Understanding Visual Metaphors in Ads

By Tahereh Arabghalizi

Advisors: Dr. Hwa, Dr. Kovashka

CS2002 – Spring 2018

1. Introduction

1.1 Context

Nowadays advertising campaigns heavily use visual rhetoric in ads to make a powerful persuasive influence that can affect purchasers' attention and interest. Visual rhetoric in advertising can take different forms, which in turn may range in complexity. Visual metaphor is a common form of visual rhetoric in which two divergent objects are compared by which the characteristic(s) of one object, the source, are attributed to the other object, the target [1]. The source domain and target domain need at least one property in common as the basis of the implied comparison [2].

Phillips and McQuarrie (2004) introduced a framework of visual metaphors which makes a difference between three types of visual structures: Juxtaposition (when the source and the target are presented alongside), Fusion (when the source and the target are merged into one image) and Replacement (only the image of the source is present, pointing to the absent image). According to their proposed model, the metaphorical Juxtaposition is less complex than Fusion and Fusion is less complex than Replacement. They assume that the more complex the visual structure, the more processing demands are required [3].

In addition to the efforts about visual metaphor, the study of verbal metaphors has been taken up as a subject of research in a variety of fields under the framework of natural-language processing in recent years. Yoshimura et al. introduce a sensory association method using commonsense human knowledge that is an effective technique in simile-recognition systems [4]. Moreover, Meta4meaning, is a metaphor interpretation method that uses word associations extracted from a corpus to retrieve an approximation to properties of concepts. They obtain the interpretations as an aggregation or difference of the saliences of the properties to the source and the target [5].

One of the main problems that computational models based on natural language processing encounter for metaphor comprehension, lies in the nature of the similarity that allows two concepts to be aligned in a metaphor. Two concepts that are aligned in a metaphorical comparison need to have something in common for them to be comparable. However, the similarity that characterizes two metaphor terms is not as consistent and stable as the similarity that characterizes, for example, two synonyms. The type of semantic information encoded in the shared features, which accounts for the similarity between two metaphor terms, is still unexplored territory. Bolognesi presents a study that aims at modeling and contrasting the semantic similarity between metaphor terms in visual vs. linguistic metaphors. Three different types of similarity are investigated in this work: Attributional similarity, relational similarity, and linguistic similarity using a corpora that encode streams of semantic information [6].

Various research works have been conducted in metaphor identification and interpretation, specifically in the field of Natural Language Processing. However, the number of studies on understanding visual metaphor particularly in advertisements is rather very limited. Considering this, we propose a novel approach to predict the common sense between the source and target.

1.2 Problem Statement

Visual advertisements rely on many strategies to convey their message. One strategy is for physical content inside the image to refer to other concepts or other content outside the image. In this study, our goal is to automatically understand this metaphorical strategy such that given the target (what is seen) and the source (what is implied), we can predict the best possible common sense between the source and the target in an ad (Figure 1).

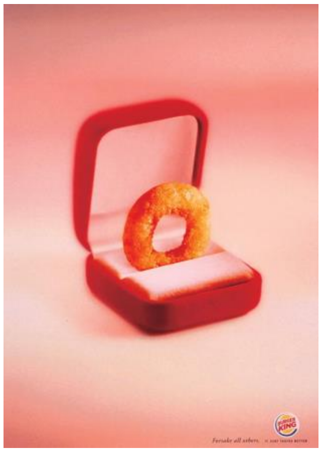

Figure 1: Source: wedding ring – target: onion ring. "Onion ring is precious"

1.3 Proposed Solution

The dataset which is used in this work consists of 25 ads with visual metaphors about food, which were annotated by Amazon MTurk[1]. In order to reach our goal, we use the given source and target in each annotated image and find all the possible common properties between them with the help of semantic similarity methods and rank the desired properties based on their association significance with the source and target. We then evaluate our results by comparing the given properties as the ground truth, with the predicted properties and return the best ranking of the best found property.

2. Approach

The approach presented in this study defines a process that comprises data enrichment and post-processing steps, followed by the evaluation phase. We propose three baseline methods to retrieve properties for each source and target. These methods use different knowledge bases namely Word Association Network[2], datamuse[3] and ConceptNet[4] to extract the most relative words (maximum 300), in particular "adjectives" to the source and the target.

2.1 Knowledge Bases

Word Association Network is inherently an ideographic dictionary or thesaurus. The associative dictionary, or analogical dictionary groups the words of the language by psychological perception, sense and meaning. It uses three types of associations including similarity, contrast and contiguity. Association by similarity is based on the fact that the associated phenomena have some common features. Association by contrast is explained by the presence in phenomena of opposite features. And association by contiguity comes into existence when events are situated close together in time or space [7].

datamuse leans on many freely available data sources such as Google Books Ngrams data set that is used to build the language model that scores candidate words by context, and also for some of the lexical relations. Word2vec is used for re-ranking result sets by topic [8].

ConceptNet is a multilingual knowledge base, representing words and phrases and the common-sense relationships between them. The knowledge in ConceptNet is collected from a variety of resources, including crowd-sourced resources (such as Wiktionary), games with a purpose (such as Verbosity), and expert-created resources (such as WordNet) [9].

[1] https://people.cs.pitt.edu/~xiaozhong/mturk_metaphor/mturk_meta.html
[2] https://wordassociations.net/en
3 https://www.datamuse.com/api/
4 http://conceptnet.io/

2.2 Extract Desired Candidates

After retrieving the properties for the source and the target using the mentioned knowledge bases (separately), we find the intersection and also the union of the extracted words for each ad. The reason for using intersection is obvious since it refers to the common properties but the reason for using union of the properties is to have a richer bag of words which contains the properties of both source and target, because for some metaphorical cases, the common sense is an explicit property of one of the metaphor terms and an implicit property of the other one, so union of properties will consist of the common sense in any case.

2.3 Ranking of Candidates

Since all the properties in the intersection or union word lists are not proper candidates for metaphorical relationship between each source and target, we employ two techniques namely Word2Vec and WordNet to measure the semantic similarity between the extracted properties and the source and target and therefore find their association significance. Both techniques use "cosine similarity" as their distance function. Having this association strength for each property, now we can rank the properties by a rank function.

We used the rank function proposed in [5] which first takes the product of the association strengths of the property to the source and target that emphasizes properties that are strongly associated with both and second takes difference of the same association strengths based on this hypothesis that the properties highlighted by a metaphor are among the common properties of the target and the source and are more salient to the source than to the target. So the larger the difference, the higher the metaphor aptness [10]. These two weights measure different complementary aspects of metaphor properties. The final ranking function is introduced as a combination of these two measures associate the property with the better of these. More details are explained in the Implementation section.

At the end of this phase, we have a list of desired candidates (maximum 100) ranked based on their similarity strength with the source and the target.

2.4 Evaluation

In order to evaluate the introduced approach, we need to compare the highlighted properties (maximum 3 for each ad) in the ground truth with the predicted ones. However, since an exact property (annotated by human) might not be found in the predicated candidate list but its similar words might be, we add at most 10 more words that are most similar to each ground-truth property, using the knowledge bases and rank them by the introduced ranking function. So after this step, for each ad, there will be three separate word lists for the ground-truth properties and one word list for the predicated properties in our dataset.

Now it is time to compare the ground truth properties and the predicated ones. Two approaches are used for this comparison. The first approach is to compare the exact words in the ground-truth properties with the exact words in the predicated list and return the average/median of the ranks of the common words found in two lists (if there is more than one common word). In this case the similarity measure between each pair of common words is one, because they are the same.

The second approach is to find the most similar words (existed in the predicated list) to the words in the ground-truth properties and return the words (and their ranks) with the highest similarity strengths (less than or equal to one). Furthermore, a similarity threshold can be used to limit the number of the most similar words found in the predicated list.

After the comparison phase, we can introduce a ground-truth property with the minimum average/median rank as the best ground-truth property and also the best property with the best rank in the predicted list.

3. Implementation

The implementation of this project is done using Python and its useful packages such as gensim.models, nltk, numpy, etc.

3.1 Datasets[5]

- *annotations.csv*: contains the data (imageURL, target, source, p1, p2, p3) collected from 25 annotated ads. This dataset has been used as a baseline for preparing the results.
- *wan_props*.csv, wan_props_union.csv: contain intersection and union of properties of the source and target using Word Association Network. There are also similar datasets for datamuse and ConceptNet.
- *wan_ranked_props.csv*: contains top1, top10 and top100 of ranked properties in the intersection/union lists. There are also similar datasets for datamuse and ConceptNet.
- *wan_GTProps_props.csv*: contains the most similar words to the ground-truth properties. There are also similar datasets for datamuse and ConceptNet.
- *wan_dataset.csv*: the final dataset(s) created after the evaluation. There are different versions for different methods (wordnet/word2vec, intersection/union and two evaluation approaches). The same for datamuse and ConceptNet.

[5] Available at: https://drive.google.com/open?id=1uaLOXrc0eEotFhmHP5AnvNvEQEM8G8Ql

3.2 Code[6]

The following files and their belonging functions have been implemented and run to obtain the final outcomes:

- *wan_getCandidates.py*: to extract the properties of each source and target and their intersection/union using Word Association Network.
- *wan_getGTprops*.py: to extract the most similar words for each ground-truth property using Word Association Network.
- *dm_getCandidates.py*: to extract the properties of each source and target and their intersection/union using datamuse.
- *dm_getGTprops*.py: to extract the most similar words for each ground-truth property using datamuse.
- *cn_getCandidates.py*: to extract the properties of each source and target and their intersection/union using ConceptNet.
- *cn_getGTprops*.py: to extract the most similar words for each ground-truth property using ConceptNet.
- *SimUtil.py*: contains all required functions to compute the similarity strength between each property and the source and the target. Besides, functions about finding the average/median of ranks of two word lists and a few more useful functions are implemented here. Word2vec Produce word vectors with deep learning via word2vec's "skip-gram and CBOW models", using either hierarchical softmax or negative sampling [11]. Wordnet is a NLTK corpus reader in which synsets of a word can be looked up and wordnet similarity returns a score denoting how similar two word senses are, based on the shortest path that connects the synsets in the is-a taxonomy [12]. It is worth saying that all the words are lemmatized before being used in similarity functions.
- *rank_candidates.py*: to compute the ranks of words in a list using the ranking function and based on their similarity measures.
- *evaluation.py*: to compare the ground-truth properties and the predicated ones and find the best rank and best property using average/median. The similarity threshold that is used in the second evaluation approach is set to 0.7 which means we only consider the predicted properties that their similarity strength to the ground-truth properties is greater than or equal to 0.7.

[6] Available at: https://drive.google.com/open?id=1FOnb66cvpgEyylsKAw4GhgPVB0XVoJHF

4. Results and Discussion

The results that are obtained are different depending on the employed knowledge base (wan, dm or cn), the method for finding the common properties (intersection or union), the applied technique for similarity (word2vec or wordnet) and the approach used for evaluation.

Since our data does not have a normal distribution, we only report the median of the ranks instead of mean/average in the final results. Moreover, since the results for word2vec and wordnet similarities were almost the same, we only report the results for wordnet. A summary of the final outcomes for all 25 ads is provided in Table 1:

Table 1: Summary of the results

Knowledge Base	Method (inter/union)	Evaluation approach	Min rank of GT[7]	Max Sim.[8]
wan	intersection	(1)	88	0.23
wan	union	(1)	91	0.18
wan	**intersection**	**(2)**	**10**	**0.40**
wan	union	(2)	33	0.39
dm	intersection	(1)	92	0.21
dm	union	(1)	99	0.21
dm	intersection	(2)	18	0.40
dm	union	(2)	34	0.40
cn	intersection	(1)	101	0.06
cn	union	(1)	94	0.29
cn	intersection	(2)	95	0.01
cn	union	(2)	35	0.39

As it can be seen in the table, using Word Association Network as the knowledge Base and having the intersection of properties between the source and the target returns the best rank and similarity. In overall, the three knowledge bases do not make a big difference in the final results but the second approach for evaluation makes a difference compared to the first approach. More details such as the best ground-truth and also predicted property for each ad can be found in the datasets links. (See section 3.1)

[7] Minimum rank of the ground-truth properties using top100 predicated properties. This value can be varied between 1 (best rank) and 101 (worst rank which means the word was not found in the candidate list)

[8] Maximum similarity measure computed between ground-truth properties and top100 predicted properties

5. Summary of Computer Vision Project

The goal of this project as a complementary work for the current project, is to predict the metaphorical mapping(s) between the source and the target with the help of concepts, logo or brand and the text in the ad. In this work, we assume that concepts and logo correspond to the target (what is seen) in a visual metaphor and text corresponds to the source (what is implied) in a visual metaphor. The intuition is clear about concepts because they are directly related to what you can see in the image. Regarding the logo or brand, it implicitly relates to the concepts that are seen in the image (e.g. Burger King is related to fast food, burger, onion ring, etc.) and in the property extraction phase, we will extract more words in the logo's domain such as its products which can also support this institution. Regarding the text, as Alousque explained in their research [13], the verbal element (text) often helps to determine the metaphoricity of the image and it usually contains some words that are explicitly/implicitly related to the missed concept or the source in the visual metaphor.

The proposed approach is illustrated in the following figure:

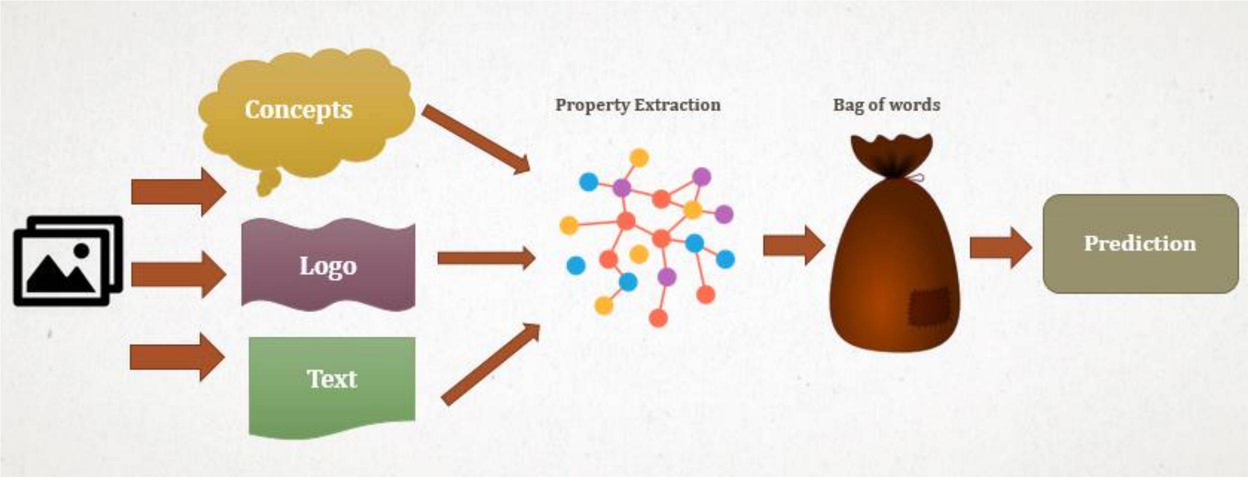

Figure 2: the proposed approach

As it can be seen this approach consists of three main phases:

1. *Knowledge extraction*: the first phase of the approach is to detect and extract concepts, logo/brand and text in ads. For this purpose, Google Vision API that can provide us with the most steady and predictable performance for vision tasks such as concept detection, logo detection, text detection, face detection, etc. is used.
2. *Property extraction:* to make a larger scope/domain for the source and target, we use another knowledge base called "Word Association Network" to extract the most similar/related concepts

to the source and the target. This API provides us with different associations including similarity, contrast and contiguity.

3. *Prediction:* To predict the desired mapping candidates between the source and target, we have to find the common concepts in both source and target domains and then rank these candidates based on their similarity with source and target. The method used for finding the similarity of words is WordNet that uses cosine as it distance function. The ranking function is the same as the one used in the current project.

Then in order to evaluate our approach, we compare the ranked candidates with the properties in the ground truth (25 ads annotated by Amazon MTurk) and report the rank of a candidate if there is a match between this candidate and a ground truth property (or a similar word to that).

6. Conclusion and Future work

In this study we proposed an approach by using different knowledge bases and applying different methods and techniques to tackle the problem of finding the metaphorical relationship between the metaphor terms (source and target) in advertisements with visual metaphor. The final results show that the best ranking of a ground-truth property can be found in an intersection of properties between the source and the target.

One of the big challenges of this project was the small training dataset containing only 25 annotated ads. We tend to collect a much bigger dataset containing visual metaphor or no metaphor that can be labeled by Amazon MTurk. Having such dataset will help us to try more methods such as visual-semantic embeddings (VSE++).

Bibliography

[1] P. a. J. P. D. Sopory, "The persuasive effects of metaphor: A meta-analysis.," Human Communication Research , pp. 382-419, 2002.

[2] M. A. v. H. a. U. N. Van Mulken, "Finding the tipping point: Visual metaphor and conceptual complexity in advertising," Journal of Advertising, pp. 333-343, 2014.

[3] B. J. a. E. F. M. Phillips, "Beyond visual metaphor: A new typology of visual rhetoric in advertising.," Marketing theory , pp. 113-136, 2004.

[4] E. M. I. S. T. a. H. W. Yoshimura, "A Simile Recognition System using a Commonsense Sensory Association Method," Procedia Computer Science, pp. 55-62, 2015.

[5] P. K. A. M. G.-W. K. A. a. H. T. Xiao, "Meta4meaning: Automatic metaphor interpretation using corpus-derived word associations," in In Proceedings of the 7th International Conference on Computational Creativity (ICCC)., Paris, 2016.

[6] M. Bolognesi, "Modeling Semantic Similarity between Metaphor Terms of Visual vs. Linguistic Metaphors through Flickr Tag Distributions," Frontiers in Communication , 2016.

[7] 2006-2018. [Online]. Available: https://wordassociations.net/en.

[8] "datamuse," 2016. [Online]. Available: https://www.datamuse.com/api/.

[9] "ConceptNet," GitHub Inc., [Online]. Available: http://conceptnet.io/.

[10] A. Ortony, "The role of similarity in similes and metaphors," in Metaphor and thought, 1979.

[11] R. Řehůřek, 2009. [Online]. Available: https://radimrehurek.com/gensim/models/word2vec.html.

[12] "nltk," 2017. [Online]. Available: http://www.nltk.org/howto/wordnet.html.

[13] I. N. Alousque, "The role of text in the identification of visual metaphor in advertising," in Procedia-Social and Behavioral Sciences 212, 2015.