The Dawn of Metamaterial Engineering Predicted via Hyperdimensional Keyword Pool and Memory Learning

Young-Ok Cha and Yang Hao*

Metamaterials research has been ongoing for more than 20 years and it has gained much public and scientific interest. There have been many experts forecasting on the road ahead for metamaterials, notably, in lieu of “knowledge tree” in 2010. Ten years on, it is proposed to re-examine these claims by using automated computer tools, such as natural language processing (NLP), to extract research information for processing and analyzing from unstructured texts in publications. In this study, a fully auto-generated database of 43,678 abstracts related to metamaterials published between 2000 and 2021 using Scopus Search API (Application Programming Interface) is built. Applying word embedding, each keyword is studied in a hyperdimensional vector space and clusters so that their relationships can be visualized for assessing the popularity and trends of research themes. A neural network model developed based on the encoder–decoder long short-term memory (LSTM) architecture is finally trained to predict future directions and theme evolutions in the next four years for selected topics. This study not only provides vital information in terms of impact of metamaterials research but also lays down a solid foundation for the development of future metamaterial research roadmap in the form of Gartner hype cycle.

1. Introduction

In scientific research, there is no doubt that setting a correct path is of great significance and importance to all key stakeholders in the scientific community, as it will ensure the proposed research to be effective and productive while reducing its cost by effective forward planning. This could be accomplished by the analysis of the plethora of data of relevant and past studies using the predictive analytics. Current human-expert-based approaches however could only be performed within one’s area of expertise. Unlike the systematic review, where a formulated research question shall be provided with a collection of secondary data and meta-analysis, objectivity is often considered for scientific inquiry, as a good reason for valuing scientific knowledge, and as the basis of the authority of science in society.[1] The prediction made by a human expert is likely to be influenced by his or her particular scientific perspectives or personal interests leading to biased objectivity. In addition, the performance of such predictions is difficult to be quantitatively measured. Quantification in science prediction has a number of advantages such that it can lead to direct comparisons, time-saving, and large-scale analysis. Many researchers have therefore believed that every aspect of science can, and in fact should be quantified.[2]

Thanks to electronic and open access publications, scientific data can now be made available in the task of automated information extraction (IE) from unstructured or semi-structured electronic documents, including articles, tables, and figures etc. This is particularly useful for researchers not only to retrieve an increasing amount of digital data from literature, but also search information from patents, papers, and theses.[3]

Recently, Cole et al. proposed a new approach that enables materials discovery via a four-step “design-to-device” pipeline that entails 1) data extraction, 2) data enrichment, 3) material prediction, and 4) experimental validation, which has dramatically reduced the average molecule-to-market lead time.[4] Court et al. automatically built a database of 39,822 records from a corpus of 68,078 journal articles containing chemical compounds and their associated Curie and Néel magnetic phase transition temperatures.[5] In 2016, Swain et al. developed a toolkit named ChemDataExtractor that can build a chemical database from a large number of literature automatically,[6] which paves the way for the automated extraction of physical/chemical entities and their associated material properties, measurements, and relationships that can be used to guide new scientific studies. Tshitoyan et al. proposed an approach which applies unsupervised word embedding models for collecting and processing approximately 3.3 million scientific abstracts published between 1922 and 2018 in more than 1000 journals deemed likely to contain materials-related research, resulting in a vocabulary of approximately 20,000 words.[7] The authors used Elsevier and Springer Application Programming Interface (API)
and abstracts related to inorganic material science. Finally, 20,000 words were embedded into 200-dimensional vector spaces and these embeddings made it possible to perform word analogies.\textsuperscript{[6,9]} They suggested that latent knowledge captured by natural language processing (NLP) can predict future discoveries to a large extent based on the information embedded in past publications. In 2020, Krenn and Zeilinger demonstrated a semantic network for predicting future research trends of quantum physics from 750,000 scientific papers as well as information from books and Wikipedia and confirmed high-quality predictions using historic data.\textsuperscript{[10]} In this study, the abstracts of 100,000 arXiv articles in quantum physics generated a list of quantum physics concepts using one of the NLP tools named rapid automatic keyword extraction (RAKE).\textsuperscript{[11]} This approach, however, is not fully automated as the authors used the RAKE algorithm as well as the keywords which were manually extracted by human experts based on information on quantum physics from books and Wikipedia. They have demonstrated that the proposed network can predict future trends in the field of quantum physics, particularly by suggesting pairs of disjointed and unexplored concepts, independent from the research agendas of an individual scientist.

In this paper, we will turn our focus on the research of metamaterials, which one of the co-authors has witnessed and actively participated in its extraordinary development. Metamaterials were originally defined as “a new class of ordered composites that exhibit exceptional properties not readily observed in nature”.\textsuperscript{[12]} While conventional, isotropic, “positive” index materials and composites have provided the necessities for designing electromagnetic devices, metamaterials have certainly extended our palette of electromagnetic materials to include non-isotropic, multilithic materials, with effective material parameters (permittivity and permeability) that simply do not exist in nature, and in which the electromagnetic properties vary in at least one dimension in a controlled and pre-designed way, both spatially and temporally. Most materials found in nature are highly polarisable on application of an electric field, and by contrast, most naturally occurring materials respond weakly to magnetic fields. Metamaterials are finely structured (on a scale less than the electromagnetic wavelength but larger than individual atoms) composites whose geometrical patterning results in specific (often resonant) interactions with electromagnetic fields leading to artificial, effective material parameters such as permittivity and permeability. This leads, for example, to structures such as the “Swiss roll” which respond to time-varying magnetic fields but not to static magnetic fields, a desirable property for magnetic resonance imaging machines.\textsuperscript{[13]} Another much sought after metamaterial has negative refraction, corresponding to a material in which the power flow is in the opposite direction to the wave vector.\textsuperscript{[14]} Such unconventional material characteristics are inherent in the design of electromagnetic cloaks and perfect lenses and clearly demonstrated the opportunities that novel electromagnetic materials may provide. Further developments have also included the so-called metallic “fishnet” structure,\textsuperscript{[15]} and all-dielectric metamaterials comprised of spherical and disk resonators,\textsuperscript{[16]} but all of these resonant metamaterials show relatively narrow bandwidth and significant loss. Creating artificial electromagnetic media without these two severe constraints is still very challenging, despite the fact that active/tunable materials may be incorporated. Originally, metamaterials were specifically referred to as a class of artificial materials which, have simultaneous negative permittivity and permeability and are also known as left-handed materials (LHMs). Present researchers tend to expand the concept of metamaterials so as to make it as broad as possible.\textsuperscript{[17]} Recent trends are pointing to design and fabrication of metamaterials, bringing together disciplines such as the physics and chemistry of condensed matter, quantum, electromagnetic theory and computational methods ranging from optics to microwaves, and micro- and nanofabrication, as well as electronic, mechanical, thermal and optical of materials. This multidisciplinary approach has unlocked the full power of electromagnetics and make new practical applications a realizable target. In particular, there have recently been many new developments in material fabrication techniques for “designer materials”, including inkjet printing, selective laser sintering etc., but these processes offer neither the scale nor the speed to fabricate the more ambitious demonstrators. New approaches to large area prototype manufacturing, and successful demonstration of metamaterials and their applications will require tight integration of the expertise in the physics of materials with the team specializing in innovative manufacturing strategies.

In 2010, Zeludev predicted the road ahead for metamaterials with a “Tree of Knowledge” describing the negative-index media as “forbidden fruit”.\textsuperscript{[18]} He articulated in the paper that the study of chiral, negative index and artificial magnetic metamaterials has matured, while the theory and technologies behind transformation optics, materials with high/low epsilon and designer dispersion have been “ripe”. He reported that switchable metamaterials based on arrays of micro- and nano-electromechanical devices were also being developed and their research was highly interesting. This is also true with active metamaterials, such as those based on non-Foster’s theorem,\textsuperscript{[19]} which loss and bandwidth issues of metamaterials may be overcome. Most metamaterials exhibit strong local field enhancement near their resonances and it makes metamaterials attractive for non-linear optical applications. Sensor applications were another growth area in metamaterials research where 2D materials such as graphene and TMD (transition metal dichalcogenides), which a single molecular layer of carbon can induce a multi-fold change in the transmission of metamaterials.\textsuperscript{[20]} Finally, he predicted that superconducting metamaterials would find their applications in exploiting quantum coherence with a multilevel quantum structure replacing the classical plasmonic resonators.

Meanwhile, ten years on, there have been many bodies of review articles summarizing recent research and commercial developments of metamaterials.\textsuperscript{[21,22]} These papers and reports are often focused on specific subject areas and fail to present a whole picture of recent metamaterial research. None of these studies has validated previous predictions and research roadmap development with a systematic study of a large volume of historical and recent research outcomes. More importantly, many reviews and technical reports were written in line with author’s area of expertise.

In this work, we introduce a novel framework that combines automated data extraction from scientific databases with NLP for intelligent forecasting. As the scientific findings in this domain are published in a variety of academic venues, this study uses
Elsevier’s Scopus API which allows access to 50 million abstracts of over 20,500 peer-reviewed papers from more than 5,000 publishers, capturing articles published in virtually all scholarly journals of any significance in the world, including the American Association for the Advancement of Science, Springer-Nature, the Institute of Electrical and Electronic Engineers (IEEE) and the American Institute of Physics (AIP).\(^{23}\) Our work differs in several aspects from the previous studies. First, all keywords of metamaterials research are extracted automatically using the RAKE algorithm, we then categorize these keywords into objects and properties depending on their attributes. We then vectorize the list of 3,187 keywords as unstructured natural texts not only for visual analytics but also future trend forecasting based on their frequencies appeared in the search as well as a sequential machine learning (ML) algorithm. We conduct a clustering analysis of all keywords via unsupervised mapping and labeling. As a result, 10 mostly appearing keywords from 8 clusters in the field of metamaterials have been obtained. This study also demonstrates the feasibility of using a modified recurrent neural network (e.g., encoder–decoder LSTM) for predicting research trends for the next four years from a sequence of published data collected between 2000–2021. Finally, a hype cycle is automatically generated to trace the evolution of metamaterials research as they pass through successive stages pronounced by the peak, disappointment, and recovery of expectations.

2. Results

An overall architecture of the proposed approach can be broadly described including four key steps, namely, data extraction, hyperdimensional keyword pool building (KPB), keywords relationship visualization, and future trend forecasting as shown in Figure 1. We anticipate that this framework is not designed simply to assist a human expert but rather substitute the conventional literature review and roadmap development process via its four aforementioned novel concepts used within the framework, and it alone can perform any prediction tasks accurately on any subject of scientific research while reducing associated project costs.

2.1. Data Extraction

An abstract of published research articles is a self-contained, short and powerful statement that describes or summarizes a whole paper.\(^{24}\) Learning from abstracts for keyword prediction offers a number of benefits: First, it helps text pre-processing and learning by reducing the complexity that comes from irrelevant information potentially contained in other parts of a paper. Second, the computational cost of text preprocessing and learning could be significantly reduced due to a smaller amount of texts in an abstract than those in a whole paper. This study thus not only improves the efficiency of the required NLP and learning processes but also leads to accurate learning for forecasting.

Scopus client is one of the Elsevier’s APIs which allows an access to its largest database of abstract and citation of the literature and relevant web sources. Scopus API particularly indexes “metadata” from abstracts and references of thousands of publishers including Elsevier. For our data extraction task, connecting the Scopus repository\(^{25}\) with our own

---

Figure 1. The architecture of the proposed keyword prediction system. The keyword prediction system is composed of four main components, namely, information extraction (IE), keyword pool building (KPB), visualization, and forecasting. The keyword pool (KP) is built as a result of the rapid automatic keyword extraction (RAKE) algorithm and author keywords in the KPB phase that processes 43k abstracts. The visual analysis of KP is performed to seek the relationships among the selected keywords in KP. The normalized frequencies (NFs) of keywords in KP are fed into the encode-decoder long short-term memory (LSTM and the learned models are validated for further processing such as magnitude analysis. The future materials research trends are finally predicted and published.
X-ELS-APIKey, we downloaded multiple JavaScript Object Notation files that contain the information of each paper related to metamaterial studies such as title, publication date and abstract from the year 2000 to 2021.

2.2. Hyper-Dimensional KP

To build a domain-specific KP, we have used various techniques of locating and defining keywords. RAKE finds the most relevant words or phrases in a piece of text. Using such key phrases is of particular importance as we can identify object-property pairs. One potential issue however is that too many candidate keywords could be selected due to the way to find keywords-pairs, which is mainly based on word occurrence or frequency. For example, some of the less meaningful keywords such as “huge potential”, “grown rapidly”, “current issues” could be selected as they have been frequently used in the literature.

To address this issue, the less-meaningful keywords were filtered out by using the words in the title and the list of the author’s keywords as the reference. Some of the published papers accessible via Elsevier’s API do not have the author’s keywords. In this case, the words appear in both the titles of the literature and the RAKE generated word list were selected. If the paper has the author’s keywords, the words appear in both the list of author’s keywords and the RAKE generated word list were selected. Here, RAKE was set to choose up to 2-words phrase. Setting RAKE to choose 3 or more word phrases may bring an additional complexity in searching candidate keywords. The example of keyword selection from a typical paper published in 2020 is shown in Figure 2a.

This is a fully automated process that creates a KP which contains 3187 keywords. A simple statistical analysis of the KP shows that the most frequently occurring keywords over the most recent 20 years are “transformation optics”, “photonic crystal”, “negative refraction”, “mutual coupling”, and “negative permittivity”.

For further analysis, all the selected words in the KP were vectorized onto the hyper-dimensional space using Mat2Vec, a pre-trained word embedding model which extracts multidimensional vector representations of each word. Mat2Vec is built from an unsupervised word embedding using 3.3 million scientific abstracts. Each word in KP is however either 1-word or 2-word. While the 1-word keyword could easily be vectorized however in the case of 2-word keyword, each word in the 2-word keyword was vectorized then added together by using the word embedding analogy which allows vector-oriented reasoning based on the offsets between the pair of words. For example, the male/female relationship is automatically learned, and with the induced vector representations, “King − Man + Woman” results in a vector very close to “Queen”.

Likewise, we could calculate a vector using the analogy for each 2-word keyword. For example, “acoustic cloaking” was embedded onto the sum of the vectors of “acoustic” and “cloaking” as shown in Figure 2b. Each keyword in the KP was respectively vectorized into 200-dimensional embeddings as shown in Figure 2c, which represents the process of how the KP is built and visualized.

3. Clustering Keywords

To gain a high-level view of all relevant topics in metamaterials research, we have created keyword clusters, by merging words with similar meaning. At this stage, we had 3187 embedded keywords in 200-dimensional spaces with no label. A clustering analysis was thus used in order to perform an unsupervised mapping and label each word based on pre-identified clusters in the vector space. For the clustering analysis, the high dimensionality of the words in the KP was undesirable. The curse of dimensionality theory demonstrates that as the data is moving into higher dimensions, the sparsity of data and statistical error will grow exponentially. We thus reduced the current number of dimensions, as illustrated in Figure 3, by computing the principal components that simplify the complexity in high-dimensional data while increasing interpretability. It draws on a distance-based clustering algorithm, k-means, which aims to minimize the sum of squared Euclidean distances to each cluster mean (centroid): $\min \sum\sum (x - \mu_i)^2$, where $\{\mu_i\}_{i=1}$ are the cluster means and $S_c$ are all vectors assigned to cluster $c$. The algorithm alternates between reassigning vectors to the closest cluster means, and then updating the means. The number of clusters, $k$, is found using the elbow method ($k = 8$).

The resulting plots are depicted in Figure 3 demonstrating labeled clusters and 10 most frequently appeared keywords for each cluster.

3.1. Building Time-Series Data for KP

To analyze the trend of word occurrences over the years, we start with building time-series data related to appearance frequencies of each keyword. We first recorded the frequencies of each keyword per year over the duration of the first quarter of 2000 to the second quarter of 2021. To reduce the bias in time period selections, we normalized the value of frequencies for each keyword using Equation (1), which corrects all data in the time series to a common scale. As in Equation (1), NF is a normalized value that is computed as the frequency of the given word frequency (WF) divided by the total number of literatures from certain period.

$$NF_i = \frac{WF_i}{\text{total no. of literatures}} \times 100$$

3.2. Keyword Forecasting

The NFs were now fed into a sequential memory-based ML algorithm for the prediction of the frequency of each word. To date, many different types of time-series analysis algorithms have been proposed. Methods for time series analysis may be categorized into two classes, namely, frequency-domain and time-domain methods. The former includes wavelet analysis while the latter comprises autocorrelation. The approach for time series analysis could also be categorized into parametric and non-parametric methods. Parametric methods rely on assumptions about the shape of the distribution in the
underlying population while non-parametric methods presume that the data distribution cannot be defined in terms of such a finite set of parameters thus taking flexible and model-free approaches. This study uses the LSTM, a time-domain and non-parametric method given the complex and temporal nature of input data. We believe that the future research trend could be predicted via the analysis of predicted keyword frequency and its rate of changes in a sequential format. We thus adopted the encoder–decoder architecture of LSTM for this task.

The model consists of three main components, namely, encoder, intermediate (encoder) vector and decoder. Encoder and decoder use a multi-layered LSTM unit to map the input sequence to a vector of a fixed dimensionality, and then another LSTM is used to decode the target sequence from the vector.[37–39] These are trained using the input data

Figure 3. The labeled clusters using k-means and 10 most frequently appearing keywords from each cluster. The data space is split into 8 clusters and each cluster is labeled. In this way, the keywords could be analyzed as a group instead of individuals. Green cluster seems to be focused on applications category, however, “pin diode” shall be in red cluster for fabrication and enabling materials category. “liquid crystal” in pink cluster which seems to focus on material constitutive parameters, however, shall be in red cluster. Blue cluster may represent modeling and optimization category, yet “energy harvesting” from blue cluster shall be in green cluster for applications. Purple cluster is focusing on wave phenomenon; yellow cluster focuses on antennas while orange cluster is for optical properties; black cluster seems to focus on guided wave devices category.

Figure 2. The process of selecting keywords and embedding. a) The words in blue circle are the ones selected by RAKE using the abstract while the words in red circle are from the paper’s title and author’s keywords. The words in the intersection are the selected keywords. b) In word embedding, each word is represented in vector space thus the words could be added or subtracted for analogy. For example, “acoustic cloaking” can represent the sum of “acoustic” and “cloaking”, and “acoustic cloaking” is also embedded with amount of a distance between “metasurface” and “cloaking” from “acoustic metasurface”. c) From 42k literatures, the overlapping words between RAKE and author’s keyword and title are selected as the keywords in KP. The KP has 3187 keywords which are then vectorized into 200-dimensional spaces for the visualization on 3-dimensional spaces.
while maximizing the conditional probability of the target sequence for a given sequence. As our data is in a time-sequential format, the current keyword frequency has been predominantly calculated from the legacy data in the past years. We designed a many-to-one architecture model so that it was able to look back several sequences and predict a new sequence ahead. To avoid the data scarcity in training, especially for new keywords from new emerging research in metamaterials, we applied the data computed quarterly thus we have in total of 78 data points until the second quarter of 2021, instead of using the frequency of keywords calculated on the annual basis. These 78 data points were then used as an input to the many-to-one encoder–decoder model as shown in Figure 4. The learned model was fed with an input data of a sequence of every 6 frequencies in the form of sliding window for predicting the next data point. To predict the research trend in the next 4 years, we repeat the above process 16 times. The modified architecture of the encoder–decoder model has two LSTM units (e.g., encoder and decoder) and each unit is composed of 300 layers with rectified linear unit (ReLU).

To understand the past and current trends of keywords in metamaterial research, we have presented them respectively with an increasing, decreasing, leading, or emerging trend, together with our own reasoning. We noticed that keyword frequencies of “plasmonic nanoantenna”, “graphene”, and “terahertz wave” have increased steadily over the past 20 years. On the other hand, the keywords such as “negative refraction” and “effective permeability” fall under the typical decreasing trend. The keywords such as “coding metasurface”, “2D materials”, and “additive manufacturing” have emerged over the recent 5 years truly reflecting the current research trend. Figure 5 shows a 100% stacked area chart for overall leading keywords and heatmaps for the other three trends.

To analyze the relationships between the labeled clusters and each type of trends, we have selected and followed one hundred keywords showing the greatest SF difference from each type of trend by counting the number of keywords in its corresponding clustering category. Figure 6 shows that the words in the “modeling and optimization (blue)” cluster have the majority portion in the overall leading and decreasing trends. Figure 6 illustrates the keywords which are related to “applications (gray)” and “fabrication and enabling materials (green)” have large proportions in the trends of emerging and increasing. This indicates that the number of studies which gained popularity in the early era of metamaterial research (e.g., keywords in “modeling and optimization” as well as “optical properties” categories) has decreased while the interest in translating metamaterials into practical applications (e.g., keywords in “fabrication and enabling materials” and “applications” categories) have increased. Besides, the cluster of “antennas” shows less than 10% in all trends. Figure 7 visualizes 30 keywords with the emerging trend in 8 clusters, each of which shows the most popular research area over the recent years in each labeled category.

Given the small number of sequences, the LSTM models were trained with the objective of minimizing mean square error (MSE) which is defined by \( \text{MSE}(\hat{x}_{T+h}) = E[(x_{T+h} - \hat{x}_{T+h})^2] \). The optimal h-step-ahead forecast of \( x_{T+h} \) at time \( T \) is the conditional expectation, \( E[x_{T+h} | \Omega_T] \). Therefore, if \( \hat{x}_{T+h} \) is any h-step predictor at time \( T \), we have that \( \text{MSE}(\hat{x}_{T+h}) \geq \text{MSE}(E[x_{T+h} | \Omega_T]) \).

To build an optimal model, the MSE between the predicted sequence and the target sequence has been minimized. The validated models were used for 16 steps ahead predictions for the period of the third quarter of 2021 to the second quarter of 2025 based on the data collected for the duration between the first quarter of 2002 and the second quarter of 2021. As a result of prediction (Figure 9a), we now have estimated frequencies of appearance for every keyword in the KP. The trend of scientific research and technology development often follows an S curve which measures the effort (e.g.,
Figure 5. The 100% stacked area chart and heatmaps for each trend. a) This 100% stacked area chart shows how high-frequency keywords (the overall leading trend) of a whole KP have changed over 20 years. The y axis scale is 100%. Each area of color represents one keyword of the whole KP. The parts are stacked up, vertically. The height of each colored stack represents the percentage proportion of that keyword at a given point in time. b) Increasing trend, c) Decreasing trend, d) Emerging trend: Each heatmap shows a 2D graphical representation of the scaled frequency (SF) of each keyword in a data matrix. Each cell reports a numeric frequency; however, the numeric frequency is accompanied by a color, with larger frequencies associated with darker colorings.

time and money) on its X-axis against technical performance on its Y-axis. It tells that at the beginning of technology lifecycle, there is a great deal of investment with relatively little performance improvement. However, when it reaches some tipping point, its performance increases significantly, while towards the end, the performance becomes less distinguishably improved.
and reaches a plateau. Gartner’s Hype Cycle, a graphical depiction of a common pattern that arises with each new research and technology, was proposed in 1995. It graphically visualizes the technology lifecycle, however, in different dimensions with time on its X-axis and expectations on its Y-axis. Its emphasis is particularly on the expectations of technical performance in the marketplace, and it is generally accepted that although it focuses on specific technologies, the same pattern of hype and disillusionment applies to more high-level concepts.\(^{[43]}\) The Hype Cycle usually starts when an event generates public interest in a technology innovation. The expectation increases dramatically after the stages of technology trigger (TT) and peak of inflated expectation (PI). In the phase of trough of disillusionment (TD), as many experiments and implementations fail to deliver, expert groups react negatively. The technology is then more widely understood so that more industries fund pilots cautiously in the phase of slope of enlightenment (SE) followed by the phase of plateau of productivity (PP) when mainstream adoption starts to take off. It is undoubtedly that the hype cycle also applies to the research of metamaterials, it would be therefore of great help in setting a correct path for future research ensuring it to be efficient and cost-effective. We noticed, however, that in metamaterial research, some research topics do not slide into SE or PP phases, but they rather disappear. When the inflated expectations begin to die down via the phase of TD, they start to decrease, and this

---

**Figure 6.** The number of keywords from each cluster in four different types of trends.

**Figure 7.** The 30 emerging trend keywords from each cluster and their extent. This 3D plot shows 30 words with the largest increase in frequency over the past 5 years. The size of sphere represents the frequency growth rate. Based on the size of sphere, we can identify the emerging keywords. Emerging trend keywords are “dielectric metasurface”, “coding metasurface”, “2D materials”, and “acoustic metasurface”. As its color represents, each of these keywords belongs to the cluster of antenna and material constitutive parameters.
trend continues over the phases of SE and PP as indicated with a red line in Figure 9b. Among the four trend types that we defined, it is obvious that emerging and increasing keywords fall into TT and PI phases. However, decreasing keywords should be divided into the words sliding into SE and PP phases (continue following blue line) or the words disappearing (following the red line) based on its prediction results. The order of each keyword in the Hype Cycle is decided within each phase, based on the difference of corresponding SF values.

On the basis of the preceding fully automated analysis of the past metamaterial research trend and its forecasting, it can be said that, for the past 20 years, we have witnessed substantial effort in translating metamaterial research from fundamental studies to engineering practices including industrial applications, large scale manufacturing and multi-material integration etc. Aligning with the argument of Zheludev,[46] the research on “graphene”, “plasmonic metasurface”, and “nonlinear metasurface” has also gained significant attention among metamaterial researchers. Although he stated in his article[18] that the switchable metamaterials will bring major benefits via material’s properties (e.g., vanadium oxide (VO₂)), its role and significance in coding metasurface or programmable metasurface to harness the power of both computer science and metamaterials were not anticipated. Most relevant research is still based on the use of III-V semiconductor-based diodes and varactors. On the “forbidden fruit” of negative-index media, although the concept was widely accepted in the scientific community, research activities on “negative refraction”, “double negative media”, and “negative index media” have significantly decreased.[45] It is partially due to the difficulty in fabricating low loss materials and demonstrating ground-breaking applications such as “perfect lens”, and, meanwhile, “metals” consisting of millions of meta-atoms with nanoimprint lithography process has been successfully demonstrated and commercialized.[45] Another cornerstone of metamaterial research is “transformation optics”, a.k.a. “coordinate transformation”, where the concept of “cloaking”[46] has drawn significant attention from the public and academic community. Despite being an important mathematical tool in the design of electromagnetic devices, and, more broadly any device with operating principles governed by partially differential equations, the volume of its research in metamaterials has never been significant, and also seen the decreasing trend. Coupled tightly with the concept of “transformation optics” is “cloaking”, which still remains to be an active subject of research.[47] far from being “ripe” as predicted in article of Zheludev,[18] due to the complexity of material designs and inherent loss/bandwidth limitations, leading to plateau of productivity (Figure 9b). Technological trigger or user pushing may be needed to regain some momentum of basic research on “negative refractive index” materials and “perfect lens”. “nonlinear metamaterials”, “amplifying metamaterials”, and “switchable metamaterials” can be broadly classified as “active metamaterials” have seen continuous growth in subject areas ranging from microwave to optics in topical areas of “asymmetric transmission”, “broadband absorption”, and etc.[48] They are largely driven by applications in all optical modulation and switching enabled by silicon and graphene-based nanophotonics. Metamaterials have their uniqueness to manipulate polarization states, modulation depths, and absorption of propagating waves as well as radiation from plasmonic nanoantennas. The development of “quantum metamaterials” is accelerated by a vast amount of recent new funding invested by major agencies worldwide, in line with the growth of strong research activities in “quantum emitter”, “topological phase”, and “superconducting” etc. “Sensor metamaterials” maintain a strong growth in areas such as “remote sensing” and “THz imaging”. Two distinctive subjects missing from the “Knowledge of Trees”[18] are “acoustic metamaterials” and “thermal metamaterials”. both of them witness a strong growth in topics around “thermal expansion”, “sound transmission”, “transformation thermodynamics”, and “ultrasonic waves”, especially, for the latter, the focus of study shifts from local resonances to broadband sensing performance, and the transmission of longitudinal waves. “Mechanical metamaterials” such as those based on “Kirigami structures” have a capability of being hyper-elastic and possessing on-demand auxetic behavior. Three key topics, namely, “metasurfaces”[49], “graphene”[50], and “surface plasmon” dominate the current metamaterial research, it can be evident by the fact that “new” emerging subjects emerge in “nonlinear metasurfaces”, “plasmonic metasurface”, “quantum photonics”, and “topological photonics”, all of which are enabled by “wonder” materials such as graphene, graphene oxide, black phosphorus, and nanoporous gold, etc. “Programmable
metasurface” and “coding metasurface” have drawn attention from microwave and communication engineers, with analogical terms such as “reflecting intelligent surface (RIS)” etc., which are largely made of conventional PIN diodes controlled via FPGA (field-programmable gate array), a semiconductor-integrated circuit where a large majority of the electrical

Figure 9. The results of prediction for next 4 years using encoder–decoder LSTM models and the hype cycle in year 2025.
functionality inside the device can be changed. Discoveries of novel tunable and phase changing materials are urgently needed and accelerated by ML and artificial intelligence.[10] Hologram metasurfaces seem to be common for all spectra for applications including camouflage, even for acoustic waves. Finally, “additive manufacturing” ranging from 3D/4D printing to roll-to-roll (R2R) processing has become a key technological enabler, which sustains the growing trend of metamaterial research.

Looking into the future, for the next four years, we will continue to see the growth in studies of “metasurfaces” including “plasmonic metasurfaces” (Figure 9a), “acoustic metasurfaces” and “nonlinear metasurfaces” (Figure S5, Supporting Information). The popularity of “cloaking” and “additive manufacturing” studies will reach its peak of recent cycle and start to decline (Figure 9a). It is interesting that despite a declining trend overall in “2D materials” research, studies on “graphene” and “black phosphorus” will remain strong for the next four years (Figure 9b). Metamaterials applied to antenna applications including wearable antennas will decline but “photonic crystals” research will regain its momentum, by focusing manipulating light in deep 3D structures consisting of a large number of nanopores etc.

4. Concluding Remarks

This study proposed a state-of-the-art method that enables an automatic extraction of the information from unstructured texts, analyses of publicly available publications and prediction of future research trends in materials science. This is of paramount importance for those areas where setting a correct future research path ensures the research to be effective while reducing its cost. As it is well believed today that the most respected sources of scientific data are coming from scientific publications such as journals and patents, this study automatically built a KP from over 43 000 abstracts of various scientific publications and all the words in KP were vectorized into hyper-dimensional spaces for a mathematical algorithm to understand the meaning of word via correct representations. The proposed keyword prediction framework not only effectively visualizes the trends and relationships of each word in KP but also forecasts the future research trend in the form of hype cycle until 2025. Our results demonstrated that the proposed approach is valid and versatile, and arguably it can be applied to any research field of interest. To the best of our knowledge, this is the first study that proposes a fully automated future keyword prediction framework in materials science and provides useful benchmark to future metamaterial research based on newly built KPs.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

forecasting, machine learning, metamaterials, metasurfaces, natural language processing

Received: November 10, 2021
Revised: February 3, 2022
Published online: March 18, 2022

[1] J. Reiss, J. Sprenger, in The Stanford Encyclopedia of Philosophy (Ed: E. N. Zalta), Stanford University, Stanford, CA, 2014.
[2] Y. B. Zhou, L. Lü, M. Li, New J. Phys. 2012, 14, 033033.
[3] A. C. Vaucher, F. Zipoli, J. Geluykens, V. H. Nair, P. Schwaller, T. Laino, Nat. Commun. 2020, 11, 3601.
[4] M. C. Cole, Acc. Chem. Res. 2020, 53, 599.
[5] C. J. Court, J. M. Cole, Sci. Data 2018, 5, 180111.
[6] M. C. Swain, J. M. Cole, J. Chem. Inf. Model. 2016, 56, 1894.
[7] V. Tshitoyan, J. Dargdelan, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, A. Jain, Nature 2019, 571, 95.
[8] T. Mikolov, K. Chen, G. Corrado, J. Dean, ICLR 2013.
[9] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Advances in Neural Information Processing Systems (NIPS) 2013, 2, 3111.
[10] M. Krenn, A. Zeilinger, Proc. Natl. Acad. Sci. USA 2020, 117, 1910.
[11] S. Rose, D. Engel, N. Cramer, W. Cowley, M. W. Berry, J. Kogan, Text Mining: Applications and Theory, John Wiley & Sons, Inc., Hoboken, NJ 2010, pp. 1–20.
[12] V. M. Browning, M. H. Tanielian, R. W. Ziolkowski, N. Engheta, D. R. Smith, MRS Online Proc. Libr. 2006, 919, 102.
[13] D. R. Smith, J. B. Pendry, M. C. K. Wiltshire, Science 2004, 305, 788.
[14] K. G. Balmain, A. A. E. Luttgen, P. C. Kremer, IEEE Trans. Antennas Propag. 2003, 51, 2612.
[15] M. Kafesaki, I. Tsipa, N. Katsarakis, Th. Koschny, C. M. Soukoulis, E. N. Economou, Phys. Rev. B 2007, 75, 235114.
[16] S. Jahani, Z. Jacob, Nat. Nanotechnol. 2016, 11, 23.
[17] Y. Hao, R. Mittra, FDTD Modeling of Metamaterials: Theory and Applications, Artech House, Norwood, MA 2008.
[18] N. J. Zheludev, Science 2010, 328, 582.
[19] K. Z. Rajab, Y. Hao, D. Bao, C. G. Parini, J. Vazquez, M. Philippakis, J. Appl. Phys. 2010, 108, 054904.
[20] E. Jan, N. A. Kovota, Nano Lett. 2007, 7, 1123.
[21] Z. Wang, F. Cheng, T. Winsor, Y. Liu, Nanotechnology 2016, 27, 412001.
[22] S. Walia, C. M. Shah, P. Gutruf, H. Nili, D. R. Chowdhury, W. Withayachumnankul, M. Bhaskaran, S. Sriram, Appl. Phys. Rev. 2015, 2, 011303.
[23] S. Elsevier, Scopus Content Coverage Guide, Elsevier BV, Amsterdam 2016.
[24] W. L. Belcher, Writing Your Journal Article in 12 Weeks, SAGE Publications, London 2009.
[25] Elsevier Developer Portal, https://dev.elsevier.com/.
[26] C. W. Lee, H. J. Choi, H. Jeong, Nano Convergence 2020, 7, 3.
[27] T. Mikolov, W. T. Yih, G. Zweig, Proceedings of the 2013 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 2013, pp. 746–751.
[28] T. Bolukbasi, K. W. Chang, J. Zou, V. Saligrama, A. T. Kalai, Advances in Neural Information Processing Systems 2016, 29, 4349.
[29] R. E. Bellman, Adaptive Control Processes, Princeton University Press, Princeton, NJ 1961.
[30] A. E. Ezugwu, A. K. Shukla, M. B. Agbaje, O. N. Oyelade, A. José-García, J. O. Agushaka, Neural Computing and Applications 2021, 33, 6247.
[31] A. Likas, N. Vlassis, J. J. Verbeek, Pattern Recognit. 2003, 36, 451.
[32] A. M. Fahim, A. M. Salem, F. A. Torkey, M. Ramadan, J. Zhejiang Univ., Sci., A 2006, 7, 1626.
[33] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, J. Liu, EURASIP Journal on Wireless Communications and Networking 2021, 2021, 31.
[34] R. Nainggolan, R. Perangin-angin, E. Simarmata, A. F. Tarigan, J. Phys.: Conf. Ser. 2019, 1361, 012015.
[35] L. Cohen, Time-Frequency Analysis, Vol. 778, Prentice Hall, Hoboken, NJ 1995.
[36] G. E. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, Time Series Analysis: Forecasting and Control, John Wiley & Sons, Hoboken, NJ 2015.
[37] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, arXiv:1406.1078, 2014
[38] K. Park, Y. Choi, W. J. Choi, H. Y. Ryu, H. Kim, IEEE Access 2020, 8, 20786.
[39] M. Roondiwala, H. Patel, S. Varma, International Journal of Science and Research (IJSR) 2017, 6, 754.
[40] G. P. Zhang, Neurocomputing 2003, 50, 159.
[41] R. Adhikari, R. K. Agrawal, arXiv:1302.6613 2013.
[42] A. Bejan, S. Lorente, J. Appl. Phys. 2011, 110, 024901.
[43] M. Campani, R. A. Vaglio, Scientometrics 2015, 103, 75.
[44] W. J. Padilla, D. N. Basov, D. R. Smith, Mater. Today 2006, 9, 28.
[45] V. J. Enick, M. Torfeh, A. McClung, D. E. Jung, M. Mansouree, A. Arbabi, J. J. Watkins, ACS Photonics 2021, 8, 2400.
[46] R. C. Mitchell-Thomas, T. M. McManus, O. Quevedo-Teruel, S. A. R. Horsley, Y. Hao, Phys. Rev. Lett. 2013, 111, 213901.
[47] M. McCall, Contemp. Phys. 2013, 54, 273.
[48] S. Xiao, T. Wang, T. Liu, C. Zhou, X. Jiang, J. Zhang, J. Phys. D: Appl. Phys. 2020, 53, 503002.
[49] H. Chu, H. Zhang, Y. Zhang, R. Peng, M. Wang, Y. Hao, Y. Lai, Nat. Commun. 2021, 12, 4523.
[50] B. Wu, H. M. Tuncer, M. Naeem, B. Yang, M. T. Cole, W. I. Milne, Y. Hao, Sci. Rep. 2014, 4, 4130.
[51] A. Ihalage, Y. Hao, npj Comput. Mater. 2021, 7, 75.

Young-Ok Cha is currently a Ph.D. candidate in the School of Electronic Engineering and Computer Science at Queen Mary University of London, United Kingdom. Prior to this, she was a Research Engineer at Samsung (Mobile Display Lab) in South Korea. In her career, she has amassed more than eleven international patents, has received a number of awards and recognition, including the Best Patent award from Samsung. Her general research interests are NLP and machine learning with particular focus on material science.

Yang Hao is a Professor at Queen Mary University of London and also serves in the management team of Cambridge Graphene Centre. His research includes computational electromagnetics, metamaterials and transformation optics, functional nanomaterials for antenna, and wearable applications. He won 2015 IET AF Harvey Research Prize and is a co-recipient of BAE Chairman’s Silver Award in 2014 and the Royal Society Wolfson Research Merit Award. Prof. Hao is an elected Fellow of Royal Academy of Engineering and IEEE. Professor Yang Hao received the Ph.D. degree from the Centre for Communications Research (CCR) at the University of Bristol, UK in 1998.