

専門語彙グラフに基づく自動用語生成 枠組・手法・評価

藤井俊英¹ Amir Hazem² 影浦峽²

¹ 東京大学大学院学際情報学府 ² 東京大学大学院教育学研究科

fujit1234@e.c.c.u-tokyo.ac.jp amir.hazem@gmail.com kyo@p.u-tokyo.ac.jp

概要

本稿では自動専門用語取得 (Automatic Term Acquisition: ATA) へのアプローチとして専門語彙の構造に着目した自動用語生成 (Automatic Term Generation: ATG) を導入する。与えられた専門語彙を構成する要素 (単語) を頂点とし用語内依存関係と用語完結関係という2種類の辺を持つグラフを構成し、用語生成問題を辺推定問題として定義する。グラフ・ニューラル・ネットワーク (GNN) の手法を用い、実験では極めて優れた結果を得た。実験評価と ATA としての実世界応用との関係も検討する。

1 はじめに

自動専門用語獲得 (ATA) 研究は主にテキスト・コーパスからの用語抽出 (Automatic Term Extraction: ATE) として発達してきたが [1, 2, 3], 先端的手法は実用にはあまり使われていない。原因の一つは専門語彙を考慮していないことにある。例外的に多言語抽出の一環として専門語彙の拡張を扱う研究があるが [4, 5], 語彙の構造を十分に考慮できていない。

本研究では単言語の ATA に、与えられた専門語彙に基づく自動用語生成 (ATG) の観点からアプローチする。用語の構成要素 (単語と呼ぶ) を頂点とし、用語内依存関係および冒頭要素と末尾要素を繋ぐ用語完結関係という二種類の辺を持つ専門語彙グラフを定義し、用語生成を辺推定問題とみなす。推定に GNN を利用し実験評価で優れた結果を得た。

2 提案手法

専門語彙グラフ 専門語彙グラフを、用語を構成する単語を頂点とし、2種類の辺を有するグラフとして定義する。第一は用語内の単語の依存関係を示す辺 (type 1: 依存関係辺)¹⁾, 第二は冒頭と末尾の

単語を結ぶ辺 (type 2: 用語完結関係辺) である (図 1)。単純語の用語完結関係辺はループとなる。

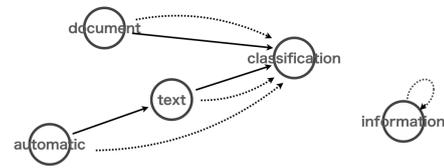


図 1 information, text classification, automatic text classification, document classification の4用語から作られた専門語彙グラフ。実線矢印は type 1, 破線矢印は type 2.

専門語彙グラフの探索 ATG を専門語彙グラフ上の辺予測問題として定式化する。最初に type 1 と type 2 それぞれの辺予測を行い、次にそれらの結果を統合し用語を再構成する。最後に用語を再構成した結果を、用語予測タスクの結果とする。これに対し、様々な辺予測手法を適用することができる。

3 実験データ構築とタスクの設定

3.1 データセット

データセットとして、農学 [6], 計算機科学 [7], 歯学 [8], 物理学 [9] の用語集英語見出しのうち ASPEC-JE Parallel corpus²⁾ [10] の Agriculture, Computer Science, Dentistry, Physics の文書に出現しているものを用いた。データセットの用語数を表 1 に示す。

分野	用語集	∩ ASPEC
農学	15,618	5,921
計算機科学	14,186	3,708
歯学	14,480	6,316
物理学	9,984	5,533

表 1 4分野のデータセットの用語数

1) ただし後に見るように実験では曖昧性を解消せず隣接単語間に辺を張る。

2) <https://jipsti.jst.go.jp/aspec>

実験では、各データセットを訓練、検証、テストに7:1:2で分割する。

3.2 専門語彙グラフの構築

3.1のデータに対して、専門語彙グラフを以下の手続きで構築した。はじめに、データセットに含まれる用語について、空白および記号を区切りとしたトークン化を施し、トークンの異なりをグラフの頂点とする。節2で定義したように、グラフの頂点は専門語彙語彙中の専門用語を構成する単語に対応する。頂点の特徴量として、one-hotベクトルを各頂点に割り当てる場合、事前学習済みモデルを用いて各頂点に単語ベクトルを割り当てる場合の2つの設定を用意した。事前学習済みモデルにはFastText[11]を使用している。

3単語以上の複合専門用語では依存関係の曖昧性が生ずるが、依存関係の解消はせず隣接する単語に辺を張り依存関係辺とした。多重辺と辺の重みを導入することができるが、採用していない。

3.3 辺予測タスク

本研究における辺予測タスクはグラフにおける一般的な辺予測タスクと同一のものである。ここでテストにおける辺の作成の際には、節3.1で構築したテストセットに含まれる用語によって張られるtype 1とtype 2の辺を正例のテスト辺としている。辺予測では、このテスト辺をあらかじめ削除し、それらの存在を適切に予測できるかどうかを評価する。同様に、本来は辺が存在しない箇所について、テスト用の負例辺を用意し、その負例辺について適切に辺の不在を予測できるかを評価する。これは標準的な手続きであるが、ATGの観点からの課題が残る。これについては後述する。

辺予測には多くの応用でSOTAの結果を出しているGNNを採用する[12]。GNNで辺予測を行うためにはエンコーダーによってグラフ構造を学習し、デコーダーを用いて辺の存在確率を算出する必要がある。本研究ではエンコーダーにR-GCNs[13]を採用する。R-GCNsは関係タイプごとに異なる重み行列を学習し、異種グラフの多様な関係パターンを捉える。デコーダーにはComplEx[14]を用いる。ComplExは頂点および関係の埋め込みを複素空間で表現し、三つ組 (s, r, o) ³⁾に対してスコア $\text{Re}(\langle s, r, \bar{o} \rangle)$

3) s は始点となる頂点を、 o は終点となる頂点を、 r は辺の種類を示す。

を計算する。これによって本研究で扱う有向グラフのような非対称な関係も適切にモデル化できる。

各正例辺 (s, r, o) に対し、始点ならびに終点となる頂点をランダムな頂点 s' 、 o' で置換した負例 (s', r, o') あるいは (s, r, o') を生成する。置換頂点は50%を一様分布、50%を次数比例分布からサンプリングし、予測困難な負例を混合する。生成した負例が訓練、検証、テストいずれの正例集合にも存在しないことを保証しリークを防止する。正例と負例の比率は1:1とし、モデルによる負例の暗記を予防するため、5エポックごとに負例を更新する。

推論時には、検証セットでF1値を最大化する閾値を探索し、テストセットの二値分類に適用する。

3.4 専門用語予測タスク

専門用語予測は専門語彙に基づくATGのポイントであり、本研究独自のタスクである。本タスクでは辺予測で予測された辺を用いて構築した用語が、テスト辺を作成するもととなった、用語レベルでのテストセットを再現するかどうかを検討する。

グラフからの用語構築は、まず予測されたtype 2辺のみで有向グラフを構築し、次にtype 2辺の始点頂点と終点頂点間にある、予測された各タイプ1辺の始点から終点への単純パスを探索する。このとき計算量の問題からカットオフ長を設定し、長すぎるパスを除外している。発見された各パスについて、経路頂点列を単語に変換し連結することで専門用語候補を生成する。この手続きにより、学習された単語間の構成関係に基づき新規専門用語候補を体系的に抽出することができる。

4 評価方法

4.1 評価指標

評価指標は二つのタスクに対して以下を用いる。まず辺予測の性能評価には、標準的に用いられているAUC[15]に加えAverage Precision, Accuracyを採用する。専門用語予測の評価にはPrecision (P), Recall (R), F1-score (F1)を用いる。評価指標を複数用いることによって、辺の有無を当てる性能と、用語予測の性能をそれぞれ定量的に評価する。

4.2 ベースライン

本提案手法の用語予測性能を比較評価するために、グラフを用いた用語予測手法とLLMを用いた

手法をベースラインの手法として導入する。

4.2.1 Graph Attention Networks

グラフを用いたベースラインとして Graph Attention Network (GAT) [16] による用語予測を考える。GAT では 1 種類の辺のみを扱うため、提案手法における異種グラフとは対照的な比較となりうる。GAT については type 2 辺を考慮できないため、用語が 3 語以上の単語から構成される場合、復元時に単語の対応づけができず再構成が難しい。このことから GAT によるベースラインは bigram の評価にのみ適用可能である⁴⁾。

4.2.2 LLM

第二のベースラインとして LLM を用いる。LLM は多くのタスクで高い性能を示し、また生成モデルであるため用語生成と相性が良いと考えられる。比較の公平性を保つため、本提案手法が入力として用いる頂点に対応する形で、LLM に学習用の単語リストを与える。LLM による用語予測は二つのプロンプトのもとおこなう。一つはオープンクエスチョンによる用語予測である。オープンクエスチョンでは分野名のみで用語を予測させる場合と、分野名に加えて訓練用のデータセットから選んだ用語例を与える場合の 2 ケースを比較する。もう一つのプロンプトでは用語拡張を指示する。すなわち単語列と対応する用語の例を提示し、例示に近い条件で用語を拡張させる。後者は本提案手法に大枠の設定が近く、(a) unigram の候補リストから unigram を識別する課題と、(b) 単体では用語にならない構成要素をシャッフルしたリストから bigram と trigram を構成する課題を含んでいる。LLM について本研究では llama3.3 と Qwen2[17] の結果を報告する。

5 結果

表 2 に本実験の結果を示した。表から観察されるように、本研究で提案した手法はベースラインとして設定した GAT ならびに LLM の手法の結果をすべての場合で上回った。これは本研究で提案した手法の優位性を明確に示すかたちとなった。

さらには 2 種類の辺を導入した効果として、用語予測においては R をほとんど損なわずに P が大きく向上する点を確認できる。これは 2 つの辺タイプ

4) なお、GAT を用語生成に適用する試みにも新規性があり、本ベースラインの設定も補助的な貢献点としても位置づけられる。

を設定したことに強く起因するものであると考えられる。すなわち辺予測から用語予測へとタスクが移る際に、二重のフィルタリングが働いているのである。

節 3.4 にて示したとおりグラフから用語を構築する際には type 2 辺の有無を確認し、存在したのちに type 2 辺が張られた頂点間の type 1 辺の存在を確認する。この二段階を踏むことによって、仮に type 1 辺が誤った辺を予測していたとしても type 2 辺の不在が誤っていないければ問題はない。逆についてもまた同様である。

つまり、本提案手法はランダムな FP を除去することができるということである。今回の実験では transductive な条件で辺予測をおこなっているが、今後は inductive な条件で（つまり全く未知の語群に対して）ATG の実験をおこなうことも考えられる。その場合であっても、2 つの辺はグラフ内で役割が異なり、訓練される構造的な文脈も別であるため、FP の出現傾向が異なることが予想される。したがって 2 種類の辺を組み合わせることは FP の効果を抑制することにつながる。この点で本提案手法はより堅牢で強力であるといえる。

さて、ここからは対照的な点を観察する。まずは頂点の特徴量に FastText による単語ベクトルを用いたかどうかである。表が示すように、用語予測には FastText による埋め込みは影響を与えなかった。一方で単純な辺予測タスクの結果については影響があったことが表 5 から示唆される。すべてのケースにおいて埋め込みがあった方が辺予測の結果は向上していることが分かる。これは GAT の結果を観察しても同じことがいえる。GAT における bigram の実験結果は実質的に辺予測である。そのことを踏まえると埋め込み有の場合の結果の方が埋め込み無の場合の結果を上回っていることも自然に解釈できる。これは、単語埋め込みが特に専門語彙グラフ構造の探索に有効であることを示す。

次に bigram のみで実験をおこなった場合と、3-gram までの全用語を用いて実験をおこなった場合について比較する。用語レベルの結果については bigram よりも全用語を用いた場合の結果の方が全体的に悪い傾向にある。この結果は n-gram が多様になればなるほどグラフから用語を構築することが難しくなったからだと考えられる。一方で辺予測では all ngram の結果の方が全体的に向上している。辺予測の結果が向上したのは単純に学習データが増えた

Methods	Bigram Evaluation											
	Agriculture			Computer			Dentistry			Physics		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
llama3.3	3.6	6.42	4.61	4.75	12.28	6.85	41.18	1.16	2.26	24.06	11.67	15.72
Qwen2	17.65	0.55	1.07	13.16	8.95	10.65	5.26	0.17	0.33	23.43	14.21	17.69
GAT	55.95	66.01	60.56	57.36	57.36	57.36	52.33	61.61	56.59	59.87	74.29	66.31
GAT (embed)	63.91	90.47	74.91	66.47	83.54	74.03	62.33	89.68	73.54	61.55	92.19	73.81
Ours	100.00	93.76	96.78	99.36	79.28	88.19	100.00	83.14	90.79	99.28	77.64	87.13
Ours (embed)	100.00	85.69	92.29	99.38	82.61	90.22	100.00	79.34	88.48	99.16	83.26	90.52

Methods	All ngram Evaluation											
	Agriculture			Computer			Dentistry			Physics		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
llama3.3	8,50	15,45	10,96	40,10	30,99	34,96	18,77	16,91	17,79	48,50	30,64	37,55
Qwen2	15,19	1,27	2,34	44,86	37,95	41,11	30,79	2,17	4,05	43,81	39,50	41,54
Ours	98.28	82.51	89.71	94.05	79.27	86.03	98.94	82.20	89.80	95.93	77.44	85.70
Ours (embed)	99.24	77.33	86.93	97.76	70.87	82.17	99.48	77.02	86.82	95.32	74.34	83.53

表 2 4 分野における専門語彙予測の結果。上表は bigram の、下表は全 ngram の結果を示している。

ためであると思われる。

LLM の結果と比較しても本提案手法の結果が総じて上回っていることが確認される。LLM の出力はプロンプトの設計や few-shot 例の選択とその順序に大きく影響されることが分かっている [18, 19]。これらは主に分類タスクでの知見であるが ATG にも当てはまる可能性がある [19]。また用語生成向けに調整されていない LLM は、生成が少数に偏るか、逆に多く生成しても分野の特徴が見られにくい傾向が実験で観察されている。また GNN は学習が速く推論のコストも低いため、用語生成の面では全体的に LLM より適していると結論づけられる。

以上よりさまざまな観点から本研究で提案した手法の優位性を示した。2 link 辺にすることによって、用語予測タスクに対する結果が向上し、またいかなる場合でも安定してその結果を得ることができるとわかった。

6 まとめと課題

本稿では、従来の ATA・ATE が十分に扱ってこなかった専門語彙の構造を直接利用して、専門語彙グラフ上の辺予測として用語候補を生成する自動用語生成 (ATG) の新規手法を提案した。その要点は二つに分けられる。一つ目は用語を単語に分解してそれを頂点し、また用語内の単語の結合を表す type 1 辺と、用語単位を表す type 2 辺を併用する専門語彙グラフを提案したことである。これにより ATG を辺予測のタスクとして定式化することができる。二つ

目は R-GCNs を ATG のタスクに適用し、設定したベースラインを大幅に上回る良好な結果を得たことである。これにより本提案手法を実利用につなげる可能性が示された。

今後の課題として技術的な観点からは大きく二つが挙げられる。一つは辺や頂点の重みづけによるグラフの精緻化である。本研究は ATG を辺予測問題として定式化した最初の実験を紹介したため、もっとも単純な構成による専門語彙グラフの構築をおこなっている。二つ目は 4 語以上で構成される用語に対する用語予測である。本研究では trigram までのみを扱っているため、4 語以上の単語の接続関係については調査できていない。

応用との関係では、本実験の評価に対応する ATG のタスク設定を行うために、正解を前提として構築した本実験のテストセットと類似の探索空間を定義できなくてはならない。そのためには、グラフの構造から得られる情報を利用して対象頂点集合を分割すること、分野文書等の外的情報から探索空間を絞り込むことといった対応が必要となる。

また、本研究では GAT による手法と LLM による手法の二つのベースラインを導入したが、これらについても ATG として展開の可能性はある。

謝辞

本研究は JSPS 科研費 24H00736 の助成を受けた。本稿の内容を検討する上で、有益なコメント・示唆をくださった宮田玲氏と矢田竣太郎氏に感謝いたします。

参考文献

- [1] Kyo Kageura and Bin Umino. Methods of automatic term extraction: A review. **Terminology**, Vol. 3, No. 2, pp. 259–289, 1996.
- [2] Hanh Thi Hong Tran, Matej Martinc, Jaya Caporusso, Antoine Doucet, and Senja Pollak. The recent advances in automatic term extraction: A survey. **arXiv preprint arXiv:2301.06767**, 2023.
- [3] Kang Xu, Yifan Feng, Qiandi Li, Zhenjiang Dong, and Jianxiang Wei. Survey on terminology extraction from texts. **Journal of Bid Data**, Vol. 12, No. 29, pp. 1–40, 2025.
- [4] Koichi Sato, Koichi Takeuchi, and Kyo Kageura. Terminology-driven augmentation of bilingual terminologies. In **MT Summit XIV**, pp. 3–10, Nice, 2013.
- [5] Miki Iwai, Koichi Takeuchi, Kyo Kageura, and Kazuya Ishibashi. A method of augmenting bilingual terminology by taking advantage of the conceptual systematicity of terminologies. In **Computerm 2016**, pp. 30–40, Osaka, 2016.
- [6] 文部省, 日本造園学会. 学術用語集 農学編. 日本学術振興会, 東京, 1986.
- [7] 情報処理用語大事典編集委員会. 情報処理用語大事典. オーム社, 東京, 1992.
- [8] 文部省. 学術用語集 歯学編 増訂版. 日本歯科医師会, 東京, 1992.
- [9] 文部省. 学術用語集 物理学編 増訂版. 培風館, 東京, 1990.
- [10] Toshiaki Nakazawa, Manabu Yaguchi, Kiyotaka Uchimoto, Masao Utiyama, Eiichiro Sumita, Sadao Kurohashi, and Hitoshi Isahara. Aspec: Asian scientific paper excerpt corpus. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, **Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)**, Paris, France, may 2016. European Language Resources Association (ELRA).
- [11] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. **arXiv preprint arXiv:1607.04606**, 2016.
- [12] Takoua Ben Smida, Riadh Bouslimi, and Hadhemi Achour. A comprehensive survey on link prediction: From heuristics to graph transformers. **The Journal of Supercomputing**, Vol. 81, No. 1, pp. 1–42, 2025.
- [13] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In **The Semantic Web**, pp. 593–607. Springer International Publishing, 2018.
- [14] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction, 2016.
- [15] Tao Zhou. Discriminating abilities of threshold-free evaluation metrics in link prediction. **Physica A: Statistical Mechanics and its Applications**, Vol. 615, No. 1, p. 128529, 2023.
- [16] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. **6th International Conference on Learning Representations**, 2017.
- [17] An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, Jianxin Yang, Jin Xu, Jingren Zhou, Jinze Bai, Jinzheng He, Junyang Lin, Kai Dang, Keming Lu, Keqin Chen, Kexin Yang, Mei Li, Mingfeng Xue, Na Ni, Pei Zhang, Peng Wang, Ru Peng, Rui Men, Ruize Gao, Runji Lin, Shijie Wang, Shuai Bai, Sinan Tan, Tianhang Zhu, Tianhao Li, Tianyu Liu, Wenbin Ge, Xiaodong Deng, Xiaohuan Zhou, Xingzhang Ren, Xinyu Zhang, Xipin Wei, Xuancheng Ren, Xuejing Liu, Yang Fan, Yang Yao, Yichang Zhang, Yu Wan, Yunfei Chu, Yuyang Liu, Zeyu Cui, Zhenru Zhang, Zhifang Guo, and Zhihao Fan. Qwen2 technical report, 2024.
- [18] Sewon Min, Xinxin Lyu, Ari Holtzman, Mikel Artetxe, Mike Lewis, Hannaneh Hajishirzi, and Luke Zettlemoyer. Rethinking the role of demonstrations: What makes in-context learning work? In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 11048–11064, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [19] Yao Lu, Max Bartolo, Alastair Moore, Sebastian Riedel, and Pontus Stenetorp. Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 8086–8098, Dublin, Ireland, May 2022. Association for Computational Linguistics.

A データ分割

データ分割に関連するデータを表3と4にまとめた。表3は専門語彙レベルでのデータ分割，表4は辺レベルでのデータ分割をそれぞれおこなったときのものである。

ngram	Agriculture				Computer				Dentistry				Physics			
	All	Train	Dev	Test	All	Train	Dev	Test	All	Train	Dev	Test	All	Train	Dev	Test
1	2,896	2,027	290	579	1,429	1,000	143	286	2,816	1,971	282	563	1,457	1,020	146	291
2	2,725	1,908	272	545	1,952	1,366	195	391	3,022	2,115	302	605	3,558	2,491	356	711
3	270	189	27	54	307	215	31	61	427	299	43	85	466	326	47	93
all	5,891	4,124	589	1,178	3,688	2,581	369	738	6,265	4,385	627	1,253	5,481	3,837	549	1,095

表3 専門語彙レベルにおける訓練・検証・テストデータセット。

link type	Agriculture				Computer				Dentistry				Physics			
	All	Train	Dev	Test	All	Train	Dev	Test	All	Train	Dev	Test	All	Train	Dev	Test
type 1	2764	1938	276	550	1974	1380	198	396	3085	2165	305	615	3680	2584	369	727
type 2	2723	1908	272	543	1948	1363	195	390	3021	2114	302	605	3544	2482	355	707
all	5487	3846	548	1093	3922	2743	393	786	6106	4279	607	1220	7224	5066	724	1434
type 1	3181	2248	311	622	2362	1697	233	432	3698	2630	361	707	4289	3060	417	812
type 2	5867	4115	586	1166	3630	2551	360	719	6230	4366	621	1243	5426	3809	542	1075
all	9048	6363	897	1788	5992	4248	593	1151	9928	6996	982	1950	9715	6869	959	1887

表4 辺レベルにおける訓練・検証・テストデータセット。上表はbigramにおけるデータセット，下表は全用語を用いたときのデータセットである。

B 辺予測の結果

表5に辺予測タスクの結果をまとめた。

Methods	RGCN											
	Agriculture			Computer			Dentistry			Physics		
	AUC	AP	Acc	AUC	AP	Acc	AUC	AP	Acc	AUC	AP	Acc
bigram	62.55	58.63	53.89	60.15	56.50	56.04	64.08	59.84	58.36	66.34	63.18	61.58
bigram (embed)	71.24	71.85	60.16	69.35	68.11	59.99	76.83	77.75	68.36	72.88	71.09	62.76
all	64.78	65.23	56.12	67.50	66.12	57.25	62.12	59.81	59.03	64.69	62.35	58.29
all (embed)	84.26	85.81	77.10	78.49	77.99	72.15	86.67	87.13	78.67	76.82	75.56	67.94

表5 4分野における辺予測の結果。