

Title	ホットペーパーの検討
Author(s)	山下, 泰弘; 吉田, 秀紀
Citation	年次学術大会講演要旨集, 36: 677-680
Issue Date	2021-10-30
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/17800
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

ホットペーパーの検討

○山下泰弘, 吉田秀紀 (JST)
yasuhiro.yamashita@jst.go.jp

1. はじめに

毎年膨大な研究論文が発行される現状において、真にウォッチすべき研究トピックをもれなく把握するためには、情報の適切な要約が不可欠となる。JST 経営企画部エビデンス分析室では、これまでは主としてクラリベイト社の **Research Front (RF)** を用いて重要な研究トレンドのフィルタリングを行い、人手で分野ごとのトレンドの分析を行うとともに、新たな分析手法の開発を行ってきた [1-3]。RF は、研究論文をクラスタリングした「リサーチフロント」単位でトピックを表し、リサーチフロント間の関連性の強さについてもデータで示されているため、より確立されたトピックの把握に適している。しかしながら、論文の観測期間が 6 年間にわたるため、直近に注目を集めるようになった研究を常に迅速に把握できるとは限らない。また、「領域 (リサーチフロント)」単位での把握となるため、領域を形成する以前のフェーズにある研究については漏れている可能性が否めない。

上記のような RF を補完するため、当室ではクラリベイト社の **Hot Paper (HP)** を導入した。本稿では、当室の HP を用いた研究動向把握の取り組みについて報告する。

2. HP の特徴とエビデンス分析室の取り組み

HP は、RF と同様にクラリベイト社の **Essential Science Indicators (ESI)** をデータソースとするが、抽出方法が異なる。RF は過去 6 年間に発行された論文の累積被引用数に基づく **トップ 1% 論文** を、共引用によってクラスタリングすることによって構成される。それに対し、HP は、過去 2 年間に発行された論文から直近 2 か月の被引用数が **トップ 0.1%** に入るものを抽出したものである。両者とも ESI に合わせて更新されるため、2 か月ごと (年 6 回) に更新される。

RF の分析において、我々は独自に人手での研究トピックのラベリングを行っているが、HP についても同様にラベリングを行い、トレンドの把握を試みている。例えば、図 1 は、ライフサイエンスについての分析例である。

グループ分類	論文数 世界(日本)	日本著者 比率(%)	責任著者日本 比率(%)	RFの概要	RF数	コア論文数 世界(日本)	平均 出版年	日本著者 比率(%)	責任著者日 本比率(%)
解析技術	イメーシング(21, 脳電位(2), 光遺伝学(8), 脳刺激(7), 単細胞解析(1), 電気生理(1), 脳波(1), その他(5))	46(3)	7	2	解析技術 (イメーシング, 光遺伝学など)	4	7(0) 6(1) 6(1) 5(0)	2018 2016 2016 2017	0 17 17 0
神経栄養因子・ホルモン		9(0)	0	0	神経細胞・グリア細胞	1	6(1)	2017	17
神経細胞・グリア細胞	神経細胞(36), グリア細胞(16), ミクログリア(22) (1細胞解析を含む報告: 6/74)	74(2)	3	1	神経回路	7	9(0) 12(0) 9(0) 7(0) 6(0) 5(0) 5(0)	2018 2016 2016 2015 2015 2015 2014	0 0 0 0 0 0 0
神経回路		66(1)	2	0	高次中枢機能	4	23(0) 10(1) 8(0) 5(0)	2018 2016 2017 2017	0 10 0 0
高次中枢機能		31(0)	0	0	発生・発達・老化・再生	4	6(0) 5(0) 6(0) 5(0)	2018 2015 2017 2015	0 0 0 0
コネクトーム		10(0)	0	0	脳オルガノイド				
睡眠・覚醒		10(1)	10	10	ブレイン・マシン・インターフェース				
発生・発達・老化・再生		22(0)	0	0	神経免疫				
脳オルガノイド		5(0)	0	0					
ブレイン・マシン・インターフェース		4(0)	0	0					
神経免疫		4(0)	0	0					

(a) HP

(b) RF

図 1 HP と RF の分析例

HP は厳選された論文群であるが、2 か月ごとに選出される論文の数は 3,000 報以上となり、その全てを短期間に精査することは困難である。そこで、特に直近に注目された研究を重点的に取り上げるために、前のバージョンまでは出現しなかった新出の HP に注目して分析を行っている。

HP は RF 以上に厳選された論文群と言えるが、引用の観測期間が 2 ヶ月と短いため、比較的引用数が多い分野においては、特定の国や研究グループに偏った引用に影響されるケースもみられる。そ

のような論文が含まれる分野については、下記の二つの手法を併用してフィルタリングを行っている。

(1) 複数バージョンに継続して出現する HP の抽出

短期間での評価では身内の引用が過大に評価されるリスクが大きくなることは否めないため、最近初出し、複数のバージョンに継続して出現する HP に注目する。

(2) 掲載ジャーナルの限定

ローカルなジャーナルに掲載された論文は、その地域の研究コミュニティで認知され、短期的に多くの引用を集める場合もあるが、それらのほとんどは日本が注目する必要性が薄いテーマである。国際的に注目を集める研究に絞るため、ローカルジャーナルの影響が大きい分野（工学等）では、分析対象を Nature Index ジャーナル+ 5 大医学雑誌に限定する。

3. HP と RF の関係

HP と RF は、特に対象論文と引用のタイムスパンの面で抽出方針が異なるが、抽出元を同じくするため、相互の重なりは大きい。例えば、材料科学分野について、RF2019v5（2013年11月～2019年10月の論文を収録）と HP2014v1～HP2019v6 の重なりを見ると、RF2019v5 の新出材料 RF は 148 で、そこに含まれるコアペーパー数は 1,345 報であるが、そのうち 397 報（29.5%）は HP として選出されている。RF 単位でみた場合、148RF 中 111RF（75.0%）が既出 HP を含む。このことから、HP が RF の生成核として機能している可能性が示唆される。

RF2019v5 において、特に HP 数が多い（コアペーパー45 報中 29 報）「ツイスト二層グラフェン」について、RF の生成過程に注目しよう。図 2 に RF2019v5 編纂以前の状況を示す。この HP 中で最も古いものが、二層のグラフェンシートを 1.1° ずらして重ねたときに超電導を発現することを実験的に示した MIT のヘレロとハーバード大のカジラスらの論文であり、2018 年 4 月に発行された Nature に掲載された。HP2019v2 で一時減少したものの、それ以降 HP 数は大幅に増加し、RF2019v5 編纂直前の HP2019v4 では 15 報に及んでいる。RF 全般について同様に HP を中心に RF が生成される傾向が見られるか否かについては、今後精査を進める必要がある。

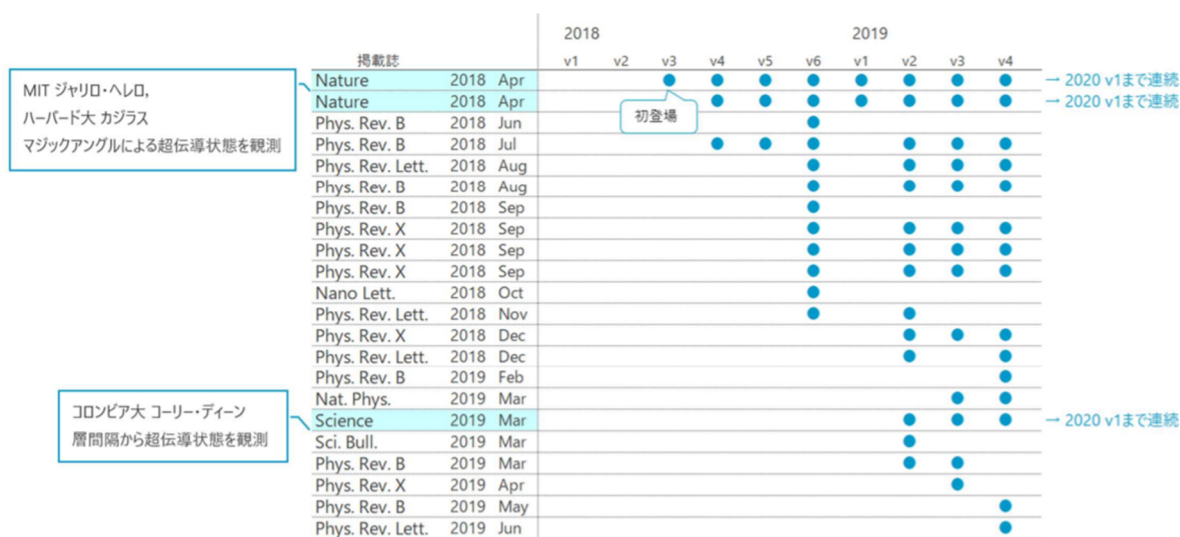


図 2 マジックアングルグラフェンコアペーパーの HP 選出時期

4. 研究サイクルに応じた HP 再定義の試み —COVID-19 研究動向の把握—

2019 年末の COVID-19 パンデミックにより、論文出版のサイクルが大幅に短縮された。短期的なトレンド捕捉には HP が重要な役割を負うことになる。しかしながら、HP も配信 3 か月前の論文が最新であり、論文の多くが投稿から 1 ヶ月未満で出版されるようなサイクルにおいて最新論文をリアルタイムで捕捉できる設計とはなっていない。そこで当室では、COVID-19 をテストベッドとして、より短期的な論文出版サイクルを想定した HP（以下 COV-HP と呼称）の定義の再検討を行った。

4.1 COV-HP 抽出条件の検討

論文は、COVID-19 に関する Web of Science 収録論文、観測期間は論文を 3 か月、引用を 1 か月と

し、観測の時期については、引用の閾値設定、実際の観測実験それぞれのための期間を表1のように設けた。相互の比較・検証を容易にするため、論文・引用の観測期間の終端は、既存のHP(それぞれv4、v5)に揃えた。

表1 COV-HP 抽出元論文の観測条件

HP 抽出用データセット	論文観測期間	引用観測期間	論文数	抽出する HP の名称
1. 引用閾値設定用データ (D-COV-HP1)	2020年6~8月	2020年8月	13,605	COV-HP1
2. 観測実験用データ (D-COV-HP2)	2020年8~10月	2020年10月	17,603	COV-HP2

ここから閾値を設けて COV-HP (COV-HP1、COV-HP2) を抽出することとなるが、抽出元の論文数が小さいため、通常の HP と同じ 0.1% では COV-HP は 20 件弱しか抽出されないこととなる。新たな研究トピックについて、目視チェック可能な範囲で網羅的に捕捉することを目途として、以下のように閾値の検討を行った。

COV-HP は COVID-19 に関する HP をより早期かつ確実に捕捉することを目途としている。そこで、HP2020v1~v4 に出現せず、D-COV-HP1 発行時点でまだ発行されていない HP2020v5 (2018年11月~2020年10月論文を収録) に初出した HP を効率的に抽出できるよう閾値設定を行った。

D-COV-HP1 から被引用数上位より順に COV-HP を抽出した場合、抽出件数によってどの程度の割合の HP2020v5 初出論文を含む COV-HP1 を抽出できるかを示したものが図3である。ここでは、HP2020v4 以前の HP に含まれる論文は既知の情報として母数から除いている。曲線は単調減少ではなく、27位をピーク (COV-HP1 中に 33.3%、9報の HP2020v5 を含む) とし、減少に転ずる形となっている。このような形状になる理由については今後検証が必要である。HP をより多く先取りしたいというニーズを鑑みた場合、9報では少なすぎるため、ある程度 2020v5 初出論文含有率の低下を許容し、ここでは変曲点となる 228位 (12.3%、28報の 2020v5 初出論文を含有) を閾値の候補とした。このラインは、D-COV-HP1 全体でみると、上位 3.2% に相当するので、上位 3% を COV-HP 抽出の閾値として採用した。

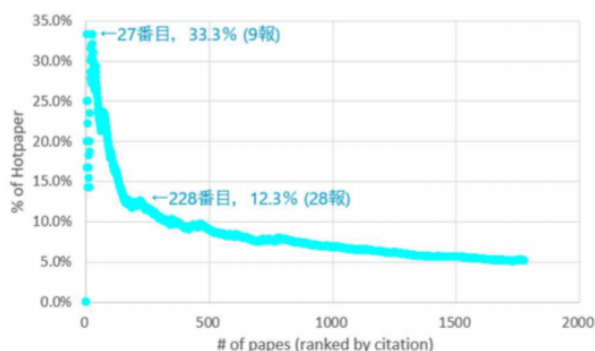


図3 COV-HP1 に含まれる HP2020v5 の割合 (被引用数降順)

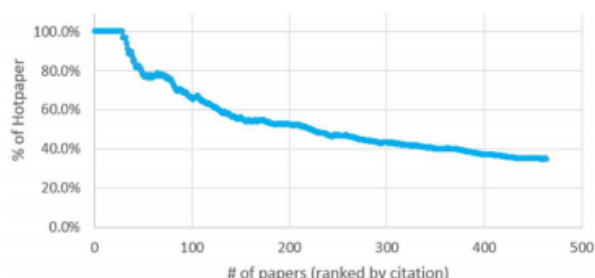


図4 COV-HP2 に含まれる HP2020v5 までの HP の割合

4.2 COV-HP 抽出実験

4.1 節で決定した閾値を D-COV-HP2 に適用し、464 報の COV-HP2 を抽出した。COV-HP2 を図3と同様に被引用数順に並べ、その件数を抽出した場合にデータセットに含まれる v5 までの既存 HP の割合を示したものが図4である。上位 29 報はすべて既存 HP に含まれている。464 報中に 162 報 (34.9%) が HP2020v1~v5 までに既出の HP である。

COV-HP の主たる目的は、COVID-19 に関する重要論文をより早期かつ網羅的に抽出することにある。それは HP を先取りするのみならず、HP では抽出できない重要論文を抽出できることが好ましいことを意味している。以下では COV-HP2 に含まれる、既存 HP では抽出されなかった論文 (引用上位) について検討を行う。

既存 HP に含まれない論文のうち、ネイチャーインデックスジャーナル (NIJ) または 5 大医学雑誌に掲載された論文、それ以外のジャーナルに掲載された論文を、それぞれ表2、表3に示す。NIJ 及び 5 大医学雑誌

論文では、免疫に関する論文が半数を占める他、「アカゲザル」を用いた論文が 4 報あり、ワクチン開発が動物実験段階まで進んでいることを反映している。表中で最多引用を得た論文は、英・独・仏・

伊など 11 か国が導入したロックダウンやソーシャルディスタンスなどの行動制限に関するもので、感染抑止に効果があったと結論付けている。

NIJ 及び 5 大医学雑誌以外についてみると、症例に関する論文が多く、約 1/3 を占める。特に胃・肝臓などの消化器や腎臓における症状に関する論文が目立つ。初期には重症肺炎に関する症例研究が中心であったが、気道以外からも感染する可能性についても多様な観点から研究されていると考えられる。

表 2 既存 HP に含まれない COV-HP2 (被引用上位 15 件、NIJ 及び 5 大医学雑誌)

citation Topic	Journal	Published
40 欧州における新型コロナ感染伝播はロックダウンでどのように変化したか	Nature	2020 AUG 13
34 抗体が低レベルであっても新型コロナに再感染しない可能性	Nature	2020 AUG 20
31 抗がん剤治療は新型コロナ死亡率に影響せず	Lancet	2020 JUN 20
29 新型コロナ患者B細胞からSARS-CoV-2特異的中和抗体を産生	Nature	2020 AUG 6
29 ニューヨーク州の小児における多臓器炎症候群	N Engl J Med	2020 JUL 23
28 新型コロナウイルスに対するT細胞記憶	Nature	2020 AUG 20
27 ヒト中和抗体の標的はSARS-CoV-2受容体結合部位	Nature	2020 AUG 6
26 イタリア・パドヴァのウォーにおけるロックダウンの疫学調査	Nature	2020 AUG 20
22 新型コロナ禍におけるがん患者ケア	Nat. Med.	2020 MAY
22 新型コロナ陰性ドナーの34%がスパイクペプチドに対するT細胞反応性あり	Nature	2020 NOV 12
20 新型コロナウイルスはヒトからイヌへ感染する—香港事例	Nature	2020 OCT 29
19 免疫の誤作動が新型コロナ重症化につながっていた	Nature	2020 AUG 20
18 新型コロナウイルスを植菌されたアカゲザルの呼吸器疾患	Nature	2020 SEP 10
15 回復期の新型コロナ患者から分離したモノクローナル抗体が複数の標的を捉える	Science	2020 AUG 7
15 レニン-アンジオテンシン-アルドステロン系 (RAAS) 阻害薬は他の降圧薬に比べ入院リスクは低い	Lancet	2020 MAY 30
14 新型コロナ感染したアカゲザルは中和抗体ができて再感染しない	Science	2020 AUG 14
14 小動物モデルから分離した強力な新型コロナ中和抗体とその疾患防御機構	Science	2020 AUG 21
14 新型コロナウイルスのスパイクタンパク質をコードするアデノウイルスベクター ChAdOx1 nCoV-19がアカゲザルの肺炎を予防	Nature	2020 OCT 22
13 単一抗体よりも二重抗体薬 (抗体カクテル) の方が治療抵抗性変異ウイルス株の出現しにくい	Science	2020 AUG 21
13 ジョンソン・エンド・ジョンソンのワクチン、アカゲザルに新型コロナウイルス中和抗体を誘導、感染を抑制	Nature	2020 OCT 22

表 3 既存 HP に含まれない COV-HP2 (被引用上位 15 件、NIJ 及び 5 大医学雑誌以外)

citation Topic	Journal	Published
36 中国の症例患者391人の濃厚接触者3410人を対象とした報告	Lancet Infect Dis	2020 AUG
35 喫煙は新型コロナ重症化との危険因子：系統的文献レビューとメタ分析アナリシス	J Infect	2020 AUG
30 新型コロナの軽症患者における消化器症状	Am J Gastroenterol	2020 JUN
27 COVID-19に伴う肝機能異常の臨床的特徴	Clin Gastroenterol	2020 JUN
27 重度のCOVID-19肺炎に関連する臨床および胸部CTの特徴	Invest Radiol	2020 JUN
25 新型コロナウイルスに対する免疫応答と免疫病理の変化メカニズム	Allergy	2020 JUL
23 新型コロナウイルスに祖先であるサルベコウイルス系統は何十年にもわたってコウモリ群を循環	Nat Microbiol	2020 NOV
21 COVID-19肺炎患者の腎病変と早期予後	J Am Soc Nephrol	2020 JUN
21 新型コロナパンデミック下の胃腸手術に関するAGA迅速ガイドライン	Gastroenterology	2020 AUG
21 新型コロナ重症患者は炎症性サイトカインの過剰産生によってCRPが上昇する	J Allergy Clin Immunol	2020 JUL
20 COVID-19における心血管合併症	Am J Emerg Med	2020 JUL
18 COVID-19に対するヒドロキシクロキノンとアジスロマイシン併用の有用性	Int J Infect Dis	2020 AUG
18 電顕像から確認された新型コロナの直接腎感染	Clin J Am Soc Nephrol	2020 AUG
17 新型コロナウイルスの進化	J Med Virol	2020 APR
17 新型コロナ重症化の高リスク国・地域を推定：モデリング研究	Lancet Glob Health	2020 AUG

5. まとめ

本稿では、JST 経営企画部エビデンス分析室での HP 分析の取り組みについて報告した。HP は RF を補完する重要なツールであるが、分析は端緒についたところであり、その特性について未知の部分が多い。HP の検討は継続中の取り組みであるため、口頭発表の際に最新の状況について報告したい。

参考文献

- [1] 田中珠, 藤沢仁子, 迎祐介, 吉田秀紀, 材料科学リサーチフロントの体系化と著者所属国割合の比較, 第 34 回研究・イノベーション学会年次学術大会講演要旨集: 144-147 (2019)
- [2] 山下泰弘, 吉田秀紀, 計量書誌学分析によるホットトピック抽出の試み: ポストグラフィック研究を事例として, 第 34 回研究・イノベーション学会年次学術大会講演要旨集: 148-151 (2019)
- [3] 藤沢仁子, 迎祐介, 山下泰弘, 吉田秀紀, 最新論文クラスタとのキーワード比較によるリサーチフロント進展状況の把握: 材料科学“二次元物質”におけるケーススタディ, 第 35 回研究・イノベーション学会年次学術大会講演要旨集: 717-722 (2020)