Visual creativity through concept combination using quantum cognitive models

M. Ahrabi Tabriz, T. Rafiei Atani, M. Ashtiani*, and M.R. Jahed-Motlagh

Complex Systems (Chaos) Laboratory, School of Computer Engineering, Iran University of Science and Technology, Tehran, Postal Code: 16846-13114, Iran.

Received 24 November 2022; received in revised form 19 May 2023; accepted 15 August 2023

KEYWORDS
Computational creativity; Episodic memory; Concept combination; Quantum cognition; Artificial intelligence.

Abstract. Computational creativity modeling, including concept combination, enables us to foster deeper abilities of Artificial Intelligence (AI) agents. Although concept combination has been addressed in a lot of computational creativity studies, findings show incompatibility amongst empirical data of concept combination and the results of the used methods. In addition, even though recent neuroscientific studies show the crucial impact of retrieving concepts’ relations explicitly stored in episodic memory, it has been underestimated in modeling creative processes. In this paper, a quantum cognition-based approach is used to consider the context more effectively and resolve logical inconsistencies. Also, episodic memory is leveraged as the basis for the concept combination modeling process based on the created context. The result of the proposed process is a set of meaningful concepts and expressions as a combination of stimuli and related episodes, which are used to depict a visual collage as an image. The significant improvement in the quality of results in comparison with the existing methods suggests that quantum-like modeling can be considered the foundation for developing AI agents capable of creating artistic images or assisting a person during a creative process.

© 2021 Sharif University of Technology. All rights reserved.

1. Introduction

It can be said that modeling human creativity as a cognitive phenomenon in Artificial Intelligence (AI) will be a subset of thinking humanly, one of the approaches to modeling AI presented by Russell and Norvig [1]. Creative thinking enables us to come up with new and potentially valuable products [2]. Computational creativity [3,4] improves AI agents’ performance in a wide range of creative domains, such as visual artworks, music production, and inferences that lead to advancement and discovery in various sciences, such as physics [2]. For a better understanding, learned patterns from concepts and logical rules shape the knowledge that in a human counterpart has emerged as associative relations in semantic memory [5]. However, the knowledge base for a creative entity to produce an artifact is based on associative and non-associative descriptions in the explicit memory, including both semantic and episodic memories [5,6].

Creativity is a cognitive phenomenon resulting from divergent thinking that eventually finishes with convergent thinking [7–9]. However, in most modeling methods for cognitive systems, the focus is on modeling through semantic memory [10]. Semantic memory is considered the associative relationship database for concepts and how they interact with each other ontologically. Therefore, semantic memory has been recog-
nized as a cognitive component where new relationships emerge [11]. Empirical evidence shows that creativity and divergent thinking are strongly associated with individuals’ episodic memory.

The existence of two levels of explicit memory (i.e., associative and non-associative) has been pointed out in previous cognitive frameworks as well as in the Clarion cognitive architecture [12, 13], which by many researchers is considered the most suitable architecture for modeling creativity [14]. In addition, reviewing the studies on mental representations (i.e., the way that our mind represents information to be processed and transformed later on) [15, 16] show that classical models have difficulty in considering concepts as the primary way of representation and taking into account the existence of fallacies that lead to differences in the meaning of combined concepts compared to rational criteria [17, 18]. Recent quantum-like cognitive models show that taking advantage of powerful quantum mathematics can result in better models to represent both the concept combination and model the existing biases and fallacies [19, 20]. Therefore, the main effort in this research is to propose a cognitive framework based on quantum mathematics for modeling concept combinations used in creative processes. This modeling should be such that the effect of the thematic context of the stimulus concepts on the event’s context in the episodic memory can be represented by Quantum Information Retrieval (QIR). The main reason for retrieving information from episodes through a quantum version is based on three important general features [21].

The first one is the ability of such a framework to preserve the context of any observable. Second, such a framework can interpret and describe inconsistencies based on the order of observation. This phenomenon is known as fallacies in concept combination. The third and most important one is the ability to predict potential states in which two concepts are highly correlated in the context of a particular episode. The former enables us to both filter and rank the episodes based on their potentiality.

In this paper, computational creativity has been studied using concept combinations through episodic memory. In addition to using episodic memory to find relationships between concepts, retrieving information has been performed based on a QIR framework. The fundamental reason to take advantage of such an information retrieval method is that the quantum framework is capable of preserving the context of each concept within each episode. Also, axioms such as quantum interference and quantum entanglement boost the process of resolving classical inconsistencies (i.e., combinational logical fallacies) and determining the potentiality of a new concept emergence. Utilization of this method across episodic memory makes it possible to find the most correlated events with the input stimuli where the combination of stimuli concepts and the occurrence of new concepts indicate a high correlation between the two basic concepts. This, in turn, leads to quantum entanglement based on quantum cognition theories. Previous works based on classical cognitive frameworks find concepts related to a query using only one concept. In contrast, in this work, an approach for combining two concepts in a query through a cognitive framework is presented that forms the ideas as conceptual combinations and expressions to be used in the construction of an image. The whole process can be represented as shown in Figure 1.

The main contributions of this work can be summarized as follows:

1. Modeling the cognitive process of creativity through concept combination by taking episodic memory into account alongside semantic memory;
2. Utilizing QIR to capture the possibility of the user’s prompt;
3. Consideration of quantum entanglement within the context of episodic memory as a measure of new concept emergence (i.e., contextualizing triggered episodes of memory with the user’s prompt using QIR);
4. Proposing a transparent cognitive process creating a collage of images which, compared to its counterpart (i.e., DARCI explained in Section 2.2), creates the collage based on the combination of two words as input, not a single word.

The rest of this paper is organized as follows. In Section 2, the related work of this research is reviewed. Then, in Section 3, the problem statement is given, which is followed by Section 4, in which the proposed approach is introduced by explaining the general framework and its components. In Section 5, considering

![Figure 1](image-url). An overview of the process is presented in this article.
the input stimulus samples, the system outputs are analyzed. Finally, Section 6 concludes the paper.

2. Related work

In this section, cognitive modeling perspectives are discussed.

2.1. Quantum cognition and related concepts

Quantum cognitive methods integrate heuristic-based approaches [22], where a decision-maker is bound to limited rationality, and rational approaches, where basic axioms under a probabilistic theory are defined to let the decision-maker perform inferences. The key point is that classic statistics and probability are based on Bayesian probability, which is the source of conflicts. Thus, in this new method, the axioms are defined under the quantum probability framework [22].

One of the inconsistencies between the empirical data derived from human experiments and classic mathematics is the conjunction and disjunction of fallacies through the combination of concepts discovered by Osherson and Smith [23]. Then, supplementary studies were conducted by Hampton [24,25]. Based on these findings, Gabara and Aerts proposed a general quantum framework representing the semantic space of concepts and removing the barriers to describing any kind of concepts and their combinations [26]. Also, Tvensky and Kalmanen proposed a quantum-like approach for a cognitive retrieval bias named conjunction fallacy. The conjunction fallacy refers to the tendency of people to judge a conjunction of two events as more likely than one of the events occurring alone, despite it being logically impossible. Through their experiments, they found that people tend to rely on intuitive reasoning rather than logical extensional reasoning when making probability judgments [27]. Then, the document that hits the peak of the defined metric, based on the Bell inequality applied on its state vector, is considered the document that is highly correlated to the query in the minimum amount of window size. It is worth mentioning that emerging new concepts through combining two concepts are defined based on quantum interference and quantum entanglement in the existing literature [22,23,28,29].

2.2. DARCI

DARCI, or the Digital Artist Communicating Intention, is an example of a system proposed by Ventura [30] that exploits a human's semantic memory mechanism to find related concepts to a single word prompt to finally generate a collage that conveys a message. Relying on a semantic network obtained from Corpus-Based Associations (CBAs) built out of the English text of Wikipedia and Free Association Norms (FANs) used from the Edinburgh associative thesaurus [31] and the University of Florida's Word Association Norms [32], DARCI retrieves related nouns to the single word query. In addition, DARCI can also describe images by tagging appropriate adjectives to related pairs of nouns and images-this can be further investigated in the main work done by Ventura [30]. To do so, it is equipped with trained neural networks called appreciation networks that create a mapping between images and adjectives. Training of this neural network is based on human volunteers’ feedback in a way that, after showing a random image to the user, the system would ask them to describe the image with adjectives. If the user’s entered adjective has different meanings (e.g., bright can mean (1) light, (2) sunny, (3) intelligent, (4) promising) it extracts a list of different meanings available from WordNET [33] and asks the user to select the best one.

2.3. Recent studies

As synthetic image generation has been at the center of attention in recent years, many works have been performed to study the problem and propose techniques for creating images out of textual prompts. Since GANs [34] are well-known in such tasks, many early works are trained on image captioning datasets to find a solution for text-based image generation, including [35–39]. Another work is performed by Ramesh et al. [40] (known as DALL-E) by training autoregressive transformers on a series of text tokens followed by their counterpart image tokens. This work is an adaptation of VQ-VAE [41] to a text-based image generation task. Also, Razavi et al. [42] by training a multi-layer discrete autoencoder, leveraged sampling coarse-grained latent codes to be used as conditioning information later. This information will be used by sampling higher-resolution latent codes. Diffusion-based models are of high interest due to their ability to create natural images. Amongst those in the work performed by Preecehakul et al. [43], two diffusion models are used as the autoencoder's components. The first one acts as images by rendering the latent variables, and the second one generates the latent layers. Moreover, the method proposed by Rombach et al. [44] encapsulates latent layers generated by a VQGAN-like [45] autoencoder in diffusion models.

The introduction of CLIP [45,46] enabled the community to propose many different models based on it. Some of these works are based on diffusion models, which are trained either on noisy or clean images to guide text-conditional models (known as GLIDE) [47] or unconditional-text/class-conditional models [48,49]. Also, Wang et al. [50] trained an autoregressive generative model directly conditioned on CLIP image embedding, resulting in a model of text-conditional image synthesis. Likewise, in LAFITE, Zhou et al. [51] showed that a generalization of GANs can be obtained by randomly perturbed CLIP image
embedding, resulting in text-conditional image production. In addition, Sauer et al. [52] proposed StyleGAN-T, which changes their baseline architecture of class-conditional ImageNet synthesis [53] to a text-conditional model. This has been performed by embedding text prompts using a pre-trained CLIP ViT-L/14 text encoder. Tao et al. [54] proposed generative adversarial CLIPs, known as GALIP, which uses a pre-trained CLIP model in both the discriminator and generator.

The parametrical summarization of the reviewed approaches is shown in Table 1. The process introduced by each method should be interpretable cognitively-assessed by **cognitively transparent**: their learning process is such that they can expand their knowledge and have access to the Internet or not represented by the Dynamic/Online and Static/Offline parameters in the table, in addition to whether they need the supervision of an external expert or not-supervised versus unsupervised learning. As we mentioned before, one crucial factor in combining two concepts is taking the potentiality of inconsistencies for granted. These inconsistencies are resolved by preserving the context and asymmetrical features of the meaning between two concepts, and it is considered the **preserving asymmetrical meaning features** factor. Some of these methods provide only a description of concepts’ representations and/or their combinations, while others can generate new concepts-known as **generative methods**.

3. Problem statement

Considering the introduction section and what has been reviewed in the related work, the problem targeted in this article can be defined as follows:

“**How a computational model can be designed in a way that by getting two words as stimuli, a combination of them can be created in a way that new concepts can emerge within their context and finally an image could be made based on the emerged concepts?**”

We have thoroughly discussed that concept combination is mostly performed by creating conceptual embedding from the perspective of semantic memory in existing approaches. Recent neuroscientific studies demonstrate the indisputable role of episodic memory in combining concepts and generating new ones in the human mind. Therefore, this work proposed an approach to make a bridge between both episodic and semantic memory based on quantum cognitive models. Also, to the best of the authors’ knowledge, most AI-based studies have no major resolution to propose a cognitively transparent image generator. While the final artifact of these works could be far different from what is proposed in this article, the problem we insisted on resolving here is to propose an approach that conforms to the cognitive theories of mind, making it cognitively understandable.

4. The proposed approach

The proposed approach is based on episodic memory and QIR from the perspective of quantum cognitive modeling to retrieve concepts related to the stimulus by considering the context of events where two concepts can be contextually combined.

4.1. The general framework

Table 2 and Figure 2 show an overview of the perspectives and steps used in the proposed approach. Each step is a cognitive mapping between an agent and the factors forming the basis of emergent creativity. The first part is divergent thinking, and the second part is convergent/rational thinking. Amongst the different modeling approaches for semantic memory, retrieving information from episodic memory is used. The reason for such selection, as mentioned in the introduction, is that neuroscientific studies on divergent thinking, the emergence of creativity, and finding relations among concepts show that retrieving information from episodic memory profoundly impacts them. Algorithm 1 represents the creation of a knowledge-base from the perspective of the episodic memory. Also, using prevalent modeling of semantic memory, such as Bayesian modeling, alongside episodic memory
<table>
<thead>
<tr>
<th>Step</th>
<th>Perspective</th>
<th>Information representation</th>
<th>Component</th>
<th>Information retrieval approach</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Verbal</td>
<td>Visual</td>
<td>Episode/event including concepts</td>
<td>Obtaining co-occurrence matrix using HAL</td>
</tr>
<tr>
<td>1</td>
<td>Episodic memory</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Bayesian modelling</td>
<td>✓</td>
<td>x</td>
<td>Including images and their mapping to verbal concepts. Each image is a context for concepts</td>
<td>Documents as episodes (including text-based documents and verbal mappings extracted from images)</td>
</tr>
<tr>
<td>3</td>
<td>Two-layer thinking</td>
<td>✓</td>
<td>x</td>
<td></td>
<td>Classic statistics and probability rules</td>
</tr>
<tr>
<td>4</td>
<td>Rational thinking</td>
<td>✓</td>
<td></td>
<td>Retrieved episodes</td>
<td>Using lingual rules and external references</td>
</tr>
</tbody>
</table>

**Table 2.** The abstract view of the perspectives, components, and outputs of the proposed approach.

**Input:** Episodes or Events of the environment including documents, images, videos, audio; all as verbal information i.e. concepts as words.

**Output:** A verbal memory of the environment’s episodes

```plaintext
1 episodicMemory ← array();
2 while Observing the Environment do
3   if Type(scene) = Episode then
4     scene ← ExtractVerbalConceptsFromNonVerbalScene(scene)
5   episodicMemory ← episodicMemory.insert(scene)
```

**Algorithm 1.** AI agent’s knowledge base formation from the perspective of episodic memory with an emphasis on verbal episodes.

can be a richer model to find novel relations among concepts. In previous methods of combining concepts using quantum cognition, the substantial criterion in the emergence of new combinations of two concepts is based on quantum entanglement and quantum interference. Also, there are QIR methods that evaluate the existence ratio of two concepts in the context of a document. In other words, instead of mapping the existence of a word within a document to Boolean logic, a ratio is calculated to measure how related the word is to the given document. Thus, based on these works in the field of quantum cognition, we can propose our approach as follows:

1. Retrieving information from episodic memory focuses on finding $K$ episodes that query concepts that have been entangled within their context;
2. Finding topic-based similarity between the selected $K$ episodes and query concepts;
3. Selecting $N$ episodes satisfying the criterion of both
Bell inequality and topic-based similarity defined as a voting system;

4. Using the criteria presented by Aerts and Gabora [29] to distinguish whether a referent of the combination of two concepts can introduce a new concept itself (i.e., being inconsistencies within their logical combination). Then, the exemplars that are present in the conceptual context of the combination of the query’s concepts through an episode are included. The frequency of these exemplars is added to the candidate extracted combinations score in the next level;

5. Converging the ideas within the context of episodes that have been retrieved before, according to the perspective of convergent or rational thinking. Episodes are narrowed down using keyword extraction methods such as Rapid Keyword Extraction (RAKE) [55]. Then, each extracted word or expression is transformed into a word vector obtained by FastText trained model on English text on the Internet to evaluate the cosine distance of each candidate expression vector and the stimulus. The combination of the obtained cosine distance and the score of the expression reported by RAKE is used to select $M$ expressions and/or concepts with the least similarity, meaning these keywords have semantic relations with stimulus but cannot be easily discerned.

4.2. Quantum episode retrieval perspective

Since the proposed method is mainly based on accessing semantic memory through episodic memory, any information in the format of the text, image, or sound which have an appropriate mapping to a verbal description of concepts can be considered events or episodes that an agent, by accessing them, realize the context of a query stimulus and extract the related concepts in the identical context within it. Having said that, concepts are the basis of mental representations, so being exposed to an event, an agent extracts concepts that are describable using natural language as a verbal representation. This process corresponds to collecting new information in episodic memory—which is performed through Algorithm 2. QIR, proposed by Barros et al. [56], is used to filter the episodes related to the stimuli. The first reason behind using QIR is that, in the past decade, quantum cognition
frameworks have made a breakthrough in describing unexpected mental phenomena in comparison to the classic approaches. Episode filtration is achieved by (1) transforming each episode to the episode matrix, including a co-occurrence matrix obtained from the HAL method and its transpose, (2) extracting stimuli vectors from the episode matrix and obtaining episode state vector from the accumulation of the episode matrix’s rows, (3) calculating the correlation between the episode state vector and stimuli vectors according to the Bell test proposed by Barros et al. [56], and finally (4) selecting K episodes with the highest value corresponding to the query entanglement degree in the context of the episode state vector (i.e., $S_{query}$).

The output of this step includes K number of episodes, attentional window sizes passed to the HAL algorithm for each episode, and the amount of $S_{query}$.

4.3. Topic model-based semantic memory perspective

This algorithm learns the relationship between random variables generating words under a particular topic and uses the episodes stored in episodic memory to estimate the algorithm parameters, including $\phi$, $\alpha$, and $\beta$. Also, Bayes’ rule is bidirectional while maintaining asymmetrical lingual features. So, it can readily estimate the topic of a group of words [57].

Therefore, according to Algorithm 3, in the first step, bypassing any episode of the episodic memory or a new stimulus, including a set of verbal concepts (i.e., words), to the topic model and calculating the marginal distribution of the joint distribution of topic mixture, $\theta$, a set of $N$ topics $z$, and a set of $N$ words $w$, the joint probability of a document’s words and the topics can be calculated [57] (topic model of episodic memory). Then, by extracting the topic of the K episodes selected in the previous section, cosine similarity between any episode and stimulus topic vectors is computed. As a result, each episode’s score includes $S_{query}$ and episode-stimulus topic-based similarity.

4.4. Voting-based episode selection

In the previous steps, the related episodes for the primary stimuli have been selected. In this stage, a voting system is proposed to integrate the opinion of two previous components such that it combines the first stage criterion based on the Bell test and the second stage criterion based on the similarity of each episode’s topic to the stimulus topics. Then, filter the top N episodes with maximum similarity as candidate episodes. Considering two weighting parameters, and
input: \( K \) Episodes or Events in which \( \text{Stimulus}_1 \) and \( \text{Stimulus}_2 \) have been entangled in their context

output: \( K \) ranked episodes based on the topic similarity to the Stimuli

1. \( \text{topicModel} \leftarrow \text{loadTrainedModelOnEpisodicMemory}() \);
2. \( \text{index} \leftarrow 0; \)
3. while \( \text{index} < K \) do
   4. \( \text{episode} \leftarrow \text{retrievedEpisodes}[\text{index}]; \)
   5. \( \text{episodeStimuliTopicBasedSimilarity} \leftarrow \text{topicModel.getTopicBasedSimilarity}(\text{Stimuli}, \text{episode}); \)
   6. \( \text{retrievedEpisodes} \leftarrow \text{retrievedEpisodes.insert}(\text{index}, \text{episodeStimuliTopicBasedSimilarity}); \)
   7. \( \text{index} += 1; \)

Algorithm 3. Finding topic-based similarity between text-based episodes and the stimuli (user’s prompt).

input: \( K \) Episodes or Events with \( S_{\text{query}} \) and Topic Based Similarity according to Stimuli, \( \alpha, \beta, N \)

output: \( N \) ranked episodes based on weighted voting of previous scores

1. \( \text{retrievedEpisodes} \leftarrow S_{\text{query}} \text{MinMaxScaler}(\text{retrievedEpisodes}); \)
2. \( \text{index} \leftarrow 0; \)
3. while \( \text{index} < K \) do
   4. \( \text{episode, episode } S_{\text{query}}, \)
   5. \( \text{episodeTopicBasedSimilarity} \leftarrow \text{retrievedEpisodes}[\text{index}]; \)
   6. \( \text{episodeScore} \leftarrow \alpha * \text{episode}_{s_{\text{query}}} + \beta * \text{episodeTopicBasedSimilarity}; \)
   7. \( \text{retrievedEpisodes} \leftarrow \text{retrievedEpisodes.insert}(\text{index}, \text{episodeScore}); \)
   8. \( \text{index} += 1; \)
9. if \( N > K \) then
10. \( N \leftarrow K; \)

Algorithm 4. Selecting \( N \) episodes based on weighted voting between retrieved episodes of QIR and the selected ones from topic modeling.

\( \beta \), and the number of candidate episodes, voting is defined as below:

\[
\arg \max_{\alpha, \beta} \alpha \times QIR_{eScore} + \beta \times \text{DA}_{eScore}
\]

\( \alpha + \beta = 1. \) (1)

According to Eq. (1) and Algorithm 4, the top \( N \) documents with the maximum integrated scores are selected.

4.5. Two-layer thinking

In this phase, using the Hampton dataset reported for some conjunction and disjunction combinations and WordNET as a representation of semantic memory, according to the criteria presented by Aerts and Gabora [29], verbal concepts in each episode have been analyzed. All the exemplars of the Hampton data not presented in an episode are excluded, while those that have happened within the episode are linked to at least one concept in the Synset of the stimulus (obtained from WordNET) and are extracted. Thus, if there is a connection between an exemplar and the stimulus, the exemplar with its frequency in the retrieved episodes will be stored to influence the score of the ultimate candidate combinations found in the next step. The corresponding algorithm is shown as Algorithm 5.

4.6. Convergent/rational thinking

According to the creativity theories reviewed in Sections 1 and 2.2, creativity has two significant phases: (1) Divergent thinking and (2) Convergent thinking. Till now, an agent’s episodic memory is searched to find the episodes that lay the foundations for the stimulus’s two concepts.

In fact, in this step, an agent, without any particular purposes, delves into the events to find the ones
that potentially create a context for the entanglement of stimulus concepts. All of these can be grouped as divergent thinking. Convergent thinking is modeled by filtration and extracting expressions and concepts lying down in the context of retrieved episodes. Finding expressions and concepts in the conceptual space of the combination of two words in the stimulus is performed with the RAKE algorithm introduced in [55].

The score of each candidate keyword is calculated based on the following criteria: (1) the score that the RAKE algorithm has computed based on the appearance of each keyword within the document, (2) the score of each retrieved episode. Each candidate keyword appears in an episode with a specific score calculated before. So, this score impacts the importance of a keyword, (3) the frequency of exemplars that appeared in the candidate keyword. Each episode has a list of exemplars found in the previous phase. If there is an intersection between the exemplars of this list and the candidate keyword, the exemplar’s frequency is added to the candidate keyword score. Since name entities do not play a substantial role in keyword extraction, they have been removed from the episodes while passing to the RAKE algorithm. To normalize the scores, we have scaled them using the min-max scaling approach.

Now, each candidate keyword’s similarity and commonality compared to the stimulus are evaluated in the context of the general knowledge of an omniscient agent. It is the trained model of FastText on English webcrawl and Wikipedia [58]. Then, by transforming each candidate keyword and the stimulus to the corresponding vectors in the general semantic space of FastText, the cosine distance between each candidate keyword vector and the stimulus vector is calculated. Finally, a measure is defined as a weighted combination of the general distance and the keyword score itself based on Eq. (2). Then, a set of M candidate keywords with the minimum similarity score based on Eq. (3) are returned as the output, including expressions and concepts representing the combined space of two concepts in the stimulus. Algorithm 6 represents the explained rational thinking process.

\[
gDistance (Keyword, Stimulus) = 1 - \frac{\|Keyword\_Stimulus\|_2}{\|\|Keyword\|_2\|\|Stimulus\|_2}. \tag{2}
\]

\[
similarity (Keyword, Stimulus) = \lambda \ gDistance (Keyword, Stimulus) + \eta \ keyword\_Score \quad \lambda + \eta = 1. \tag{3}
\]

### 4.7. Visualizer component

Drawing an image or visual collage of the set of retrieved concepts and expressions is performed through a mapping between each entry of the retrieved set to an image. To do so, in the simplest way, a query consisting of each entry and the concepts of the stimulus is passed to the Google Custom Search JSON API [59] through the Python library, Google-Images-Search [60] and some random images with the corresponding captions to the query are returned. Then, the retrieved images are resized using

\[
\text{Algorithm 5. Using two-layered thinking and prioritizing episodes having exemplars of the Hampton dataset.}
\]
the OpenCV-python library [6] to make all of them the same size to be representable in a visual collage. Finally, the images' matrices are stacked in the shape of (rows and columns) using the numpy library, and the final matrix is exported as an image (The source code for the developed software is available online at https://github.com/mozhiani/VisualRECOMBINANT). In this paper, the elaboration phase of a creative process to produce the visual output is not thoroughly developed. Therefore, the visual collage here is created to manifest the output of the extracted set of concepts related to the combinational semantic space of the stimulus as a prototype on a small scale, and it is clear that the last phase of this process (i.e., the elaboration stage), should be developed in supplementary studies.

5. Evaluations and comparisons

One criterion to evaluate the presented approach is episodic memory's effectiveness in finding remote creative relations between concepts. In this section, we will perform various evaluations to demonstrate the effectiveness of the proposed approach compared to other existing approaches in terms of topic modeling as well as effective concept combinations.

5.1. Dataset (episodic memory)

Episodic memory is considered to be verbal-conceptual episodes represented as text-based documents. Any available mapping of non-textual conceptual entities, including images, videos, and audio, to text-based representations would be easily developed in this framework.

5.2. QIR assessment

This component has been developed according to the method proposed by Barros et al. [56]. Our implementation has been assessed using the four articles on English Wikipedia, including Reagan Administration Scandals, Ronald Reagan, Iran Contra Affair, and Iran. To better compare the implementation results with the original method, these four articles have been retrieved through Archive.org within a time close to the original article's publication in July 2013.

The simulation results are shown in Figure 3. This figure demonstrates that by searching two words, Reagan and Iran, amongst four presented articles, the relationship between the query concepts has been maximized within the Iran Contra Affair document. The result of the Bell test on the two query concepts hits the peak sooner (i.e., with a smaller attentional window in each document) in the semantic and contextual context of the Iran Contra Affair document, indicating a high correlation between these two. It should be noted that differences in the content of documents extracted from Wikipedia (due to the exact time when Wikipedia articles have been accessed) used in the original article and this paper evaluation have caused slight changes in
the growth ratio of the chart compared to the original article. However, the final result of both evaluations is the same.

5.3. Topic modeling assessment

Topic models are evaluated using different methods. However, the two mainly used approaches are based on perplexity and topic coherence.

Comparing these two assessment methods through the literature recommends the use of topic coherence instead of perplexity-based assessment. Studies on the latter’s comprehensibility indicate clear distinctions to the point that the better a model performs based on this metric, the less interpretable the topics are based on human judgments. Therefore, topic coherence being plausibly compatible with human interpretations is the basis of assessing the topic model trained on the articles extracted from Scientific American. Multiple metrics defined to assess topic coherence according to [62] consider the $C_v$ metric as the best assessment metric. Thus, the trained topic model on episodic memory is evaluated using the $C_v$ metric. Topic coherence evaluation of topic models by seeding the number of expected topics in the range of [2, 40] is shown in Figure 4. The best model has been chosen based on the maximum topic coherence in various training scenarios. This figure shows that among models trained on the articles of the Scientific American website, the model with the numbers 822 and 34 extracted topics is the best model to use.

5.4. Process assessment

Three queries have been considered to evaluate the process and the final output. The first stimulus constituting concepts ‘war’ and ‘kid’ is chosen to establish the basis of comparison with DARCI’s output by the stimulus ‘war.’ The two others are randomly generated by the authors’ curiosity as potential queries. Text-based episodes with the maximum score calculated based on Eq. (1) in the context of each query concept are retrieved as given in Table 3. Then, the parameter $\lambda$ in Eq. (3) is set to be $\lambda \in [0, 1]$. Extracted combinations based on Eq. (3) indicate that the less $\lambda$ is, the more remote selected concepts and expressions are in the context of the two concepts in the stimulus. This can be interpreted by the Remote Associative Test (RAT) [11]. These combinations are not limited to only the query concepts, but some represent the entangled space of query concepts. This can be justified by decreasing the impact of general knowledge using the $\lambda$ parameter in comparison to the importance of the agent’s ideas, the parameter $\eta$, which is equivalent to $\eta = 1 - \lambda$ based on Eq. (3), final concepts and expressions are mainly selected based on the agent’s specialist knowledge through quantum episode retrieval and topic modeling.

<table>
<thead>
<tr>
<th>Query 1</th>
<th>Query 2</th>
<th>Query 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>War, kid</td>
<td>Smartphone, mars</td>
<td>Microbe, travel</td>
</tr>
</tbody>
</table>

*Table 3. Sample queries to evaluate the proposed process.*

![Figure 3. Assessing quantum information retrieval based on the work of Barros et al. [56].](image1)

![Figure 4. Topic modeling assessment based on the $C_v$ metric.](image2)
<table>
<thead>
<tr>
<th>Query 1</th>
<th>λ</th>
<th>Extracted concept/expression</th>
<th>Score</th>
<th>Query 1</th>
<th>λ</th>
<th>Extracted concept/expression</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>War, kid</td>
<td>0.1</td>
<td>Kid</td>
<td>1.007755</td>
<td>War, kid</td>
<td>0.1</td>
<td>Tween</td>
<td>1.722463</td>
</tr>
<tr>
<td>War, kid</td>
<td>0.1</td>
<td>Young kid</td>
<td>1.175520</td>
<td>War, kid</td>
<td>0.1</td>
<td>Young player</td>
<td>1.746632</td>
</tr>
<tr>
<td>War, kid</td>
<td>0.1</td>
<td>War plane</td>
<td>1.288060</td>
<td>War, kid</td>
<td>0.1</td>
<td>Whose mom</td>
<td>1.778314</td>
</tr>
<tr>
<td>War, kid</td>
<td>0.1</td>
<td>War shortage</td>
<td>1.328281</td>
<td>War, kid</td>
<td>0.1</td>
<td>Grandparent</td>
<td>1.780024</td>
</tr>
<tr>
<td>War, kid</td>
<td>0.1</td>
<td>Kid protest</td>
<td>1.366631</td>
<td>War, kid</td>
<td>0.1</td>
<td>Young</td>
<td>1.780975</td>
</tr>
<tr>
<td>War, kid</td>
<td>0.1</td>
<td>Child</td>
<td>1.521401</td>
<td>War, kid</td>
<td>0.1</td>
<td>Teen perception</td>
<td>1.791293</td>
</tr>
<tr>
<td>War, kid</td>
<td>0.1</td>
<td>Soldier</td>
<td>1.535120</td>
<td>War, kid</td>
<td>0.1</td>
<td>Suicidal thought</td>
<td>1.792821</td>
</tr>
<tr>
<td>War, kid</td>
<td>0.1</td>
<td>Dad</td>
<td>1.611621</td>
<td>War, kid</td>
<td>0.1</td>
<td>Young people</td>
<td>1.793872</td>
</tr>
<tr>
<td>War, kid</td>
<td>0.1</td>
<td>Fight</td>
<td>1.674464</td>
<td>War, kid</td>
<td>0.1</td>
<td>Old animal</td>
<td>1.801590</td>
</tr>
<tr>
<td>War, kid</td>
<td>0.1</td>
<td>Paint damage child</td>
<td>1.705422</td>
<td>War, kid</td>
<td>0.1</td>
<td>Deal nation</td>
<td>1.843097</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 2</th>
<th>λ</th>
<th>Extracted concept/expression</th>
<th>Score</th>
<th>Query 2</th>
<th>λ</th>
<th>Extracted concept/expression</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Orion spacecraft</td>
<td>1.800484</td>
<td>Smartphone, Mars</td>
<td>0.1</td>
<td>Technology</td>
<td>2.000808</td>
</tr>
<tr>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Rocky planet</td>
<td>1.818617</td>
<td>Smartphone, Mars</td>
<td>0.1</td>
<td>Astronaut</td>
<td>2.020534</td>
</tr>
<tr>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Alien planet</td>
<td>1.820173</td>
<td>Smartphone, Mars</td>
<td>0.1</td>
<td>Exoplanet</td>
<td>2.039023</td>
</tr>
<tr>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Planet</td>
<td>1.848448</td>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Mankind</td>
<td>2.048402</td>
</tr>
<tr>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Alien world</td>
<td>1.910573</td>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Powerful new planet hunter</td>
<td>2.074385</td>
</tr>
<tr>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Rocket</td>
<td>1.942105</td>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Launch</td>
<td>2.104849</td>
</tr>
<tr>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Orbit space station</td>
<td>1.969961</td>
<td>Smartphone, Mars</td>
<td>0.1</td>
<td>Commercial space company</td>
<td>2.113670</td>
</tr>
<tr>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Small planet</td>
<td>1.977437</td>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Tiny brightness dip</td>
<td>2.119300</td>
</tr>
<tr>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Lunar neighbor</td>
<td>1.999410</td>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Unit telescope</td>
<td>2.131721</td>
</tr>
<tr>
<td>Smartphone, mars</td>
<td>0.1</td>
<td>Human space exploration</td>
<td>1.990820</td>
<td>Smartphone, Mars</td>
<td>0.1</td>
<td>Space</td>
<td>2.138118</td>
</tr>
</tbody>
</table>

5.5. Interpreting Query 1: War and kid

The output results for the first query, 'war' and 'kid', are shown in Table 4. Through descriptive-verbal assessment, the extracted concepts and expressions are analyzed. Some selected keywords have a direct cut relationship with either concept in the query, such as 'war plane' and 'soldier' connected to the concept 'war', 'kid', and 'child' connected to the concept 'kid'. Some other expressions and concepts have a remote association with the query. For instance, the word 'dad' appears in the list of selected keywords, while the word 'mother' does not appear solely. However, the expression 'whose mom' has been selected. To provide an interpretation of these selections, it would be comprehensible that, firstly, the semantic meaning of punishment is further in the relationship of father and child instead of mother and child [51]. Also, 'whose mom', a part of 'Whose mom is she?' is reminiscent of siblings' rivalries and their fights over their mother by saying, "She is the mother of only one of them." Also, fathers go to war more than mothers. Thus, it is clear that the war's adverse effect on kids could be understandable for losing their fathers. Children damaged by paint in the context of 'war' and 'kid' are reminiscent of games like paintball or color war. Statistical reports show that suicide among 'children' and 'young people' is nearly 10% on a global scale (26 countries involved in this study), and the ratio is rising gradually. Also, serious 'suicidal thoughts' among children and young people reached 17% [63]. The 'tween' age is when the children usually experience their most rebellious period as well as an internal war with the environment. On the other hand, younger players are more likely to fight with their teammates and rivals at a younger age. In the meantime, with further investigations of the 'deal nation', the authors realized that "I Deal" is a group that designs creative skills and activities for war-torn children to bring them back to everyday pre-war life. The only example in which the authors have not found any connection between 'kid' and 'war' is the expression 'old animal'.

5.6. Remaining queries

The results of the two remaining queries, 'Smartphone, mars' and 'Microbe, Travel' are depicted in Tables 5
<table>
<thead>
<tr>
<th>Query 3</th>
<th>( \lambda )</th>
<th>Extracted concept/expression</th>
<th>Score</th>
<th>Query 3</th>
<th>( \lambda )</th>
<th>Extracted concept/expression</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Organism</td>
<td>1.616344</td>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Human respiratory tract</td>
<td>1.800832</td>
</tr>
<tr>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Live organism</td>
<td>1.644703</td>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Cell biology</td>
<td>1.807604</td>
</tr>
<tr>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Viable pathogen</td>
<td>1.702779</td>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Flu receptor</td>
<td>1.809410</td>
</tr>
<tr>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Animal health expert</td>
<td>1.749415</td>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Flu strain</td>
<td>1.817697</td>
</tr>
<tr>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Insect</td>
<td>1.723677</td>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Flu state</td>
<td>1.819802</td>
</tr>
<tr>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Origin flu</td>
<td>1.745103</td>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Infected bird</td>
<td>1.821500</td>
</tr>
<tr>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Origin virus</td>
<td>1.750249</td>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Spread infection</td>
<td>1.828541</td>
</tr>
<tr>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Primitive organism</td>
<td>1.758923</td>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Live creature</td>
<td>1.841208</td>
</tr>
<tr>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Air particle</td>
<td>1.761411</td>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Lethal virus</td>
<td>1.841929</td>
</tr>
<tr>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Seasonal virus</td>
<td>1.788260</td>
<td>Microbe, travel</td>
<td>0.1</td>
<td>Planet</td>
<td>1.849845</td>
</tr>
</tbody>
</table>

Table 7. Comparing a creative agent’s characteristics between the proposed method and DARCI.

<table>
<thead>
<tr>
<th>Creative agent characteristics inspired by CLARION</th>
<th>DARCI</th>
<th>The proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background knowledge</td>
<td>Semantic network/symbolic noun representations (The noun project)</td>
<td>Episodic memory/online search on the Internet is applicable</td>
</tr>
<tr>
<td>Learning</td>
<td>Constructing appreciation neural network and semantic network</td>
<td>Constructing topic model based on background knowledge</td>
</tr>
<tr>
<td>Intention</td>
<td>Aiming to communicate one concept through an artistic artifact</td>
<td>Aiming to find concepts and expressions in the context of two concepts, as the basis of creating an image</td>
</tr>
<tr>
<td>Conceptualization</td>
<td>Using semantic memory retrieval</td>
<td>Using quantum information retrieval from episodic memory</td>
</tr>
<tr>
<td>Aesthetic function</td>
<td>Using fitness function for style and adjustment</td>
<td>Maximizing the extracted combinations score</td>
</tr>
<tr>
<td>Artifact</td>
<td>A collage based on an evolutionary mechanism</td>
<td>A collage of extracted concepts and expressions</td>
</tr>
</tbody>
</table>

and 6 correspondingly. In the same way as the previous query’s result, these results are interpretable.

5.7. Comparison with DARCI

Newer versions of DARCI have tried to expand their semantic memory through a skip-gram model to extract related adjectives to a concept. However, this attempt is only made to impact the images’ adjustment style, and the extracted concepts related to a stimulus, such as ‘war,’ are only retrieved through a semantic network representing semantic memory. The basis of the comparison between the proposed approach and DARCI is the characteristics of a creative agent, as shown in Figure 5. The results of the comparison itself are shown in Table 7. The output collages of both the proposed method and DARCI are shown in Figures 6 and 7. A picture is selected to produce a collage based on the retrieved concepts and expressions by searching
the extracted combinations alongside the concepts in the stimulus.

6. Conclusion

In this article, a method for extracting meaningful concepts and expressions manifesting the creative combination of two query words is proposed. The difference between the current work and the methods of producing images based on text is that these methods operate essentially without considering the cognitive process and rely on new tools that are mainly based on deep learning. In contrast, in the current method, the cognitive process of creativity can be traced and analyzed. One topic that seems to increase the quality of the extracted expressions and concepts is applying...
human feedback to the learning method. It is possible to assess the quality of the extracted verbal-conceptual combinations using human volunteer feedback, resulting in changes in the search space to find new combinations. Also, it may be possible that, as Digital Artist Communicating Intention (DARI) does, we ask human volunteers to assign different verbal adjectives to each set of retrieved combinations concerning a specific stimulus by assigning a degree of membership. It also seems that the utilization of image-generating models based on generative adversarial networks or deep convolutional networks can lead to the generation of an artistic image based on the corresponding images to the selected expressions and concepts. In this regard, some works have been performed that are worth mentioning. These works are based on either the production of images through textual descriptions or the creation of an image through concept combination using a deep convolutional network. By receiving two input images and being inspired by the theory of conceptual blending, they consider one image as impressionable and the other as influential. For example, one strand of the latter transfers the influential image style to the impressionable image, resulting in a combination of two images. These examples can be a source of inspiration to develop the visualizer unit using expressions and concepts derived from the combination of stimulus concepts. The authors have also recently found two interesting studies. One is based on computational and data-driven methods, taking advantage of combining concepts through designing and creating images. The other is a simple decoder-only transformer, which receives texts and images in a single stream and models them autoregressively, claiming to compete with other methods based on generative adversarial networks.

Compliance with ethical standards

This study has received no funding from any organization.

Conflict of interest

All of the authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References


**Biographies**

Mozhdeh Ahrabi Tabrizi received both her BSc degree in Computer Engineering-Software and her MSc degree in Computer Engineering-Artificial Intelligence and Robotics from Iran University of Science and Technology, Tehran, Iran, in 2016 and 2021, respectively. Her major interests include Human-Centered and Cognitive Sciences.

Tayebe Rafiei Atani received her BSc and MSc degree in computer engineering from Iran University of Science and Technology, Tehran, Iran in 2012 and 2014 respectively. She is currently a PhD candidate at Iran University of Science and Technology. She is also an instructor at Iran University of Science and Technology. Her major interests include Creativity, Philosophy, and Cognitive Science.

Mehrdad Ashtiani received his BSc degree in software engineering from Iran University of Science and Technology (2009), Tehran, Iran, his MSc degree in software engineering from Iran University of Science and Technology (2011), Tehran, Iran, and finally his PhD degree from Iran University of Science and Technology (2015), Tehran, Iran. His main research interests include trust modeling and the applications of uncertainty modeling to the domain of computer science.

Mohammad Reza Jahed-Motlagh received his BSc in Electronic Engineering (1978) from the Sharif University of Technology, Tehran, Iran, and both his MSc and PhD in Control Engineering (1986 and 1990, respectively) from the University of Bradford, Bradford, England. His research interests include modeling and evaluation of complex systems and cognitive science. Dr. Jahed-Motlagh is currently a Full Professor at the School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.