UA.PT Bioinformatics at ImageCLEF 2019: Lifelog Moment Retrieval based on Image Annotation and Natural Language Processing

Ricardo Ribeiro, António J. R. Neves, and José Luis Oliveira

IEETA/DETI, University of Aveiro, 3810-193 Aveiro, Portugal
{rfribeiro,an,jlo}@ua.pt

Abstract. The increasing number of mobile and wearable devices is dramatically changing the way we can collect data about a person's life. These devices allow recording our daily activities and behavior in the form of images, video, biometric data, location and other data. This paper describes the participation of the Bioinformatics group of the Institute of Electronics and Engineering Informatics of University of Aveiro in the ImageCLEF lifelog task, more specifically in the Lifelog Moment Retrieval sub-task. The approach to solve this sub-task is divided into three stages. The first one is the pre-processing of the lifelog dataset for a selection of the images that contain relevant information in order to reduce the amount of images to be processed and obtain additional visual information and concepts from the ones to be considered. In the second step, the query topics are analyzed using Natural Languages Processing tools to extract relevant words to retrieve the desired moment. This words are compared with the visual concepts words, obtained in the pre-processing step using a pre-trained word2vec model, to compute a confidence score for each processed image. An additional step is used in the last two runs, in order to include the images not processed in the first step and improve the results of our approach. A total of 6 runs were submitted and the results obtained show an evolution with each submission. Although the results are not yet competitive with other teams, this challenge is a good starting point for our research work. We pretend to continue the development of a lifelogging application in the context of a research project, so we expect to participate in the next year in the ImageCLEFlifelog task.

Keywords: lifelog · moment retrieval · image processing

1 Introduction

In the past few years, with the increase of wearable and smart technologies, the term lifelogging has received significant attention from both research and

commercial communities. There is no general definition for lifelogging, but an appropriate definition is given by Dodge and Kitchin [3] as "a form of pervasive computing, consisting of a unified digital record of the totality of an individuals experiences, captured multi-modally through digital sensors and stored permanently as a personal multimedia archive" [4]. In a simple way, lifelogging is the process of tracking and record personal data created by our activities and behaviour.

The number of workshops and tasks for research has increased over the last few years and among them are the main tasks of ImageCLEF 2019 lab [6]: lifelogging, medicine, nature, and security. The lifelogging task aims to bring the attention of lifelogging to an as wide as possible audience and to promote research into some of the key challenges of the coming years [2].

Our motivation for this work is the great potential that personal lifelogs have in numerous applications, including memory and moments retrieval, daily living understanding, diet monitoring, or disease diagnosis, among others. For example: in Alzheimer's disease, people have memory problems and using a lifelog application the person with the disease can be followed by a specialist or can help the person to remember certain moments or activities of her last days or months.

This paper is organized as follows: the paper starts with an introductory section. Section 2 provides a brief introduction to the ImageCLEF lifelog and the sub-task Lifelog Moment Retrieval. The proposed approach used in our best run is described in Section 3. In Section 4, the results of all submitted runs obtained in the LMRT sub-task are presented and described the differences of each run compared to the implementation of our best result. Finally, a summary of the work presented in this paper, concluding remarks, and the future work are presented in Section 5.

2 Task Description

The ImageCLEFlifelog 2019 task [1] is divided into two different sub-tasks: the Lifelog moment retrieval (LMRT) and the Puzzle sub-task. In this work, we only addressed the LMRT sub-task, as a starting point for a research work that we intend to develop with the aim of helping in some problems that exist around the world.

In the LMRT sub-task, the participants have to retrieve a number of specific predefined activities in a lifelogger's life. For example, they should return the relevant moments for the query "Find the moment(s) when I was shopping". Particular attention should be paid to the diversification of the selected moments with respect to the target scenario. The ground truth for this sub-task was created using manual annotation [1].

ImageCLEFlifelog dataset is a completely new rich multimodal dataset which consists of 29 days of data from one lifelogger, namely: images (1,500-2,500 per day from wearable cameras), visual concepts (automatically extracted visual concepts with varying rates of accuracy), semantic content (semantic locations,

semantic activities) based on sensor readings (via the Moves App) on mobile devices, biometrics information (heart rate, galvanic skin response, calorie burn, steps, continual blood glucose, etc.), music listening history, computer usage (frequency of typed words via the keyboard and information consumed on the computer via ASR of on-screen activity on a per-minute basis) [1]. However, In this work we use the images, the visual concepts and the semantic content of the dataset.

3 Proposed Method

In this sub-task, we submitted 6 runs. Although our results have a lower F1-measure@10 in the test topics, this sub-task has become a good starting point for our research work in lifelog. In this section, we present the proposed approcah used in the last submission (run 6 of Table 1) that was our best result. However, some more details about the other runs are mentioned in Section 3.

The final approach used to the LMRT task is divided into three stages, respectively:

- **Pre-Processing:** The large amount of data (images) are analyzed and some images images are excluded based on a low-level image analysis algorithm proposed by the authors in order to reduce the search time that will be needed to analyze each topic. The images that are considered to be valuable are then processed using several state-of-art algorithms to extract information from the images.
- **Retrieval:** The relevant words of the query topic are extracted using tools of Natural Language Processing (NLP). These words are compared with the information obtained from the images in the pre-processing stage, through a state-of-art model used to produce word embedding. Finally, we assign a score to each analyzed image for the query topic.
- **Post-Processing:** Images that were skipped in the pre-processing step are reused depending on a defined distance from the images selected in the retrieval step in order to fulfill with the goal of the sub-task, where all the images was used and annotated.

3.1 Pre-Processing

We consider the pre-processing of the image dataset in a lifelogging application a very important stage, in order to select the relevant images and reduce the processing time and the errors in the annotation, extracting only relevant information from the lifelog images.

In this step, we proposed a method for automatic selection of the lifelog images that contain relevant information using a blur/focus measure operator, called modified Laplacian. We use this method to extract low-level features and machine learning algorithms, namely k-nearest neighbors, to classify these features and decide if an image is valuable in this context. Figure 1 shows a block

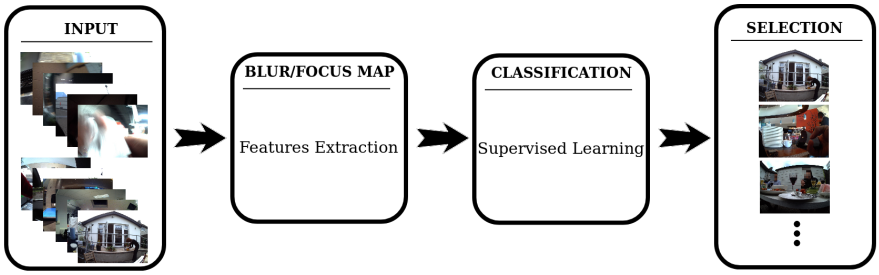

Fig. 1. Block diagram of the main steps implemented in the proposed method.

diagram presenting the steps used in the proposed method. This proposal is described in a manuscript submitted by the authors but not yet published.

Images that are not selected in this step are not processed in the retrieval step, and can be reused in the post-processing step to fulfill with the sub-task expected results.

In a lifelogging application, the most important characteristics that we can extract from images are the objects and the elements that a certain environment contains. Some content of the selected images were extracted using the label detection of Google Cloud Vision API, YoloV3 [8] and the information provided by the organizers (location, activity and visual concepts). The data associated with each image is stored into JSON files of each day of the lifelogger.

3.2 Retrieval

In the retrieval step the images are selected according to the query topic entered for the desired moment search. We use the SpaCy library [5] to analyze the topic narrative and extract relevant words. These words are divided into five categories, among them “activities”, “locations”, “relevant things”, “other things”, and “other words”.

In order to assign words to each category we define some linguistic rules, such as semantic and syntactic rules. Semantic rules build the meaning of the sentence from its words and how words combine syntactically. Syntax refers to the rules that govern the ways in which words combine to form phrases, and sentences. For example: if the sentence has an auxiliary verb, the main verb usually corresponds to an activity and the words that follow the main verb may be things or locations involved in this activity. Figure 2 presents linguistic annotations generated by SpaCy library for topic number 10 narrative of the test topics. The words extracted from this topic are ”attending”, ”meeting” and ”China”, and then divided into the categories “activities”, “relevant things” and “locations”, respectively.

A comparison is made between the words extracted from the narrative of the query topic and the concept words of each image selected in pre-processing, using a word2vec pre-trained model. We used a model trained on part of Google

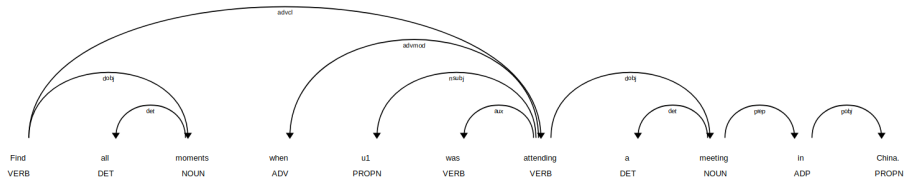

Fig. 2. Linguistic annotations generated by SpaCy library [5].

News dataset (about 100 billion words). This model contains 300-dimensional vectors for 3 million words and phrases. The gensim libary [9] allows us to load the word2vec model and compute the cosine similarity between words.

For each category defined previously we have a similarity (values from 0 to 1). As the concepts that we have for each image are not very large and accurate to decide if the image correspond to the query topic or not, we use the sequence of images in a certain distance of the image that is being analyzed and their similarities to assign a score to the category of those words. These categories have different weights associated (the sum of all categories weight is equal to 1), therefore, the confidence score of each image is computed using these weights and the scores of categories.

Finally, we determine a general threshold to select the images for the query topics. Images with confidence score above the threshold are selected for the query topic.

3.3 Post-Processing

In the last step, some images that were not analyzed in the pre-processing step are reused to increase the performance of the result images for each topic. Thus, the images between two selected images in the retrieval step, are re-selected if the sequential distance between these two images doesn't exceed a certain threshold obtained experimentally.

Figure 3 shows three images selected for query topic number 10 as final output of the proposed approach. The image (b) was rejected in the pre-processing stage and the images (a) and (b) were selected in the retrieval stage. Then, in the post-processing stage, since image (b) is sequentially between the images (a) and (b), image (b) was reselected.

4 Results

We submitted 6 runs on the LMRT sub-task. In this sub-task, the final score is computed as an arithmetic mean of all queries. The ranking metrics was the F1-measure@10, which gives equal importance to diversity (via CR@10) and relevance (via P@10), Cluster Recall and Precision at top 10 results, respectively.

(a) u1_20180523_0022_i05

(b) u1_20180523_0022_i06

(c) u1_20180523_0022_i07

Fig. 3. Some selected images for the query topic number 10.

We describe the last submission (run 6) in Section 3 and the other submissions follow the same pre-processing approach. However, we made changes in the retrieval step and added the post-processing step in the implementation of the other submissions. The post-processing step was only implemented in the two last submissions (run 5 and run 6).

4.1 Run 1

In the first submission (run 1), the query topic is analyzed using the title and narrative. From the title, stop words have been removed using the scikit-learn tools [7]. From the narrative, the nouns and the main verbs associated to an auxiliary verb were extracted using the SpaCy library [5] without rules. Then, we used the same word2vec model, described for run 6, to obtain the similarity of the topic words and concept words extracted in pre-processing step.

In this run, the topic words were not divided into categories. We check if the topic words had a similarity above a threshold and if two or more topic words score above that threshold. In this case, the sequence of images in a certain distance is selected. The confidence score is the same for all selected images.

4.2 Run 2

In the second submission (run 2), we used the same approach of the run 1, however in the query topic analysis we only used the title.

4.3 Run 3

In this submission we only analyze the narrative of the query topic and we define linguistic rules to extract relevant words. The words were divided into four categories, that is, the same categories of run 6 but without the "other words" category. We check if the words in the categories had a similarity above a threshold and if two or more scores of the categories are above that threshold. In this case, the sequence of images is selected. The confidence score is assigned by the number of categories that had the score above that threshold.

4.4 Run 4

This submission follows the same approach of run 3, however some thresholds were adjusted and it is used the fifth category described in the run 6.

4.5 Run 5

In this submission, we reorganize the implementation of the run 4 and we defined different weights for each category of words in order to calculate the confidence score of each image. Then, we added the post-processing step to our implementation.

4.6 UA.PT Bioinformatics Results

The results obtained are shown in Table 1, along with the best result in this task, for comparison. The results of all participating team can be found in [1]. We can observe that our results are still far from the best ones on this task but we consider that our first participation in the ImageCLEF lifelog 2019 was an excellent starting point for our research work. Moreover, the best team already participated in the past.

One of the main problems in our approaches is the low information and visual concepts extracted from the images of the lifelog data, for example the word "toyshop" never appears associated to an image. For this reason, half of the analyzed query topics did not obtain any result in the evaluation metrics. Using other state-of-art algorithms and API's to obtain a more rich description of the images may increase the performance.

Some visual concepts of the images are in form of bigrams or trigrams, for example "electronic device", "ice cream", "cell phone", "car interior", among others. As in our approach we only compute cosine similarity between two words, some of the visual concepts are lost and the result of our approaches decrease.

Table 1. F1-measure@10 of each run submitted by us and the best team run in the LMRT task.

Team	Run Name	F1-measure@10 (%)
Our	Run 1	1.6
	Run 2	2.6
	Run 3	2.7
	Run 4	2.7
	Run 5	3.6
	Run 6	5.7
HCMUS	Run 2	61.04

In order to solve this problem, the identification of bigrams and trigrams is one of the future implementations for our application.

Another way to increase the performance of our work is the development of new linguistic rules, in order to analyze the description of the query topic and obtain more information for the retrieval step. For example: identify negative sentences and exclude some objects or environments that are not relevant.

5 Conclusion and Future Work

The Lifelog Moment Retrieval (LMRT) sub-task of ImageCLEF lifelog 2019 was an excellent starting point for our research work in lifelogging. Although our results have a low score in this sub-task, we observe an evolution in each submitted run. After these results, our goal is to continue the development and improvement of our implementation.

For future work, we intend to do more state-of-art research for the recognition of visual concepts and text mining methods. In order to develop an efficient application, we are going to create ontologies for daily activities and create hierarchical relationships for the words that can appear in the visual concepts.

Developing a user interface is also one of our priorities for user interaction and visualization of search results.

6 Acknowledgments

Supported by the Integrated Programme of SR&TD SOCA (Ref. CENTRO-01-0145-FEDER-000010), co-funded by Centro 2020 program, Portugal 2020, European Union, through the European Regional Development Fund.

References

1. Dang-Nguyen, D.T., Piras, L., Riegler, M., Tran, M.T., Zhou, L., Lux, M., Le, T.K., Ninh, V.T., Gurrin, C.: Overview of ImageCLEFlifelog 2019: Solve my life puzzle and Lifelog Moment Retrieval. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Lugano, Switzerland (September 09-12 2019)

2. Dang-Nguyen, D.T., Piras, L., Riegler, M., Zhou, L., Lux, M., Gurrin, C.: Overview of imagecleflifelog 2018: daily living understanding and lifelog moment retrieval. In: CLEF2018 Working Notes. CEUR Workshop Proceedings, CEURWS. Avignon, France (2018)
3. Dodge, M., Kitchin, R.: 'outlines of a world coming into existence': pervasive computing and the ethics of forgetting. Environment and planning B: planning and design **34**(3), 431–445 (2007)
4. Gurrin, C., Smeaton, A.F., Doherty, A.R., et al.: Lifelogging: Personal big data. Foundations and Trends® in Information Retrieval **8**(1), 1–125 (2014)
5. Honnibal, M., Montani, I.: spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. To appear (2017)
6. Ionescu, B., Müller, H., Péteri, R., Cid, Y.D., Liauchuk, V., Kovalev, V., Klimuk, D., Tarasau, A., Abacha, A.B., Hasan, S.A., Datla, V., Liu, J., Demner-Fushman, D., Dang-Nguyen, D.T., Piras, L., Riegler, M., Tran, M.T., Lux, M., Gurrin, C., Pelka, O., Friedrich, C.M., de Herrera, A.G.S., Garcia, N., Kavallieratou, E., del Blanco, C.R., Rodríguez, C.C., Vasillopoulos, N., Karampidis, K., Chamberlain, J., Clark, A., Campello, A.: ImageCLEF 2019: Multimedia retrieval in medicine, lifelogging, security and nature. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the 10th International Conference of the CLEF Association (CLEF 2019), LNCS Lecture Notes in Computer Science, Springer, Lugano, Switzerland (September 9-12 2019)
7. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
8. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
9. Řehůřek, R., Sojka, P.: Software Framework for Topic Modelling with Large Corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. pp. 45–50. ELRA, Valletta, Malta (May 2010), http://is.muni.cz/publication/884893/en