

Trending Topics on Science, a tensor memory hypothesis approach

Felipe Torres^{1*}

*For correspondence:
felipe.torrese@sansano.usm.cl

¹Universidad Técnica Federico Santa María

Abstract The current human knowledge is written. Documenting is the most used manner to preserve memories and to store fantastic stories. Thus, to distinguish the reality from fiction, the scientific writing cites previous works moreover than become form experimental setups. Books and scientific papers are only a small part of the existent literature but are considered more thrust as information sources. It is useful to find more relations and to know where to focus the lookup of a topic using the information about the authors and the keywords on the titles and abstracts. This is possible using relational databases or knowledge graphs, a semantic approach, but with the tensor memory hypothesis, that adds a temporal dimension, is possible to process the information with an episodic memory approach. If well, knowledge graphs are of extended use on question answering and chatbots, they need a previous relational schema generated automatically or by-hand and stored in an easy-to-query file format. I use JATS, a standard format that allows integrating scientific papers in semantic searches but is not spread on all scientific publishers, to extract the markup tags from PDF files, current year journal articles of one particular topic, and then construct the tensors memory with their references to extract relations and predictions with statistical relational learning techniques.

Introduction

Memory is defined as the ability to record information and after recall it. Writing is a human invention that facilitates this capacity in particular for declarative memories that are facts or events that can be expressed with language and it could be of two types: semantic or episodic (*Tresp et al., 2017*).

The memories and knowledge of humanity are stored on written documents, getting more reliability if they include references to previous works from others authors. Scientific articles are the model of well-structured presentation and storage of information, each one of them with an own title, explicit authorship, and references to information related to other documents or within the same document. But, what almost always is relevant for the consideration of reading them, the retrieval action, is their publishing year. Thus, their ordered structure makes possible to use them as a representation of global human episodic knowledge and memories. Also, scientific publication as a human activity could be modeled as a social network. From this kind of networks the expression “trending topic” emerged to call the more frequent term or word used in a specific temporal window and it is understood as the principal theme or main subject that is related to the information described in a piece of content.

In a mathematical and computational framework, semantic memories could be represented as knowledge graphs, where the entities are nodes and the links are relations between them. A relation between entities is then possible to define as a triple (s, p, o) or as a simple sentence subject-predicate-objective. An episodic memory adds a time marker, thus a temporal prepositional phrase is added to the simple sentence: subject-predicate-objective-temporal_preposition

or a quad (s, p, o, t) . This approach is widely used on semantic web technologies under the *Linked Data* methodology (Bizer et al., 2011).

Thus, it is plausible to use complex networks analysis tools to search for the most relevant relations between authors, paper titles or keywords. The scientific publication databases can easily contain millions of authors, papers and their respective citations. A reduced number of relevant documents is expected from a specific topic query, and not thousands of results that search engines like Google Scholar or publisher's own engines could generate for a given chain of words. The field of science of science studies these relations and the former works were realized using knowledge graphs, that are expressed as adjacency matrices. If the temporal dimension and various types of relationships are considered, then its possible to form tensors of fourth order. A matrix X of the network could be bipartite ($X \in \mathbb{R}^{n \times m}$) if there are two types of nodes (authors-articles, authors-words, articles-words) or monopartite ($X \in \mathbb{R}^{n \times n}$); unweighted ($x_{ij} \in \{0, 1\}$) or weighted ($x_{ij} \in \mathbb{R}$), directed or undirected ($X^T = X$) (Zeng et al., 2017).

(Tresp and Ma, 2017) introduced the Tensor Memory Hypothesis, where a knowledge graph is represented by a Tucker decomposition of the tensors. It is based on representational learning, i.e, a discrete entity e is associated with a vector of real numbers \mathbf{a}_e called latent variables. (Tresp and Ma, 2017) also argue that representational learning might also be the basis for perception, planning and decision making. From a physiological point of view, there is evidence that the hippocampus plays a central role in the temporal organization of memories and supports the disambiguation of overlapping episodes (Eichenbaum, 2014a), then in the standard consolidation of memory theory (SCT), the episodic memory is a neocortical representation that arises from hippocampal activity while in the multiple trace theory (MTT) the episodic memory is only represented on the hippocampus and is used to form semantic memories on the neocortex. Also, there is evidence of the existence of "place cells" and "time cells" in the hippocampus and that these support associative networks that represent spatiotemporal relations between the entities of memories (Eichenbaum, 2014b).

Table 1. PCA variance for the number of latent components.

Latent Components	PCA variance (%)
3	2.93
5	4.3
10	7.32
15	10.03
20	12.5
25	14.8
50	24.99
100	41.88
200	63.9

Memory Hypothesis belongs to the use of matrix decomposition to reduce the rank of the matrix. (Alshareef et al., 2018) indexes based on cosine similarity to estimate a score that represents the anticipation of a prospective relationship between authors. They used two subsets of the *IEEE digital library* containing the keywords "database" and "multimedia".

Results

The quantity of latent components is not associated with a specific statistical measure of data. However, to have an approach, table 1 presents the correspondent percentage of variance if the same number of PCA components were employed.

There are some previous works on trending or hot topics in science: (Griffiths and Steyvers, 2004) used Latent Dirichlet Allocation (LDA) to analyze the abstracts from *Proceedings on the National Academy of Sciences (PNAS)* from 1991 to 2001. (Wei et al., 2013) performed a statistical analysis to find if scientists follow hot topics on their investigations, they used published papers from the *American Physical Society (APS) Physical Review* journals beginning in 1976 and ending in 2009. (Kang and Lin, 2018) used non-smooth non-negative matrix factorization (snNMF) to extract the more prominent topics from a dataset of keywords from scientific articles related to "Machine Learning" from 2014 to 2016 in *arXiv.org stat.ML*, the similarity of this work with the Tensor

Table 2. Most probable words for the query with an entity type.

Latent Components	Authors	Entity Type	
		Articles	Words
3	neuromodulation	neuromodulation	neuromodulation
5	stimulus, presented	stimulus, presented	stimulus, technique
10	presented	presented	presented
15	sleep, memory	sleep	sleep
20	stimulus, memory	stimulus, cued	stimulus, cued
25	memory, sws	memory, spatial, sws	memory, sws
50	sleep, stimulus	sleep, stimulus	sleep, stimulus
100	assr, memory	assr, memory	assr, memory
200	wireless, monitoring	sleep, slow	sleep, slow

Table 3. Most probable word with NMF decomposition.

Latent Components	Authors	Entity Type	
		Articles	Words
3	slow, sleep, auditory	stimulation, sleep	sleep, memory
5	spindles, auditory, sleep	sleep	sleep
10	sleep, stimulation	sleep, stimulation	sleep, memory
15	sleep, memory	brain, consolidation	sleep, memory
20	sleep, memory	oscillations, sleep	sleep, memory
25	sleep, stimulation	activity, memory	sleep, memory
50	sleep, memory	oscillations, humans	sleep, memory
100	sleep, role	reactivation, slow-wave	sleep, memory
200	sleep, slow	sleep, brain	sleep, memory

The words with more relations in the complete tensor, before decomposition, are `sleep`, `memory`, `stimulation`, `slow`, `brain`, `consolidation`, `auditory`, `spindles`, `reactivation`, and `activity`. Table 2 is populated using a selection strategy of most frequently word from queries of the type

$$word_i = \operatorname{argmax}_o \{P(s, o, t)\}, \quad (1)$$

where s is each author, paper title or word in the database, o a word, t a year and, i is the index of a entity .

The most probable words, from the same queries, using more latent components are more than using a few latent variables. For example, there are 21 different words from query results using 200 latent components. In the other hand for few latent components, the results of queries are only the words shown in table 2.

Table 3 is populated using the of NMF decomposition in the collapsed on time matrix, adding the weights of each year. The more frequently words are selected from which are maximum for each topic or k -row in the matrix H of the decompositions. The same processing using nsNMF decomposition results with the words `sleep` and `memory` as the most probable for all the cases.

The analysis of relationships between entities needs a metric of distance. Each entity is represented by latent vectors, then one metric selection could be the Euclidean distance but given this particular type of data, content from documents, the usual metric employed is the cosine similarity. However, the use of distances on the original data space demand high computational costs, the use of a reduced space alleviates the computational cost of calculating distances but requires a previous high cost of space transformation. Figure 1 is an example of the Euclidean

distance and cosine similarity that was extracted from the \mathbf{R} tensor of the RESCAL factorization. The difference between the years of sources and the years of only cited papers is most evident with less latent components. Moreover, the similarity is greater, then lesser Euclidean distance, between the entities of the previous years.

Discussion

There are scientific papers meta-data databases or it is possible to extract article's meta-data from a specific journal or publisher. But in practice, it is usual to have few references from a previous search and they are from different journals or publishers, then to extract the meta-data I used the JATS format ¹, a semantic web standard format for scientific papers popularized by National Center for Biotechnology Information (NCBI). A most popular format is the Resource Description Framework (RDF) and various scientific publishers are adopting this one.

The analysis of the statistical features of the tensor without any other process could give information of the most related entities, as the most cited author, most cited article or most used word in each slice of time. However, employing a tensor decomposition technique allows the use of a latent components space, where more information could be extracted given that the relationships are expressed in fewer variables, thus, clustering some properties of data. This work is an example of how from a small sample of documents with a known relationship between them, the topic was already known, some words that are not the most frequent could be extracted and provide a new perspective of the topics covered on the documents. The figure 1 is an example of extracted information that is not easy to visualize in the original space of the data.

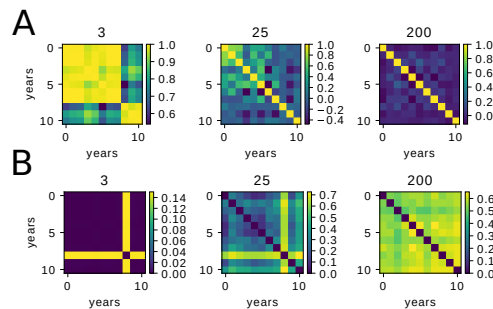


Figure 1. Distance metrics on the latent space. **A.** Cosine similarity between years with 3, 25, and 200 latent components. **B.** Euclidean distance between years with 3, 25, and 200 latent components.

The comparison of different tensor decompositions and the search for the optimum number of latent components is work to be done to take advantage of relational data, that due to semantic web technologies is not restricted only to formal scientific documents and it is available for various type of data. Also, the proposal of (Tresp *et al.*, 2017) of considering the knowledge graphs as semantic and episodic memories allows having a framework that links computational memory with the biological one. Its capacities and defects need to be explored. Curiously, the etymology of “topic” comes from the Greek *topos* or place, that as memory is other of the known hippocampus cognitive functions.

Finally, from the results obtained is evident that `sleep` and `memory` are the most relevant words of the selected papers, these words and `slow` are the few words that are the result from queries too with RESCAL decomposition. The nsNMF decomposition gives for any number of components the same words, then it is more robust to the change in the number of components.

Methods and Materials

Data extraction

The meta-data from 11 articles from different publishers (Table 4) related to “Stimulation during NREM sleep” in PDF files was obtained using the software *CERMINE* (Tkaczyk *et al.*, 2015) and stored in JATS format. After, with a Python script, the own title, authors and abstract were extracted and also the title and authors of references inside the time range 2008-2018. Later, the titles and

¹<https://jats.nlm.nih.gov/>

Table 4. Most probable word with NMF decomposition.

Year	Authors	Title
2016	Batterink et. al	Phase of Spontaneous Slow Oscillations during Sleep Influences Memory-Related Processing of Auditory Cues
2016	Weigenand et. al	Timing matters: open-loop stimulation does not improve overnight consolidation of word pairs in humans
2017	Besedovsky et. al	Auditory closed-loop stimulation of EEG slow oscillations strengthens sleep and signs of its immune-supportive function
2017	Lafon et. al	Low frequency transcranial electrical stimulation does not entrain sleep rhythms measured by human intracranial recordings
2017	Leminen et. al	Enhanced Memory Consolidation Via Automatic Sound Stimulation During Non-REM Sleep
2017	Lustenberger et. al	High-density EEG characterization of brain responses to auditory rhythmic stimuli during wakefulness and NREM sleep
2017	Oyarzun et. al	Targeted Memory Reactivation during Sleep Adaptively Promotes the Strengthening or Weakening of Overlapping Memories
2017	Kinzing et. al	Odor cueing during slow-wave sleep benefits memory independently of low cholinergic tone
2018	Ashton et. al	No effect of targeted memory reactivation during slow-wave sleep on emotional recognition memory
2018	Debellemaniere et. al	Performance of an Ambulatory Dry-EEG Device for Auditory Closed-Loop Stimulation of Sleep Slow Oscillations in the Home Environment
2018	Ezzyat et. al	Closed-loop stimulation of temporal cortex rescues functional networks and improves memory

abstracts were tokenized and semantically tagged, using *nltk* library, to extract the adjectives and nouns that are considered the principal terms of the articles. For de-duplicating authors, all names are formatted to “(Last name) (First name initial.) (Middle name initial.)” For de-duplication of titles and words, all words were transformed to lowercase and special characters were eliminated.

For each year, a zeros square matrix $X_k \in \mathbb{R}^{(n_a+n_i+n_w) \times (n_a+n_i+n_w)}$ was populated with weighted and directed values using the next rules only in the k -year correspondent to relations:

$$\begin{aligned}
 \text{Author}_i \text{ co-wrote with } \text{Author}_j & \quad x_{k_{ij}} + = 2 \\
 \text{Author}_i \text{ cited } \text{Author}_j & \quad x_{k_{ij}} + = 1 \\
 \text{Author}_i \text{ wrote } \text{Article}_j & \quad x_{k_{ij}} + = 2 \\
 \text{Author}_i \text{ cited } \text{Article}_j & \quad x_{k_{ij}} + = 1 \\
 \text{Author}_i \text{ wrote } \text{Word}_j & \quad x_{k_{ij}} + = 1 \\
 \text{Article}_i \text{ cited } \text{Article}_j & \quad x_{k_{ij}} + = 1 \\
 \text{Article}_i \text{ contained } \text{Word}_j & \quad x_{k_{ij}} + = 1 \\
 \text{Word}_i \text{ is in the same document of } \text{Word}_j & \quad x_{k_{ij}} + = 1
 \end{aligned}$$

This approach for expressing the relations simplifies the tensor representation because the dimension correspondent to `predicate` are intrinsic on the weighted values and allows the use of RESCAL factorization. (Ma et al., 2018) explain other tensor decomposition methods that could be used to get the latent components.

Cosine similarity

The cosine similarity is an adequate distance metric for vectors where the magnitude is dependent on the size of the sample, as the frequency of words in a document.

$$s = \frac{\sqrt{x \cdot y}}{\|x\|^2 \|y\|^2} \quad (2)$$

Tensor memory hypothesis

A fourth order tensor could be decomposed as

$$X \approx G \times_1 A_s \times_2 A_p \times_3 A_o \times_4 A_t \quad (3)$$

The probability of the existence of the relationship between the entities of a quad is given by

$$P((s, p, o, t)) = \text{sig}(\theta_{s,p,o,t}), \quad (4)$$

Where

$$\text{sig}(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

$$\theta_{s,p,o,t} = f^e(\mathbf{a}_{e_s}, \mathbf{a}_{e_p}, \mathbf{a}_{e_o}, \mathbf{a}_{e_t}), \quad (6)$$

$$f^e(\mathbf{a}_{e_s}, \mathbf{a}_{e_p}, \mathbf{a}_{e_o}, \mathbf{a}_{e_t}) = \sum_{r_1=1}^{\bar{F}} \sum_{r_2=1}^{\bar{F}} \sum_{r_3=1}^{\bar{F}} \sum_{r_4=1}^{\bar{F}} a_{e_s, r_1} a_{e_p, r_2} a_{e_o, r_3} a_{e_t, r_4} g^e(r_1, r_2, r_3, r_4). \quad (7)$$

The analysis of tensors, as for matrices, is possible to perform using a reduced form obtained by factorization. One popular factorization method of tensors is the Tucker representation, however, there are other matrices and tensor decomposition algorithms. Here, I used RESCAL and the construction of the tensor with weighted values allow to omit the predicate dimension, then the characteristic function becomes

$$f^e(\mathbf{a}_{e_s}, \mathbf{a}_{e_o}, \mathbf{a}_{e_t}) = \sum_{r_1=1}^{\bar{F}} \sum_{r_2=1}^{\bar{F}} \sum_{r_3=1}^{\bar{F}} a_{e_s, r_1} a_{e_o, r_2} a_{e_t, r_3} g^e(r_1, r_2, r_3). \quad (8)$$

RESCAL

This tensor decomposition was proposed by (*Nickel, 2013*). The decomposed tensor needs to have two dimensions of the same size, i.e., $\mathbf{X} \in \mathbb{R}^{n \times n \times m}$ and the results are a matrix $A \in \mathbb{R}^{n \times r}$ and a tensor $\mathbf{R} \in \mathbb{R}^{r \times r \times m}$.

$$\begin{aligned} X &\approx R \times_1 A \times_2 A, \\ X_k &= A r_k A^T. \end{aligned} \quad (9)$$

The algorithm is an alternating least squares (ALS) procedure where the outputs are updated with:

$$A \leftarrow \left(\sum_{k=1}^m X_k A R_k^T + X_k^T A R_k \right) \left(\sum_{k=1}^m R_k A^T A R_k^T + R_k^T A^T A R_k + \lambda_A I \right)^{-1}, \quad (10)$$

$$R_k \leftarrow V (P * U^T X_k U) V^T, \quad (11)$$

Where R_k is a slice of the tensor \mathbf{R} and for optimization a singular value decomposition of matrix A is employed. P is the matrix such that $\text{diag}(\text{vec}(P)) = \hat{S}$, which can be constructed by rearranging the diagonal entries of \hat{S} via the inverse vectorization operator $\text{vec}_r^{-1}(\cdot)$

$$A = U \Sigma V^T, \quad (12)$$

Then, for regularization, the Kronecker product of the diagonal matrix is employed.

$$S = \Sigma \otimes \Sigma \quad (13)$$

$$\hat{S}_{ii} = \frac{S_{ii}}{S_{ii}^2 + \lambda_R} \quad (14)$$

Non-negative Matrix Factorization (NMF)

This matrix factorization method finds two matrices $W \in \mathbb{R}^{n \times r}$ and $H \in \mathbb{R}^{r \times m}$ which multiplication minimizes the Froebenius norm with the original matrix $X \in \mathbb{R}^{n \times m}$.

$$X \approx WH, \quad (15)$$

The updates using the algorithm proposed by (*Lee and Seung, 2001*) are:

$$W \leftarrow W \frac{X H^T}{W H H^T + \delta}, \quad (16)$$

$$H \leftarrow H \frac{W^T X}{W^T W H + \delta}. \quad (17)$$

Non-smooth Non-negative Matrix Factorization (nsNMF)

This decomposition is a modification of NMF proposed by (**Kang and Lin, 2018**).

$$X \approx WSH, \quad (18)$$

Where

$$S = (1 - \theta)I + \frac{\theta}{k}11^T, \quad (19)$$

$$D = \left(\sum_{j=1}^m H_{i,j} \right) I, \quad (20)$$

And using

$$W = W_n D^{-1} S^{-1}, \quad (21)$$

Finally, the matrix decomposition could be expressed as

$$X \approx W D^{-1} S^{-1} S D H. \quad (22)$$

Funding

This work was supported by Beca Doctorado Nacional Conicyt, Folio No 21180640.

References

- Alshareef AM**, Alhamid MF, El Saddik A. Recommending Scientific Collaboration Based on Topical, Authors and Venues Similarities. 2018 IEEE International Conference on Information Reuse and Integration (IRI). 2018; p. 55–61. <https://ieeexplore.ieee.org/document/8424687/>; doi: 10.1109/IRI.2018.00016.
- Bizer C**, Heath T, Berners-Lee T. Linked data: The story so far. In: *Semantic services, interoperability and web applications: emerging concepts* IGI Global; 2011.p. 205–227.
- Eichenbaum H**. Memory on time. *Trends in Cognitive Sciences*. 2014; 17(2):81–88. doi: 10.1016/j.tics.2012.12.007.Memory.
- Eichenbaum H**. Time cells in the hippocampus: A new dimension for mapping memories. *Nature Reviews Neuroscience*. 2014; 15(11):732–744. doi: 10.1038/nrn3827.
- Griffiths TL**, Steyvers M. Finding scientific topics. *Proceedings of the National academy of Sciences*. 2004; 101(suppl 1):5228–5235.
- Kang Y**, Lin Kp. Topic Diffusion Discovery based on Sparseness-constrained Non-negative Matrix Factorization. . 2018; doi: 10.1109/IRI.2018.00021.
- Lee D**, Seung H. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*. 2001; (1):556–562. <http://papers.nips.cc/paper/1861-algorithms-for-non-negative-matrix-factorization>, doi: 10.1109/IJCNN.2008.4634046.
- Ma Y**, Tresp V, Daxberger E. Embedding Models for Episodic Memory. . 2018 jun; <http://arxiv.org/abs/1807.00228>.
- Nickel M**. Tensor Factorization for Relational Learning. . 2013; p. 161. <http://nbn-resolving.de/urn:nbn:de:bvb:19-160568>.
- Tkaczyk D**, Szostek P, Fedoryszak M, Dendek PJ, Bolikowski Ł. CERMINE: automatic extraction of structured metadata from scientific literature. *International Journal on Document Analysis and Recognition (IJ DAR)*. 2015; 18(4):317–335.
- Tresp V**, Ma Y. The Tensor Memory Hypothesis. . 2017; <http://arxiv.org/abs/1708.02918>.
- Tresp V**, Ma Y, Baier S, Yang Y. Embedding learning for declarative memories. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2017; 10249 LNCS:202–216. doi: 10.1007/978-3-319-58068-5_13.
- Wei T**, Li M, Wu C, Yan XY, Fan Y, Di Z, Wu J. Do scientists trace hot topics? *Scientific Reports*. 2013; 3:3–7. doi: 10.1038/srep02207.
- Zeng A**, Shen Z, Zhou J, Wu J, Fan Y, Wang Y, Stanley HE. The science of science: From the perspective of complex systems. *Physics Reports*. 2017; 714-715:1–73. <https://doi.org/10.1016/j.physrep.2017.10.001>, doi: 10.1016/j.physrep.2017.10.001.