

直近1年の動向を考慮した最新論文のインパクト予測

平子 潤 笹野 遼平 武田 浩一

名古屋大学大学院情報学研究科

hirako.jun.e5@s.mail.nagoya-u.ac.jp

{sasano,takedasu}@i.nagoya-u.ac.jp

概要

発表される論文の増加に伴い、新たに発表された論文の将来的なインパクトを自動で予測する技術の重要性は増している。本研究では、論文のインパクト予測を発表から1年後の被引用数予測として定式化し、学習データとしてそのまま用いることができない、発表から1年未満の論文から得られる情報を活用する手法を提案する。また、arXivから収集した約17,000の論文を用いた実験を通して、直近1年以内の論文を活用することの有効性を示す。

1 はじめに

近年、自然言語処理を含む人工知能分野の論文数は大きく増加しており、興味があるトピックに絞ったとしても全ての論文に目を通すことは困難になっている。このため、同時期に発表された論文の中から、将来的なインパクトが大きくなる重要性の高い論文を選択的に自動提示する技術が求められている。本研究では、このようなインパクトの指標として発表から1年後の被引用数を採用し、新たに発表された論文の1年後の被引用数予測に取り組む。

発表から1年後の被引用数予測において、学習データとしてそのまま使用できるのは1年後の被引用数が既知の論文、すなわち、発表から1年以上経過した論文のみである。しかし、人工知能分野の技術革新の速度は増しており、直近1年以内に発表された論文から得られる情報を用い、最新の動向を捉えることが重要であると考えられる。そこで本研究では、1年後の被引用数予測において、直近1年以内に発表された論文を活用する2つの手法を提案する。1つ目の**被引用数補完**では、過去の論文の被引用数の経過月数ごとの推移を利用し、発表から1年未満の論文の1年後の被引用数を予測することで、発表から1年未満の論文を学習データとして利用する。2つ目の**早期導入度**の利用では、被引用数予測対象の論文が、直近1年以内に発表された被引用数

の高い論文を引用しているかどうかの情報(早期導入度)を利用する。これは発表後、短期間で多く引用されている論文は重要な研究成果を報告している可能性が高く、そのような重要な研究成果をいち早く取り込んでいる研究も、新規性が高く技術的貢献が大きい可能性が高いという直観に基づいている。

被引用数予測に取り組んだ研究はすでに多く存在する。初期の研究としてFuら[1]は、被引用数予測を分類と回帰で定式化し、論文のアブストラクトやジャーナルのインパクトファクターなどの被引用数予測において有効な特徴量を調査した。Liら[2]は、論文に対応した複数の査読テキストを被引用数予測に活用した。Dongenら[3]は、より大規模なデータセットを使いつつ、BERTを用いて効率的に論文の本文をエンコードし、被引用数予測を行った。しかし、これらの研究では、予測対象の論文が発表された時点では利用できない情報、具体的にはその論文と近い時期に発表された論文の将来的な被引用数を利用しており、発表されてから間もない論文の将来的な被引用数予測には適用できない設定が採用されている。これに対し、本研究では被引用数予測の対象論文が発表された時点の情報だけを利用するという設定を採用し、そのような設定における直近1年以内の論文の活用方法を提案する。

2 最新論文を活用した被引用数予測

2.1 問題設定

本研究では、論文の著者情報とタイトル、アブストラクトから、その論文の発表から1年後の被引用スコアを予測するタスクに取り組む。被引用スコアは、Wennigerら[4]に倣い、被引用数の対数、すなわち論文の発表から1年後の被引用数 c に対し、 $\log(c+1)$ と定義する。

学習に使用できる情報は、被引用数予測の対象論文が発表された時点で入手可能な情報に限定する。すなわち、本研究で予測対象とする1年後の被引用

表 1 n ヶ月以内に発表された被引用数が上位 $k_1 \sim k_2\%$ の論文を引用している論文の発表 1 年後の被引用数の平均。「引用なし」は、 n ヶ月以内に発表された論文を引用していない論文、括弧内の数字は各グループに属する論文の割合を示す。

	上位 0~1%	1~2.5%	2.5~5%	5~10%	10~25%	25~100%	引用なし
3ヶ月以内	15.5 (4.6%)	14.3 (3.8%)	10.6 (3.6%)	8.8 (5.0%)	7.5 (6.6%)	6.5 (4.9%)	5.0 (71.5%)
6ヶ月以内	14.3 (9.6%)	12.6 (7.3%)	9.8 (6.2%)	7.6 (8.8%)	6.3 (10.7%)	4.6 (9.4%)	3.7 (48.1%)
9ヶ月以内	13.8 (15.4%)	11.2 (10.3%)	7.7 (7.8%)	6.6 (10.1%)	5.3 (12.9%)	3.5 (12.4%)	2.5 (31.0%)
12ヶ月以内	12.7 (21.6%)	10.1 (12.0%)	6.4 (9.1%)	5.7 (10.5%)	4.4 (13.5%)	2.5 (12.9%)	2.0 (20.6%)

数情報が利用できるのは、それより 1 年以上前に発表された論文のみであり、直近 1 年に発表された論文についてはその時点の被引用数のみを利用する。

2.2 ベースライン手法

Transformer [5] ベースの事前学習済みモデルである BERT [6] を fine-tuning することで、被引用数予測を行う。入力に用いる情報は、著者情報、タイトル、アブストラクトとし、1 文目として著者情報とタイトルを繋げたもの、2 文目としてアブストラクトを入力する。ただし、著者情報には著者名ではなく、以下の手順で生成する特殊トークンを用いる。まず、論文の各著者がその論文発表時から見て 1 年以上前かつ 5 年以内に発表した論文の発表 1 年後の被引用数を収集する。続いて、各著者ごとに平均被引用数を算出し、さらにその値を全著者で平均した値をその論文の著者スコアとする。このような著者スコアを学習用論文すべてに対し算出し、スコアが高いものから順に 10% ごとにグループに分け、上位 10% のグループであれば [著者 0~10%]、50%~60% のグループであれば [著者 50~60%] といった特殊トークンを生成する。

BERT の出力として、特殊トークンである [CLS] に対応するベクトルを利用し、CLS ベクトルを線形変換することで、被引用スコアの予測値を得る。学習中には CLS ベクトルにドロップアウト [7] を適用し、予測値と被引用スコアの平均二乗誤差 (MSE) が最小になるように学習を行う。

2.3 被引用数補完

予測対象である 1 年後の被引用数情報が利用できるのは、発表から 1 年以上経過した論文のみであるが、最新の動向を反映したモデルを構築するためには、発表から 1 年未満の論文も考慮することが重要であると考えられる。そこで、過去の論文の被引用数の経過月数ごとの推移を利用し、発表から 1 年未満の論文の 1 年後の被引用数を予測することで、被引用数の補完を行い、発表から 1 年未満の論文を学

習データとして利用する。被引用数の補完には次の事例ベースの手法と比率ベースの手法を併用する。

事例ベース 学習用論文において、発表から n ヶ月後に被引用数が c の論文の 1 年後の被引用数を全て抽出し、その中央値を利用

比率ベース 発表から n ヶ月後の被引用数の平均と発表から 1 年後の被引用数の平均の比を計算し、被引用数 c に掛けた値を利用

事例ベースの補完は、同じ被引用数をもつ論文が多数存在しうる“被引用が少ない論文”に対しては正確な補完が期待できる一方で、同じ被引用数をもつ論文がない、または、少ないと考えられる“被引用が多い論文”には適していないと考えられる。そのため本研究では、発表から n ヶ月後の被引用数の中で、被引用数 c が上位 10% に属する場合は比率ベースの手法で補完を行い、下位 90% に属する場合は事例ベースの手法で補完を行う¹⁾。

2.4 早期導入度

発表から短い期間に多く引用されている論文は重要な研究成果を報告している可能性が高く、そのような重要な研究成果をいち早く取り込んでいる研究も、新規性が高く技術的貢献が大きい可能性が高いと考えられる。この仮説を検証するため、“ n ヶ月以内に発表された被引用数が上位 $k_1 \sim k_2\%$ の論文”を引用している論文の、発表から 1 年後の被引用数の平均を調査した。

調査には、3.1 節で述べる学習用の論文集合の 1 つである 2015 年 6 月から 2020 年 5 月に arXiv²⁾ 上で発表された 15,962 論文を用い、上位 $k\%$ の計算には、発表から m ヶ月後の論文の場合、論文集合から発表から m ヶ月後時点の被引用数の分布を作成し、その分布内での順位を用いた。また、 n ヶ月以内に発表された論文を複数引用している場合、その中でもっとも被引用数の順位が高い論文の順位を利用した。

調査結果を表 1 に示す。全体的な傾向として、よ

- 1) 被引用数補完の性能については付録 A を参照。
- 2) <https://arxiv.org/>

り被引用数が高い論文を、より早期に引用しているほど、平均的な被引用数が高いことが確認できる。たとえば、発表から3ヶ月以内に、被引用数が上位1%以内の論文を引用している論文の平均被引用数は15.5であり、これは全論文の平均被引用数6.5の約2.4倍であった。

以上の調査結果を踏まえ、本研究では、論文が早期に被引用数が高い論文を引用しているという情報(早期導入度)を被引用数予測に利用する。入力方法としては、著者情報と同様に、早期導入度を特殊トークン化して1文目の先頭に挿入する。本研究では、「上位0~1%」、「上位1~2.5%」、「上位2.5~5%」、「上位5~10%」、「上位10~25%」、「上位25~100%」、「引用なし」をそれぞれ表す7つの特殊トークンを作成し、予測対象の論文がそれぞれ3, 6, 9, 12ヶ月以内に引用している論文の中で最も被引用数の順位が高い論文に対応する特殊トークン4つを1文目の先頭に挿入することにより早期導入度を考慮する。たとえば、ある論文が3ヶ月以内に発表された論文を引用しておらず、6, 9ヶ月以内に発表された被引用数が上位5~10%の論文を引用しており、12ヶ月以内に発表された被引用数が上位0~1%の論文を引用している場合に入力する特殊トークン列は、「[引用なし][上位5~10%][上位5~10%][上位0~1%]」となる。

3 実験

3.1 データセット

被引用数予測において直近1年以内の論文を活用することの有効性を確認するため、arXivで発表された論文を対象とした評価実験を行った。arXivの論文を対象としたのは、大量の論文が発表される一方で、無査読であり品質が保証されていないことから、論文の質の自動推定技術の重要性が極めて高い論文集合であると考えたためである。

本研究では、2019年7月~2020年6月の各月にarXiv上でcs.CL(計算言語学)カテゴリで発表された論文集合をそれぞれ評価セットとする12個のデータセットを用いて実験を行った。各データセットにおいて、評価対象の論文の発表から過去5年間に発表された論文を学習用論文として用いた。たとえば、2019年7月に発表された論文集合を評価セットとする場合、2014年7月~2019年6月に発表された論文が学習用論文となる。学習用論文数の平均、学

習用論文のうち1年以上経過した論文数の平均、評価用論文数の平均はそれぞれ13,457, 8,714, 490であった³⁾。発表時期は月単位で扱い、過去のある時点における被引用数を算出するための引用情報は、Semantic Scholar⁴⁾から収集した。

3.2 比較手法

本研究では以下の5つのモデルを比較する。

全1年後被引用数使用 過去5年間に発表された全論文の発表から1年後の被引用数を用いベースライン手法で学習したモデル

直近1年論文不使用 発表から1年以内のものを除く過去5年間に発表された論文を用いベースライン手法で学習したモデル

+被引用数補完 被引用数補完を行うことで過去5年間に発表された全論文を用いベースライン手法で学習したモデル

+早期導入度 直近1年論文不使用モデルに早期導入度を追加したモデル

+補完+早期導入度 被引用数補完モデルに早期導入度を追加した提案手法

このうち、全1年後被引用数使用モデルは、評価用論文の発表時に取得不可能な情報を用いており、最新論文の被引用数予測には利用できない。

3.3 評価方法

評価指標には大きく分けて2つの指標を用いる。まず、全体的な被引用数予測性能を評価するため、スピアマンの順位相関係数(ρ)を用いる。さらに、将来的にインパクトが大きくなる論文を上位にランキングできているか評価するため、実際の被引用数が上位n%となる論文が、予測値の上位k%にどのくらい含まれるかという指標(n%@k%)を用いる。

また、3.1節で述べた通り、データセットごとの評価用論文の平均論文数は490と小さいことから、データセットごとに評価した場合、スコアが安定しない可能性が考えられる。そこで、より大規模な論文集合で評価を行えるように、評価値の算出は全データセットの評価用論文を用いて月横断的に行う。具体的には、各データセットで学習を行い、評価用論文の被引用数スコアの予測値を出力し、全12か月分の予測値を合わせて評価値を算出する⁵⁾。

3) データセットごとの論文数は付録Bを参照。

4) <https://www.semanticscholar.org/>

5) データセットごとの実験結果は付録Cを参照。

表 2 被引用数予測の実験結果 (評価値 × 100)

初期重み	手法	ρ	5%@5%	5%@10%	5%@25%	10%@10%	10%@20%	10%@50%
BERT	全 1 年後被引用数使用	43.5 ± 0.3	30.5 ± 1.2	49.2 ± 0.9	75.5 ± 1.2	38.4 ± 0.4	57.8 ± 0.5	88.0 ± 0.4
	直近 1 年論文不使用	40.1 ± 0.2	22.7 ± 2.2	38.4 ± 2.7	69.7 ± 1.7	31.2 ± 1.2	53.2 ± 0.5	85.6 ± 0.8
	+ 被引用数補完	42.3 ± 0.3	31.5 ± 0.9	47.4 ± 0.2	72.8 ± 0.0	36.7 ± 1.3	56.4 ± 0.7	86.2 ± 0.4
	+ 早期導入度	42.9 ± 0.2	25.5 ± 1.6	44.1 ± 0.5	72.2 ± 2.6	34.5 ± 0.4	54.6 ± 0.3	87.9 ± 0.4
	+ 補完 + 早期導入度	44.1 ± 0.2	30.5 ± 0.4	50.9 ± 1.2	77.9 ± 0.7	38.1 ± 0.6	59.5 ± 0.4	88.7 ± 0.2
SciBERT	全 1 年後被引用数使用	43.9 ± 0.3	33.3 ± 1.2	52.0 ± 1.2	76.0 ± 2.0	39.2 ± 0.5	58.3 ± 0.4	88.8 ± 0.8
	直近 1 年論文不使用	41.1 ± 0.6	29.8 ± 0.9	43.9 ± 2.2	71.8 ± 1.2	33.8 ± 1.9	53.6 ± 0.5	85.5 ± 0.4
	+ 被引用数補完	42.8 ± 0.2	34.5 ± 0.5	51.0 ± 0.9	75.4 ± 1.4	37.8 ± 0.9	57.5 ± 0.4	87.0 ± 0.5
	+ 早期導入度	43.8 ± 0.2	29.6 ± 0.6	47.2 ± 1.9	74.3 ± 1.6	35.5 ± 0.9	57.0 ± 0.0	88.4 ± 0.4
	+ 補完 + 早期導入度	44.9 ± 0.0	34.8 ± 2.2	51.8 ± 0.9	78.3 ± 0.9	39.2 ± 0.4	59.3 ± 0.3	89.6 ± 0.7

3.4 実験設定

BERT の初期重みには, Transformers [8] で公開されている, Wikipedia などの一般ドメインのコーパスで事前学習を行なった BERT (bert-base-uncased) と, 大規模な論文で構築された科学分野のコーパスで事前学習を行なった SciBERT [9] (allenai/scibert_scivocab_uncased) を用いた. 全ての手法において, エポック数は 3, バッチサイズは 32, 最適化関数は AdamW [10] を用い, 全体の学習ステップの 10% で warm-up し, 残りのステップで線形に減衰する学習率スケジューリングを行った. 学習率は, 2019 年 6 月に発表された論文を評価用論文として作成したデータセットを用いて, Devlin ら [6] に倣い, $2e-5$, $3e-5$, $5e-5$ で探索を行い, 全ての手法において最も順位相関係数が高くなった $2e-5$ を使用した. また, 各手法でランダムシードで 3 回実験を行って評価値の平均と標準偏差を算出した.

3.5 実験結果

実験結果を表 2 に示す. 全体的に SciBERT の性能が BERT の性能を上回っており, 科学分野のコーパスでの事前学習が, 被引用数予測タスクにおいて有効であることが確認できる. また, 被引用数補完, 早期導入度をそれぞれ用いることで, 性能が向上する傾向が確認できる. 特に, 被引用数補完と早期導入度を組み合わせた提案手法は, 予測対象の論文の発表時に利用可能な情報しか用いていないにも関わらず, 全 1 年後被引用数使用モデルと同等以上の性能となった. 被引用数補完は, 全 1 年後被引用数使用モデルで使用している情報を近似的に再現しているに過ぎないことから, 性能向上は主に, 早期導入

度を考慮することで, 最新の技術動向をいち早く取り込んでいる研究を評価できるようになったためであると考えられる. 実際の予測性能に着目すると, 被引用数補完と早期導入度の両方を用いた SciBERT ベースのモデルは, 5%@25% で 78.3 という値であった. これは, 与えられた論文集合のうち 25% を読むことで, 将来の被引用数が上位 5% となる論文の 78.3% をカバーできるということであり, 実用面において有用性の高い手法であるといえる.

4 おわりに

本論文では, 与えられた論文の発表 1 年後の被引用数予測タスクにおいて, 最新の動向を捉えることの重要性に着目し, 1 年後の被引用数が未知であることからそのままでは学習データとして使用できない直近 1 年以内に発表された論文の活用法を提案した. 具体的には, 過去の論文の被引用数の推移から 1 年後の被引用数を推定し学習データとして利用する手法と, より被引用数が高い論文をより早期に引用している論文は平均的な被引用数が高くなるという傾向に着目した早期導入度を利用する手法の 2 つ手法を提案した. arXiv から収集した約 17,000 の論文を用いて行った実験の結果, 直近 1 年以内に発表された論文を活用することで, 被引用数予測の性能が向上することが確認できた. 特に, 早期導入度を用いて, 最新の技術動向をいち早く取り込んでいる研究を評価することで, 論文発表時点では取得不可能な未来の被引用数を利用した手法と同等以上の性能が達成できた. 今後の課題としては, 自然言語処理以外の人工知能分野の論文を含めた評価や, 提案手法では活用していない本文や図, 表などを考慮したモデルの構築が挙げられる.

参考文献

- [1] Lawrence D. Fu and Constantin F. Aliferis. Models for predicting and explaining citation count of biomedical articles. In **AMIA Annual Symposium Proceedings**, pp. 222–226, 2008.
- [2] Siqing Li, Wayne Xin Zhao, Eddy Jing Yin, and Ji-Rong Wen. A neural citation count prediction model based on peer review text. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 4914–4924, 2019.
- [3] Thomas van Dongen, Gideon Maillette de Buy Wenniger, and Lambert Schomaker. SChuBERT: Scholarly document chunks with BERT-encoding boost citation count prediction. In **Proceedings of the First Workshop on Scholarly Document Processing (SDP)**, pp. 148–157, 2020.
- [4] Gideon Maillette de Buy Wenniger, Thomas van Dongen, Eleri Aedmaa, Herbert Teun Kruitbosch, Edwin A. Valentijn, and Lambert Schomaker. Structure-tags improve text classification for scholarly document quality prediction. In **Proceedings of the First Workshop on Scholarly Document Processing (SDP)**, pp. 158–167, 2020.
- [5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, **Advances in Neural Information Processing Systems (NeurIPS)**, Vol. 30, 2017.
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (NAACL)**, pp. 4171–4186, 2019.
- [7] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. **Journal of Machine Learning Research (JMLR)**, Vol. 15, No. 56, pp. 1929–1958, 2014.
- [8] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. Transformers: State-of-the-art natural language processing. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP): System Demonstrations**, pp. 38–45, 2020.
- [9] Iz Beltagy, Kyle Lo, and Arman Cohan. SciBERT: A pretrained language model for scientific text. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 3615–3620, 2019.
- [10] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In **International Conference on Learning Representations (ICLR)**, 2019.

A 被引用数補完の性能

被引用数補完性能の評価結果を表 3 に示す。評価は、被引用数補完前、被引用数補完後のデータセットごとに、全論文の 1 年後の被引用数を正解データセットした場合の、被引用数のスピアマンの順位相関係数 (ρ)、および、正解データにおいて上位 10%に含まれる論文がどのくらい上位 10%に含まれるかの値 (10%@10%) により行った。評価データには 3.1 節で述べた、12 個のデータセットを用い、各評価指標の平均値と標準偏差を算出した。

表 3 被引用数補完の評価結果 (評価値 $\times 100$)

データセット	ρ	10%@10%
被引用数補完前	88.1 \pm 1.5	81.5 \pm 2.0
被引用数補完後	92.4 \pm 0.4	89.6 \pm 1.9

B 各データセットのサイズ

2019 年 7 月 ~2020 年 6 月の各月に arXiv 上で cs.CL カテゴリで発表された論文集合をそれぞれ評価セットとする 12 個のデータセットの統計値を表 4 に示す。

表 4 各データセットの論文数

論文集合	2019/7	2019/8	2019/9	2019/10	2019/11	2019/12	2020/1	2020/2	2020/3	2020/4	2020/5	2020/6
学習用論文 (すべて)	11026	11404	11854	12529	13031	13552	13820	14049	14339	14617	15305	15962
学習用論文 (1 年以上経過のみ)	7044	7436	7806	8098	8475	8660	8779	8991	9204	9609	9956	10520
評価用論文	404	479	720	550	564	345	260	326	334	747	713	440

C 各データセットにおける実験結果

3 節で行った実験の、各データセットごとの実験結果を示す。各評価指標のうち、表 5 にはスピアマンの順位相関係数、表 6 には 10%@10%の結果を示す。

表 5 各データセットにおける実験結果 (スピアマンの順位相関係数 $\times 100$)

初期重み	手法	2019/7	2019/8	2019/9	2019/10	2019/11	2019/12	2020/1	2020/2	2020/3	2020/4	2020/5	2020/6
BERT	全 1 年後被引用数使用	49.7 \pm 0.9	52.9 \pm 1.3	51.9 \pm 0.8	37.6 \pm 0.9	40.6 \pm 0.9	44.3 \pm 2.0	36.3 \pm 0.4	46.9 \pm 0.5	35.5 \pm 0.4	36.7 \pm 1.1	43.4 \pm 1.0	31.9 \pm 1.1
	直近 1 年論文不使用	46.5 \pm 1.6	50.0 \pm 1.4	47.1 \pm 0.2	34.3 \pm 1.7	39.9 \pm 0.3	42.1 \pm 0.4	34.8 \pm 1.2	46.8 \pm 0.6	35.1 \pm 0.9	34.0 \pm 0.4	40.7 \pm 0.7	30.3 \pm 0.6
	+ 被引用数補完	47.1 \pm 0.4	50.5 \pm 1.7	49.6 \pm 1.1	36.1 \pm 0.7	40.3 \pm 1.1	43.0 \pm 1.9	35.7 \pm 1.1	46.0 \pm 0.9	32.7 \pm 1.1	35.8 \pm 0.4	42.8 \pm 0.9	32.7 \pm 0.8
	+ 早期導入度	50.5 \pm 0.3	53.6 \pm 0.4	49.9 \pm 0.9	40.0 \pm 0.9	44.2 \pm 0.6	47.3 \pm 1.6	37.8 \pm 0.7	50.6 \pm 0.7	37.1 \pm 1.7	35.6 \pm 0.7	43.9 \pm 0.8	30.6 \pm 1.2
	+ 補完 + 早期導入度	48.9 \pm 0.6	52.5 \pm 0.5	50.5 \pm 0.1	40.1 \pm 0.8	44.1 \pm 1.1	47.5 \pm 1.4	37.4 \pm 0.5	48.8 \pm 1.9	35.1 \pm 2.3	37.5 \pm 0.4	45.1 \pm 0.2	31.9 \pm 0.6
SciBERT	全 1 年後被引用数使用	50.0 \pm 1.0	53.6 \pm 0.8	52.1 \pm 1.0	39.5 \pm 0.7	42.4 \pm 1.1	42.9 \pm 0.5	39.0 \pm 1.4	47.8 \pm 0.5	33.3 \pm 1.9	37.5 \pm 0.6	45.2 \pm 0.7	28.1 \pm 1.0
	直近 1 年論文不使用	47.2 \pm 0.9	51.2 \pm 0.7	48.1 \pm 1.9	34.1 \pm 0.6	40.0 \pm 1.3	40.8 \pm 1.4	38.4 \pm 1.3	46.6 \pm 0.8	32.1 \pm 0.4	35.3 \pm 0.2	43.1 \pm 0.1	27.6 \pm 0.2
	+ 被引用数補完	48.4 \pm 1.1	51.0 \pm 0.7	50.7 \pm 0.9	37.4 \pm 0.5	41.9 \pm 0.8	42.2 \pm 0.4	37.3 \pm 0.8	46.5 \pm 0.9	30.8 \pm 1.9	36.4 \pm 1.2	45.0 \pm 0.1	27.7 \pm 0.6
	+ 早期導入度	51.3 \pm 0.9	54.8 \pm 0.3	50.4 \pm 1.1	40.3 \pm 1.5	43.5 \pm 0.9	45.7 \pm 0.6	39.6 \pm 0.5	49.8 \pm 0.5	34.7 \pm 0.4	36.7 \pm 0.6	46.5 \pm 0.6	29.8 \pm 0.3
	+ 補完 + 早期導入度	51.3 \pm 1.6	54.4 \pm 0.2	53.1 \pm 0.9	41.0 \pm 0.7	44.9 \pm 1.0	45.5 \pm 1.2	39.4 \pm 0.9	49.8 \pm 0.6	34.3 \pm 1.1	37.7 \pm 0.5	46.7 \pm 0.3	28.6 \pm 1.5

表 6 各データセットにおける実験結果 (10%@10% $\times 100$)

初期重み	手法	2019/7	2019/8	2019/9	2019/10	2019/11	2019/12	2020/1	2020/2	2020/3	2020/4	2020/5	2020/6
BERT	全 1 年後被引用数使用	54.2 \pm 1.4	44.7 \pm 3.7	38.9 \pm 1.4	29.7 \pm 1.0	32.1 \pm 0.0	49.0 \pm 3.4	34.6 \pm 3.8	34.4 \pm 3.1	34.3 \pm 1.7	33.8 \pm 5.9	31.0 \pm 1.4	31.1 \pm 2.6
	直近 1 年論文不使用	47.5 \pm 0.0	32.6 \pm 2.5	31.5 \pm 1.6	24.8 \pm 2.8	31.0 \pm 4.5	32.4 \pm 5.9	37.2 \pm 2.2	36.5 \pm 3.6	30.3 \pm 0.0	34.7 \pm 0.8	25.4 \pm 1.4	26.5 \pm 3.5
	+ 被引用数補完	50.0 \pm 2.5	42.6 \pm 4.3	36.1 \pm 2.4	29.1 \pm 4.8	31.5 \pm 5.7	51.0 \pm 1.7	42.3 \pm 3.8	32.3 \pm 7.9	31.3 \pm 1.7	34.2 \pm 3.4	30.5 \pm 2.2	26.5 \pm 5.7
	+ 早期導入度	48.3 \pm 1.4	38.3 \pm 2.1	34.3 \pm 1.6	35.8 \pm 3.8	36.3 \pm 3.7	39.2 \pm 3.4	37.2 \pm 2.2	42.7 \pm 1.8	30.3 \pm 3.0	32.9 \pm 1.6	30.0 \pm 1.6	26.5 \pm 3.5
	+ 補完 + 早期導入度	52.5 \pm 2.5	40.4 \pm 0.0	39.8 \pm 2.9	32.1 \pm 2.1	39.3 \pm 3.6	56.9 \pm 1.7	41.0 \pm 2.2	35.4 \pm 3.6	31.3 \pm 3.5	40.1 \pm 0.8	30.5 \pm 2.9	27.3 \pm 3.9
SciBERT	全 1 年後被引用数使用	54.2 \pm 1.4	44.7 \pm 2.1	42.1 \pm 0.8	25.5 \pm 1.8	38.7 \pm 5.5	53.9 \pm 6.1	41.0 \pm 8.0	45.8 \pm 1.8	30.3 \pm 0.0	33.8 \pm 4.9	32.9 \pm 1.6	26.5 \pm 3.5
	直近 1 年論文不使用	51.7 \pm 2.9	36.9 \pm 1.2	33.3 \pm 4.8	26.7 \pm 2.1	31.5 \pm 4.5	48.0 \pm 3.4	33.3 \pm 5.9	44.8 \pm 1.8	28.3 \pm 1.7	33.3 \pm 1.6	28.2 \pm 1.4	22.0 \pm 2.6
	+ 被引用数補完	53.3 \pm 1.4	41.1 \pm 5.4	42.1 \pm 0.8	26.7 \pm 2.8	37.5 \pm 3.6	55.9 \pm 0.0	42.3 \pm 3.8	44.8 \pm 1.8	29.3 \pm 1.7	32.4 \pm 2.7	34.3 \pm 1.6	25.0 \pm 3.9
	+ 早期導入度	52.5 \pm 2.5	39.7 \pm 4.9	37.0 \pm 4.9	33.9 \pm 4.2	32.7 \pm 1.0	54.9 \pm 1.7	43.6 \pm 2.2	42.7 \pm 1.8	26.3 \pm 4.6	31.5 \pm 2.1	30.0 \pm 2.2	28.0 \pm 1.3
	+ 補完 + 早期導入度	50.0 \pm 2.5	41.8 \pm 3.3	41.2 \pm 0.8	32.7 \pm 4.8	42.3 \pm 2.1	60.8 \pm 1.7	43.6 \pm 2.2	45.8 \pm 1.8	32.3 \pm 1.7	31.5 \pm 0.8	35.7 \pm 2.2	25.0 \pm 2.3