

Avoiding Green and Colorless Ideas – – Text-based Color-Related Knowledge Acquisition for Better Image Understanding

Rafal Rzepka, Keita Mitsuhashi, and Kenji Araki

Hokkaido University, Sapporo, Kita-ku, Kita 14 Nishi 9, 060-0814, Japan
{rzepka,mitsuhashi,araki}@ist.hokudai.ac.jp,
WWW home page: <http://arakilab.media.eng.hokudai.ac.jp/>

Abstract. In this paper we introduce a simple text mining method which could be helpful for automatic image understanding process, especially in object recognition. By using colors as an example of vision-related feature category, we describe how word frequencies, dependency parsing and quasi-semantic filtering help to acquire more accurate knowledge which usually requires costly and time consuming annotation process to be properly obtained. We describe retrieved data, preliminary experimental results and after analyzing errors we suggest possible solutions for improvement. We conclude the paper with a discussion about using the retrieved knowledge in fields like image processing or metaphor generation and understanding.

1 Introduction

Machines’ lack of outside world knowledge is one of the biggest problems on our way to achieving a human-level artificial intelligence. Even if we can equip robots with sensors that humans do not possess, cognitive architectures are far from acquiring the same quality of perception. When we read the famous example of a grammatically proper English sentence “Colorless green ideas sleep furiously” [2] we effortlessly realize what is wrong with its meaning, but for artificial agents it is not so obvious because they lack common sense. For Japanese language, which we use in our applications, there is not much choice when it comes to common sense knowledge resources. The only database available is ConceptNet ontology [8] which growth relies mostly on knowledge from <http://nadia.jp> service where users “teach” a child named Nadya through a guessing game. The problem with this approach is that users get bored quickly and try to be original or humorous which leads to assertions as “music *CreatedBy* me” or “words *CreatedBy* expectations” which are not completely wrong in a semantic way but difficult to be utilized for reasoning about physical world. For that reason we try to automatize the knowledge acquisition process [17] and concentrate on acquiring semantically new entries in opposition to methods utilizing similarity [9]. Our approach to this problem is to support sensing technology with text-mining techniques [16] and applying them for various AI tasks as moral reasoning [12,

15], artificial therapists [13] or poetry generation[14]. One of the biggest problems with “textual sensing” approach is that words used in text are used both literally and figuratively. In this paper we use colors as an example of physical features of objects which, when shallow NLP techniques are used, can lead to learning wrong commonsense knowledge which often happens in automatic approaches [1]. Automatic color recognition method in text was introduced in [3], where English ConceptNet was used. Because of a very small size of Japanese version, implementing the same vector-based method was impossible, therefore we decided to develop an original method utilizing a corpus. Another difference is that the work on English language has limited single word to single color, but in our method one noun can have multiple colors.

2 System Overview

Colors in ConceptNet are usually stored under *HasProperty* relation, for example “ink *HasProperty* black” or “cayenne pepper *HasProperty* red”, however they can be also found as edges of other relations as in “blood *SymbolOf* red” or “apple *IsA* edible red fruit”. In our experiments we concentrated on *HasProperty* relation which comprises only 3.6% of Japanese ConceptNet 4 and needs to be extended. There are 140 entries regarding 6 colors (red, blue, yellow, brown, white and black) which we chose for being an adjective (for example “green” is expressed by a noun in Japanese language), while the YACIS blog corpus we developed [10], contains 748,078 sentences using these colors. We retrieved these sentences under the condition that color adjectives are accompanied by nouns labeled by the morphological analyzer [6] as “usual nouns” (this helps excluding proper nouns as names) and that are not noun phrases consisting more than one noun. Then the acquired nouns are counted and if they appear less than twice in the whole set, they are deleted as rather unusual ones. Finally, the filtered sentences are used for retrieving nouns stored in six colors categories generating a database. For example, “There was a red apple shaped key case in my bag” put “apple”, “key case” (one word in Japanese) and “bag” as the related nouns candidates under the “red” category. When we analyzed the most frequent 10 nouns for each color, except natural associations (“blue sky”, “white rice” or “red flower” we found numerous errors as expected (“brown part”, “yellow color”, “white lover”, which is a popular cookie brand name or “yellow elephant” which is a cartoon character). Error analysis showed that there are many examples which are true but hard to be unequivocally categorized as common sense about the physical world as “black feeling”. Many nouns appeared in different color categories (“eyes”, “flowers”, “birds”, “men”, etc.) but we believe they should not be treated as errors and might be stored as separate concepts. To determine which noun is described by a given color in a sentence more accurately, we implemented dependency parsing with CaboCha tool [5]. If a color adjective was not immediately followed by a noun, we also collected nouns from dependency chunk ending by a particle suggesting existence of subject (*wa*, *ga*, *mo*¹, etc.) For

¹ Original Japanese words are written in italic throughout the paper.

example from *Tomato-wa mi-mo akai* (“Tomatoes, and also their seeds, are red”) both “tomatoes” and “seeds” are retrieved because subject indicating particle *wa* followed “tomato”. This addition allowed to eliminate some ambiguous nouns but many problematic pairs still existed. Next we decided to extend our method with corpus frequency checks and eliminated pairs with significantly smaller hit rates, however the overall quality did not improved as much as expected. For that reason we decided to limit nouns to these which exist in Japanese ConceptNet 4, what decreased number of color related words (see Table 1) but visibly improved the quality of retrievals.

Table 1. Number of nouns before and after ConceptNet filtering.

	<i>Baseline</i>	<i>Dependency</i>	<i>Dependency + Frequency</i>
Before filtering	67,577	13,955	12,592
After filtering	12,497	5,043	4,659

3 Experiments and Results

To investigate how the above mentioned quality improved, we performed an experiment with nouns chosen randomly from the ConceptNet and to showed them to five judges (3 male and 2 female Japanese students in their twenties). They labeled the words with six previously mentioned colors and if the majority agreed on the same colors for a noun, it became the part of “color” set, giving 60 nouns in total. To prepare balanced counter-set, we randomly chose 60 other nouns for which the judges did not choose a color. For example, judges agreed that roses are red, corn is yellow, snowman is white and pianos are black and white. Words as “book”, “skirt”, “fruit” or “scarf” were marked with all six colors.

Then we input the nouns into the system and generated five sets of results for 1, 5, 10, 15 and 20 retrievals as thresholds to see how strict the process should be. The results (see Tables 3) show that 10 retrievals threshold is the most balanced one, however if we focus on precision, for example when gathering new concepts for the ontology, it seems that the higher threshold the higher quality noun-color(s) associations can be obtained. Adding dependency parsing was obviously effective but limiting output to high occurrences improved the results only minimally.

In the second part of our experiment we wanted to see how good is our system in recognizing non-physical nouns without colors (e.g. “youth”, “leadership”, “energy-efficiency”, “universal gravitation” or “disapproval”). Experimental results for this recognition are presented in Table 3. This time high precision was achieved and the method showed capability of distinguishing objects with and

Table 2. Precision (upper), Recall (middle) and F-score (lower) depending on thresholds (retrievals after filtering).

Threshold	<i>Baseline</i>	<i>Dependency</i>	<i>Dependency + Frequency</i>
Precision			
1	0.199	0.373	0.382
5	0.248	0.510	0.527
10	0.278	0.569	0.588
15	0.313	0.615	0.634
20	0.335	0.638	0.633
Recall			
1	0.982	0.874	0.829
5	0.919	0.703	0.703
10	0.829	0.631	0.631
15	0.802	0.577	0.577
20	0.730	0.541	0.514
F-score			
1	0.331	0.523	0.523
5	0.390	0.591	0.602
10	0.416	0.598	0.609
15	0.451	0.595	0.604
20	0.459	0.585	0.567

without colors (physical and non-psychical objects), though there are transparent physical object which should be dealt with separately.

Table 3. Precision depending on thresholds for recognizing words without colors.

Threshold	<i>Baseline</i>	<i>Dependency</i>	<i>Dependency + Frequency</i>
1	0.117	0.783	0.800
5	0.317	0.917	0.917
10	0.517	0.950	0.950
15	0.583	0.950	0.983
20	0.600	0.967	0.983

4 Error Analysis and Discussion

The first problem we have spotted with automatic color assignment was that proper names often had unproportionally high frequencies in the corpus. Japanese language does not have any equivalent of big letters, so to tackle this problem, we

added a step to the algorithm which checks if there is a Wikipedia page for any color-noun pair. If it is found, this straight phrase is then not counted, leaving only counts of retrievals like “noun was color”. For example it eliminated rather unnatural in Japanese language associations as “White Lover” (cookie brand) or “Red Fox” (instant noodles brand), but also “White Scarf” (music piece title), “Yellow Book” (comic book title), “White Book” (cartoon title) or “Blue Sky” (book title), which represent natural colors. It appears that a deeper semantic processing is needed, because there are much more natural color-noun phrases in proper names than we expected. Another concern was to see if setting more strict thresholds will further improve the results but it appeared that after 40 retrievals limit recall drops causing f-score to decrease below the baseline level because high recall of the baseline is not limited by occurrences and accepts even very peculiar color-noun pairs (see Figure 1).

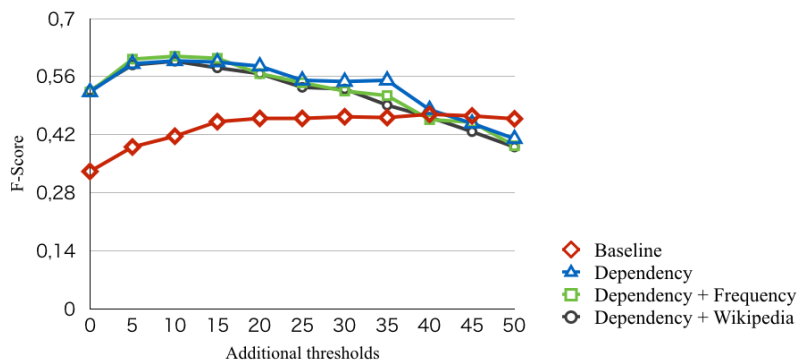


Fig. 1. Changes in results after adding thresholds and Wikipedia filtering.

5 Conclusion and Future Work

Automatic color recognizing is not a new research topic but image processing field tends to concentrate on specific tasks as licence plates recognition [18]. It is often a part of wider applications as locating faces in complex backgrounds [7] but until recently gathering knowledge through images was difficult. The latest image understanding tasks using Deep Learning [11, 4] open a wide range of possibilities for enriching common sense knowledge bases. However, statistical approaches need often costly annotation process and any additional support is needed. We think that commonsense knowledge ontologies as ConceptNet could help with adding weights to such algorithms. For example Karpathy’s experiments with Deep Learning² (see Figure 2), suggest that statistical methods

² cs.stanford.edu/people/karpathy/linear_imagenet/

are not perfect (“green gorilla”, “pink ambulance”, “pink bucket”, etc.). In our opinion combining both ontological and stochastic approaches could decrease numbers of classification mistakes as background knowledge might provide simple rules as leaves are green / yellow / red and branches are brown / black to avoid mixing both colors as a “tree” representation. Our preliminary tests showed that without laborious and costly annotations it is possible to predict colors of objects and distinguish concepts with color features from ones which do not possess them. For example, for our metaphor generation project we need a solid knowledge base about physical and abstract words and their features. Preliminary tests presented in this paper show promising level of precision for text-based approach in this task, however there are obvious problems remaining when it comes to proper color assignment. There are two basic directions in which the method could be used, to concentrate on precision given by threshold and to minimize manual check before adding data to the ConceptNet or maximizing the number of automatically acquired concepts for further automatic refining. In the next step we plan to experiment with image processing and to extend the method to adjectives representing different types of cognitive perceptions as shapes, sizes or sounds.



Fig. 2. Example linear classifiers for a few ImageNet classes categorization.

References

1. Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Jr., E.R.H., Mitchell, T.M.: Toward an architecture for never-ending language learning. In: Proceedings of the Twenty-Fourth Conference on Artificial Intelligence (AAAI 2010) (2010)

2. Chomsky, N.: Three models for the description of language. *IRE Transactions on Information Theory* 2(3), 113–124 (September 1956)
3. Havasi, C., Speer, R., Holmgren, J.: Automated color selection using semantic knowledge. In: *AAAI Fall Symposium: Commonsense Knowledge* (2010)
4. Karpathy, A., Fei-Fei, L.: Deep visual-semantic alignments for generating image descriptions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3128–3137 (2015)
5. Kudo, T.: Cabocha: Yet another japanese dependency structure analyzer. Tech. rep., Technical report, Nara Institute of Science and Technology (2004)
6. Kudo, T.: Mecab: Yet another part-of-speech and morphological analyzer (2005), <http://mecab.sourceforge.net/>
7. Lee, C.H., Kim, J.S., Park, K.H.: Automatic human face location in a complex background using motion and color information. *Pattern recognition* 29(11), 1877–1889 (1996)
8. Liu, H., Singh, P.: Conceptnet: A practical commonsense reasoning toolkit. *BT Technology Journal* 22, 211–226 (2004)
9. Makabi, A., Yamamoto, K.: Automatic acquisition of commonsense expressions for creating large scale common sense database (in japanese). In: *Proceedings of the 20th Annual Conference of Association for Natural Language Processing* (2014)
10. Ptaszynski, M., Dybala, P., Rzepka, R., Araki, K., Momouchi, Y.: Yacis: A five-billion-word corpus of japanese blogs fully annotated with syntactic and affective information. In: *Proceedings of The AISB/IACAP World Congress*. pp. 40–49 (2012)
11. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* 115(3), 211–252 (2015)
12. Rzepka, R., Araki, K.: Automatic reverse engineering of human behavior based on text for knowledge acquisition. In: N. Miyake, D. Peebles, .R.P.C. (ed.) *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. p. 679. Cognitive Science Society (2012)
13. Rzepka, R., Araki, K.: ELIZA fifty years later: An automatic therapist using bottom-up and top-down approaches. In: van Rysewyk, S.P., Pontier, M. (eds.) *Machine Medical Ethics, Intelligent Systems, Control and Automation: Science and Engineering*, vol. 74, pp. 257–272. Springer International Publishing (2015)
14. Rzepka, R., Araki, K.: Haiku generator that reads blogs and illustrates them with sounds and images. In: *Proceedings of the 24th International Conference on Artificial Intelligence*. pp. 2496–2502. AAAI Press (2015)
15. Rzepka, R., Araki, K.: Rethinking Machine Ethics in the Age of Ubiquitous Technology, chap. *Semantic Analysis of Bloggers Experiences as a Knowledge Source of Average Human Morality*, pp. 73–95. Hershey: IGI Global (2015)
16. Rzepka, R., Krawczyk, M., Araki, K.: Replacing sensors with text occurrences for commonsense knowledge acquisition. In: *Proceedings of IJCAI 2015 Workshop on Cognitive Knowledge Acquisition and Applications (Cognitum 2015)* (2015)
17. Rzepka, R., Muramoto, K., Araki, K.: Generality evaluation of automatically generated knowledge for the japanese ConceptNet. In: *Springer-Verlag Lecture Notes in Artificial Intelligence (LNAI) 7106, AI 2011: Advances in Artificial Intelligence (Proceedings of 24th Australasian Joint Conference)*. pp. 648–657 (2012)
18. Wang, F., Man, L., Wang, B., Xiao, Y., Pan, W., Lu, X.: Fuzzy-based algorithm for color recognition of license plates. *Pattern Recognition Letters* 29(7), 1007–1020 (2008)