



Jimmy Lin :
The Neural Hype, Justified! A Recantation



ACM SIGIR Forum Vol.53 No.2 (Dec. 2019)

情報検索コミュニティにおける ニューラルネットワーク

本記事で紹介する記事は The Neural Hype, Justified! A Recantation です。これは ACM の SIG (Special Interest Group) のうち情報検索 (Information Retrieval) に興味を持つ人たちのコミュニティである SIGIR が発行しているニュースレターである SIGIR Forum^{☆1} の 2019 年 12 月号に、Jimmy Lin によって寄稿された Opinion (意見) です (コーナー名は『有名論文』ナナメ読み) ですが……。本文は 3 ページ程度であり、その内容は後述のとおり非常にシンプルです。しかしその内容は端的に情報検索コミュニティにおける近年の研究動向そのものを表している非常に示唆的であるため、背景や経緯にも触れつつ記事を紹介したいと思います。

さて、まずはタイトルを見てみましょう。一単語目の neural とは近年コンピュータ・サイエンスやその融合分野にて爆発的な流行を見せているニューラルネットワークのことでしょう。二単語目の hype の日本語訳を調べると「誇張」や「興奮」と載っていて少しニュアンスが分かりづらいですが、ハイブ・サイクル (hype cycle) を知っていれば多少はイメージしやすいかもしれません。ハイブ・サイクルは、新しい技術が出現すると世間で大きく着目され期待値が上がりすぎて、実際の技術の到達度よりも高い評価を得てしまい、結果

^{☆1} IR コミュニティが関心を持つ事柄に関する一般的な情報や意見、会議やワークショップの報告、論文や書評、博士論文の要旨など幅広い情報を発信する媒体です。

的に誇大広告になる状況を説明しています^{☆2}。これを踏まえると、neural hype とは、「ニューラルネットワークへの期待による興奮」くらいに認識していればよさそうです。そして justified と recantation の日本語訳はそれぞれ「正しいと証明される」と「取り消し」です。一見、「正しいと証明されたにもかかわらず取り消されるとはどういうこと?」となるかもしれません。実はこの Opinion には同じ著者である Jimmy Lin による前日譚 (ACM SIGIR Forum 2018 年 12 月号に寄稿された Opinion である The Neural Hype and Comparisons Against Weak Baselines¹⁾) があり、justified と recantation はその内容に対しての自己回答なのです。justified と recantation の意図を理解するために、まずは 2018 年の記事公開当時に情報検索コミュニティで大きな注目を集めた前日譚の内容をご紹介します。

Jimmy Lin の分析と主張とは?

前日譚¹⁾ の冒頭では、機械学習コミュニティの主要な国際会議である The International Conference on Learning Representations (ICLR) 2018 にて発表された Sculley ら²⁾ の主張を引用して、学術研究が過度に leaderboard chasing^{☆3} に陥り、「勝

^{☆2} 続いて、技術の到達度が過度に高められた期待値に及ばないと気付いた人々の関心が薄れる幻滅期が訪れますが、その後期待値と技術力が歩み寄って最終的には安定状態になります。

^{☆3} リーダーボード (leaderboard) とは、同一のデータセットを用いた評価実験やコンペティションで好成績を示したチームや手法が掲載されたリストのことです。つまり leaderboard chasing とは性能競争のことですが、ここでは、ほかの先行研究より少しでも高い性能を達成することが研究の主眼となっていることに対する批判のニュアンスを含みます。

つ (= 先行研究よりも高い性能を示す)」ことが重視された結果、実際の技術の進捗具合よりも誇張されて技術の進歩が報告される可能性があることに言及しています。そして情報検索コミュニティにおいて大流行中のニューラルネットワークでもそのような事態 (これがまさに neural hype です)、が起こっていないかコミュニティ内部での調査が必要だと述べています。

さて、どのような方法で neural hype の真偽を検証しようとしたのでしょうか。実は、そもそも情報検索コミュニティにおいてこの種の注意喚起がなされたのは初めてのことはありません。Armstrong ら³⁾が行った、1998年から2009年までに発表された論文の調査によって、その10年間で検索技術が向上した根拠は見つけれず、その原因として、弱いベースライン手法 (適切にハイパーパラメータ・チューニングがされていない古典的手法のこと) が設定されているため提案手法の性能が改善したという錯覚を起こしているところを挙げました。注意すべきことは、弱いベースライン手法に対して性能改善を実現した手法は必ずしも強いベースライン手法 (適切にハイパーパラメータ・チューニングがされた古典的手法のこと) の性能を改善するとは限らないということです。であればこそ、提案手法が真に

情報検索技術の性能を改善するかどうかを評価するためには、強いベースライン手法を設定する必要があります。

Jimmy Lin は Armstrong らの流れを汲み、ニューラルネットワークを用いた情報検索手法 (ニューラル検索モデル) で採用されているベースライン手法が適切にチューニングされているか (強いベースライン手法かどうか) を調査しました。具体的には、まずは十分にチューニングを行って強いベースライン手法を自前で構築して、ニューラル検索モデルを提案した論文中で用いられているベースライン手法と比較を行いました。その結果、論文中のベースライン手法は十分にチューニングされておらず、驚くべきことにニューラル検索モデルの性能は強いベースライン手法と同等もしくは劣ることを示しました (参考に、Jimmy Lin らの研究グループにて行った調査⁴⁾の結果を図-1に示します)。この結果を踏まえて、ベースライン手法として「最も良い実装^{☆4)}」を選択することを提案しています。脇道に逸れるためここでは紹介しませんが、そのほかにも情報検索コミュニティの在り方について建設的な提案をしているので、興味のある方はぜひ文献1)を参

☆4 一例として、Jimmy Lin らによってメンテナンスされている Anserini project (<https://github.com/castorini/anserini>) が挙げられています。

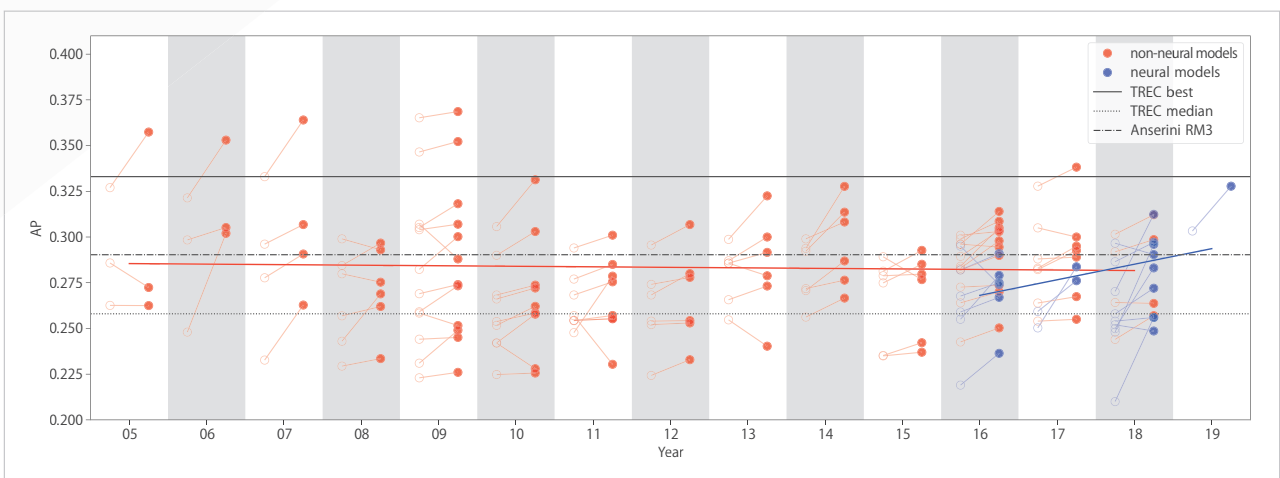


図-1 (文献4)のFigure 1):各線は各論文のベースライン手法 (線の左側の丸の内部が透明) と提案手法 (線の右側の丸の内部が色付き) の性能改善を表現。赤線は非ニューラル検索モデルであり、青線はニューラル検索モデル。黒線のTREC bestは強いベースライン手法を表し、多くのベースライン手法はその性能を下回る。また、いずれのニューラル検索モデルも強いベースラインを上回ることはなかった。

照く下さい。

何が「justified」され、 何を「recantation」したのか

ようやく本題に入ります。1年のときを経て Jimmy Lin は再び SIGIR Forum に Opinion を寄稿します。何が契機となったのでしょうか。2018年から2019年の間に起こったニューラルネットワークに関する重大なトピックとは何でしょうか？と質問すると察する読者もいらっしゃるかもしれませんが、2018年10月に登場したBERT^{☆5}です。Jimmy Lin は記事の冒頭で自らの文献1) (前日譚)での発言を「撤回 (recant)」し、BERT やその関連モデル (Transformer モデルと大規模事前学習による事前学習) によって情報検索手法の性能は著しく、かつ、実質的に向上したことを認めました。つまり、neural hype は誇張や錯覚ではなく「正しいと証明された (justified)」ということです。Jimmy Lin は、BERT のソースコードや大規模事前学習モデルを惜しげもなく公開した Google 社に向けて感謝を述べ、情報検索研究者たちに向けてはこの心躍る時代 (post-BERT ニューラル検索モデルの時代) を楽しもうと呼びかけています。余談ですが、当初俯瞰した立場でニューラル検索と対峙していた Jimmy Lin は、今やニューラル検索分野を牽引する存在の一人です。

情報検索コミュニティを取り巻くニューラルネットワークの象徴的な記事である The Neural Hype, Justified! A Recantation の内容は以上です。しかしながら、(平素の「ナナメ読み」のような) もう少し技術的なトピックにも興味を持たれる読者がいらっしゃるかもしれないため、おまけとして検索モデルのパラダイムと BERT 以降のニューラル検索モデルの概要に触れて、本記事を締めたいと思います。

☆5 BERT については過去のナナメ読み⁵⁾にて非常に分かりやすい解説がされています。

おまけ1：検索モデルのパラダイム

Jimmy Lin はなぜ現在を「心躍る時代」と表現したのでしょうか？それを説明するにあたり、これまでの情報検索研究の歴史を紹介します。

まず、情報検索研究の歴史は、特に機械学習技術の切り口では下記の4つの時代に分けられます。Jimmy Lin らが2020年10月に公開した、情報検索の基本的な事項から post-BERT ニューラル検索モデルまでの網羅的なサーベイ論文⁶⁾も参考にしつつ、下記に要約します。

- a. 伝統的な語のマッチング・重み付け手法 (BM25, 適合性フィードバックによるクエリ拡張など) : ユーザの検索質問 (クエリ) 語と文書中の語のマッチングを行う。その際、文書中の語の出現頻度や語の特定性 (希少性), 文書長 (文書中の語の個数) など、多様な統計量を用いて語の重み付けを行う。また、当時から、クエリと文書の語の mismatches を解消するための技術は継続的に議論されており、自然言語処理分野の技術 (語義の曖昧性解消やシソーラス) を活用する取り組みも多数存在するが、安定した性能向上は達成できなかった。
- b. ランキング学習 (LambdaMART, RankNet など) : 適合度付きのクエリ-文書ペアを用いてモデルを学習する教師あり学習の設定。クエリに対する文書の適合度を推定し、適合度の降順に検索結果のリストを提示する (ランキング問題は文書の適合度の順序のみに着目した回帰問題の一種とも見なされる)。代表的なランキング学習モデルの多くは1990年代に提案され、2000年代以降は特微量エンジニアリングがよりフォーカスされた。なお、ユーザの行動ログを使うことで検索性能を向上することが可能であると一般的に知られていたが、それらを利用できるのは Web 検索エンジンを運営する一部のテック・ジャイアントのみであるため、多くの研究者にとってユーザの行動録ロ

グの恩恵を受けることは困難という認識であった。この時期に提案されたモデルの中にもニューラルネットワークに基づくモデルが存在するものの、ランキング学習モデルに分類されることが一般的である。また、厳密には後述の深層学習ベースの手法もランク付けのための機械学習手法であるが、「ランキング学習」と表現する場合にはこの時期に提案された機械学習ベースのモデルを指すことが一般的である。

- c. pre-BERT ニューラル検索モデル (DSSM, DRMM など) : 画像処理コミュニティや自然言語処理コミュニティにおける深層学習モデルの活躍を受け、情報検索コミュニティでも深層学習に基づくモデルが多数提案される。ランキング学習と異なり、特徴量設計が不要となるため人手の負荷軽減が期待された。また、疎ベクトルで表現される従来の語のマッチングと比較して、高い表現力を持つ密ベクトルの特徴を活かしてクエリと文書の語の mismatches を解消することを目指した研究も多く取り組まれた。
- d. post-BERT ニューラル検索モデル : Transformer モデルと大規模テキストによって学習される高品質な大規模事前学習モデルを用いることで、軽微なファインチューニングを行うだけで高い検索性能を達成できる^{☆6}。また、文脈を考慮した分散表現が得られるため、可変長のテキストの情報の埋め込みが容易となる。

上記の流れをまとめてみましょう。ランキング学習全盛期には、一部の研究者のみが利用できるユーザの行動ログがなければ、当時の state-of-the-art な検索性能のシステム構築は困難でした。その後に出現した深層学習によってゲーム・チェンジが起こることが期待されたものの、実際はブレイクスルーとはなりませんでした。そんな折に登場した

^{☆6} The Neural Hype, Justified! A Reantation では、(従来の例とは異なり) 自然言語処理コミュニティのブレイクスルーが情報検索研究に有益であることが判明した喜びが述べられています。

BERT のおかげで、リソースの限られた研究者でも高性能な検索システムを構築することができるようになりました。著者自身、ランキング学習の高性能化に苦勞した経験を持つため、情報検索研究者にとって現在が心躍る時代という表現に大いに賛同しています。

おまけ2：BERTの登場によって大きく前進したニューラル検索モデル

BERT の登場以降、ニューラル検索モデルの研究の方向性は多岐に及んでいます。本記事の結びとして、そのうちのごく一部を紹介したいと思います。

- a. BERT の推論結果 (適合度) を用いるモデル : monoBERT, Birch, BERT-MaxP など
- b. BERT 最終層の埋め込み表現を後段のニューラル検索モデルへの入力として用いるモデル : CEDR, PARADE など
- c. クエリと文書の語の mismatches を解消するために BERT (Transformer) を利用するモデル : doc-2query, DeepCT, HDCT
- d. BERT (Transformer) を用いて密ベクトル表現の学習を行うモデル : Sentence-BERT, ANCE, ColBERT⁷⁾ など

また、興味深いことに、BERT の登場以来、従来の情報検索コミュニティではあまり着目されてこなかったクエリ処理速度に関する議論が盛んになっていると感じます。BERT による推論時間が従来の語のマッチングベースの手法と比較しても圧倒的に遅いことが理由ですが (図-2 参照)、逆にいうと高速化に取り組んでも使う価値のある、実用性の高いニューラル検索モデルが実現したことを示唆しているように思われます。この調子で実用に耐え得る高速検索も実現すれば、BERT ベースの検索モデルは産業界にも広まるのではないかと考えていま

す。いずれにせよ、筆者も情報検索に取り組んできた研究者として、この時代に心躍らせながら引き続き研究に邁進していきたいと思えます。

参考文献

- 1) Lin, J. : The Neural Hype and Comparisons Against Weak Baselines, SIGIR Forum, 52 (2) : 40-51 (2018).
- 2) Sculley, D., Snoek, J., Rahimi, A., and Wiltchko, A. : Winner's Curse? On Pace, Progress, and Empirical Rigor, In Proceedings of the 6th International Conference on Learning Representations, Workshop Track (ICLR 2018) (2018).
- 3) Armstrong, T. G., Moffat, A., Webber, W. and Zobel, J. : Improvements That Don't Add Up: Ad-Hoc Retrieval Results Since 1998, In Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 1998), pp.601-610 (2009).
- 4) Yang, W., Lu, K., Yang, P., and Lin, J. : Critically Examining the "Neural Hype": Weak Baselines and the Additivity of Effectiveness Gains from Neural Ranking Models, In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019), pp.1129-1132 (2019).

- 5) 柴田知秀 : 5分で分かる!? 有名論文ナメ読み : Jacob Devlin et al. : BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding, 情報処理, vol.61, No.3, pp.294-295 (Feb. 2020).
- 6) Lin, J., Nogueira, R., Yates, A. : Pretrained Transformers for Text Ranking: BERT and Beyond, arXiv(2020), <https://arxiv.org/abs/2010.06467>
- 7) Khattab, O. and Zaharia, M. : ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT, In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020), pp.39-48 (2020).

(2022年6月21日受付)



榎 惇志 (正会員)

a.keyaki@r.hit-u.ac.jp

一橋大学ソーシャル・データサイエンス教育研究推進センター准教授。博士(工学)。2014年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了。東京工業大学, (株) デンソーアイティラボラトリーを経て, 2022年より現職。情報アクセス技術に関する研究に従事。

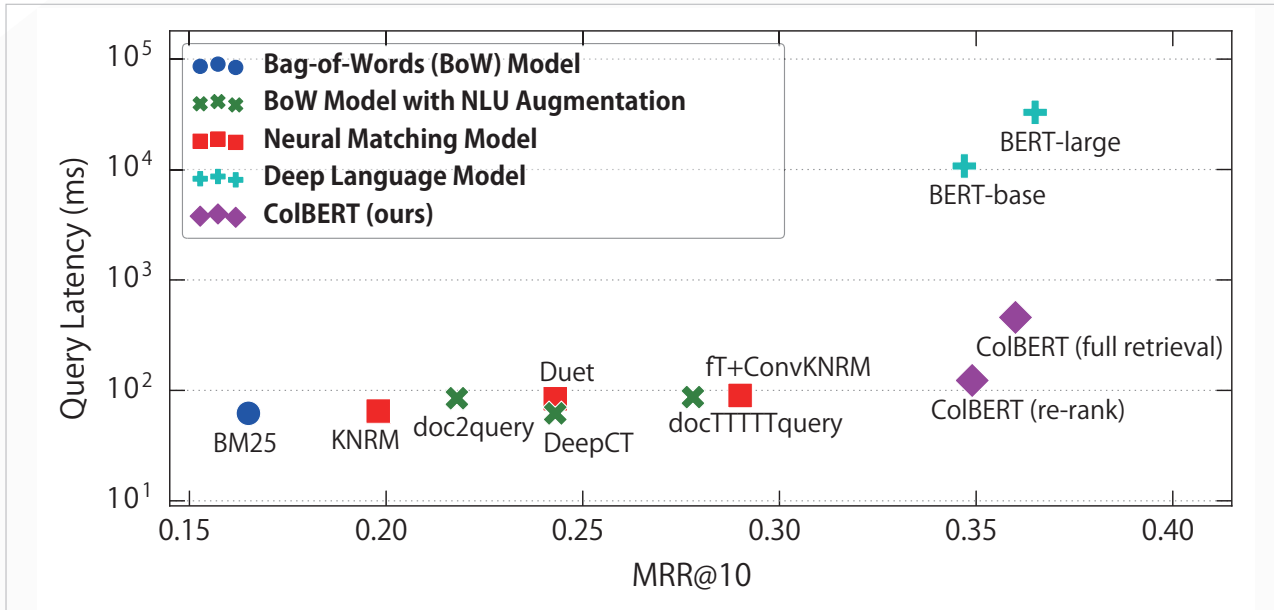


図-2 (文献7)のFigure 1) : BM25と比較してBERTのクエリ処理時間は100倍から1,000倍の間である。高速化の工夫を取り入れたColBERTでは、検索性能は同程度を維持しつつ、BM25やそのほかのニューラル検索モデルに迫るクエリ処理時間を達成した。