumbiaおよびUChicagoの面接ではこれまでの経験や博士課程で研究したい分野についてそれなりに詳しく質問されました。その一方で、CMUの面接では開始直後にほぼ確実に合格である旨を伝えられ、面接というよりも研究室紹介のような雰囲気でした。

大学	コンタクト	面接	結果
Columbia	あり	あり	合格
CMU - ECE	あり	あり	合格
CMU - CS	なし	なし	不合格
Cornell	あり	なし	不合格
UChicago	あり	あり	合格

Princeton	なし	なし	不合格
UC Berkeley	あり	なし	不合格
Stanford	なし	なし	不合格
MIT	なし	なし	不合格
Harvard	あり	なし	不合格

出願の過程でやってよかったことは、出願先の研究室が近年発表している論文を読むだけでなく、実際に自分で再現実装して実験したことです。論文の提案手法への理解も深まりますし、コンタクトのためのメールのやり取りや面接でも、自分で再現実装までしたというエピソードはかなり受けが良かったように感じました。また、研究室の教授だけでなく、現役の院生の方々にもコンタクトを取り、様々なアドバイスをもらうことができました。

逆に後悔していることとしては、同じ大学の複数の学部に出願する際(CMUのCSとECEに出願など)、別々に推薦状を提出しなければいけない場合があり、推薦者の先生方に混乱を招いてしまうことがありました。

3. 各学年で取り組んでいたこと

学年	やっていたこと・レベル感
中学生	<ul style="list-style-type: none"> ● 物理部無線班で電子工作に取り組み、C言語の簡単なif文やfor文と気合でロボットやゲームの制御をしていた。気合が9割。 ● ポインタや構造体などは全く使いこなせていない。
高校生	<ul style="list-style-type: none"> ● 帰宅部に転部し、Haskellでゲームを実装して遊んでいた。 ● モナドや意味論など、プログラミング言語の理論面がちょっと分かる気がしていた(大学の関数型言語の授業で良(B)しか取れなかったのでおそらく勘違い)。 ● 文理選択に悩み、ペンを転がした結果、文系で大学受験した。

<p>大学1年 (前半)</p>	<ul style="list-style-type: none"> ● 空前の機械学習ブームに流され、色々教科書を読み、web上の初心者向け教材 (Coursera等) でPython・機械学習を勉強した。 ● 理論面はほとんど理解できていなかった。 ● 文科一類という法学部に行く人が多い学科だったこともあり、AIの法規制やAIが抱えるセキュリティ上の問題点に興味を持ち始めた。 ● 力試しに受けたTOEFLの点数は90点台後半だった。
<p>大学1年 (後半)</p>	<ul style="list-style-type: none"> ● ビジネスコンテストを運営する学生団体に協賛金集めを行うために、営業のまねごとをしていた。この国の気候でスーツを着ることの苦しみを知り、スーツを着なくてよい仕事に就こうと決めた。 ● 夏休み後半からFRONTEOで機械学習系の長期インターンを始めた。Pythonで多少複雑なデータ加工・機械学習モデルの学習学習パイプラインを組めるようになったが、理論面の理解はあまりできておらず、論文も特に読んでいなかった。 ● 1年生向けのプログラミング・統計の授業は全て優 (A、おおむね上位3割) 以上だったが、優上 (A+、おおむね上位1割) はあまりない。
<p>大学2年 (前半)</p>	<ul style="list-style-type: none"> ● パンデミックが始まり暇になったので、インターネットで公開されている大学の授業動画などで、線形代数・微積分や機械学習の理論部分などを自習した。 ● 経済学部の研究室でデータセット構築のバイトをし、研究者って楽しそうだなという気持ちが芽生えた。 ● 夏休みはリクルートのデータサイエンス系長期インターンに参加した。SQLやSwiftを使う機会もあり、スキルセットが広がった。
<p>大学2年 (後半)</p>	<ul style="list-style-type: none"> ● Kaggle (国際AIコンペティションの代表的プラットフォーム) でソロ銅メダルが取れるようになった。 ● UTEconという計量経済学・データ分析などを用いるコンサルティング会社のリサーチアシスタントを始めた。 ● PySyftというプライバシー保護型データ分析・機械学習を行うためのOSSに貢献するようになり、開発チームの週次ミーティングにも参加するようになった。英語のミーティングにある程度慣れた。

<p>大学3年 (前半)</p>	<ul style="list-style-type: none"> ● 学際科学科B群総合情報学コースという、情報系学科に進学した。 <ul style="list-style-type: none"> ○ 以下のようなレベル感のプログラムを授業で作っていた。 https://github.com/Koukyosyumei/TenToHako ● パンデミックに伴う自粛に飽きてしまい、上半期はずっと遊んでしまった。ほんとに何にもしていない。全部新宿と六本木のせいだ。 ● 夏休みは奈良先端科学技術大学院大学で、Federated Learning (分散機械学習の一種)を研究し、情報処理学会にて研究会報告を行った。
<p>大学3年 (後半)</p>	<ul style="list-style-type: none"> ● 論文の再現実装にそれなりに慣れた。 ● Kaggleでソロ銀メダルが取れるようになった。 ● 2年後半 ~ 3年生までの成績は、約5割が優、約4割が優上(A+)だった。ただし、前述の通り関数型言語に関する授業は良かった。傲慢になったらそこでおしまい。 ● TOEFLの点数が100点を超えた。
<p>休学 (1年間)</p>	<ul style="list-style-type: none"> ● 中国の清華大学がFederated Learning・プライバシー保護型機械学習を研究するインターンを募集しており、一年休学して参加した。 ● パンデミックのため、オンラインインターンとなってしまったが、主著1本がCVPRに採択された。指導の先生方から研究の進め方・論文の書き方など、本当に多くのことを学べた。 ● 自分でOSSを運営してみたくなり、機械学習モデルに対する様々な攻撃・防御手法を試すAIJack²というツールをC++とPythonで作りはじめた。この報告書の執筆時点で、GitHubのスターが300を超え、書籍や論文でも使用されるなど、少し世界に役立てた気がしている。
<p>大学4年 (前半)</p>	<ul style="list-style-type: none"> ● 卒業単位を埋めるために、工学部のシステムセキュリティ系の授業を4つ受講し、3つが優、1つが優上であった。 ● 配属先の指導教員の勧めで、米国の博士課程に出願することにした。比較的海外大学院に進学する人が多い研究室であり、指導の先生には、SOPの書き方や志望校選びなどでとてもお世話になった。 ● 7月頃から興味のある研究室の先生方にコンタクトを取り始めた。基本的に、メールを送る前に最低一本は論文を再現実装し、思いついた改善案などをメールに盛り込むようにした。

² <https://github.com/Koukyosyumei/AIJack>

大学4年 (後半)	<ul style="list-style-type: none"> ● 8月頃には卒論がだいたい形になっていた。 ● 8月～9月頃、奨学金に応募した。 ● 10月ごろからSOPの原稿を書き始め、指導教員や船井財団の先制点先輩方からフィードバックをいただいた。 ● 12月頭までに、各大学への出願を済ませた。 ● 12月末に、卒論がAAMASに口頭発表枠で採択された。 ● 1月に、清華大学時代の共著論文がICLRに採択された。 ● 1月に、Appleで長期インターンを始めることが決まった。 ● 1～2月にかけて、各大学の面接を受けた。 ● 2～3月にかけて、各大学の合格発表があった。
------------------	---

4. 役に立った勉強方法

コンピューターサイエンス専攻博士課程受験の競争率は年々激化しており、基本的には、有力な教授からの強い推薦状・トップ国際会議に論文採択・高GPAなどを満たす必要があります。それぞれの条件を満たすために様々なテクニックが日々考案されているようですが(内容の近い授業を大量に履修して負担なくGPAを上げるなど)、結局のところ一番大事なものは、自分の実力を上げることでしょう。そこで、筆者がコンピューターサイエンス・プログラミングの勉強をするにあたり、有効であると感じた方法をまとめてみました。

- 自分が普段使っているツール・アルゴリズムなどを自作してみる。例えば、深層学習の研究をしているなら自動微分・深層学習ライブラリを自作してみたり、バイナリ解析をしているなら逆アセンブラを実装してみたりすると、多くの学びがあると思います。
- 教科書や論文に書いてある疑似コードを実際に実装して実験してみる。数式や疑似コードを追うだけでは分からなかった気づきがあるかもしれません。
- 自分がよく使うツールの公式ドキュメントをきちんと読んでみる。今まで知らなかった機能や使い方を発見できるかもしれません。
- 自分が興味のある分野のOSSや自分がよく使うOSSに積極的に貢献してみる。Googleが毎年開催しているGSoCや類似イベントなどいい機会だと思います。
- 自分でOSSプロジェクトを運営し、他の人に使ってもらえるようになることを目指す。この時、ドキュメントを英語でしっかり整備したり、チュートリアルを用意しておく他の人が使いやすく

なると思います。できる範囲でテストコードを書いたりコード整形ツールを導入しておく、継続的な開発・メンテナンスがより楽になります。RedditやHacker Newsなどで宣伝して認知度向上を図るのもいいですね。

- OSS・インターン等でチーム開発の経験をつむ。他の人のコードをレビューしてもらったり、逆にコードをレビューする経験はとても役に立つと思います。
- 基礎は何回だって繰り返していい。例えば、基本的な線形代数や微積・確率の教科書は何度やりなおしてもいいなと強く思います。
- YoutubeやCoursera等で公開されている他の大学の授業で勉強する。自分の大学で開講されていないピックでも無料もしくは低額で学べるなんて素晴らしい時代ですよ。海外大の授業を視聴すれば、英語の勉強にもなるかもしれませんね。
- 変な人に搾取されない。プログラミングが少しできるようになると、色々な人に仕事を依頼されたり、起業の手伝いを頼まれることがあるかもしれません(別にプログラミング・情報系に限った話ではないですが)。自分の時間を使う価値はあるのか、正当な報酬を受け取れるのかどうかは絶対に考慮すべきです。世の中には、世間知らずの学生を搾取しようとする人たちが沢山います。自分の身は自分で守ろう。
- 時給にとらわれ過ぎない。上記と矛盾するようですが、逆に目先の時給にとらわれ過ぎるのも良くないと思います。ある仕事が自分の長期的な目標に対して有用であるなら、時給が下がるとしても、生活に無理のない範囲でそれをやった方がいいと思います。
- いじけない & 焦らない。大学やSNSでは、同年代で自分より遥かに優秀そうな人を簡単に見つけることができてしまいますよね。しかし、そこで卑屈になって勉強をやめてしまったり、逆に焦り過ぎて睡眠時間を削って勉強したりすることはあまりおすすめできません。月並みな表現ですが、隣の芝生は青く見えるものです。特定の分野の勉強を淡々と無理のない範囲で続けていくことが、結局一番の近道であるように感じます。つぶれていった人を何人も見きました。ちゃんと寝てね。
- 見下さない & 奢らない。大学やSNSでは、同年代で自分より遥かに愚鈍そうな人を簡単に見つけることができてしまいますよね。なんかずっとダラダラしているように見えたり、バイトやサークルに明け暮れて勉強なんか全然していない人たち。「自分はあーいう奴らとは違う」、なんて思っていないですか？みーんな同じ穴のムジナ。自分は意識が高いと自負していても、もっと頑張ってる人からは怠けているように見えたりするもんです。傲慢な人ほど、自分が見下していた人にすぐ追いつかれていたり、そもそも自意識と実力乖離がある気がします。人間なんてみんな等しく馬鹿なんです。マサカリ投げても誰も幸せにならないよ。他人に優しく、自分にも優しく。
- EmacsではなくVimを使う。個人の信条です。別にエディタなんて何でもいいと思います。それでも俺はVimを使う。