Journal of Informetrics 14 (2020) 101047

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Regular article

Understanding hierarchical structural evolution in a scientific discipline: A case study of artificial intelligence

Yue Qian[a], Yu Liu[a,*], Quan Z. Sheng[b]

[a] School of Software, Dalian University of Technology, Dalian, Liaoning 116620, China
[b] Department of Computing, Macquarie University, Sydney, NSW 2109, Australia

ARTICLE INFO

Article history:
Received 15 August 2019
Received in revised form 10 April 2020
Accepted 13 April 2020

Keywords:
Topic evolution
Artificial intelligence
Hierarchical knowledge structure
Nonnegative matrix factorization
Evolutionary patterns
Visual analysis approach

ABSTRACT

Detecting what type of knowledge constitutes a discipline, tracking how the knowledge changes, and understanding why the changes are triggered are the key issues in analyzing scientific development from a macro perspective, which is usually analyzed by the topic of evolution. However, traditional methods assume that the disciplinary structure is flat with only one-layer topics, rather than a tree-like structure with hierarchical topics, which leads to the inability of existing methods to effectively examine the details of the evolution, such as the interactions between different research directions. In this paper, we take artificial intelligence (AI) as a case in which we study its hierarchical structural evolution, more precisely inspecting disciplinary development, by analyzing 65,887 AI-related research papers published during a 10-year period from 2009 to 2018. From a hierarchical topic model that can construct a topic-tree with multi-layer organizations, we design a visual analysis model for the topic-tree to systematically and visually investigate how knowledge transfers along the topic-tree and how the topic-tree changes over time. Moreover, some assistant indicators are employed to help in the exploration of the complicated structural evolution. Then, we discover the latent relationship between the sub-structures within a topic as well as the triggering reason for the knowledge migration. Based on these results, we conclude that different topics have different development patterns and that the recent artificial intelligence revolution stems from the interactions among the different topics.

1. Introduction

With the fast-moving nature of the scientific community, how to understand the topic evolution of scientific literature has become an interesting and important problem, including when new research topics emerge, why they fade out, how they develop, and what factors affect their evolution (Chen, Tsutsui, Ding, & Ma, 2017). Answering these questions can help researchers, policy makers and funding agencies to effectively and efficiently capture the full picture of a scientific discipline, especially interdisciplinary areas, with a complex knowledge structure. Indeed, modelling, tracking and understanding the topic evolution has received much attention in recent years (Song, Heo, & Kim, 2014). Most of the existing work has been confined to single-layer topic evolution (e.g., Chang, Huang, & Lin, 2015; Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011; Li, Qiao, & Wang, 2017; Neff & Corley, 2009; Ronda-Pupo & Guerras-Martin, 2012; Tuomaala, Järvelin, & Vakkari, 2014), which assumes that a discipline's topics have no child-topics attached to them. However, scientific knowledge has distinct

* Corresponding author.
E-mail addresses: y.tsien@gmail.com (Y. Qian), yuliu@dlut.edu.cn (Y. Liu), michael.sheng@mq.edu.au (Q.Z. Sheng).

https://doi.org/10.1016/j.joi.2020.101047
1751-1577/

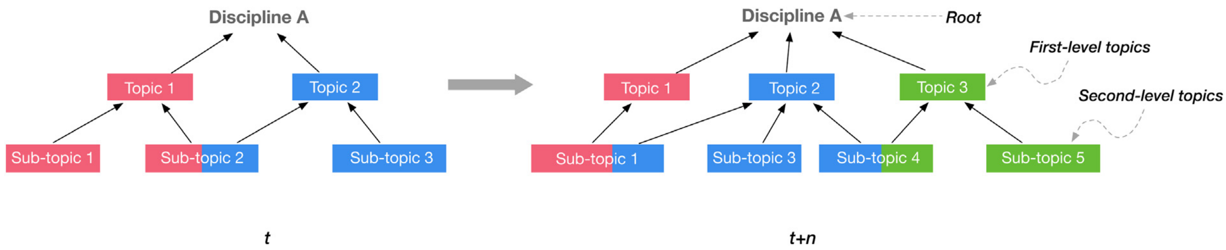

Fig. 1. An example of an evolving topic-tree.

and definite hierarchical structures. A discipline usually includes some professional research fields, and each field consists of research topics. Each topic similarly consists of child-topics and so on. This knowledge hierarchy can be naturally visualized by a multi-branch topic-tree, in which the root is a discipline and each layer covers a set of topics. An evolving topic-tree delivers an interpretable way to discover how the disciplinary topics update, flow and interact, from a global overview to local details.

We give an example in Fig. 1 to show the different structural states of a two-layer topic-tree at period t and period $t+n$. In this topic-tree, each node denotes a topic and a branch represents the affiliation between two topics. In the first layer, each colour represents a topic, and in the second layer, the length of the colour indicates how much weight the lower topic holds for the upper topic marked by the same colour.

From the figure, we can observe many meaningful and unique problems compared with single-layer evolution:

- *Changes in the topic-tree's content*: As seen from Fig. 1, the tree's nodes have changed significantly in the two time slices t and $t+n$. For example, compared with the first time slice, at time $t+n$, the first-level adds *topic-3* and *sub-topic-2* disappears in the second-level. Analyzing the content changes will provide a convenient way to inspect how the topic-tree is formed.
- *Changes in the member weights*: Topics in adjacent layers have apparent owner-member relationships, such as *topic-1* and *sub-topic-1*, which are called the parent-topic and child-topic, respectively. In this paper, we assume that child-topics synthesize a parent-topic by their proportions, which are called member weights. For example, at time $t+n$, *sub-topic-5* has the largest member weight for *topic-3*, followed by *sub-topic-4*. Quantifying the member weight in different periods can directly identify the major vs. minor child-topics for a parent-topic and can also discover the changes in the parent-topic's components.
- *Evolutionary patterns*: In single-layer topic evolution, the life-cycle events of a topic usually evolve by emerging, disappearing, merging, and splitting. However, a child-topic in a topic-tree has another specific evolutionary pattern: migration, which refers to the changes in its parent-topics over time. For example, *sub-topic-1* belongs to only one parent-topic at time t, whereas it has two parent-topics at time $t+n$. This evolutionary pattern can be seen as a type of knowledge flow at a fine granularity, which specifies what external factors trigger and contribute to the development of a research topic.

Compared to the traditional topic of evolution analysis, the hierarchical structure evolution approach provides richer and more detailed information, which is critical to obtaining a multi-level picture to better understand the knowledge evolution of a discipline over time. However, there are two major challenges for addressing the above problems. The first challenge centres around discovering the hierarchical topics. Existing hierarchical topic models (e.g., Ba, Cao, Mao, & Li, 2019; Dou, Yu, Wang, Ma, & Ribarsky, 2013; Griffiths, Jordan, Tenenbaum, & Blei, 2004; Song et al., 2016; Tu, Chen, Lv, Shi, & Chen, 2018; Wang, Zhang, Zhou, Li, & Yang, 2017) can construct a topic-tree, but they can only determine the location of a topic in the tree, and they cannot directly estimate the member weights between parent-child topics. Although similarity estimation, such as Kullback-Leibler divergence, can estimate the parent-child weight based on the results of the aforementioned models, they do not actually reflect the knowledge integration from the child-topic to the parent-topic. The second challenge involves analyzing topic evolutionary patterns. Because there are richer life-cycle events for hierarchical topics, the current analysis models, which assume a flat topic evolution, are far from adequate for studying how a topic-tree changes over time.

To address the first challenge, we propose a novel hierarchical topic model based on non-negative matrix factorization (NMF) (Lee & Seung, 1999) that is specifically designed for the metrics of structural evolution, called hierarchical NMF (HNMF). The approach is based on an intuitive assumption that a parent-topic is a linear combination of its child-topics in the lower layer, along with the member weights. For example, as shown in Fig. 1, *topic-1* (period t) is the weighted sum of *sub-topic-1* and *sub-topic-2*. HNMF is a cascade algorithm that runs layer by layer to generate an entire topic-tree. The topics in the bottom-layer are generated directly by standard NMF, and then, the topics in the other layers are represented as linear combinations of the child-topics in the lower layer using the member weights learned by HNMF.

For the second challenge, we design a visual analysis approach to discover the patterns on the basis of the generated topic-tree and the member weights. This approach visually displays how knowledge flows along the branches of the topic-tree, which leads to a more definitive revelation on how and why the multi-layer topic structures evolve. Moreover, we introduce two auxiliary indicators to assist in the discovery of evolutionary patterns in the topic-tree.

The remainder of this paper is organized as follows. In the next section, a brief review of the related work is given. Section 3 explains our model and the analysis method. Section 4 reports the experimental results and analysis. Finally, the concluding remarks and suggestions for future work are provided in Section 5.

2. Related work

This section briefly reviews two categories of related work, namely, topic evolution models and topic evolution visualization.

2.1. Topic evolution models

In bibliometrics, research on topic evolution has a long and rich history, and the most common methods are based on co-citation analysis, co-word analysis and topic models. Co-citation analysis (Small, 1973) considers that pairs of documents that are cited together commonly have similar content. The more co-citations of the two documents there are, the closer the relationships that they have. Based on this principle, the method discovers the latent structure of co-citation networks in a research field through, e.g., clustering (Wang & Koopman, 2017) and community discovery (Moed, Glänzel, & Schmoch, 2005), to explore the evolution of the documents during different time slices (e.g., Chang et al., 2015; Chen, 1999; Hou, Yang, & Chen, 2018; Li et al., 2017; White & Griffith, 1981). A shortcoming of this method that it is sensitive to the preprocessing of the vectorization and the measurement of similarities. Moreover, the interpretability of the results of this method is not intuitive (Eck & Waltman, 2009). Co-word analysis (Callon, Courtial, Turner, & Bauin, 1983) employs term occurrence statistics to determine the latent semantics in the textual data. Then topic evolution in a discipline can be analyzed by the semantic communities that are distributed along a time-line (e.g., Cobo et al., 2011; Neff & Corley, 2009; Rip & Courtial, 1984; Ronda-Pupo & Guerras-Martin, 2012; Song et al., 2014; Tuomaala et al., 2014).

Topic models represented by the LDA algorithm (Blei, Ng, & Jordan, 2003) simultaneously generate the topics of words and documents, effectively addressing polysemy. Since its inception, LDA has been widely used in the area of scientific topic evolution (e.g., Jeong & Song, 2014). Compared to co-word analysis, the accuracy of the results of LDA topic models is remarkably increased because of the two-way estimation for both documents and words and the introduction of the Dirichlet distribution. Time-evolving topic models based on LDA can be divided into three major categories. The first group treats time information as continuous observable variables, such as the Topic Over Time model (TOT) (Wang & McCallum, 2006), in which the generation of each word is not only based on its topics but also influenced by its timestamp. The second category calculates the per-document topic distribution on the entire text collection by topic models, and then, it uses the timestamps of the documents to re-calculate the topic distribution over discrete time-series to measure the evolution, which is widely used in scientific evolution analysis (e.g., Griffiths & Steyvers, 2004). On the basis of this scheme, Chen et al. (2017) analyze topic evolution in the field of information retrieval. Wang, Zhai, and Roth (2013) propose Citation-LDA, which extends this approach to citation networks and considers scientific articles to be “bag of citations” rather than “bag of words”.

However, the above two categories can only measure the intensity change of the topic distribution; they cannot measure the content change. To tackle this obstacle, a third group of methods that pre-discretize text collection is proposed, in which the full time period is first discretized into time windows, and then, the text in each time window is processed in turn, such as the Dynamic Topic Model (DTM) (Blei & Lafferty, 2006) and Dynamic Mixture Models (DMM) (Wei, Sun, & Wang, 2007). These methods consider the posterior distributions of the model parameters at the current time window to be the conditional distributions of the model parameters at the next time window. In addition, some methods developed from Matrix Factorization have also been studied (e.g., Kalyanam, Mantrach, Saez-Trumper, Vahabi, & Lanckriet, 2015; Vaca, Mantrach, Jaimes, & Saerens, 2014), which introduce an evolution matrix to connect the current time window to the next time window. At the same time, the non-parametric topic evolution model, which can automatically determine the number of topics, has also been developed (Ahmed & Xing, 2010) for detecting topic evolution. Beykikhoshk, Arandjelović, Phung, and Venkatesh (2018) and Wang, Zhang, et al. (2017) extend the algorithm to discover the life cycle of a topic along a time route.

Hierarchical structure is considered to be an effective way to organize large-scale topics because it is easy to understand and makes it easy to search topics. Hierarchical Latent Dirichlet Allocation (hLDA) (Griffiths et al., 2004) is designed for hierarchical topics. However, the higher layer topics generated by the hLDA usually contain more meaningless words, such as stopwords, which weakens the interpretability of the topic-tree. Song et al. (2016) propose a model for the structural evolution of hierarchical topics based on the distance-dependent Chinese Restaurant Process (ddCRP). One problem with this model is similar to that of the TOT model in that it cannot capture the content evolution of topics. Moreover, one subtopic generated by the model belongs to only one parent topic, which means that the branches of its hierarchical topic-tree cannot be crossed.

There are some non-Bayesian frameworks for multi-level evolution analysis. Topic-Rose-Tree is proposed to construct the topic taxonomy by a visual framework (Dou et al., 2013). Jensen, Liu, Yu, and Milojevic (2016) use the unrestricted meta-path to construct and analyze hierarchical structures. Tu et al. (2018) propose a NMF-based method to decompose documents from top to bottom to build the hierarchy. Moreover, the CH-NMF yields meaningful and interpretable clusters by automatically learning the internal structures of a dataset (Kersting, Wahabzada, Thurau, & Bauckhage, 2010). This algorithm adapts a linear combination method similar to our method, but it has not been extended to hierarchical clustering.

Compared to the above-mentioned hierarchical topic evolution models, our method has the following advantages:

- Our method infers a member weight between the child-topic and its parent-topic, which opens a range of insights into the analysis of evolutionary patterns.
- A child-topic in the generated topic-tree can have multiple parent-topics with different member weights, which enables robust exploration of the knowledge structure of the topics.
- The method automatically filters the topics that are not important, which preserves large-scale structures in the topic-tree, while obscuring unimportant details.

2.2. *Topic evolution visualization*

Discovering the topic evolution of a large document collection requires the help of a visualization method, which is a useful tool for understanding complex and high-dimensional data (Iwata, Yamada, & Ueda, 2008). Based on the co-citation method, Moya-Anegón et al. (2007) draw a scientific map of the world in 2002. Chen, Cribbin, Macredie, and Morar (2002) use PCA to discover the thematic areas in the co-citation networks, which are regarded as scientific paradigms, and visually study two disciplines. However, these methods only analyze the documents in a static time window and cannot process the dynamic evolution of stream documents over time.

One popular visualization for dynamic evolution is ThemeRiver (Havre, Hetzler, & Nowell, 2000). This approach is a type of stacked graph in which different topics are represented by tributaries with different colours and are gathered in a large river. The width of each tributary represents the intensity of a topic. This river flows from left to right, which indicates the changes in a topic's strength over time. This method cannot explore the interactions between topics.

TextFlow is an influential method for visually modelling the topics' relationships, including analyzing topic evolution trends, expressing key events, and explaining the relationships of the keywords (Cui et al., 2011). Then users can analyze complex interactions such as the topic's merging and splitting. Based on ThemeRiver and TextFlow, researchers have developed a series of extensions for analyzing the evolutionary patterns of events in various time series (e.g., Cuenca, Sallaberry, Wang, & Poncelet, 2018; Luo, Yang, Krstajic, Ribarsky, & Keim, 2010; Sun et al., 2014; Wongsuphasawat et al., 2011). However, the above methods do not have the ability to address topics with multi-level structures.

For hierarchical topics, Li et al. (2019) propose BarcodeTree, which maps the nodes in a topic-tree to parallel bars with a depth-first traversal order. This method is similar to ThemeRiver in that it can analyze only changes in the intensities of nodes. Cui, Liu, Wu, and Wei (2014) employ a tree cut technology to build a topic-tree without trivial details, and then, a visualization scheme, RoseRiver, was designed to explore the merging and splitting of nodes in the topic-tree over time. This approach is also a type of stacked graphs, in which the nodes are aligned based on their depth information.

In our topic model, there is a weight between the parent-topic and the child-topic, and thus, it is impossible to directly employ the traditional visualization method to inspect this topic-tree. We propose a variant of a stacked graph in which a topic is represented by a block instead of a river. In this way, the member weight can be identified by the colour's width in each block, and the line's colour between the blocks can indicate the strength of the relationship between the topics in adjacent time periods.

3. Methods

3.1. *Problem formulation*

In this section, we formally define key definitions for discovering the structural evolution in a document stream.

Definition 1 *(Document Matrix Stream.).* Given time slices $\{1, 2, \ldots, t, \ldots\}$ indexed by year, we have time-ordered matrices $\{D_1, D_2, \ldots, D_t, \ldots\}$, where D_t is a term-by-document co-occurrence matrix scaled by TF-IDF, in which each document is published in period t.

Definition 2 *(Knowledge Structure.).* The knowledge structure organizes research topics as an L-layer topic-tree with multi-branches where each parent-topic is more abstract and general than its child-topics. The root in the topic-tree is a virtual node that identifies the research field, and thus, the layers are numbered consecutively from the second layer to the bottom (Starting from 1), such as in Fig. 1. In this paper, we refer to the ACM Computing Classification System[1] to define a three-layer knowledge structure for the artificial intelligence (AI) discipline.

Definition 3 *(Member Matrix.).* For an L-layer knowledge structure in period t, the member weight matrix is a factor in a sequence $\{M_{t,L-1}, \ldots, M_{t,1}\}$. $M \in R^{K_{l+1} \times K_l}$ is a matrix, where K_l and K_{l+1} are, respectively, the numbers of topics of the lth layer and $l+1$th layer. An element $M_{t,l}(i,j)$ indicates the member weight between topic j in layer $l+1$ and topic i in layer l.

Definition 4 *(Topic Migration.).* For a non-first-layer topic i, at period t, its parent-topic's set in the topic-tree is denoted as $F_{i,t}$, and at period $t+1$, the set is $F_{i,t+1}$. If the similarity of topic i between the two periods is greater than u, and $F_{i,t} \neq F_{i,t+1}$, and

[1] https://dl.acm.org/ccs/ccs.cfm

$F_{i,t}$ and $F_{i,t+1}$ are in the same layer, we call this circumstance a topic migration. Here, the threshold value u (it is empirically set to 0.2 in this study) guarantees that the content of topic i has not changed greatly in adjacent periods. The second condition indicates that the parent-topics of topic i have changed, and the last condition simplifies the evolution analysis. According to this definition, topic migration can be seen as a knowledge transition between different sub-topic-trees in the same layer, which is discussed in detail in Section 4.

3.2. Hierarchical topic model

3.2.1. Standard NMF

NMF has been widely applied in many fields, such as image processing, computer vision, recommender systems and text mining (Cai, He, Han, & Huang, 2011). Suppose that we have N nonnegative data vectors denoted as the columns in matrix $X \in R^{M\times N}$. In this work, X represents a term-by-document matrix. Given a pre-specified parameter K, which is usually set to be much smaller than M or N and defines the scale of the latent space, the algorithm aims to well approximate the original matrix X as the product of two non-negative matrices $W \in R^{M\times K}$ and $H \in R^{K\times N}$:

$$X \approx WH \tag{1}$$

To learn W and H, a least square cost function is introduced as follows:

$$\begin{aligned} \min \|X - WH\|_F^2 &= \sum_{i=1}^{M}\sum_{j=1}^{N}(X_{ij} - (WH)_{ij})^2 \\ s.t. \quad W \geq 0, &\quad H \geq 0 \end{aligned} \tag{2}$$

An Iterative Multiplicative Update method is presented to optimize the function (Lee & Seung, 1999).

For text mining, a reasonable assumption could be that a document corpus is constituted by K topics. Then, these topics span a K-dimensional semantic space W that is determined by decomposing the term-document matrix. It is natural to consider that these coordinate values in the space are non-negative and can be used to decide the topic distribution, which is the theoretical basis for NMF and its derived algorithms to be applied in topic mining.

3.2.2. Hierarchical NMF

The topics generated by NMF or LDA are independent without hierarchical relationships. In this paper, hierarchical NMF (HNMF) with the cascade architecture is designed to discover multi-layer structures. First, for a time slice t, we directly employ the standard NMF given by Eq. (2) to generate the initial topics, which are denoted by matrix $W_{t,L}$ for the lowest layer of the hierarchy. Then, starting with the assumption that a topic in a non-bottom layer is composed of the next-layer topics, we give the following mathematical formulation:

$$\begin{aligned} W_{t,L-1} &\approx W_{t,L}M_{t,L-1} \\ W_{t,L-2} &\approx W_{t,L-1}M_{t,L-2} \\ &\ldots \\ W_{t,1} &\approx W_{t,2}M_{t,1} \end{aligned} \tag{3}$$

where $W_{t,l} \in R^{M\times K_l}$ is the topic matrix of the terms in the lth layer, K_l is the corresponding number of the topic, and $M_{t,l} \in R^{K_{l+1},K_l}$ is the member matrix. An item $W_{t,l}(i)$, the ith row in $W_{t,l}$, is represented as linear combinations of the ith row in W_{l+1} using weights supplied by the columns of $M_{t,l}$. In this way, new topics are synthesized from the sub-topics by the member matrix. The member weight $M_{t,l}(i,j)$ describes how much knowledge in sub-topic j flows into topic i instead of the similarity between the two topics. This algorithm is a bottom-up process that simulates how the discipline is formed from microscopic to macroscopic. Then, HMNF is presented by casting Eq. (3) into the NMF framework in Eq. (2), according to the following serial objective functions:

$$\begin{aligned} \min \|D_t - &\quad W_{t,L}M_{t,L-1}H_{t,L-1}\|_F^2 \\ \min \|D_t - &\quad W_{t,L-1}M_{t,L-2}H_{t,L-2}\|_F^2 \\ &\ldots \\ \min \|D_t - &\quad W_{t,2}M_{t,1}H_{t,1}\|_F^2 \end{aligned} \tag{4}$$

In each function, $M_{t,l}$ and $H_{t,l}$ are variables to be optimized, and they denote term topics and document topics, respectively. In period t, HNMF sequentially optimizes the above serial functions to yield the hierarchical topics $W_{t,\bullet}$ and the parent-child weight $M_{t,\bullet}$. For the full-time document stream $\{D_1, D_2, \ldots, D_T\}$, we decompose each D_t in turn, to construct the topic-tree for each period.

For a succession of optimizations in Eq. (3), we first optimize one of the sub-problems as follows:

$$\begin{aligned} &\min \|D - WMH\|_F^2 \\ &s.t. \quad W \geq 0, H \geq 0 \end{aligned} \tag{5}$$

Here, we optimize only M and H because W is a constant matrix as the input of the algorithm. This optimization problem is not convex when both M and H are taken as variables. However, when one of the two variables is fixed, Eq. (5) is convex. Therefore, we employ the Iterative Multiplicative Update method, which alternately optimizes the two variables until converging to a local minimum. Based on the Karush-Kuhn-Tucker (KKT), we have

$$L = f(M, H) + Tr(\Lambda M) + Tr(\Theta H) \tag{6}$$

where Λ and Θ are the Lagrange multiplier matrices for the constraint W and $H \geq 0$, respectively, and $Tr(\bullet)$ is a trace function. For the function in Eq. (6), its first partial derivatives are

$$\begin{aligned} \frac{\partial L}{\partial M} = -2W^T D H^T + 2W^T W M H H^T + \Lambda = 0 \\ s.t. \quad \Lambda_{i,j} M_{i,j} = 0 \end{aligned} \tag{7}$$

$$\begin{aligned} \frac{\partial L}{\partial H} = -2M^T W^T D + 2M^T W^T W M H + \Theta = 0 \\ s.t. \quad \Theta_{i,j} M_{i,j} = 0 \end{aligned} \tag{8}$$

Then, the update rules of the two variables are the following:

$$M \longleftarrow M \bullet \frac{W^T D H^T}{W^T W M H H^T} \tag{9}$$

$$H \longleftarrow H \bullet \frac{M^T W^T D}{M^T W^T W M H} \tag{10}$$

Algorithm 1 is the optimization procedure. In the experiments, the stopping threshold of the convergence criterion is set to 10^{-11}, which is the average absolute difference of the reconstruction error of the term-document matrix in each layer.

Algorithm 1. HNMF

Input: Time-ordered term-by-document matrices $\{D_1, D_2, \ldots, D_T\}$; time slices $\{1, 2, \ldots, T\}$; the layer number of topic-tree L; the topic number of each layer $\{K_1, K_2, \ldots, K_L\}$.
Output:$\{W_{t,1}, W_{t,2}, \ldots, W_{t,L}\}$ and $\{M_{t,1}, M_{t,2}, \ldots, M_{t,L}\}$ for time slice t.
for t in $\{1, 2, \ldots, T\}$ **do**
for l in $\{L, L-1, \ldots, 1\}$ **do**
if $l = L$ **then**
if $t > 1$ **then**
Initializing Standard NMF with $(W_{t-1,L}, H_{t-1,L})$;
Running Standard NMF with parameters (D_t, K_L) to get $W_{t,L}$;
else
Initializing Standard NMF randomly;
Running Standard NMF with parameters (D_t, K_L) to get $W_{t,L}$;
end if
else
Running **Algorithm 2** with parameters $(W_{t,l+1}, D_t, K_l)$ to get $M_{t,l}$;
Updating $W_{t,l} \longleftarrow W_{t,l+1} M_{t,l}$;
end if
end for
end for

Algorithm 2. Iterative Multiplicative Update

Input: The term topic matrix in the lower layer $W_{t,l+1}$; term-by-document matrices D_t; the topic number K_l.
Output: $M_{t,l}$.
Initialization: $M_{t,l}$ and $H_{t,l}$ randomly.
while not converge **do**
Fixing $H_{t,l}$, update $M_{t,l}$ by Eq. (9);
Fixing $M_{t,l}$, update $H_{t,l}$ by Eq. (10);
end while

3.3. Construction of the knowledge structure

A topic-tree for the knowledge structure consists of three parts: the nodes of each layer, the branches that link two topics of adjacent layers, and the weight of each branch. For a period t, the nodes of layer l are defined as topic indexes:

$Tree_{t,l} = \{i \mid W_{t,l}(i) > 0.1, 1 \leq l \leq L, 1 \leq i \leq K_l\}$

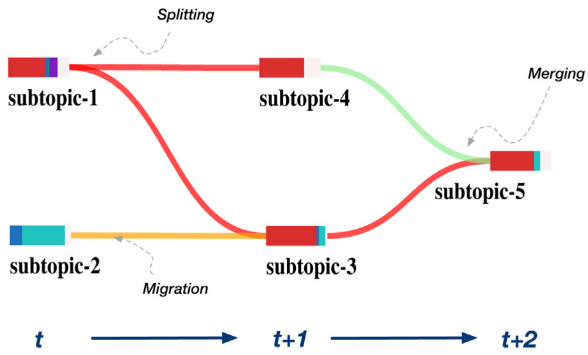

Fig. 2. An example of StructureFlow.

The branches between layer l and layer $l-1$ are given by
$Branch_{t,l,l-1} = \{(i, j) \mid M_{t,l}(i, j) > \varepsilon, 2 \le l \le L\}$
Here, ε is a filter parameter that discards trivial topics. If topic i and topic j are connected by a branch, the they have a parent-child relationship in our analysis.

3.4. Detecting the structure evolution

We propose a visual method called StructureFlow, which can be seen as an enhancement of the traditional topic evolution analysis, to visually and conveniently analyze how and why a topic's structure develops over time. Please note that StructureFlow explores only the structure of one topic, rather than the entire topic-tree, which is too complex for drawing valid conclusions. This method includes two components: the *subtopic extractor* and the *subtopic visualizer*.

3.4.1. Subtopic extractor

For a topic θ in the lth layer $(1 \le l \le L-1)$ in the topic-tree, StructureFlow first extracts the next-layer topics by chronology into $\{Subtopics(\theta_1), Subtopics(\theta_2), \ldots, Subtopics(\theta_T)\}$, where $\{1, 2, \ldots, t, \ldots, T\}$ is the time series, as in the following two steps:

- If for subtopic i, $(i, \theta) \in Branch_{t,l+1,l}$, then we add it into the collection $Subtopics(\theta_t)$;
- Starting from the second period $(t \ge 2)$, for any two topics i and j in the $l+1$th layer, where $i \in Subtopics(\theta_t)$, j is in the previous period $(t-1)$ and $j \notin Subtopics(\theta_{t-1})$, we first calculate their cosine similarity; then, we normalize the similarity scores into [0, 1]. If the similarity score is more than a threshold (in this study, the variable is 0.2 for finding an important foreign influence), then we place subtopic j in $Subtopics(\theta_{t-1})$.

Finally, each collection $Subtopics(\theta_t)$ includes two types of subtopics: the child-topics and other subtopics related to the child-topics in period $t+1$, which are the data foundation observing the knowledge flow between two adjacent stages.

3.4.2. Subtopic visualizer

The visualizer shows the distributions of member weights and the evolutionary relationship of subtopics by timeline, as shown in Fig. 2. In this figure, a block represents a subtopic, and the colour bar indicates its parent-topics, in which the width of a colour corresponds to the member weight between the topic and its parent-topics. For example, this figure demonstrates the red topic's structure, in which *subtopic-1* has three colour bars: red, blue and purple, which denote that it belongs to three parent-topics. There is the largest member weight between *subtopic-1* and the red topic, and *subtopic-1* is also a part of the blue and purple topics.

The links are empirically divided into three grades, which are expressed by green, yellow and red colours, indicating weak (similarity score $\in (0.2, 0.5]$), medium (similarity score $\in (0.5, 0.7]$) and strong (similarity score $\in (0.7, 1]$), respectively. The links with scores less than 0.2 are discarded for the purpose of highlighting the important interactions. For example, the link between *subtopic-1* and *subtopic-4* is red, which indicates that their similarity is strong. The topic blocks with the same time stamp are stacked and aligned vertically. The label under each block is summarized according to the top-50 terms in the topic. To explore more details of the evolution, we select one year as a time slice.

Moreover, due to the constant stream of topics' emergences and disappearances, how to determine the number of topics is one of the key issues in our study. Fortunately, when generating the upper-layer topics, our bottom-up algorithm can automatically assign larger weights to principal topics and smaller weights to trivial topics, which is equivalent to dynamically readjusting the number of topics in each layer. In this study, we slightly magnify the number of topics in the last layer, setting the first layer number to 5, the second to 20, and the third to 100.

From the visualizer, we can conveniently observe and explore the splitting, merging and migration of the subtopics to capture the structure evolution. For example, in Fig. 2, *subtopic-1* splits into *subtopic-4* and *subtopic-3* in period 2. In the same period, *subtopic-2*, which does not belong to the red topic in period 1, migrates to this parent-topic, which indicates that the knowledge in the blue and green topics transfer to the red topic. Then, in period 3, *subtopic-4* and *subtopic-3* merge into *subtopic-5*. In this way, we can intuitively trace the mutual impact of topics from a structural perspective to understand the reason for the evolution. Specifically, due to the migration of *subtopic-2*, the blue and green topics are most likely the reasons that cause the structural changes of the red topic in period 2.

3.5. Assistant indicators

The evolution of topic-trees is too complicated to draw accurate conclusions by relying on only a visualization method. Thus, for a more comprehensive and convenient analysis, we use two metrics as assistant indicators to help the exploration: the *topic importance* and the *richness*.

The topic importance is calculated by

$$importance(topic_i) = \sum_{j}^{N} H_{i,j,l} \tag{11}$$

where H_l is the document-topic distribution in the lth layer, N is the number of papers, and i denotes the index of the topic. This indicator is the weight of topic i in the entire dataset, which reflects the degree of attention to the topic.

The topic richness indicates the breadth of the study in topic i:

$$richness(topic_i) = \frac{s_i}{S} \tag{12}$$

where s_i is the number of its child-topics, and S is the total number of topics in topic i's next layer.

4. Results and discussion

4.1. Dataset

We collected a dataset in the field of artificial intelligence (AI) for this study, which is a good case study due to the extremely active research in the field over recent years. Indeed, AI has grown into one of the most active and influential research fields, in which new knowledge constantly emerges, gathering into new topics, forming new directions, and resulting in changing the structure of knowledge. Thus, AI is a suitable research area to be used in our study. This dataset collects research publications from 35 authoritative AI journals and conferences, after consulting various publication ranking systems such as CCF[2], CORE[3], and Google Scholar[4] (detailed venue list is given in Appendix A). We collected 65,887 papers with titles and abstracts from Scopus[5] based on the venue list from 2009 to 2018. In general, the scientific knowledge and concepts are represented by terminologies rather than words. We extracted the keywords from the titles and abstracts using the TextRank algorithm (Mihalcea & Tarau, 2004) after discarding stopwords and verbs. Then, we removed the terms that appear less than 5 times. Finally, 7901 distinct keywords were available for this experiment.

4.2. Knowledge structure

Figs. 3 and 4illustrate two AI topic-trees with three layers in 2009 and 2018, generated by the method given in Section 3.3 with filter parameter $\varepsilon = 0.15$. Due to space limitations, we list only part of topic-trees for discussion purposes, in which the sub-topic-tree *Artificial Neural Networks* is complete and other subtopics are ignored in the third layer.

In the two trees, each index of nodes is replaced by a manually summarized label based on the top-terms in the topic, to intuitively understand this topic. Starting from the second layer, the circle size of each node represents the member weight between the topic and its parent-topic. Note that the root of the topic-tree is a virtual node, and thus, there is no member weight for the first layer node. The topics linked by branches in the topic-tree have parent-child relationships. For example, in 2018, topics such as *convolutional neural networks*, *recurrent neural networks*, and *network architectures* in the second layer, were the child-topics of *Artificial Neural Networks* in the first layer. In general, the higher the level of the topic is, the broader its content, and vice versa.

Because the branches in the topic-tree require member weights to be greater than a threshold, the number of topics in each layer might be smaller than our setting. For example, in the 2009 topic-tree, there were 61 nodes in the third layer and 20 nodes in the second layer, while for the 2018 topic-tree, the second and the third layer, respectively, had 19 and 68 nodes. Some of generated topics were not in the topic-tree. For example, the topic *search algorithm* (the third layer in

[2] https://www.ccf.org.cn/xspj/rgzn/
[3] http://portal.core.edu.au/jnl-ranks/
[4] https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_artificialintelligence
[5] https://www.scopus.com/

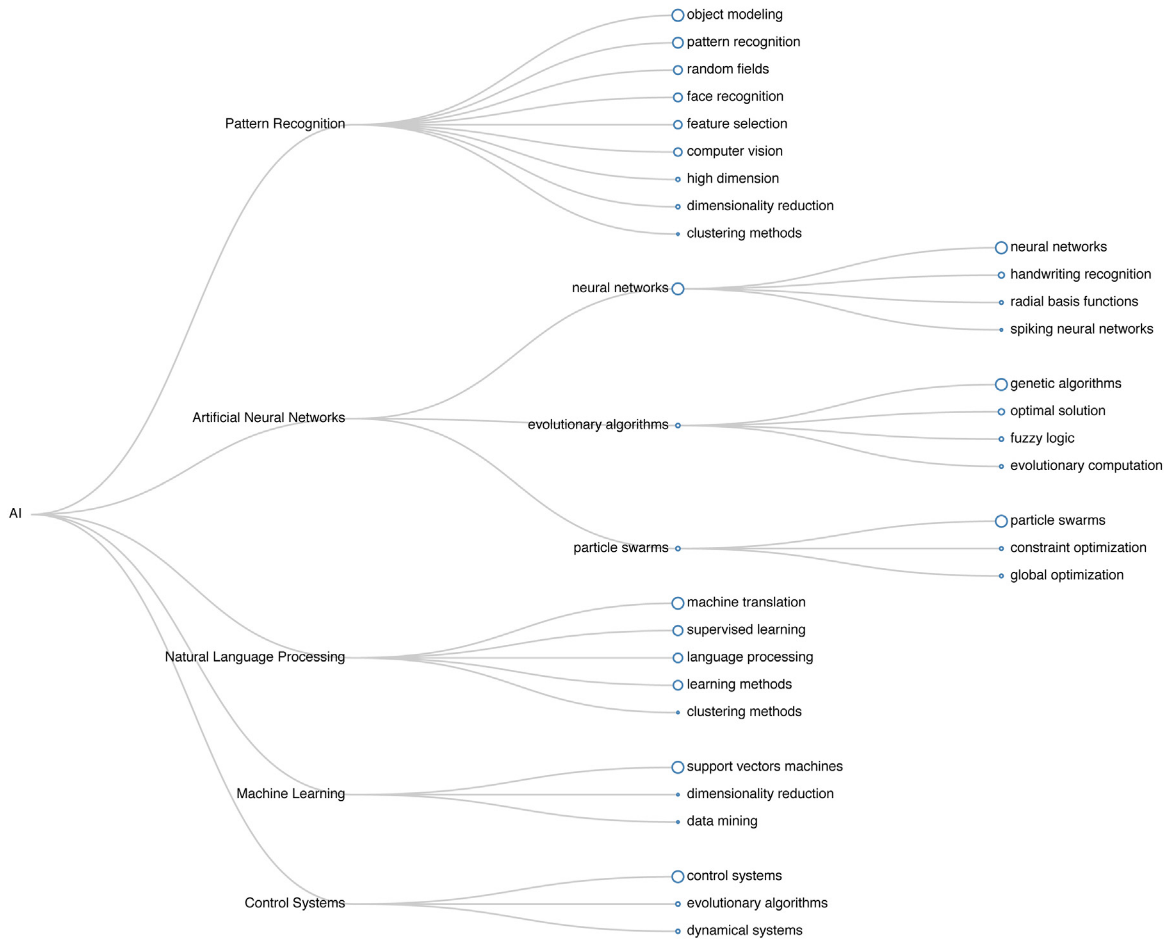

Fig. 3. The partial topic-tree of AI in 2009.

2009), which solves the search problem, had a member weight under the topic *evolutionary algorithms* (the second layer in 2009). However, its member weight was less than filter parameter ε. As a result, this topic was not included in the tree as a node. In fact, *search algorithm* had only 61 articles in 2009, which accounted for approximately 1.0% of the total number of documents in 2009; this finding indicates that it was a small-scale area for AI research during that period. Since we study the knowledge structure from a macro perspective, deleting this topic does not affect the evolutionary analysis. Clearly, due to the introduction of the member weights, our method has the ability to readjust the number of nodes in the topic-tree.

Specifically, the first layer in the two topic-trees includes the following: *Pattern Recognition* (PR), *Artificial Neural Networks* (ANNs), *Natural Language Processing* (NLP), *Machine Learning* (ML) and *Control Systems*. PR is closely related to computer vision and is identified as one research area by our algorithm. This topic focuses on automatically recognizing patterns by using machine learning algorithms, such as object representation, classification, and detection. NLP is a traditional area that is concerned with the interactions between computers and human languages, and it involves computational linguistics, natural language understanding and natural language generation. ML studies how computers simulate and implement the learning behaviours to extract knowledge from experience. ML has grown into the core of AI, almost impacting all fields in AI. ANNs are an important branch of ML that simulates the biological neural networks of human brains to realize AI. In recent years, ANNs have become one of the most attended research subjects due to its powerful performance in computer vision and object recognition. Finally, control systems (i.e., *neural control*) are generally considered to be the basis of robotics using neural networks, which is an important sub-area of AI.

From 2009, the knowledge structure of AI has undergone significant changes. On the one hand, some topics in the second layer disappeared and some topics emerged. For example, *reinforcement learning* has become a new child-topic under ML. *Support vector machines* disappeared in the second layer and merged into the third-level topic *classification/clustering*. On the other hand, the child-topics under a topic and the member weight also changed. For example, in PR, the number of child-topics had decreased from 9 to 3. The weight of *computer vision* (2018) significantly increased compared with 2009.

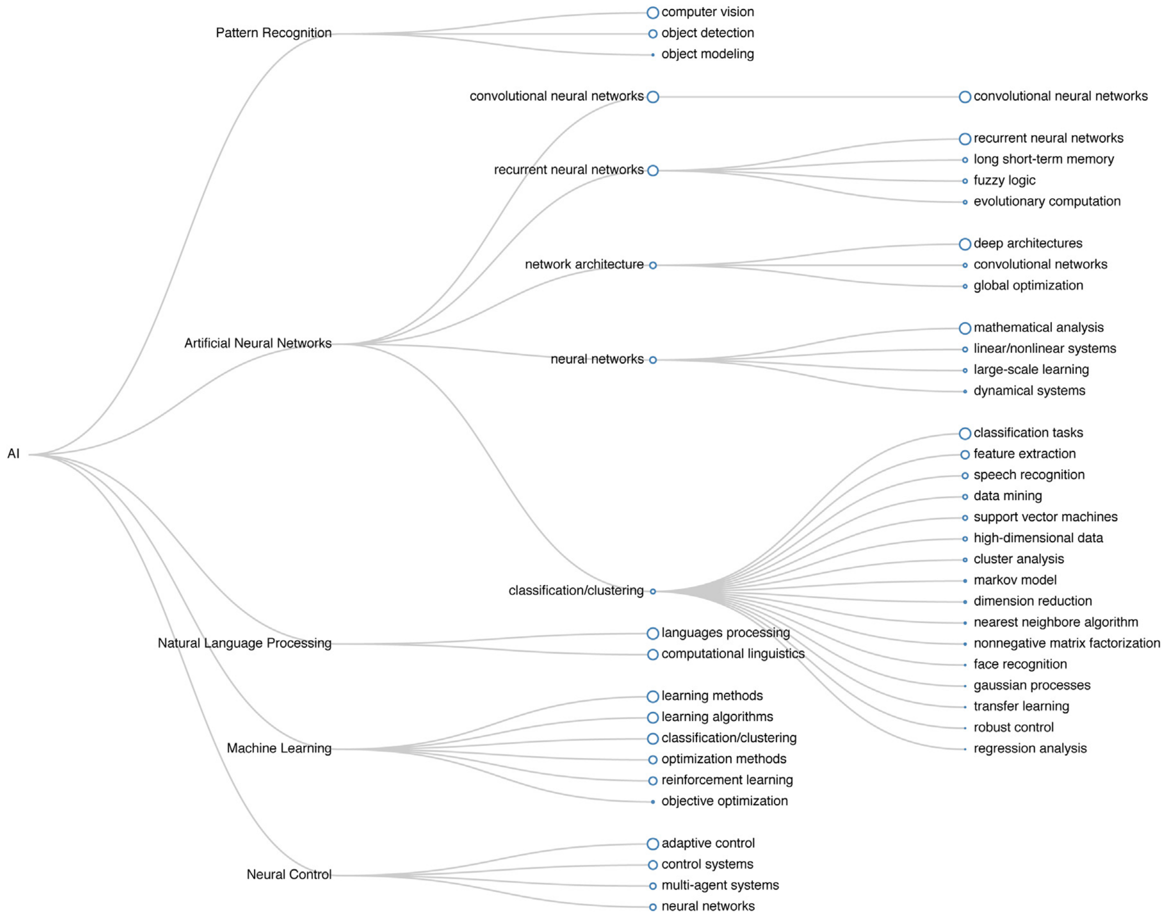

Fig. 4. The partial topic-tree of AI in 2018.

It is clear that from the two topic-trees, it can be seen that the content and composition of knowledge in a discipline are evolving and fluctuating. In the next subsection we will analyze in detail how the knowledge of AI has evolved during the 10-year period from 2009 to 2018.

4.3. Structural evolution

Four first-level topics, ANNs, NLP, PR and ML, were studied by our approach, which are the traditional and major research directions for AI, being the centre of attention in the scientific community. We set the filter parameter $\varepsilon = 0.2$ for the topic-tree to display the graph due to having limited space. With the help of the visualization tool StructureFlow, we not only analyzed the evolutionary patterns of each topic in different periods from a macro perspective but also discovered the evolutionary factors.

4.3.1. Structural evolution of artificial neural networks

Figs. 5 and 6show the structural evolution and auxiliary indicators of ANNs over 10 years. From 2009 to 2011, the evolution of the ANNs was in a “silent” state, having low scores of topic importance and only one main component, *neural networks*, during each period. During those three years, this primary child-topic was dominated by traditional shallow structure, such as RBF neural networks, time delayed neural networks and recurrent neural networks. ANNs had little interaction with other research fields during that period. The only connection took place in 2010 when the child-topic *feature selection* in PR migrated to this field (however, its weight was very small, and thus, it was not listed as a node in this figure), which is mainly used to select a subset of relevant features for learning algorithms.

However, this situation changed significantly after 2012. Since 2012, deep learning has become the mainstream of ANNs, and the scores of topic importance showed an upward trend. In 2012–2014, the multilayered neural networks and activation functions constituted the key research content for *neural networks*. However, in 2011–2012, this child-topic had high similarity in the process of evolution, linked by red lines, which suggests that the basic research activities on neural networks did not change much. In 2012, *neural control* grew into a prominent child-topic, which is widely used in intelligent control,

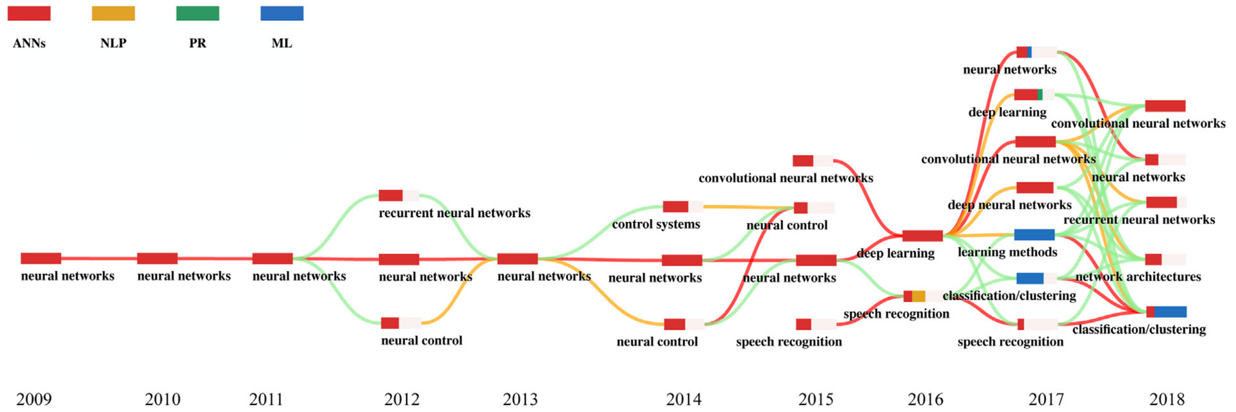

Fig. 5. Structural evolution of artificial neural networks in 2009–2018.

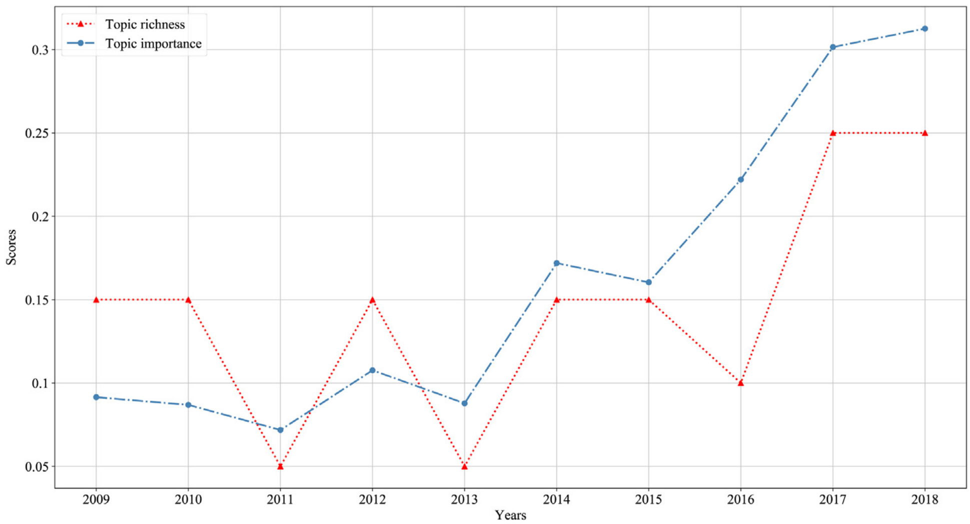

Fig. 6. Topic richness and importance in artificial neural networks in 2009–2018.

fuzzy systems, dynamic system modelling, and so on. At the same time, *recurrent neural networks* split from *neural networks* was increasingly applied in the prediction of sequence data. After a short merger of the three child-topics in 2013, *neural control* re-split in 2014. Since 2015, this research topic had moved away from *neural networks* because it was more prominently applied in robots. In the meantime, *convolutional neural networks*, which are based on deep learning, stood alone as an independent child-topic, in which object detection was an important content. Child-topic *speech recognition* (2015) became the earliest application direction for the application of deep learning.

From 2015 to 2018, ANNs were in a “flourishing” stage. During this period, the topic importance increased significantly, and the topic richness was also at a high level in 2017 and 2018, which showed that the breadth and attention of the research on this topic was increasing. Although there were few research topics in the research field in 2016, it had obvious impacts on NLP and PR (see Figs. 7 and 9). In this year, *deep learning* (mainly including convolutional neural networks and recurrent neural networks) replaced *neural networks* as the most important child-topic. During this time, the child-topic *neural networks* focused on basis theory, *deep learning* focused on feature learning, and *deep neural networks* focused on the study of training methods.

In 2018, *classification/clustering* and *learning methods*, which are the traditional research topics of ML, partly migrated into this field. The algorithms *convolutional neural networks* and *recurrent neural networks* once again became independent child-topics. The child-topic *network architectures* was about how to design neural networks. From 2016 to 2018, the frequent topic migrations between ANNs and the other three areas indicated that ANNs had attracted attention from the wider AI community.

In 2009–2018, the topic importance of ANNs was positive correlated to the time series $r = 0.67$ (Spearman’s rank correlation coefficient), $p < 0.01$, which suggests that ANNs had a research process that was gradually accumulating interest.

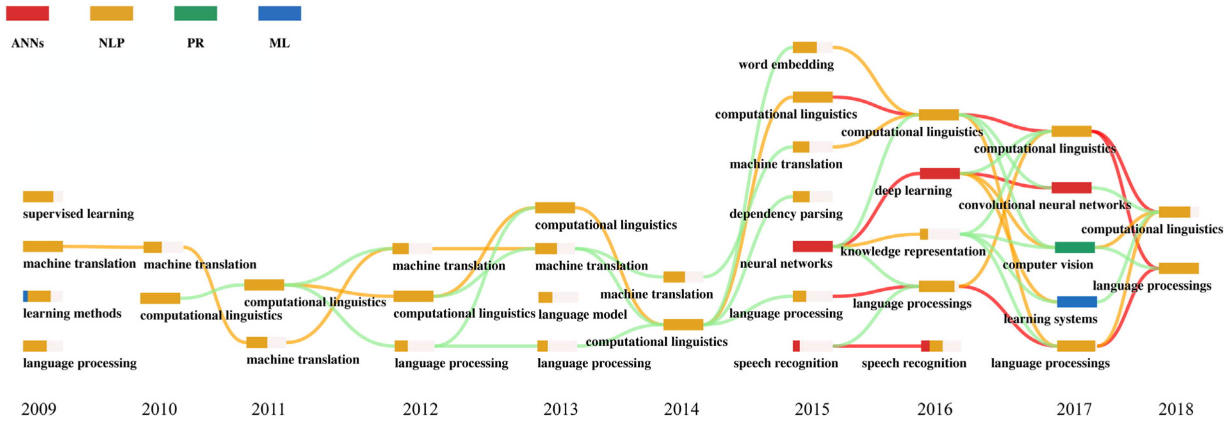

Fig. 7. Structural evolution of natural language processing in 2009–2018.

4.3.2. Structural evolution of natural language processing

As seen in Fig. 7, *computational linguistics* and *machine translation* were basic research at each stage, and *language processing* was also an important child-topic. The child-topics of *computational linguistics* and *language processing* had similar content. It is generally recognized that *computational linguistics* is biased towards linguistics and *language processing* focuses on the applications, but the boundaries between them have been increasingly blurred. Therefore, the two child-topics often split and merged during the ten years.

NLP experienced a major change from 2009 to 2010. In 2009, due to the explosive amount of information on the Internet, "information retrieval" and "information extraction" became urgent tasks. Therefore, in *computational linguistics* and *language processing*, the weights of the terms "information extraction" and "information retrieval" were increasing. These two research content areas were related to ML. Then, the entire research on NLP gradually moved towards ML, such as the child-topic *learning methods* (2009), with a part of the weight of ML. At the same time, *supervised learning*, which is an ML-related child-topic, appeared in NLP as an important method for information retrieval and extraction. After 2009, the weight of "information retrieval and extraction" began to decline, while the weights of other natural language processing questions, such as "dependency parsing" and "sentiment analysis", began to rise. Consequently, in 2010 the child-topic related to machine learning disappeared.

In 2010–2011, "dependency parsing" became one of the main research content areas in *computational linguistics*. Afterward, the relationship between this child-topic in each year became stronger, which suggests that it had established a stable development trend. The child-topics of *machine translation* had great changes in 2014 (the line between two adjacent periods was green). Before 2014, statistical translation methods were dominant in that child-topic, especially IBM Translation Models. However, from this year, the neural translation model grew into the main approach. Here, this child-topic had no obvious relation to ANNs because it also widely applied NLP-specific technologies, such as parse tree. From 2009 to 2014, the attention of the research (topic importance in Fig. 8) was relative stabilization and a breadth of that varied greatly, which indicates that NLP did not form a unified research paradigm.

We also can see that 2015 was a remarkable watershed for the evolution. Since 2015, the topic importance has increased, and the richness decreased, which suggests that more research in NLP focused on fewer questions. In this year, *word embedding* split from *computational linguistics* (2014) and emerged as an important basis for NLP. In 2016, *speed recognition* transferred into NLP with the development of language understanding in this child topic. In the same year, *computational linguistics* and *knowledge representation* were greatly influenced by *neural networks*, which can be considered to be a topic migration from ANNs to NLP. Moreover, *computational linguistics* was combined with *machine translation* (2015) and *word embedding* (2015) because of the deep learning techniques that are shared among them. In 2017–2018, since deep learning had become the public basis method for NLP, some child-topics were merged together. In 2018, there were only two child-topics in the field. In the same year, part of the knowledge about an end-to-end approach based on deep learning in *computer vision* migrated into the two child-topics, which grew up as the mainstream technology for NLP.

4.3.3. Structural evolution of pattern recognition

The child-topics in PR can be roughly divided into three categories: learning methods class (e.g., *generating models*, *support vector machines*), feature engineering class (e.g., *feature selection*, *component analysis*, *sparse representation*) and applications class (e.g., *computer vision*, *object detection*, *image processing*, *pattern recognition*, *face recognition*). Note that these categories do not have strict boundaries and can be merged together, for example, in the period of 2010–2011, *feature selection* and *face recognition* merged into *object detection*.

Before 2016, PR and ML shared a close relationship. During that period, feature engineering methods (e.g., *feature selection*, *dimension reduction*) were also important components for ML. Consequently, in this period, the topic migrations between the two fields often occurred. Due to the migrations, the topic importance and richness of PR and ML (Figs. 10 and 12)

Fig. 8. Topic richness and importance in natural language processing in 2009–2018.

also fluctuated greatly. The instability of this research structure also indicated that PR and ML were in a “turbulent” period without a research paradigm.

From 2013, *sparse representation* became the main method for feature engineering. Then, the other child-topics about feature engineering, such as *feature selection* and *component analysis*, were becoming extinct. In 2014, *humanoid robots* and *motion planning* emerged as child-topics, which marked the expansion of the influence in this area.

In 2016–2018, topics about deep learning migrated into this field and had a continuous impact. Similar to NLP, since 2016, the number of child-topics and the topic richness in PR has decreased significantly, and the child-topics related to feature engineering and learning methods have disappeared, leaving only applied child-topics. The reason is that the deep learning method with multiple hidden layers had excellent feature learning ability, and the learned features can more efficiently and effectively describe datasets, which is superior to the traditional feature selection and learning methods. Since 2017, deep learning has become the universal method in PR.

4.3.4. Structural evolution of machine learning

Similar to PR, the child-topics of ML can be roughly divided into the learning systems class (e.g., *classification/clustering*, *support vector machines*, *learning methods*), auxiliary methods class (e.g., *feature selection*, *optimization*), and applications classes (e.g., *pattern recognition*, *image processing*).

During the period of 2009–2015, the most important child-topic of the learning systems class was *support vector machines*, which is an effective classification algorithm. However, in 2016, this algorithm became a part of learning methods and its weight continued to decline. The child-topic *learning methods* was another important learning systems type, which included supervised, semi-supervised and unsupervised learning methods. These were the basic research directions for machine learning, and they did not change much over the past 10 years. In 2017, *deep learning* moved into learning methods.

In 2018, *reinforcement learning* emerged, which had a greater relationship with *Markov processes* (2017), *learning systems* (2017), and *optimization methods* (2017). In fact, this research appeared in the child-topic *machine learning* in 2009, which described the basis theory of the field, but the weight was not high. In recent years, its weight has increased and grown into a separate child-topic.

The methods of auxiliary classes before 2015 were mainly about features, such as *feature selection* and *sparse representation*. Similar to PR, after 2016, due to the development of deep learning, which can automatically learn features, child-topics related to features continued to decrease. In 2016, *optimization method*, which is mainly about gradient optimization, emerged as the main content in the auxiliary class. In terms of applications, this field’s child-topics were widely used in PR (2009–2015), but after 2016, ANNs squeezed its position in PR; 2016 was a watershed year for the research field. Before that year, the child-topics about feature, support vector machines and learning system were in the main stream. Then, in 2016–2018, the research field changed significantly, and new child-topics about optimization and reinforcement learning took up a large proportion of activity.

General speaking from 2009 to 2015, in the ML topic, which provides solutions for PR and NLP, the importance and richness (Fig. 12) swung obviously. Since 2016, the importance has increased, and the richness has stabilized, which indicates that the “flourishing” of the research was coming. Although ML had transfers of child-topics during 2016–2018, the reason for the transfer was that the content of the migration topic appeared in the target topic (which is discussed in more detail in the next section). Therefore, compared with ANNs, ML did not have much influence during this period.

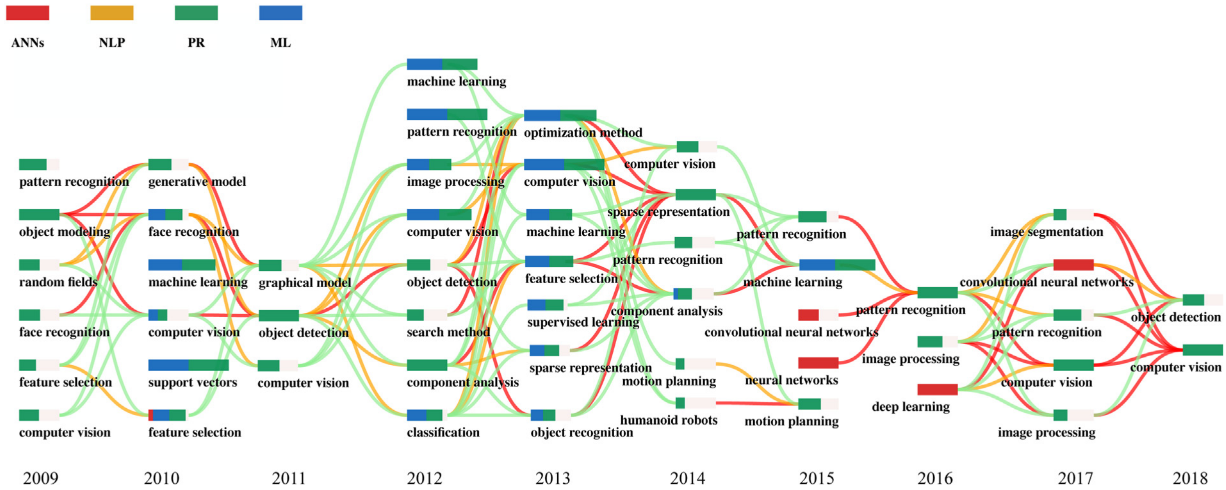

Fig. 9. Structural evolution of pattern recognition in 2009–2018.

Fig. 10. Topic richness and importance in pattern recognition in 2009–2018.

4.3.5. Discussion

The visualization method StructureFlow can distinctly demonstrate the evolution of the hierarchical structure for a discipline, from which we can conveniently observe the emergence, disappearance, merging, and splitting of child-topics within each topic. At the same time, topic migration is one of the main factors in the evolution of discipline knowledge. As seen from Figs. 5, 7, 9, and 11 , this cross-topic knowledge integration deeply affects the paradigm of the research. For example, in Fig. 5 (2018), the influx of deep learning had revolutionized the basic methods in NLP and PR.

In general, the migrated and target topics in the cross-topic integration have the following two characteristics:

- The target topics are highly complex and comprehensive, which makes the intrinsic knowledge in the topics be difficult when completing a research task. It is therefore a must to rely on the intersection and fusion of multi-topics' knowledge to promote the research development. At the same time, the migrated topics are highly integrated and can solve hard problems in the target topics. For example, in Fig. 9, *deep learning* (2015) migrated to *computer vision* (2016), which changes the fundamental research approach of the target topic.
- New research that belongs to the target topic has emerged in the migrated topic. For example, the research content of deep learning appeared in *learning methods* (2017), which re-positions the sub-topics of ANNs and NLP.

Topic migration is typically a result of long-time accumulation. For example, in 2013, *computer vision* began to have the keyword of "convolutional neural networks", but it was not until 2016 that deep learning transferred into this child-topic. This aspect shows that migration is an evolutionary pattern after forming an identifiable research trend, which indicates

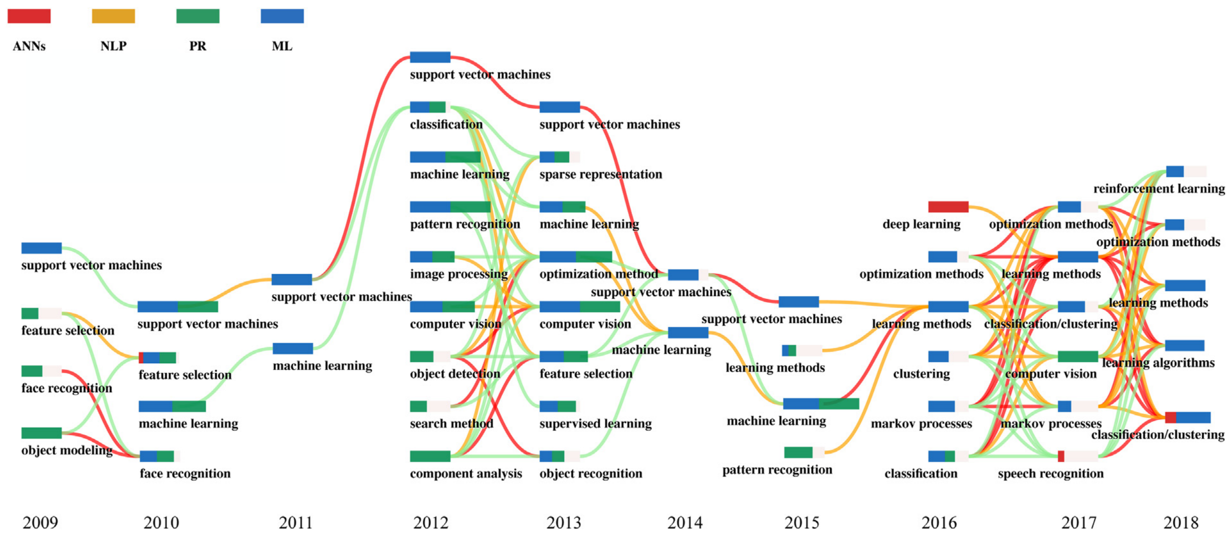

Fig. 11. Structural evolution of machine learning in 2009–2018.

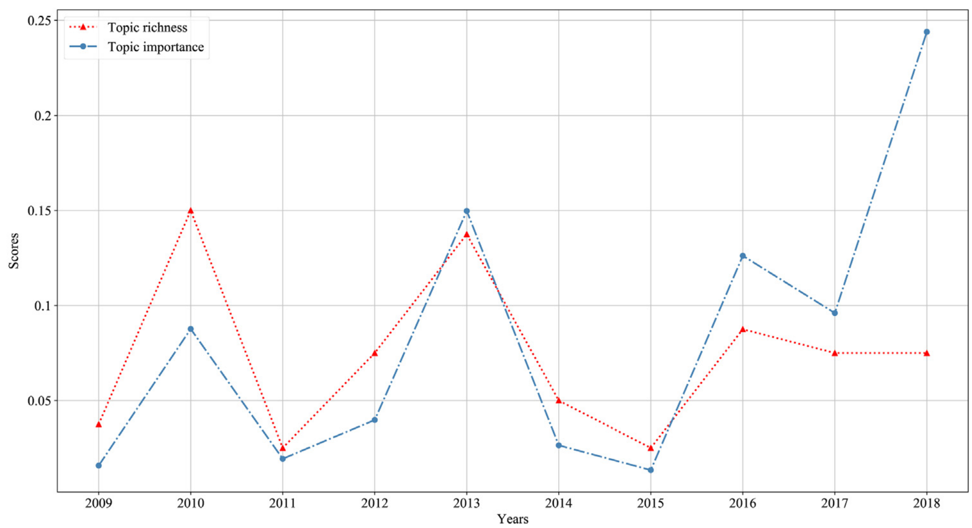

Fig. 12. Topic richness and importance in machine learning in 2009–2018.

a change in the conceptual structure and the formation of a new academic community. Topic migration also provides an effective way to investigate what knowledge is novel, which is considered to be a difficult task. Wang, Veugelers, and Stephan (2017) measure the novelty by examining whether a published paper has an innovative combination of existing knowledge. Based on this theory, the topic migration can be seen as a type of evidence of novel research, which marks an establishment of a formal relationship between the two knowledge entities.

Moreover, we can further analyze some reasons for evolution in the discipline. Overall, ANNs have a significant impact on other areas of AI in recent years, and is widely accepted and shared by the entire AI community. However, our analysis shows that the basic theory of neural networks has remained largely unchanged (the correlation between the various periods of the main child-topic in ANNs was very strong), and thus, there is no subversive scientific revolution. Neural networks is still remaining as a normal phase of a scientific revolution, according to Kuhn's theory, which considers that the development of science is an alternating process between normal science and scientific revolution (Kuhn, 1962). In fact, in addition to the basic method and theory, the development of computing resources and big data are also the key drivers for the success of deep learning.

On the other hand, for the NLP and PR fields, deep learning has promoted their revolutionary development, growing into their research paradigm. Therefore, NLP and PR are at a revolutionary phase in Kuhn's theory, which is caused by the transfer of knowledge in different fields, because a new paradigm solves puzzles that traditional methods have difficulty in addressing. Moreover, ANNs had a process of gradually accumulating interest, based on Fig. 6. However, other topics do not have this trend (their topic importance does not correlate with a time series: $p > 0.05$), which is further proof that they

Table 1
TC comparisons on three algorithms.

The first layer	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
HNMF	**7.68**	**6.57**	**7.16**	6.97	**7.01**	**6.01**	**7.84**	**6.35**	6.54	**6.99**
NMF	6.45	6.09	6.99	**7.08**	6.41	6.00	5.97	6.21	**7.68**	6.76
hLDA	1.27	1.29	1.24	1.04	1.47	1.47	1.66	1.56	1.62	1.44
The second layer	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
HNMF	**8.28**	6.33	**7.45**	**7.49**	**7.24**	**7.32**	**7.69**	7.01	**6.67**	**5.95**
NMF	8.11	**7.27**	7.11	7.25	6.78	6.75	6.73	**7.29**	5.95	5.87
hLDA	1.29	1.36	1.24	1.14	1.51	1.77	1.71	1.70	1.84	1.72

have different evolutionary patterns during the period. Large-scale topic migrations can be seen as a milestone event that triggers this AI revolution.

According to Kuhn's theory of scientific revolution, the development of normal science can make researchers form a more rigorous community. Theories and methods related to a research paradigm become a consensus or even a norm in the discipline, thus forming a scientific community. This activity has been reflected in the reduction of the topic richness of NLP and PR from 2015.

4.4. *Algorithm validity*

To prove the validity of our algorithm, as well as that of the generated topic-trees, we consider the following aspects:

- Each topic node in the topic-tree is reasonable, i.e., the topic terms can availably describe a research direction, which is validated by topic coherence.
- The member weights of the topic-tree can truly estimate the composition relationship between the parent-topic and child-topic. For this purpose, we design a novel indicator named *co-occurrence frequency*, which is based on topic terms co-occurrences, to measure the contribution of the child topic to its parent-topic.
- By studying the changes in the topic consistency with the number of topics, we verify that our method is not sensitive to the number of non-top-layer topics, and only the top-layer's number must be carefully selected.

4.4.1. *Topic coherence*

To quantify the capability of topic mining in our algorithm, we assessed the topic coherence (TC), which indicates that the terms in one topic should have a close correlation. First, for two terms w_i and w_j, we calculate their point-wise mutual information (PMI) score, as follows:

$$PMI(w_i, w_j) = \log \frac{p(w_i, w_j)}{p(w_i)p(w_j)} \tag{13}$$

where $p(w_i, w_j)$ and $p(w_i)$, respectively, denote the probability of co-occurrence of terms w_i and w_j and that of the occurrence of term w_i. For each topic, we selected the ten most representative terms to compute the average PMI. TC is the average of these PMI scores over all topics. Intuitively, for two terms, their co-occurrence contributes to the likelihood that they are assigned to the same topic. TC scores quantify the scheme on the whole dataset.

We compared our hierarchical methods against two baseline algorithms: standard NMF (Lee & Seung, 1999) and hLDA (Griffiths et al., 2004). For standard NMF, we set the number of topics in each layer to be the same as in HNMF and ran the model layer by layer to generate the topics. The hLDA learned the hierarchy based on the prior distribution of the nested Chinese restaurant process, which can automatically determine the width of the hierarchy. Table 1 shows the TC comparisons on three algorithms, in which the maximum score of each year is marked in bold. Compared with the baseline algorithms, our method achieves the best scores in almost all of the layers and periods, and the average performance is improved by 4.3%, which suggests that HNMF can learn reasonable and satisfactory results.

The topics generated by standard NMF have no hierarchy, and there are some ambiguous topics. In the topic-tree generated by hLDA, child-topics under the same parent-topic are very similar, which causes the parent-topic to lack some relevant content sometimes. For example, the SVM-related topic and its child-topics lack content about kernel methods. At the same time, too many similar child-topics also lead the topic-tree to be redundant and less interpretable.

In contrast, our algorithm employs a linear combination of topics in the lower layer to synthesize the upper topics, removing the unimportant child-topics. This algorithm is similar to feature transformation; it not only ensures the subordinate relationship of the parent-child topic but also makes the parent-topics more comprehensive and the child-topics more diverse, thus producing more understandable results with higher TC scores.

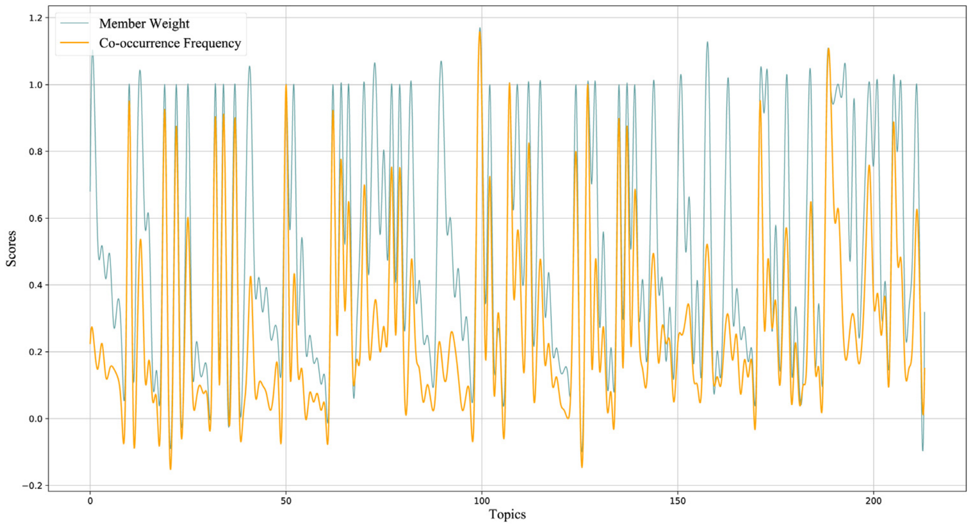

Fig. 13. Comparison of the member weight and co-occurrence frequency.

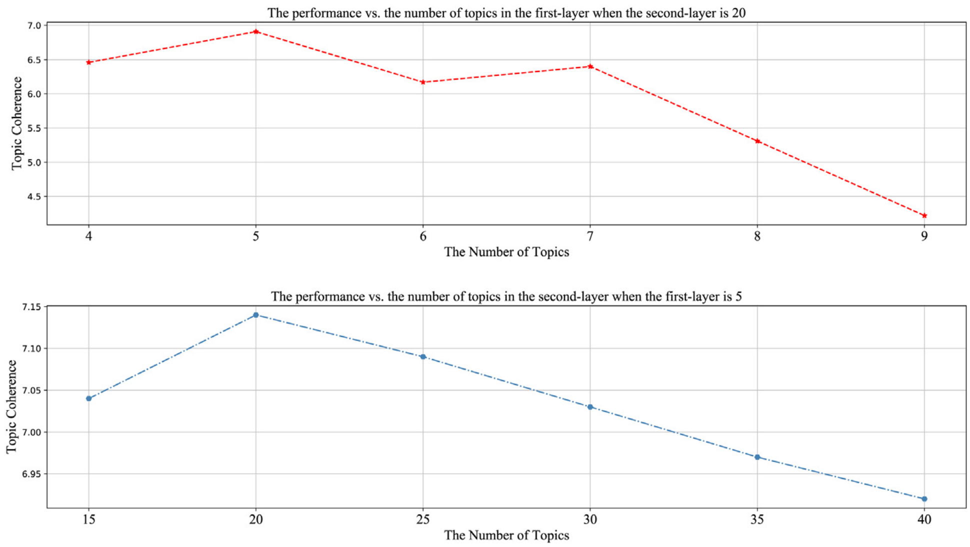

Fig. 14. The performance of HNMF vs. the number of topics.

4.4.2. Topic terms co-occurrences

In our algorithm, the member weight is the proportion parameter of the child-topic in the parent-topic, which reflects the child-topic's contribution to the parent-topic. We design an indicator, called the Co-occurrence Frequency (CF), to measure this contribution as follows:

$$CF = \frac{Size(C_{top-k} \bigcap P_{top-k})}{Size(P_{top-k})} \tag{14}$$

where C_{top-k} and P_{top-k}, respectively, denote the top-k terms in a child-topic and that in a parent-topic, and $Size(\cdot)$ is the number of these terms. Here, we set the top-k to 100.

In Fig. 13, we show 214 topics in the second layer of the topic-tree (filter parameter $\varepsilon = 0.2$) from 2009 to 2018. The x-axis is the index of the topics, and the y-axis is the scores of the member weight (blue line) and CF (yellow line). In these topics, there is a strong correlation between the member weights and CF ($r = 0.88$, $p < 0.01$), which shows that the member weights can reflect the probability that the top-terms of the child-topic appear in the parent-topic, in other words, the contribution of a child-topic to its parent-topic.

4.4.3. Topic number selection

The number of topics in each layer is the important parameter for HNMF. Fig. 14 shows how the averaged topic coherence of each layer's topics in the algorithm varies with the number of topics. In each subfigure, the x-axis represents the number while the y-axis represents the topic coherence score.

In Fig. 14, the performance of the first-layer is more sensitive to the topics' number (the scores' standard deviation is 0.98) than that of the second-layer (the standard deviation is 0.08). This finding is consistent with the theory of our model: HNMF constructs upper-layer topics by selecting important low-layer topics. Therefore, when the number of non-top-layer topics is set to be larger than the ground-truth number, HNMF can filter out some unimportant topics to build the topic-tree. For the top-layer topics, we still must carefully adjust the number of topics. Fortunately, because this number is usually small, it is relatively easy to choose.

5. Conclusions

This study expands traditional flat single-layer topic evolution analysis into multi-level structural evolution analysis, which can better understand how a large-scale discipline develops over time. First, we develop a hierarchical topic model, HNMF, that can generate a topic-tree layer by layer. In contrast to the traditional hierarchical topic model, the multi-level topics produced by our model can determine the weight between the child-topic and its parent-topic. This weight reflects how important the child-topic is in its parent-topic and greatly facilitates the analysis of structural changes. In addition, a visualization method, StructureFlow, is designed to intuitively analyze the structural changes of the topic-tree. Based on this approach, we can accurately analyze the evolutionary patterns within a topic and how the topics interact with one another. These evolutionary patterns include the emergence, disappearance, merging, splitting and migration. Taking artificial intelligence as a case study, our approach find that the development pattern of artificial neural networks is not the same as other topics, such as natural language processing. Its development pattern has a clear process of accumulation of knowledge, while the other areas have undergone revolutionary changes. This revolution is triggered by the migration of neural networks, which have become a norm for research on AI.

In future work, we plan to investigate two remaining questions:

- Our algorithm for generating the topic-tree must predetermine the number of topics in each layer, which is a human factor that affects the analysis. Thus, we will focus on investigating a nonparametric topic model that can also calculate the member weights between the parent-topic and child-topic.
- Our visualization method can only analyze the evolution of two layers under a topic, not the full sub-topic-tree in which that topic is taken as the root. We plan to extend the existing visualization methods into the multi-layer sub-topic-tree to track changes in topic knowledge from a more microscopic perspective, thereby better promoting the understanding of scientific topic evolution.

Author contributions

Yue Qian: Conceptualization; methodology; software; validation; formal analysis; investigation; data curation; writing – original draft; writing – review and editing; visualization.

Yu Liu: Conceptualization; resources; data curation; writing – original draft; supervision; project administration; funding acquisition.

Quan Z. Sheng: Conceptualization; methodology; writing – original draft; writing – review and editing; supervision.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (61672128).

Appendix A. Venue list

Table A.2 is a list of venues from which we collected papers for analysis.

Table A.2
Venue list of artificial intelligence.

1	IEEE Transactions on Pattern Analysis and Machine Intelligence
2	Artificial Intelligence
3	Machine Learning
4	Pattern Recognition
5	IEEE Transactions on Cybernetics
6	IEEE Transactions on Audio Speech and Language Processing
7	International Journal of Approximate Reasoning
8	IEEE Transactions on Neural Networks and Learning Systems
9	IEEE Transactions on Fuzzy Systems
10	Neural Networks
11	International Journal of Computer Vision
12	IEEE-ACM Transactions on Speech and Language Processing
13	Neural Computation
14	IEEE Transactions on Affective Computing
15	ACM Transactions on Applied Perception
16	Journal of Automated Reasoning
17	International Conference on Principles of Knowledge Representation and Reasoning
18	Computational Linguistics
19	Computer Vision and Image Understanding
20	IEEE Transactions on Evolutionary Computation
21	Evolutionary Computation
22	Journal of Speech Language and Hearing Research
23	Journal of Machine Learning Research
24	Journal of AI Research
25	International Conference on Computational Linguistics
26	International Joint Conference on Artificial Intelligence
27	International Conference on Automated Planning and Scheduling
28	Conference on Empirical Methods in Natural Language Processing
29	International Conference on Computer Vision
30	Annual Meeting of the Association for Computational Linguistics
31	AAAI Conference on Artificial Intelligence
32	IEEE International Conference on Robotics and Automation
33	International Conference on Machine Learning
34	IEEE Conference on Computer Vision and Pattern Recognition
35	Annual Conference on Neural Information Processing Systems

References

Ahmed, A., & Xing, E. P. (2010). Timeline: A dynamic hierarchical Dirichlet process model for recovering birth/death and evolution of topics in text stream. *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, 29–38.

Ba, Z., Cao, Y., Mao, J., & Li, G. (2019). A hierarchical approach to analyzing knowledge integration between two fields – A case study on medical informatics and computer science. *Scientometrics*, *119*, 1455–1486.

Beykikhoshk, A., Arandjelović, O., Phung, D., & Venkatesh, S. (2018). Discovering topic structures of a temporally evolving document corpus. *Knowledge and Information Systems*, *55*, 599–632.

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *Proceedings of the 23th International Conference on Machine Learning*, 113–120.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Cai, D., He, X., Han, J., & Huang, T. S. (2011). Graph regularized nonnegative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*, 1548–1560.

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Information (International Social Science Council)*, *22*, 191–235.

Chang, Y. W., Huang, M. H., & Lin, C. W. (2015). Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics*, *105*, 2071–2087.

Chen, B., Tsutsui, S., Ding, Y., & Ma, F. (2017). Understanding the topic evolution in a scientific domain: An exploratory study for the field of information retrieval. *Journal of Informetrics*, *11*, 1175–1189.

Chen, C. (1999). Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing & Management*, *35*, 401–420.

Chen, C., Cribbin, T., Macredie, R., & Morar, S. (2002). Visualizing and tracking the growth of competing paradigms: Two case studies. *Journal of the American Society for information Science and Technology*, *53*, 678–689.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*, *5*, 146–166.

Cuenca, E., Sallaberry, A., Wang, F. Y., & Poncelet, P. (2018). Multistream: A multiresolution streamgraph approach to explore hierarchical time series. *IEEE Transactions on Visualization and Computer Graphics*, *24*, 3160–3173.

Cui, W., Liu, S., Tan, L., Shi, C., Song, Y., Gao, Z., et al. (2011). Textflow: Towards better understanding of evolving topics in text. *IEEE Transactions on Visualization and Computer Graphics*, *17*, 2412–2421.

Cui, W., Liu, S., Wu, Z., & Wei, H. (2014). How hierarchical topics evolve in large text corpora. *IEEE Transactions on Visualization and Computer Graphics*, *20*, 2281–2290.

Dou, W., Yu, L., Wang, X., Ma, Z., & Ribarsky, W. (2013). Hierarchical topics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, *19*, 2002–2011.

Eck, N. J., & Waltman, v. L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, *60*, 1635–1651.

Griffiths, T. L., Jordan, M. I., Tenenbaum, J. B., & Blei, D. M. (2004). Hierarchical topic models and the nested Chinese restaurant process. *Advances in Neural Information Processing Systems*, 17–24.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 5228–5235.

Havre, S., Hetzler, B., & Nowell, L. (2000). ThemeRiver: Visualizing theme changes over time. *Proceedings of the IEEE Symposium on Information Visualization 2000*, 115–123.
Hou, J., Yang, X., & Chen, C. (2018). Emerging trends and new developments in information science: A document co-citation analysis (2009–2016). *Scientometrics*, *115*, 869–892.
Iwata, T., Yamada, T., & Ueda, N. (2008). Probabilistic latent semantic visualization: Topic model for visualizing documents. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 363–371.
Jensen, S., Liu, X., Yu, Y., & Milojevic, S. (2016). Generation of topic evolution trees from heterogeneous bibliographic networks. *Journal of Informetrics*, *10*, 606–621.
Jeong, D., & Song, M. (2014). Time gap analysis by the topic model-based temporal technique. *Journal of Informetrics*, *8*, 776–790.
Kalyanam, J., Mantrach, A., Saez-Trumper, D., Vahabi, H., & Lanckriet, G. (2015). Leveraging social context for modeling topic evolution. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 517–526.
Kersting, K., Wahabzada, M., Thurau, C., & Bauckhage, C. (2010). Hierarchical convex NMF for clustering massive data. *Proceedings of 2nd Asian Conference on Machine Learning*, 253–268.
Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago.
Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*, 788.
Li, G., Zhang, Y., Dong, Y., Liang, J., Zhang, J., Wang, J., et al. (2019). BarcodeTree: scalable comparison of multiple hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, *26*, 1022–1032.
Li, X., Qiao, H., & Wang, S. (2017). Exploring evolution and emerging trends in business model study: A co-citation analysis. *Scientometrics*, *111*, 869–887.
Luo, D., Yang, J., Krstajic, M., Ribarsky, W., & Keim, D. (2010). Eventriver: Visually exploring text collections with temporal references. *IEEE Transactions on Visualization and Computer Graphics*, *18*, 93–105.
Mihalcea, R., & Tarau, P. (2004). Textrank: Bringing order into text. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 404–411.
Moed, H. F., Glänzel, W., & Schmoch, U. (2005). *Handbook of quantitative science and technology research*. Springer.
Moya-Anegón, S. G. F., Vargas-Quesada, d., Chinchilla-Rodríguez, B., Corera-Álvarez, Z., Munoz-Fernández, E., & Herrero-Solana, F. J. V. (2007). Visualizing the marrow of science. *Journal of the American Society for Information Science and Technology*, *58*, 2167–2179.
Neff, M. W., & Corley, E. A. (2009). 35 years and 160,000 articles: A bibliometric exploration of the evolution of ecology. *Scientometrics*, *80*, 657–682.
Rip, A., & Courtial, J. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics*, *6*, 381–400.
Ronda-Pupo, G. A., & Guerras-Martin, L. Á. (2012). Dynamics of the evolution of the strategy concept 1962–2008: A co-word analysis. *Strategic Management Journal*, *33*, 162–188.
Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*, 265–269.
Song, J., Huang, Y., Qi, X., Li, Y., Li, F., Fu, K., et al. (2016). Discovering hierarchical topic evolution in time-stamped documents. *Journal of the Association for Information Science and Technology*, *67*, 915–927.
Song, M., Heo, G. E., & Kim, S. Y. (2014). Analyzing topic evolution in bioinformatics: Investigation of dynamics of the field with conference data in DBLP. *Scientometrics*, *101*, 397–428.
Sun, G., Wu, Y., Liu, S., Peng, T. Q., Zhu, J. J., & Liang, R. (2014). EvoRiver: Visual analysis of topic coopetition on social media. *IEEE Transactions on Visualization and Computer Graphics*, *20*, 1753–1762.
Tu, D., Chen, L., Lv, M., Shi, H., & Chen, G. (2018). Hierarchical online NMF for detecting and tracking topic hierarchies in a text stream. *Pattern Recognition*, *76*, 203–214.
Tuomaala, O., Järvelin, K., & Vakkari, P. (2014). Evolution of library and information science, 1965–2005: Content analysis of journal articles. *Journal of the Association for Information Science and Technology*, *65*, 1446–1462.
Vaca, C. K., Mantrach, A., Jaimes, A., & Saerens, M. (2014). A time-based collective factorization for topic discovery and monitoring in news. *Proceedings of the 23th International Conference on World Wide Web*, 527–538.
Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, *46*, 1416–1436.
Wang, P., Zhang, P., Zhou, C., Li, Z., & Yang, H. (2017). Hierarchical evolving Dirichlet processes for modeling nonlinear evolutionary traces in temporal data. *Data Mining and Knowledge Discovery*, *31*, 32–64.
Wang, S., & Koopman, R. (2017). Clustering articles based on semantic similarity. *Scientometrics*, *111*, 1017–1031.
Wang, X., & McCallum, A. (2006). Topics over time: a non-Markov continuous-time model of topical trends. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 424–433.
Wang, X., Zhai, C., & Roth, D. (2013). Understanding evolution of research themes: A probabilistic generative model for citations. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1115–1123.
Wei, X., Sun, J., & Wang, X. (2007). Dynamic mixture models for multiple time-series. *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2909–2914.
White, H. D., & Griffith, B. C. (1981). Author co-citation: A literature measure of intellectual structure. *Journal of the American Society for information Science*, *32*, 163–171.
Wongsuphasawat, K., Guerra Gómez, J. A., Plaisant, C., Wang, T. D., Taieb-Maimon, M., & Shneiderman, B. (2011). LifeFlow: Visualizing an overview of event sequences. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1747–1756.