Automatic Interpretation of Visual Metaphor For Creative Ideation

Maryam Eslami[1]

A Report of the Thesis Submitted in Partial Fulfillment of the Requirement for the Degree of Master of Science (Doctor of Philosophy) in Artificial Intelligence at the Iran University of Science and Technology (IUST).

Abstract

With the help of Machine Learning tools and algorithms, a Computational Creativity model is developed to generate creative Visual Metaphor ideas. In this paper, with the study Combination Creativity, we tried to interpret visual metaphor to train a model which can generate innovative ideas for visual metaphor. The model generates an idea for a given context based on Case-based Reasoning. By processing similar ideas with the help of Divergent Thinking, the model tries to find new connections between different concepts and generates some concept candidates. Conceptual networks and similar words and concepts based on their's embedding vectors are part of the divergent thinking module. Then a generative model generates creative ideas for each given candidate. A rating survey and a Turing test are provided to evaluate the generated ideas. The Turing test and a rating survey demonstrate that bot-generated ideas can compete with human-generated ideas.

Keywords: **Computational Creativity - Visual Metaphor - Conceptual Blending - Sentence Generation**

1 Introduction

While machines communicate with humans, they need to understand complicated products produced by human creativity. Creativity is the skill of having novel ideas in problem-solving and producing surprising products. Human creativity appears in problem-solving, music, visual art, poetry, movies, games, and any other area, and we can use computational creativity in all the mentioned fields. We need computers to be used in creative content production and understanding. One use case of these creative models is the advertisement industry, where we face visual metaphors. We tried to design a computational creativity model which generates ideas based on interpreting visual metaphors. Generating creative ideas for a given context is modeled, including three modules. The first module is case-based reasoning, where there is no grand truth, and the model provides a solution for a given problem based on similar issues and their solutions [1, 2, 3]. The second module is a divergent thinking module [4]. To understand the necessity of this module, we need to see the relationship between knowledge and creativity. The relationship between knowledge and creativity shows that depending upon the thinking style. The knowledge extent may sometimes adversely affect creativity [5]. We need to challenge conventional thinking to achieve creative thinking. Divergent thinking is a style of thinking which is used to generate ideas and lead us to creative thinking. Divergent thinking explores all the possibilities and makes unexpected connections between different

[1] Computer Engineering, Iran University of Science and Technology, Tehran, Iran
maryam_eslami@comp.iust.ac.ir

concepts. When we use our imagination for creative thinking, our thinking style is divergent [4]. Convergent thinking is the opposite of divergent thinking using facts and logical steps to converge to one or a set of correct solutions [4]. The third module is a generative language model [6], which generates a sentence for a given context. In this paper, we will address the following issues. In the related works section, we study visual metaphors and visual blending in artificial intelligence. Projects on interpreting visual metaphors and visual blends are mentioned. Visual blends are divided into categories based on different aspects, e.g., if they include conceptual level or just visual level, if they are based on photorealistic or not, and if they are assistant or not. Then in the Dataset section, we explore the data type of each dataset's column in detail and see some examples. Furthermore, the input and output of the model are mentioned. In the Modeling section, the architecture of the model is introduced. The model's architecture includes three modules: case-based reasoning, Divergent thinking, and Idea generation modules. We go through the details in each subsection. After generating some creative ideas, some metrics for product evaluation are studied; and some random bot-generated ideas based on a Turing test and a 5-star rating survey are evaluated. In the end, a conclusion on problem-solving is provided.

	[I should buy a Toyota Because they are tough inside , i should want to see this because i like the inside out, BUY THIS CAR BECAUSE ITS TOUGH, I should drive Toyota because it is reliable., I should drive Toyota Because they have spacious interiors.]	[Object with Missing Part: The objects which have missing parts are **bus**]
	[I should run Badwater Because it is more challenging than a marathon., I should take part in the Badwater Ultramarathon Because it's a particularly challenging course that only a few select people would be able to endure., I should attend the Badwater marathon Because it will be intense , I should run the Badwater marathon Because Oprah can run marathons., I should run the Badwater marathon Because it's easier than I assume.]	[Texture Replacement 1: The object which has a new texture is **shoeprint** and the new texture is **explanation of the badwater race**, Texture Replacement 2: The object which has a new texture is **shoe** and the objects that have created the texture are **words**]
	[I should buy Absolut Because I will be in paradise., I should drink Absolute Because it will be like a relaxing vacation., i should use absolut paradise because it gives me the feeling of being on an island]	[Texture Replacement 1: The object which has a new texture is **Bottle** and the new texture is **Island**, Texture Replacement 1: The object which has a new texture is **vodka bottle** and the new texture is **trees**]

Figure 1 A piece of the dataset; each row includes a picture of the visual blend, Q&A, and atypical annotation.

2 Related works

Computational Creativity is still a young interdisciplinary field, but as seen from the increasing number of publications, it is a growing field. When we look at the research and projects related to the visual metaphor field, many cases relate to our work in different aspects. The following related words can be divided into three groups. The first group is associated with the research around understanding and knowledge extraction from visual metaphors; For example, Automatic Understanding of Image and Video Advertisements [7] is one of the most motivations of this research. Some models are proposed [8, 9] for knowledge extracting and understanding advertisements based on image-text embedding.

The second group includes research related to visual metaphor and visual blend. A visual metaphor is a kind of visual blend. In the process of visual blending, there are at least two levels; the first level is the conceptual level, where we are looking for a reason for our blending; a cause should select the input spaces. The Conceptual shift project [10] can be a model based on the conceptual level because the model searches for a reason for each blend, which is the overall visual similarity between input spaces. The Boat-House visual blending experiment [11] tries to make a meaningful visual image for a given house and boat. The second level is a visual level that focuses on making visual blends with the most visual aesthetic. Another conceptual blending in the visual domain [12] proposed a model which replaces similar shape items in an image (e.g., the earth globe is replaced with a wheel). The Vismatic model [13] is proposed, which needs the user's interaction for the input space selection and the process of visual blending.

Some of these visual blenders are based on photorealistic inputs [12, 13]. Some models are not based on photorealistic inputs [10, 14, 15], and they use the Quick, Draw dataset[2], which contains non-photorealistic inputs, Or the blend in visual level using emoji inputs [16]. Some of these projects are creative systems [16], and some of them are creative support tools (assistants) [10].

On the other hand, here is the third group based on idea generation; The WHAT-IF machine [17] is a project that tries to invent fictional ideas with real cultural value. This model extracts facts from the text and then changes a small quantity of them to make fictional ideas.

With the study of research around visual metaphor and knowledge extraction, visual metaphor, visual blending, and creative idea generation, we decided to focus on understanding visual metaphor. Then use this gathered knowledge for creative ideation for visual metaphor, where we try to design a model which can be used at the conceptual level in visual blending systems.

3 Dataset

The Pitt dataset[3] [7] is used to interpret the visual metaphors for creative imagination, containing 64,832 image ads with rich annotations.

This data set proposes the novel problem of automatic advertisement understanding. This dataset contains the topic, sentiment of the ads, strategies, questions and answers(Q&A) describing what actions the viewer is prompted to take and the reasoning that the ad presents to persuade the viewer ("What should I do according to this ad, and why should I do it?"), symbolic references, and atypical annotation, which contains combination type and combination parts if the image is a visual metaphor (e.g., Texture Replacement: The object which has a new texture is deer head and the objects that have created the texture are different candies). A piece of the dataset is shown in Figure 1.

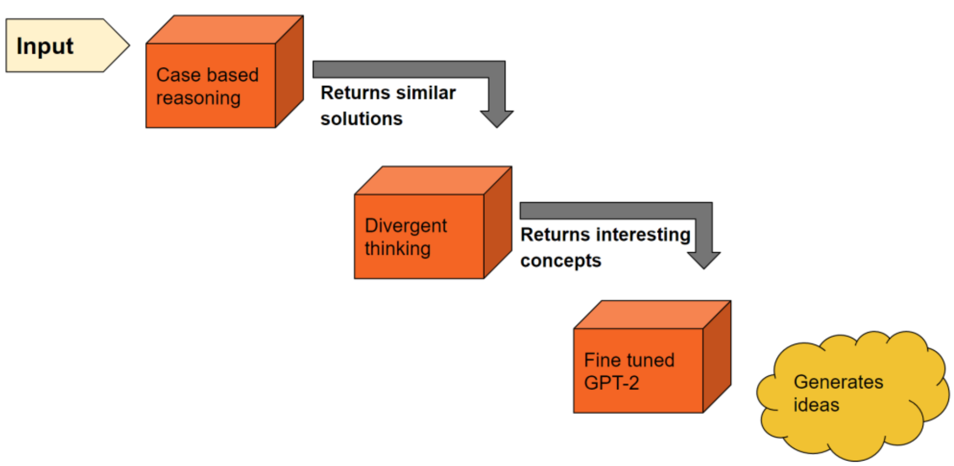

Figure 2 The architecture of the model and its three main modules: Case-based reasoning, Divergent thinking, and idea generation

[2] https://github.com/googlecreativelab/quickdraw-dataset

[3] https://people.cs.pitt.edu/~kovashka/ads/

4 Modeling

In this section, we focus on modeling, and different basic modules which get the input and produce output. In our problem-solving process, the goal is to generate a creative idea that describes a visual metaphor (similar to atypical annotation data) for a given input in Q&A shape. As shown in Figure 2, we have three main modules: case-based reasoning, divergent thinking, and generative model; In the following three sections, we go through the details.

4.1 Case-based reasoning

Case-based reasoning (CBR) is a process of solving problems that comes from philosophical concepts [1, 2]. In problem-solving in artificial intelligence systems, CBR is a good choice because it can solve new problems based on past issues [3]. In our case, where we are trying to generate some creative ideas for a given input, and there is no grand truth, CBR can lead the model to similar problems and their solutions. CBR module searches for similar problems for a given input and returns their solutions. We used a doc2vec model trained on our data to find the most similar problems to the given input. We can see the output of the CBR module in Figure 2.

4.2 Divergent thinking

As mentioned before, we need divergent thinking to challenge conventional thinking to make unexpected connections between different concepts. Here is where we start our imagination. The inputs of the divergent thinking module are similar problems' solutions to our input. The module outputs are some interesting concepts that will be the base of an idea called candidates.

We trained a word2vec model on two data segments. The first segment is the Atypical annotation of visual metaphors in the dataset that describe the combination type and parts; the second is ConceptNet description and connections related to each concept in the first segment. This word2vec model can bold and create connections between different concepts because it has learned some unexpected connections. Afterward returns similar concepts and its score as candidates for a given input.

On the other hand, a pre-trained model (word2vec-google-news-300)[4] is used to calculate the popularity of each candidate. In the end, based on scores coming from the word2ved (trained on our data) and the pre-trained word2vec, the final candidates are generated.

Table 1 The input concepts are and generated idea based on it

Input concept	Generated idea
love	Liquid Deformed Object: The liquid which has been deformed is love
cars	Texture Replacement 1: The object which has a new texture is Car and the new texture is wheels
perfume door	Texture Replacement 2: The object which has a new texture is perfume door and the objects that have created the texture are people

4.3 Idea generation

In the sentence generations phase, we used the transfer learning technique. Generative Pre-trained Transformer 2 (GPT-2)[5], created by OpenAI, is used as a pre-trained model for idea generation [6]. GPT-2 is a large-scale unsupervised language model with 1.5 billion parameters, and it is trained on 40 GB of textual data gathered from 8 M web pages. GPT-2 is trained with a simple objective: predict the next word, given all of the previous words within some text[6].

The GPT-2 is fine-tuned on a concept and its atypical annotation. An example of a given input to the model is "cookie- milk: Object Replacement: The object which is placed in the context of another object is a

[4] https://github.com/mmihaltz/word2vec-GoogleNews-vectors
[5] https://github.com/openai/gpt-2

cookie- milk and the object which is replaced by another object (expected object) person-body-hoop/net".

The strategy of sentence generation is based on Beam search instead of greedy search. The optimizer is Adam, with a learning rate of 0.0005. Some examples of generated ideas are provided in Table 1.

5 Evaluation

Evaluation of creative systems helps us to understand and better design creative systems [18]. Four main questions are presented that can guide evaluation in co-creative systems: Who is evaluating the creativity, what is being evaluated, when does evaluation occur, and how the evaluation is performed [19].

After analyzing the different aspects of evaluation creativity and looking at the related works and their evaluation [20, 17] (which was a survey), we decided to create a survey for generated ideas. The survey contains three parts. The first part is an introduction and example of a visual metaphor; the second is a Turing test; the third is a 5-star rating. In the following two sections, we go through the details.

5.1 Turing Test

In the Turing test phase, users are shown a shuffled list of ideas generated by a human or robot (6 ideas for each group) and asked to answer whether a human or a robot generates each sentence or if they are not sure (it is not distinguishable). Users' answer to each idea is shown in Figure 5, where the first six indexes belong to human and the second six sentences belong to the bot. Figure 3 shows the number of responses in each group. Overall the results show that the source of generation is not easy to recognize, and the number of correct answers is less than the number of wrong answers + not sure answers. The results in Figure 3 demonstrate that generated ideas by the model can pass the Turing test.

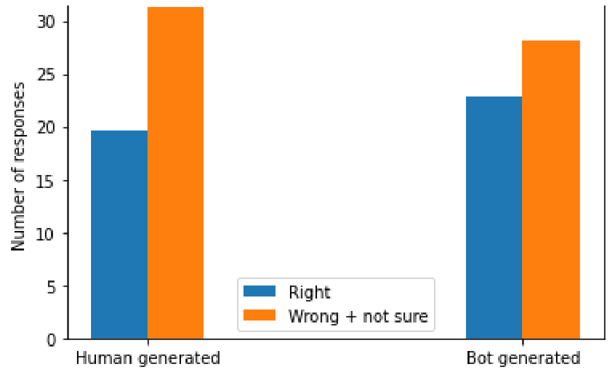

Figure 3 Creative ideas generated by Humans and Bot

5.2 5-star Rating

On the other hand, to evaluate the quality of the generated sentences, we asked users for a 5-star rating. We published five bot-generated ideas and five human-generated ideas randomly.

The given stars to the ideas are shown in Figure 6 instance-wise. The overall results in Figure 4 mean of the given stars for both groups is between 2 and 3, Where there is no significant difference.

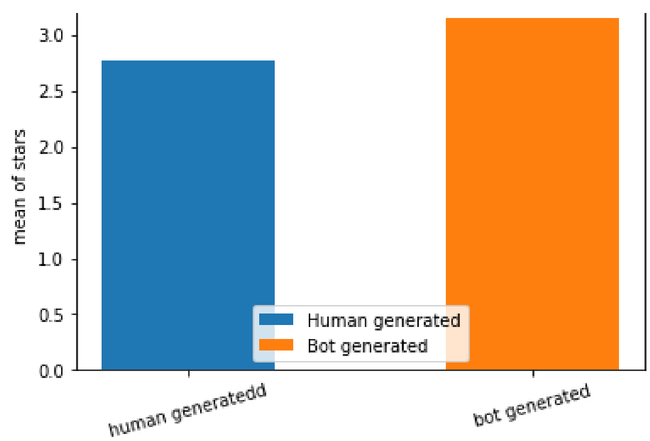

Figure 4 The 5-star rating for generated ideas by Humans and Bot

6 Conclusion

In this project, different generative models in visual metaphors are analyzed. We noticed that most models skip the first level of visual blending, the conceptual level, and focus on the second level, the visual level. We tried to design a model which focuses on the conceptual level and generates a creative idea for a visual metaphor. The model evaluation shows that these bot-generated ideas compete with human-generated ideas and are promising. The model design and outputs can be a great start, but it needs more improvements because *not all visual blends are visual metaphors*. Visual blends contain any blending with or without reason, while visual metaphor is a kind of visual blend where it includes a message, and there is a stro---ng reason for input selection for a blend. The introduced model, which contains case-based reasoning, divergent thinking, and generative modules, is based on the probability of unseen connections and relations. While for better creativity, we need more robust reasoning. We can lead our architecture for more vital reasoning in the future.

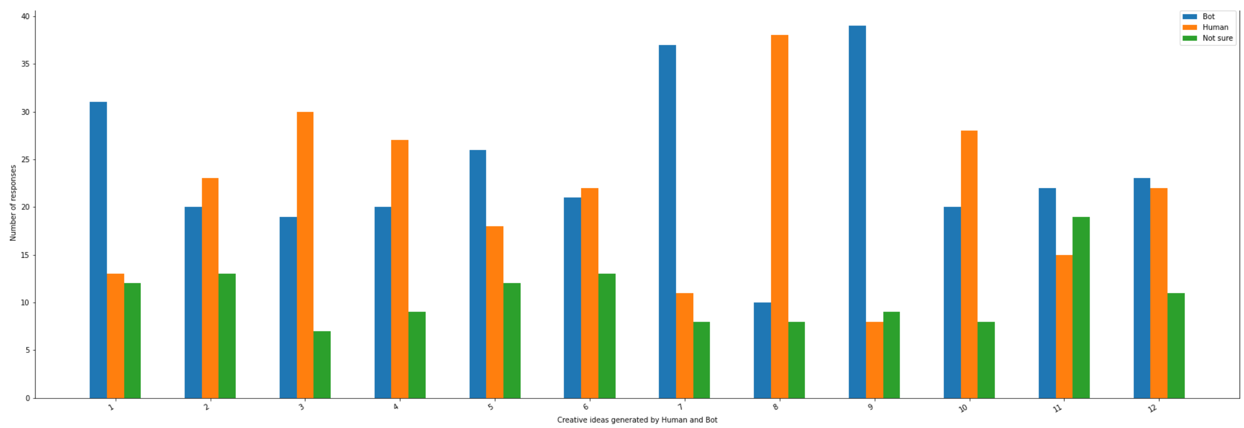

Figure 5 The result of the Turing test

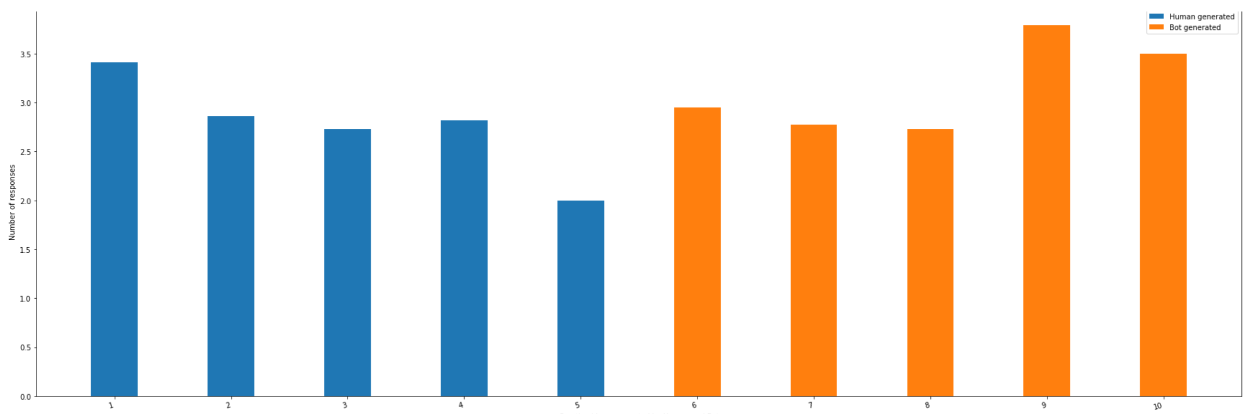

Figure 6 The result of the 5-star rating

References

[1] R. C. Schank, "A theory of reminding and learning," *A theory of reminding and learning in computers and people*, 1982.

[2] D. B. Leake, Cbr in context: The present and future. Case-based reasoning: Experiences, lessons, and future directions, 1996.

[3] R. Das, A. Godbole, N. Monath, M. Zaheer and A. McCallum, "Probabilistic Case-based Reasoning for Open-World Knowledge Graph Completion," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020.

[4] J. P. Guilford, "The Structure of Intellect," *Psychological Bulletin*, pp. 267-293, 1956.

[5] M. Sameera and S. M. Muzakkir, "Application Of Creative Design Methodology In Improved Conceptual Design Of A Top Load Washing Machine For Better And Gentle Cleaning," 2017.

[6] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei and I. Sutskever, "Language Models are Unsupervised Multitask Learners," 2019.

[7] Z. Hussain, M. Zhang, X. Zhang, K. Ye, C. Thomas, Z. Agha, N. Ong and A. Kovashka, "Automatic Understanding of Image and Video Advertisements," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[8] Y. Keren and K. Adriana, "ADVISE: Symbolism and External Knowledge for Decoding Advertisements," *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[9] K. Ye, N. Nazari, J. Hahn, Z. Hussain, M. Zhang and A. Kovashka, "Interpreting the Rhetoric of Visual Advertisements," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[10] P. Karimi, M. L. Maher, K. Grace and N. Davis, "A computational model for visual conceptual blends," *Ibm Journal of Research and Development* , 2018.

[11] F. C. Pereira and A. Cardoso, "The Boat-House Visual Blending Experiment," in *European Conference on Artificial Intelligence*, Lyon, France, 2002.

[12] A. Steinbrück, "Conceptual blending for the visual domain," Ph. D. dissertation, Masters thesis, 2013.

[13] X. Ping and L. Simo, "Vismantic: Meaning-making with Images," in *Interantional Conference on Computational Creativity*, Utah, USA, 2015.

[14] J. M. Cunha, J. Gonçalves, P. Martins, P. Machado and A. Cardoso, "A Pig, an Angel and a Cactus Walk Into a Blender: A Descriptive Approach to Visual Blending," in *International Conference on Computational Creativity*, Atlanta, Georgia, USA, 2017.

[15] D. Ha and D. Eck, "A Neural Representation of Sketch Drawings," *arXiv preprint arXiv:1704.03477*, 2017.

[16] J. M. Cunha, P. Martins and P. Machado, "How Shell and Horn make a Unicorn: Experimenting with Visual Blending in Emoji," in *International Conference on Computational Creativity*, Salamanca, Spain, 2018.

[17] "The What-If Machine project," Technologies and scientific foundations in the field of creativity, Spain, Ireland, Slovenia, United Kingdom, 1 October 2013 - 30 September 2016.

[18] A. Jordanous, "A Standardised Procedure for Evaluating Creative Systems: Computational Creativity Evaluation Based on What it is to be Creative," *Cognitive Computation*, vol. 4, pp. 246-279, 2012.

[19] P. Karimi, K. Grace, M. Maher and N. Davis, "Evaluating Creativity in Computational Co-Creative Systems," *arXiv preprint arXiv:1807.09886*, 2018.

[20] M. Pedro, U. Tanja, P. Senja, C. Amílcar and N. Lavrač, "The Good, the Bad, and the AHA! Blends," in *International Conference on Computational Creativity*, Utah, USA, 2015.