Brain Activation-based Filtering: a Scholarly Journal Recommendation Model for Human Neuroimaging Studies

Research Article

Junsol Kim¹

¹ Department of Sociology, Yonsei University, Seoul, South Korea

Corresponding Author:

Junsol Kim¹

Department of Sociology, Yonsei University

50 Yonsei-ro, Seodaemun-gu, Seoul 03722, South Korea

Email address: junsolkim@gmail.com

Brain Activation-based Filtering: an Open-Source Scholarly Journal Recommender based on Activated Brain Regions

Abstract

The development of noninvasive neuroimaging techniques such as functional magnetic resonance imaging was followed by a large volume of human neuroimaging studies of mental processes, mechanisms, and diseases. Due to the high volume of studies and the large number of journals, it is increasingly challenging for neuroscientists to review existing scholarly journals and find the most suitable journal to publish their studies. Therefore, this paper proposes a scholarly journal recommendation model for human neuroimaging studies called brain activation-based filtering (BAF). Based on the collective matrix factorization technique, BAF recommends journals relevant to the activated brain regions that are described in a given neuroimaging study. For instance, if ‘social brain’ regions such as the dorsomedial prefrontal cortex, precuneus, and temporoparietal junction are activated in a study, BAF recommends relevant social neuroscience journals (e.g., *Social Cognitive and Affective Neuroscience*). Five-fold cross-validation shows that BAF predicts journals with a reliable area under the curve score of 0.855. Furthermore, an interactive Google Colab notebook is offered to recommend relevant journals for a novel human neuroimaging study (<https://github.com/JunsolKim/brain-activation-based-filtering>).

1. Introduction

The development of non-invasive neuroimaging techniques such as functional magnetic resonance imaging (fMRI) was followed by a large volume of human neuroimaging studies of mental processes, mechanisms, and diseases (1). Every year, more than 200 journals publish approximately 6,000 human neuroimaging studies on diverse mental processes, mechanisms, and diseases (2). Therefore, it is

increasingly challenging for neuroscientists to review the large body of existing journals and find the most suitable journal to submit their studies to. To reduce the cost of finding a relevant journal, this paper proposes a scholarly journal recommendation model for human neuroimaging studies called brain activation-based filtering (BAF).

The model proposed in this study, brain activation-based filtering (BAF), recommends journals relevant to the activated brain regions that are described in a study. To date, previous studies utilized two strategies to recommend scholarly journals that are based on (1) journals relevant to the authors' previous publication history (i.e., collaborative filtering) (3–10) and (2) journals relevant to the topics appearing in the text (e.g., keywords, abstract, full paper, and bibliography), like the Elsevier Journal Finder (i.e., content-based filtering) (8,11–16). However, there are several limitations to these strategies. The first strategy is limited by the cold-start issue, which means that journals cannot be recommended to researchers without previous publications (17). The second approach relies on the content of the study. Therefore, journals cannot be recommended if the paper is not written yet. In addition, texts are naturally unstructured, so it is challenging to correctly identify the topics in the text and find relevant journals. To address the issues of these strategies, BAF recommends journals relevant to the activated brain regions that are described in the study. This information is available to researchers without a publication history before the paper is actually written (i.e., right after statistical analyses) and is based on a standardized format (e.g., Montreal Neurological Institute (MNI) template) (19).

To recommend journals based on activated brain regions, BAF uses a machine learning technique called collective matrix factorization (CMF), which is a popular technique for document recommendation models (e.g., news article recommendation models) (20–23). From the Neuroquery database (2), BAF collects the associations between studies and activated brain regions (i.e., which voxel is activated in each study) and the associations between studies and journals (i.e., which journal publishes each study). BAF then constructs a study-voxel association matrix and a study-journal association matrix, which are subsequently decomposed simultaneously via CMF. By doing so, studies and journals are embedded in a common low-dimensional

latent topic-space based on activated brain regions. Based on this, journals relevant to each study can be recommended (20–23). For example, if ‘social brain’ regions such as the dorsomedial prefrontal cortex, precuneus, and temporoparietal junction are activated in a study (Van Overwalle, 2009), the study would be related to social neuroscience journals such as *Social Cognitive and Affective Neuroscience* in the latent topic-space. Therefore, social neuroscience journals are recommended for the target study. Such an approach of embedding neuroimaging studies in the latent topic-space based on activated brain regions has long been applied in the previous meta-analysis literature (19,24,25).

This paper consists of four sections. The second section describes the BAF method. The third section presents the experimental results to verify the predictive performance of BAF. The fourth section summarizes the strength and limitation of BAF.

2. Materials and Methods

In this section, the BAF model is described. BAF recommends scholarly journals relevant to brain regions activated in a study. To implement BAF, first, two matrices are constructed based on the Neuroquery database (2): (1) the study-voxel association matrix Y that indicates which voxels are activated in each study and (2) the study-journal association matrix Z , which indicates which journal publishes each study. To evaluate the model, we constructed a training dataset Z_{train} where 80 % of randomly selected elements from Z are included and a test dataset, Z_{test} where the remaining 20 % elements were included.

Subsequently, as shown in Fig. 1, BAF factorizes the study-voxel matrix Y into two latent topic spaces having common dimensions: a study-topic space S and voxel-topic space V . By doing so, the study-topic space S represents the topics of a study and the extent to which a study belongs to these topics (26,27). Second, BAF factorizes the study-journal matrix Z_{train} into a study-topic space S and journal-topic space J . Note that the identical study-topic space S is used following the CMF technique (20). Finally, BAF trains each latent topic space S , V , and J . Following the training process, the model is validated by examining

whether the journal of each study in Z_{test} is accurately predicted by the study-topic space S and the journal-topic space J .

[Fig. 1 goes here]

2.1. Data

BAF uses the Neuroquery database (2) which is a recently published public database of activated brain regions in human neuroimaging studies. The data are collected by selecting papers where peak voxel coordinates are extractable and which were published in neuroimaging journals or with query strings such as ‘fMRI’ (2). Journals of the studies in the Neuroquery database were identified using Entrez Programming Utilities at the National Center for Biotechnology Information (28). After discarding papers in journals with fewer than three papers in the database, there were 461,135 peak voxel coordinates activated in 13,173 studies, which were published in 191 journals. Based on these data, (1) the study-voxel association matrix Y and (2) study-journal association matrix Z were constructed. N_s denotes the number of studies (i.e., 13,173) and N_v denotes the number of possible voxels in the MNI space (i.e., $91 \times 109 \times 91$; MNI-152 $2 \times 2 \times 2$ template).

First, we define the study-voxel matrix, $Y \in \mathbb{R}^{N_s \times N_v}$ representing an activated voxel v in study s as follows:

$$y_{sv} = \begin{cases} 1, & \text{if } v \text{ is activated in } s \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Here, $y_{sv}=1$ indicates that voxel v is activated in study s . We assume that voxel v is activated in study s if v is within 8 mm from the peak voxel coordinates of study s , which is in the range of smoothing kernels that are typically used for fMRI meta-analyses (1,2,29–31). Note the non-negativity of y_{sv} . A value of 0 for y_{sv} does not necessarily mean that voxel v is *not* activated in study s . Some voxels that are not in the

proximity of peak voxel coordinates may have been activated in study s (32). Matrix factorization techniques, including the CMF technique, have long been used to process such non-negative data (33).

Second, let N_j denote the number of journals (i.e., 191). We define the study-journal matrix $Z \in \mathbb{R}^{N_s \times N_j}$ representing that journal j publishes study s as

$$z_{sj} = \begin{cases} 1, & \text{if } s \text{ is published in } j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Here, a value of 1 for z_{sj} indicates that study s is published in journal j . However, a value of 0 for z_{sj} does not necessarily mean that journal j is completely irrelevant to study s (i.e., non-negativity). To evaluate the model, we constructed a training dataset $Z_{train} \in \mathbb{R}^{N_s \times N_j}$ where 80 % of randomly selected elements from Z were included and the test dataset, $Z_{test} \in \mathbb{R}^{N_s \times N_j}$ where the remaining 20 % elements were included. We repeatedly split Z into Z_{train} and Z_{test} for five times so that every element in Z is included in Z_{test} at least once (i.e., five-fold cross-validation).

2.2. Collective Matrix Factorization

To obtain the study-topic space S , the CMF technique was applied to the BAF model. Matrix factorization techniques have long been used in the field of topic modeling (22,23,26,27) and recommendation models (e.g., Netflix Prize competition) (33,34). Based on the success of matrix factorization techniques in both fields, CMF was developed to build a document recommendation model by jointly implementing topic modeling and recommendation models (20–23).

Based on the CMF technique, BAF maps both studies and activated voxels into two latent topic spaces of dimension K , which indicates the number of topics existing in neuroimaging studies and journals (20). We set the value of K to 8 to maximize model performance based on an exhaustive search method. Given the original study-voxel matrix Y , the goal was to obtain the study-topic space S ($S \in \mathbb{R}^{N_s \times K}$) and the voxel-topic space V ($V \in \mathbb{R}^{K \times N_v}$). Each row in S represents the latent topics of each study and each column in V

represents the latent topics of each voxel. S and V were optimized so that the inner product of S and V was close to the observed entries in the study-voxel matrix Y . The journal-topic space J can similarly be identified. To identify the journal-topic space J , CMF learned the journal-topic space J ($J \in \mathbb{R}^{K \times N_j}$) in the way that the inner product of matrices S and J was close to the observed entries in the study-journal matrix Z_{train} . Note that an identical study-topic space S was used (20). By doing so, each column in J represents the latent topics of each journal.

2.3. Optimization

Three latent topic spaces S , V , and J are simultaneously trained to minimize the following cost function (20). Let S_s , V_v , and J_j denote the latent topic vectors for study s , voxel v , and journal j (i.e., s^{th} row of S , v^{th} column of V , and j^{th} column of J). y_{ij} and z_{ik} denote the elements of the study-voxel matrix Y and the study-journal matrix, respectively Z_{train} . The cost function of BAF is defined as

$$\min_{M,N,O_2} \frac{1}{2} [\sum_{i=1}^{N_s} \{ \sum_{j=1}^{N_v} (y_{ij} - S_i^T V_j)^2 + \sum_{k=1}^{N_j} (z_{ik} - S_i^T J_k)^2 \} + \lambda (\|S\|_F^2 + \|(V|J)\|_F^2)] \quad (3)$$

λ is a regularization term to avoid over-fitting. $(V|J)$ is the horizontal concatenation of V and J . Before training S , V , and J , BAF implements random initialization following a Gaussian distribution with mean zero variance. Then, BAF implements the alternate least square technique to train S , V , and J (33). BAF iteratively alternates between re-computing S , V , and J in the way that each re-computing step lowers the value of the cost function. Specifically, in the first re-computing step, S is optimized as follows, while V and J are fixed.

$$S = [\{ (V|J)(V|J)^T + \lambda I \}^{-1} (V|J)(Y|Z_{train})^T]^T \quad (4)$$

In the second and third re-computing steps, V and J are optimized as follows, while M is fixed. By repeating the first, second, and third steps, the cost function is minimized.

$$V = (S^T S + \lambda I)^{-1} S^T Y \quad (5)$$

$$J = (S^T S + \lambda I)^{-1} S^T Z_{train} \quad (6)$$

2.4. Recommendation of Relevant Journals

After optimizing the study-topic space S and journal-topic space J , the inner product of the study-topic vector S_s and journal-topic vector J_j can predict the missing values in the original study-journal matrix Z_{train} . If $\hat{z}_{sj}$, the inner product of S_s and J_j , is high, BAF assumes that journal j is highly relevant to study s . We evaluated the predictive performance of BAF by examining whether $\hat{z}_{sj}$ accurately predicts the journals of the studies in the test dataset Z_{test} (i.e., five-fold cross-validation).

$$\hat{z}_{sj} = S_s^T J_j \quad (7)$$

2.5. Code Accessibility

An interactive Google Colab notebook to recommend journals for a novel neuroimaging study is freely available online at <https://github.com/JunsolKim/brain-activation-based-filtering>. In the notebook, activated brain regions in a novel study can be indicated by either peak voxel coordinates (i.e., coordinates of the most activated brain regions) or statistical maps (e.g., T map from the statistical parametric mapping (SPM) toolbox). The experiment code described in this paper is also available online.

3. Results

3.1. Journal Prediction Performance

To evaluate the performance of BAF, five-fold cross-validation was implemented. As previously mentioned, the study-journal matrix Z , which indicates journals that publish the studies, was divided into a

training (Z_{train}) and test dataset (Z_{test}). Then, the study-topic space S and the journal-topic space J were trained using Y and Z_{train} based on the CMF technique. Subsequently, BAF evaluated whether Z_{test} is accurately predicted by the inner product of S and J (i.e., $\hat{z}_{sj}$). To validate the predictive performance of BAF, we estimated the area under the curve (AUC) score. In addition, we plotted the receiver operating characteristic curve by plotting the false positive rate against the true positive rate based on various thresholds and calculated the AUC score. AUC = 1 means perfect journal prediction by a model and AUC = 0.5 indicates a random selection. The BAF model predicted journals for neuroimaging studies with a reliable AUC score of 0.855.

[Fig. 2 goes here]

3.2. Topic Modeling

In the BAF model, all neuroimaging studies and journals are embedded in the common latent topic-space of eight dimensions based on their associated brain activation patterns. By doing so, relevant journals are recommended for neuroimaging studies. Fig. 3 plots the brain regions associated with each latent topic (i.e., each row of voxel topic space V), which can be downloaded at <https://neurovault.org/collections/8788/>. In Table 1, we interpret latent topics by identifying terms associated with brain regions related to each latent topic based on Neurosynth decoder (<https://neurosynth.org>) (19). Each topic involves the information necessary to identify relevant journals. For instance, Table 1 shows that the first topic is associated with the default mode network, the posterior cingulate cortex (PCC), the medial prefrontal cortex, autobiographical memory, and theory of mind, which are related to social neuroscience. Correspondingly, as shown in Table 2, the journals most associated with the first topic are social neuroscience journals such as *Social Cognitive and Affective Neuroscience* and general human neuroscience journals which publishes a large body of interdisciplinary social neuroscience studies such as *NeuroImage*, *PLOS One*, *Neuropsychologia*, and *Frontiers in Human Neuroscience* (i.e., in the first row of the journal-topic space J).

[Fig. 3 goes here]

[Table 1 goes here]

[Table 2 goes here]

Based on topic mapping, the relations between journals can also be revealed. As shown in the dendrogram in Fig. 4, we characterized the relationships between the largest journals in the Neuroquery database by clustering the journals based on latent topics using hierarchical clustering (single-linkage clustering method). While the figure reveals the similarity between journals based on the traditional division of subjects (e.g., psychiatry), it also reveals novel and interesting similarities between journals. For instance, as expected, psychiatric journals such as *Biological Psychiatry*, *Psychiatry Research*, and *Journal of Affective Disorders* are clustered. However, surprisingly, *Social Cognitive and Affective Neuroscience* is clustered together with such psychiatric journals. In BAF models, the identification of unexpected relations between heterogeneous journals may allow researchers to search for the most suitable journals while avoiding the bias caused by researchers' previous knowledge about the traditional division of subjects. Furthermore, the dendrogram reveals the hierarchical structure of human neuroscience journals from journals related to specific topics (e.g., *Biological Psychiatry*, *Social Cognitive and Affective Neuroscience*) to the journals related to general topics (e.g., *Human Brain Mapping*).

4. Discussion

In this study, we propose a novel scholarly journal recommendation model for neuroimaging studies called BAF, which is a method to recommend journals relevant to activated brain regions in neuroimaging studies. The study shows that BAF achieves reliable journal prediction performance (AUC = 0.855) based on five-fold cross-validation. By embedding neuroimaging studies and journals in the common latent space having eight dimensions, BAF could reliably recommend relevant journals to neuroimaging studies, which range from general neuroscience journals such as *NeuroImage* to specialized journals such as *Social Cognitive and Affective Neuroscience*.

Previous neuroinformatics studies have already shown that the topics of neuroimaging studies, such as mental processes, mechanisms, and diseases, can effectively be identified based on activated brain regions. For instance, Nielsen and colleagues identified the topics of 121 neuroimaging studies based on peak voxel coordinates, using a matrix factorization model, which is well-aligned with general neuroscientific knowledge (24). Similarly, Poldrack et al. identified the topics of 5,809 neuroimaging studies based on peak voxel coordinates using a latent Dirichlet allocation model (25). More recently, Rubin and colleagues showed that the topics of neuroimaging studies can be decoded by statistical maps indicating activated brain regions based on neuroimaging databases such as Neurosynth (19). The BAF model extends the previous studies by showing that activated brain regions can be used not only for topic identification and meta-analyses but also for journal recommendations for human neuroimaging studies. In addition, by using activated brain regions, the BAF model overcomes the disadvantages of previous scholarly journal recommendation models, i.e. the author cold start issue, the incapability to recommend journals before writing papers, and the unstructured nature of text data.

However, this study has several limitations. First, as the first study that uses activated brain regions to recommend journals, this study aims to show that activated brain regions alone can be a significant predictor of relevant journals. Although brain activation alone shows reliable performance, the model would be further improved by incorporating other features used in previous scholarly journal recommendation models such as publication history, title, abstract, bibliography, and the full paper. Second, some recently published studies use deep learning-based techniques such as convolutional neural networks (CNNs) when predicting diseases based on brain regions (e.g., brain tumor classification) and achieve excellent performance (35). Future studies may use CNN instead of CMF in the BAF model to test if CNNs improve predictability (36), although matrix factorization-based techniques such as CMF may be superior to CNNs in terms of model efficiency.

247 **Acknowledgements**

248 This study was supported by the ICONS (Institute of Convergence Science), Yonsei University.

249

250 **Declaration of Competing Interest**

251 The authors report no declarations of interest.

References

1. Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. Large-scale automated synthesis of human functional neuroimaging data. Nat Methods [Internet]. 2011 Aug 26 [cited 2020 Sep 18];8(8):665–70. Available from: <https://www.nature.com/articles/nmeth.1635>
2. Dockès J, Poldrack RA, Primet R, Gözükan H, Yarkoni T, Suchanek F, et al. Neuroquery, comprehensive meta-analysis of human brain mapping. Elife. 2020 Mar 1;9.
3. Chen Z, Xia F, Jiang H, Liu H, Zhang J. AVER: Random walk based academic venue recommendation. In: WWW 2015 Companion - Proceedings of the 24th International Conference on World Wide Web. Association for Computing Machinery, Inc; 2015. p. 579–84.
4. Yu S, Liu J, Yang Z, Chen Z, Jiang H, Tolba A, et al. PAVE: Personalized Academic Venue recommendation Exploiting co-publication networks. J Netw Comput Appl. 2018 Feb 15;104:38–47.
5. Klamma R, Cuong PM, Cao Y. You never walk alone: Recommending academic events based on Social Network Analysis. Lect Notes Inst Comput Sci Soc Telecommun Eng [Internet]. 2009 [cited 2020 Sep 21];4 LNICST(PART 1):657–70. Available from: <http://www.informatik.uni-trier.de/>
6. Luong H, Huynh T, Gauch S, Hoang K. Exploiting social networks for publication venue recommendations. KDIR 2012 - Proc Int Conf Knowl Discov Inf Retr. 2012;239–45.
7. Pham MC, Cao Y, Klamma R, Jarke M. A clustering approach for collaborative filtering recommendation using social network analysis [Internet]. Vol. 17, Journal of Universal Computer Science. 2011 [cited 2020 Sep 21]. Available from: <http://www10.epinions.com/>
8. Alhoori H, Furuta R. Recommendation of scholarly venues based on dynamic user interests. J Informetr. 2017 May 1;11(2):553–63.

9. Lu Z, Xie N, Wilbur WJ. Identifying related journals through log analysis. *Bioinformatics* [Internet]. 2009 Nov 15 [cited 2020 Sep 21];25(22):3038–9. Available from: <http://www.libraryjournal.com/article/CA6606485.html>
10. Yang Z, Davison BD. Venue recommendation: Submitting your paper with style. In: *Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012*. 2012. p. 681–6.
11. Ghosal T, Chakraborty A, Sonam R, Ekbal A, Saha S, Bhattacharyya P. Incorporating full text and bibliographic features to improve scholarly journal recommendation. In: *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. Institute of Electrical and Electronics Engineers Inc.; 2019. p. 374–5.
12. Errami M, Wren JD, Hicks JM, Garner HR. ETBLAST: A web server to identify expert reviewers, appropriate journals and similar publications [Internet]. Vol. 35, *Nucleic Acids Research*. Oxford Academic; 2007 [cited 2020 Sep 21]. p. W12–5. Available from: <http://invention.swmed.edu/etblast/index.shtml>.
13. Schuemie MJ, Kors JA. Jane: Suggesting journals, finding experts. *Bioinformatics* [Internet]. 2008 Mar 1 [cited 2020 Sep 21];24(5):727–8. Available from: <http://biosemantics.org/jane>
14. Kang N, Doornenbal M, Schijvenaars B. Elsevier journal finder: Recommending journals for your paper. In: *RecSys 2015 - Proceedings of the 9th ACM Conference on Recommender Systems* [Internet]. New York, New York, USA: Association for Computing Machinery, Inc; 2015 [cited 2020 Sep 21]. p. 261–4. Available from: <http://dl.acm.org/citation.cfm?doid=2792838.2799663>
15. Küçüktunç O, Saule E, Kaya K, Çatalyürek Ü V. TheAdvisor: A webservice for academic recommendation. In: *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. 2013. p. 433–4.

16. Medvet E, Bartoli A, Piccinin G. Publication Venue Recommendation Based on Paper Abstract. In: Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI. IEEE Computer Society; 2014. p. 1004–10.
17. Boukhris I, Ayachi R. A novel personalized academic venue hybrid recommender. In: CINTI 2014 - 15th IEEE International Symposium on Computational Intelligence and Informatics, Proceedings. Institute of Electrical and Electronics Engineers Inc.; 2014. p. 465–70.
18. Feldman R, Sanger J. The text mining handbook: advanced approaches in analyzing unstructured data. Vol. 44, Cambridge university press. 2007.
19. Rubin TN, Koyejo O, Gorgolewski KJ, Jones MN, Poldrack RA, Yarkoni T. Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. Gershman SJ, editor. PLOS Comput Biol [Internet]. 2017 Oct 23 [cited 2020 Sep 19];13(10):e1005649. Available from: <https://dx.plos.org/10.1371/journal.pcbi.1005649>
20. Saveski M, Mantrach A. Item cold-start recommendations: Learning local collective embeddings. In: RecSys 2014 - Proceedings of the 8th ACM Conference on Recommender Systems [Internet]. New York, New York, USA: Association for Computing Machinery, Inc; 2014 [cited 2020 Sep 19]. p. 89–96. Available from: <http://dl.acm.org/citation.cfm?doid=2645710.2645751>
21. Singh AP, Gordon GJ. Relational learning via collective matrix factorization. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. New York, New York, USA: ACM Press; 2008 [cited 2020 Sep 20]. p. 650–8. Available from: <http://dl.acm.org/citation.cfm?doid=1401890.1401969>
22. Lee H, Yoo J, Choi S. Semi-supervised nonnegative matrix factorization. IEEE Signal Process Lett [Internet]. 2010 Jan [cited 2020 Sep 20];17(1):4–7. Available from: <http://ieeexplore.ieee.org/document/5161276/>
23. Wang Y, Zhang Y, Zhou B, Jia Y. Semi-Supervised Collective Matrix Factorization for Topic Detection and Document Clustering. In: Proceedings - 2017 IEEE 2nd International Conference on

- 323 Data Science in Cyberspace, DSC 2017 [Internet]. Institute of Electrical and Electronics Engineers
324 Inc.; 2017 [cited 2020 Sep 20]. p. 88–97. Available from:
325 <http://ieeexplore.ieee.org/document/8005460/>
- 326 24. Nielsen FÅ, Hansen LK, Balslev D. Mining for associations between text and brain activation in a
327 functional neuroimaging database. Neuroinformatics [Internet]. 2004 [cited 2020 Sep
328 18];2(4):369–79. Available from: <https://link.springer.com/article/10.1385/NI:2:4:369>
- 329 25. Poldrack RA, Mumford JA, Schonberg T, Kalar D, Barman B, Yarkoni T. Discovering Relations
330 Between Mind, Brain, and Mental Disorders Using Topic Mapping. Sporns O, editor. PLoS
331 Comput Biol [Internet]. 2012 Oct 11 [cited 2020 Sep 19];8(10):e1002707. Available from:
332 <https://dx.plos.org/10.1371/journal.pcbi.1002707>
- 333 26. Xu W, Liu X, Gong Y. Document clustering based on non-negative matrix factorization. In:
334 Proceedings of the 26th annual international ACM SIGIR conference on Research and
335 development in informaion retrieval - SIGIR '03 [Internet]. New York, New York, USA:
336 Association for Computing Machinery (ACM); 2003 [cited 2020 Sep 19]. p. 267. Available from:
337 <http://portal.acm.org/citation.cfm?doid=860435.860485>
- 338 27. Shahnaz F, Berry MW, Pauca VP, Plemmons RJ. Document clustering using nonnegative matrix
339 factorization. Inf Process Manag. 2006 Mar 1;42(2):373–86.
- 340 28. NCBI. Entrez Programming Utilities Help [Internet]. National Center for Biotechnology
341 Information (US) Bethesda. Bethesda MD; 2010 [cited 2020 Sep 23]. 1–161 p. Available from:
342 <http://www.ncbi.nlm.nih.gov/books/NBK25501/>
- 343 29. Wager TD, Jonides J, Reading S. Neuroimaging studies of shifting attention: A meta-analysis.
344 Neuroimage. 2004 Aug;22(4):1679–93.
- 345 30. Wager TD, Lindquist M, Kaplan L. Meta-analysis of functional neuroimaging data: Current and
346 future directions. Soc Cogn Affect Neurosci. 2007;2(2):150–8.

31. Turkeltaub PE, Eden GF, Jones KM, Zeffiro TA. Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *Neuroimage*. 2002;16(3 I):765–80.
32. Gorgolewski KJ, Varoquaux G, Rivera G, Schwarz Y, Ghosh SS, Maumet C, et al. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front Neuroinform* [Internet]. 2015 Apr 10 [cited 2020 Sep 20];9(APR):8. Available from: <http://journal.frontiersin.org/article/10.3389/fninf.2015.00008/abstract>
33. Hu Y, Volinsky C, Koren Y. Collaborative filtering for implicit feedback datasets. In: *Proceedings - IEEE International Conference on Data Mining, ICDM*. 2008. p. 263–72.
34. Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *Computer (Long Beach Calif)*. 2009;42(8):30–7.
35. Deepak S, Ameer PM. Brain tumor classification using deep CNN features via transfer learning. *Comput Biol Med*. 2019 Aug 1;111:103345.
36. Kim D, Park C, Oh J, Lee S, Yu H. Convolutional matrix factorization for document context-aware recommendation. In: *RecSys 2016 - Proceedings of the 10th ACM Conference on Recommender Systems* [Internet]. New York, NY, USA: Association for Computing Machinery, Inc; 2016 [cited 2020 Sep 24]. p. 233–40. Available from: <https://dl.acm.org/doi/10.1145/2959100.2959165>

Figure legends

Fig 1. Collective matrix factorization technique.

Fig. 2. ROC (receiver operating characteristic) curve. Journal prediction performance of brain activation-based filtering (BAF) model is estimated by plotting the ROC curve which plots the false positive rate against the true positive rate based on various thresholds.

Fig. 3. Brain regions associated with latent topics. Each panel presents a map of voxels that are associated with each topic (i.e., voxel latent space V).

Fig. 4. A clustering dendrogram showing the relationships between journals based on their cosine similarity in the latent topic-space. Cosine similarity was used as the distance metric for clustering. Hierarchical clustering was performed using a single-linkage clustering method.

Table legends

Table 1. Top twenty keywords associated with latent topics

Table 2. Journals associated with latent topics

Fig 1. Collective matrix factorization technique.

Fig. 2. ROC (receiver operating characteristic) curve. Journal prediction performance of brain activation-based filtering (BAF) model is estimated by plotting the ROC curve which plots the false positive rate against the true positive rate based on various thresholds.

Fig. 3. Brain regions associated with latent topics. Each panel presents a map of voxels that are associated with each topic (i.e., voxel latent space V).

Fig. 4. A clustering dendrogram showing the relationships between journals based on their cosine similarity in the latent topic-space. Cosine similarity was used as the distance metric for clustering. Hierarchical clustering was performed using a single-linkage clustering method.

Table 1. Top twenty keywords associated with latent topics

Topic 1		Topic 2		Topic 3		Topic 4		Topic 5		Topic 6		Topic 7		Topic 8	
Term	Corr	Term	Corr	Term	Corr	Term	Corr	Term	Corr	Term	Corr	Term	Corr	Term	Corr
default	0.389	task	0.384	prefrontal	0.228	auditory	0.609	anterior insula	0.31	word	0.498	motor	0.46	face	0.388
default mode	0.377	working memory	0.341	prefrontal cortex	0.218	auditory cortex	0.577	insula	0.296	reading	0.474	premotor	0.409	visual	0.388
posterior cingulate	0.366	working	0.335	medial	0.202	speech	0.575	thalamus	0.292	words	0.462	movements	0.372	fusiform	0.375
resting	0.365	parietal	0.334	cerebellar	0.202	sounds	0.564	striatum	0.272	demands	0.456	parietal	0.369	occipital	0.366
resting state	0.36	tasks	0.308	vi	0.196	sound	0.555	nucleus	0.27	inferior frontal	0.441	tasks	0.367	viewing	0.365
medial	0.344	premotor	0.288	autobiographical	0.195	listening	0.552	reward	0.269	phonological	0.44	premotor cortex	0.353	faces	0.362
medial prefrontal	0.336	parietal cortex	0.285	episodic	0.193	acoustic	0.532	task	0.248	orthographic	0.438	movement	0.349	object	0.354
dmn	0.325	motor	0.282	retrieval	0.192	superior temporal	0.531	basal	0.245	inferior	0.429	supplementary	0.347	occipito temporal	0.354
autobiographical	0.319	supplementary	0.281	resting state	0.19	heschl	0.527	reactivity	0.243	frontal	0.424	finger	0.344	objects	0.348
network dmn	0.307	supplementary motor	0.268	resting	0.18	primary auditory heschl gyrus	0.526	fearful	0.243	semantic	0.407	execution	0.336	movements	0.343
theory mind	0.258	execution	0.265	default	0.177	planum	0.523	aversive	0.242	language	0.387	intraparietal	0.331	extrastriate	0.338
mind	0.257	load	0.264	lobules	0.169	planum	0.523	amygdala response	0.239	task	0.381	supplementary motor	0.316	fusiform gyrus	0.334
episodic	0.25	anterior insula	0.26	autobiographical memory	0.169	planum temporale	0.515	anticipation	0.239	working memory	0.369	hand	0.315	occipito	0.332
autobiographical memory	0.243	premotor cortex	0.254	vermis	0.168	temporale	0.515	basal ganglia	0.239	working	0.363	intraparietal sulcus	0.305	occipitotemporal	0.33
pcc	0.242	fronto parietal	0.248	default mode	0.167	listened	0.51	neutral	0.238	lexical	0.358	tapping	0.305	hand	0.328
default network	0.241	frontal	0.246	disorder	0.164	pitch	0.497	ganglia	0.238	tasks	0.354	sensorimotor	0.303	fusiform face	0.328
social	0.239	pre sma	0.246	posterior cingulate	0.163	gyrus stg	0.478	striatal	0.238	visual word	0.349	finger movements	0.297	premotor	0.32
midline	0.233	prefrontal	0.24	hippocampus	0.158	spoken	0.47	subcortical	0.236	word form	0.325	task	0.294	ffa	0.319
mental states	0.233	memory wm	0.237	episodic memory	0.158	temporal gyrus	0.457	anterior	0.23	english	0.316	primary motor	0.281	face ffa	0.316
memories	0.231	inferior parietal	0.234	reward	0.156	hearing	0.451	happy	0.228	languages	0.314	motor cortex	0.279	motion	0.316

Table 2. Journals associated with latent topics

Topic	Journals
Topic 1	NeuroImage, PloS one, Neuropsychologia, Social cognitive and affective neuroscience, Frontiers in human neuroscience, Brain research, Psychiatry research, Biological psychiatry, NeuroImage. Clinical, Neurobiology of aging
Topic 2	NeuroImage, PloS one, Social cognitive and affective neuroscience, Pain, Neuroscience and biobehavioral reviews, Biological psychiatry, Neuroscience, Brain research, Developmental cognitive neuroscience, Frontiers in systems neuroscience
Topic 3	Social cognitive and affective neuroscience, NeuroImage. Clinical, PloS one, Psychiatry research, Biological psychiatry, Journal of affective disorders, Neurobiology of aging, Neuropsychologia, Frontiers in behavioral neuroscience, Behavioural brain research
Topic 4	NeuroImage, Brain and language, The Journal of neuroscience, Psychiatry research, Biological psychiatry, Social cognitive and affective neuroscience, Pain, Frontiers in psychology, Journal of cognitive neuroscience, Cortex
Topic 5	NeuroImage, Social cognitive and affective neuroscience, Biological psychiatry, Psychiatry research, Pain, PloS one, Biological psychology, Neuron, Human brain mapping, Behavioural brain research
Topic 6	Neuropsychologia, Brain and language, Cerebral cortex, Journal of cognitive neuroscience, Cortex, Brain research, Frontiers in human neuroscience, Frontiers in psychology, Brain research. Cognitive brain research, Human brain mapping
Topic 7	NeuroImage, Brain and language, Neuropsychologia, Magnetic resonance imaging, Brain research, Frontiers in psychology, Brain research. Cognitive brain research, Human brain mapping, Neuroscience, Brain and cognition
Topic 8	NeuroImage, Frontiers in human neuroscience, PloS one, Social cognitive and affective neuroscience, Neuropsychologia, Brain research, Frontiers in psychology, Brain and language, Cerebral cortex, Cortex

bioRxiv preprint doi: <https://doi.org/10.1101/2020.10.06.331184>; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.