Image Searching by Events with Deep Learning for NTCIR-12 Lifelog

Hsiang-Lun Lin
IDEAS, Institute for
Information Industry, Taiwan
hsianglunlin@iii.org.tw

Tzu-Chieh Chiang
IDEAS, Institute for
Information Industry, Taiwan
rubychiang@iii.org.tw

Liang-Pu Chen
IDEAS, Institute for
Information Industry, Taiwan
eit@iii.org.tw

Ping-Che Yang
IDEAS, Institute for
Information Industry, Taiwan
maciaclark@iii.org.tw

ABSTRACT

We construct a automatically system to participate Lifelog task in NTCIR-12 that find the image out correctly by events. In our system, We have employed deep learning method to approach the targets. In our processing, we use stanford parser and named-entities recognition method to process text of events. Also, we employ word to vector toolkit to transfer words to vectors.Moreover, we construct a model ,training it with word2vec, to calculate the correlation between each search task and image. By using this model, we find relevant images from every task in this topic.

Keywords

deep learning, word to vector, information retrieval, natural language processing, named entity recognition, machine learning

Team Name

III&CYUT

Subtasks

Lifelong Semantic Access Task (LSAT)

1. INTRODUCTION

Recently, lifelogging is considered as a mainstream activity[10]. Known as lifelogging, an act by using wearable camera to digitally track and record personal behavioral data. Especially, wearable camera allowing users to capture images and other sensor data continuously from a first-person perspective[8]. Through these data, analyst may find a useful information to an individual's life experience.Therefore, We construct a system to participate pilot Lifelog task[6, 7] in NTCIR-12 that find the pattern out correctly at right position and normalize to identify those lifelog data. There are two subtasks as follows:

- Task 1) Lifelong Semantic Access Task (LSAT) to explore search and retrieval from lifelog
- Task 2) Lifelog Insight Task (LIT) to explore knowledge mining and visualisation of lifelogs

We have participated the first subtasks. Many details must be considered in this task. Through fifty search tasks, each task has image information retrieval from NTCIR-lifelog. However, it's hard to split descriptions and narratives to correct semantic content, which contain both positive and negative content, makes analysis more complicated. Moreover, to find relevance between images and semantic content are not so easy. To pair image semantic content, we decided to use deep learning finding corresponding image from each semantic content we've got. There are lots of detail that have to be overcome.

2. MATERIALS

2.1 Dataset

Main dataset is provided by NTCIR-12 Lifelog task, it consist of images, visual concepts, and semantic content. In addition, we construct word2vec model from our corpus database, crawled from wikipedia, web forums, news and social media sites.

2.2 Deep learning

Deep learning[5, 2, 3, 11, 1] is a branch of machine learning based on neural network algorithms. Deep learning method has been widely used in information retrieval and natural language processing, because deep learning can handle extremely large corpus in the case of information without preannotated. In this paper, we proposed a method which approach the problem with deep learning.

2.3 Word to vector

Word to vector is a group of related models that are used to produce word embeddings. These models are two-layer neural networks, that are trained to reconstruct linguistic contexts of words: the network is shown a word, and must guess which words occurred in adjacent positions in an input text. The order of the remaining words is not important (bag-of-words assumption) Word2vec is a tool with two-layer neural net that processes text. Its input is a text corpus and its output is a set of vectors: feature vectors for words in that corpus. While Word2vec is not a deep neural network, it turns text into a numerical form that deep nets can understand.

2.4 Deep learning toolkit

Caffe[9] is a deep learning framework that provides multimedia scientists and practitioners with a clean and modifiable framework for state-of-the-art deep learning algorithms and a collection of reference models. Deeplearning4j is a open-source, distributed deep-learning library written for Java. Deeplearning4j includes both a distributed, multithreaded deep-learning framework and a normal single-threaded deep-learning framework. Training takes place in the cluster, which means it can process massive amounts of data quickly.

2.5 Language parsing toolkit

Stanford parser[4] is a natural language parser from Stanford NLP Group. It could parse part-of-speech of text and do named-entities recognition works. We have employed Stanford parser to process content of events that find the keyword out.

3. OUR APPROACH

3.1 Part I: Retrieving lifelog images

We choose CAFFE concept that consist in Lifelog dataset as our general concepts.

DEFINITION 1. *Let D be image sets, W be concept word sets, C be CAFFE concept matrix, present relationship between concept words and lifelog images. Then,*

$$R(|W| * |D|)$$

3.2 Part II: NLP on search task questions

First, we retrieve named entities from search task questions with Stanford NER system, then group these named entities to “bag of entities” for each task. Second, we train a word2vec model from our corpus, calculate semantic matrices depend on type of search task(“precision” or “recall”), for precision emphasized tasks, we detach negative entities; for recall emphasized tasks, we attach positive entities in word2vec model. After calculating cosine distance between bags and CAFFE concept words.

DEFINITION 2. *let B be bag of entities, S be Semantic matrix, present relationship between bags and concept words. Then,*

$$R(|B| * |W|)$$

3.3 Part III: combination

For convenience, we take the product of S and C as relationship between search task and concept words, Let $X = C * S$. Then, $X \in R(|B| * |D|)$, $\forall b \in B, \exists \max e \in X[b]$ as most possible classified images.

4. SYSTEM DESCRIPTION

We have three methods as follows:

4.1 Run-01: general word2vec model

In this run, we use general word2vec model as baseline and estimating word similarity distance between CAFFE concept words and keywords retrieved from lifelog tasks with word2vec model provided by Google Inc.[1] These pre-trained vectors trained on part of Google News dataset (about 100 billion words), and the model contains 300-dimensional vectors for 3 million words and phrases.

4.2 Run-02: additionally semantic analysis

In this run, we still apply stanford NLP parser on task descriptions, but add sentiment analysis as preprocess to retrieve negative feature of keywords, then combine those word vectors into a bag vector.

4.3 Run-03: additionally keyword expansion

In this run, we try to do expansion on every keywords. Therefore, we still use the same deep learning model to find the similar words out. After that, we based on the run-02 and take every keywords with their similar words to approach the task.

5. CONCLUSION AND FUTURE WORK

In this time, we used three different runs to analysis raw context. We try to use CAFFE to consist in Lifelog dataset as our general concepts, hoping to find the correct relevant image. In this method, we combine CAFFE into our training model, trying to rise the accuracy rate when raw context has been analysis by our system. However, according to the research result, it’ s hard to split descriptions and narratives to semantic content correctly. It contain both positive and negative content, making image finding more complicated and less effective. Moreover, to find relevance between images and semantic content are not so easy. It’ s hard to find the correct pattern between them. To solve this problem, we could have real image recognition to update our training model, so that system can detect more information when it analysis raw context. The training model is not good enough, there’ s more information which can make the higher accuracy rate than now. Another specific model shall be to determine in this case. As a pilot task, Lifelog consist of two well-known missions: image recognition and natural language processing. Because of team members’ experience, we focused on NLP part in this time, picked official CAFFE concepts as feature. But in experiments, we found that was difficult to construct the relations of topic question keywords and CAFFE concept words. We need to tweak CAFFE concepts to improve our model, e.g., rebuild word list, prune unmapped words, etc. On the other hand, keyword retrieval of topic question is the key of our task results, so exploring more suitable methods is reasonable.

6. ACKNOWLEDGMENTS

This study was conducted under the “Online and Offline Integrated Smart Commerce Platform (3/4)” of the Institute for Information Industry, which is subsidized by the Ministry of Economic Affairs of the Republic of China.

7. REFERENCES

[1] I. Arel, D. C. Rose, and T. P. Karnowski. Deep machine learning-a new frontier in artificial intelligence research [research frontier]. *Computational Intelligence Magazine, IEEE*, 5(4):13–18, 2010.

[2] Y. Bengio. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[1] https://www.google.com.tw/

[3] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1798–1828, 2013.

[4] M.-C. De Marneffe, B. MacCartney, C. D. Manning, et al. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, pages 449–454, 2006.

[5] L. Deng and D. Yu. Foundations and trends in signal processing. *Signal Processing*, 7:3–4, 2014.

[6] C. Gurrin, H. Joho, F. Hopfgartner, L. Zhou, and R. Albatal. Ntcir lifelog: The first test collection for lifelog research. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, 2016.

[7] C. Gurrin, H. Joho, F. Hopfgartner, L. Zhou, and R. Albatal. Overview of ntcir lifelog task. In *Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies, NTCIR-12.* National Center of Sciences, 2016.

[8] R. Hoyle, R. Templeman, S. Armes, D. Anthony, D. Crandall, and A. Kapadia. Privacy behaviors of lifeloggers using wearable cameras. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 571–582. ACM, 2014.

[9] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia*, pages 675–678. ACM, 2014.

[10] K. O' Hara, M. M. Tuffield, and N. Shadbolt. Lifelogging: Privacy and empowerment with memories for life. *Identity in the Information Society*, 1(1):155–172, 2008.

[11] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.