

HUMAN ACTIVITY RECOGNITION USING WEB BASED SEARCH ENGINES AND COMPUTER VISION API FOR SMART HOMES

¹Priya.S, ²Shreya Sonali, ³Abhijit Raj Narayan Singh

^{1,2}Final Year Student, ³Assistant Professor (Sr.G)

Department of Computer Science and Engineering,

SRM Institute of Science and Technology, Kattankulathur

¹shreyasonali2502@gmail.com, ²arnsrajawat@gmail.com,

³priya.sn@ktr.srmuniv.ac.in

Abstract: Human activity observation and identification is of high importance in the field of smart home automation. It is highly essential to maintain confidentiality in monitoring activities of daily living (ADLs) performed by people in their houses. As of now, the current approach to ADL recognition has several limitations that limit their application in real world scenarios of smart homes. In this project, we have a surrogate method, which comprises of mining the ADL models from the Web. Already existing approach in this method are focussed on Web page mining for textual content and lexical analysis of the textual content. These attempts face the persistent problem of high level of noise and unreliability of content found in textual data present on the web pages. To conquer this, we focus on the images depicting the activity being carried out, which presents a visual and more expressive description of the activity compared to textual content present on the Web pages. Hence, we aim to achieve human activity recognition for a regular household using the Web-based activity mining based on image searches through image mining from the web and computer vision APIs in order to perform an efficient and effective activity recognition in the pervasive computing domain of smart home automation. The method proposed here depends upon search engines such as (e.g. Google, Mozilla, etc.) to search and display highly related images for each queried activity. Computer vision APIs along with a lexical database is required to extract the key objects that are appearing in searched images. Probabilistic technique is used to point out the degree of relationship or relevance between the identified objects and the activity that is being performed in the image. The project is implemented and tested on a large dataset of real-world ADLs which are executed in a smart-home. Thus, the project focuses on implementing a better approach for activity recognition via Web-based activity mining, which relies on images present on the web pages rather than textual content present on the web pages.

Keywords: ADLs (Activities in daily living), Activity recognition, Web-based activity mining, Computer vision, Smart home automation.

1. Introduction

Activity recognition is an imperative requirement in detecting and categorizing abnormal activities. It is needed in a lot many places especially smart homes. Supervised learning is useful for recognizing activities using postures and occurrences. But it is difficult for supervised learning in case of complex ADLs which occur without any defined sequence (e.g. packing a gift, watering a plant). That is why unsupervised learning methods were tried to be created to recognize the ADLs from sensor data. Unsupervised learning relies on symbolic models for recognizing activities. Manually defining the models via formal ontologies presented using description logics language is a significantly popular approach [4]. These models can be defined manually but it becomes a cumbersome process. Web mining is an alternate approach to retrieve the symbolic models from the web pages. The method is widely denoted as web based activity mining. [2] The current mining technique mostly uses textual content. But there is a limitation to this mining as it involves more noise which involves unnecessary computation. So it is assumed that mining using pictures would be more efficient as it involves more concise and expressive information as compared to that of text of a web page. [1] Therefore the current system used textual based content for mining and recognizing activities while we aim to implement the system proposed of using image based mining confined for smart homes to recognize activities and find out the difference between the two systems and conclude the better system based on various parameters. [5] We use the activity labels related to smart homes as our comparison is done on the activities related to a smart home.

Major contribution of the work are as follows:

- To combine image search using web and computer vision operations to extract related objects from the images for different activities, as per our requirements.
- Using probabilistic weighting techniques to measure relevance between activities and associated objects.
- The results are presented using the large-dataset based on real-world ADLs, and to compare and prove that the existing web-mining method based on mining only textual data from the web pages is being outperformed by the proposed method.

2. Literature Survey

The supervised learning methods to perform the recognition of ADLs has various issues pertaining to applicability, cost effectiveness and maintaining privacy of subjects. Therefore, different approaches have been made to create unsupervised learning ADLs recognition methods. The approach has drawbacks in terms of defining the activity definition and manually describing the ontological axioms. Another approach focuses on to automate activity modelling and non supervision by mining activity models from the web. In [2] this is being attempted and scaled up with refining later by other approaches on top of it. The approach is shown in Fig. 1. Here, the input is taken as an activity label, say “washing the dishes”. This is used to query the web based search engine to find web pages showing results pertaining to that activity. Further, using the genre classification functionality to select pages that contain description of that activity in significantly detailed manner. [3] The text from top n pages is passed to an object identification module, which exploits a lexical database in order to extract relevant objects pertaining to that activity.

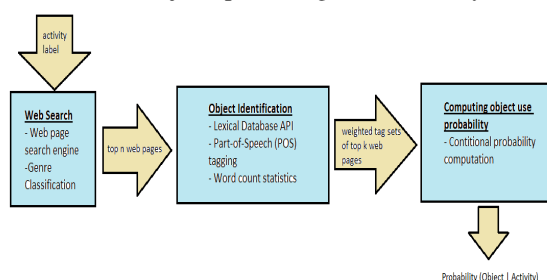


Figure 1. Activity mining based on textual data

present on the web pages. (Existing method)

Similarly, in [3] it is used for activity recognition as well as segmentation functionality. Statistical and ontological reasoning are combined and used to infer probable activities.

Different methods and techniques have been used to infer activities and objects, but to a large extent all the existing methods rely on textual content only. We have here exploited the image content present on the web pages in order to derive and define the activity models. [5]

3. Approach and Implementation

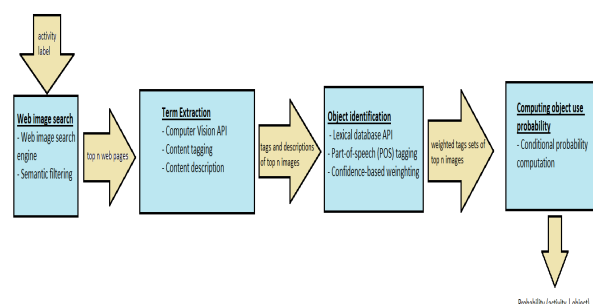


Figure 2. Activity mining based on images from the web pages and use of third party computer vision application. (Proposed method)

If we give an activity label (e.g. “watering the plants”), the model queries the web based search engine for the images and fetches the top n images for the given activity label. It should also satisfy the semantic filtering directive. Semantic filtering depends upon analysis performed on the textual data present on the web pages that also contain queried images. All the first appeared n images are passed through the term extraction module of the architecture which queries the third party supported computer vision API to retrieve the appropriate description and relevant tags of objects or elements that are identified in the image. Each relevant tag is linked with a confidence value that depicts the probability that the object actually appears in the image performing the activity. The object identification module is now given the tags and descriptions of the top n images which implies POS tagging to only save the nouns from the pool of obtained tags. Weight based on the tag’s confidence value is computed for each tag found in the image. The above method is performed for each and every activity a that belongs to the set A of activities that are taken into consideration. The objects from the considered n images are used for computation by object use probability module. This module is used to compute the correlations among the objects and activities. Let’s take for an example, for every activity a and provided object o that is observed to be in at least one of the images. Then the module estimates it to be $P(a/o)$. This is conditional

probability for an activity (under consideration) a , given that the depicted object in the activity is o .

Above method is implemented using Python 3.6 as the programming language. Packages such as nltk (natural language processing toolkit) are used to perform operations such as POS tagging, etc. The MongoDB server is used in order to create a server instance running on the same machine. During the implementation Google is used in order to query for the images on the web. Microsoft Azure based computer vision API is used for cognitive analysis of the images and to get the tags and description of the images. The CASAS dataset is used for the real-world implementation using the real-world ADLs. The MongoDB server instance is created first on the machine using the command shell. Once the server instance is created and the port is ready to listen. The program that implements the discussed method is executed. The predefined activity labels are already present in the program. Based on the activity labels, the top n images are downloaded from the web using Google image search for each activity label. For convenience, n here is taken as 15. These images are stored in Pictures folder under the respective activity label named sub-folders. Each of these images is passed through the Microsoft's computer vision API. The tags and description obtained for each of the image are stored in form of JSON output in json_output folder under the respective activity label named sub-folders. Using the probabilistic techniques, the conditional probability for occurrence of an activity when certain object is already given is calculated. Further the confusion matrix is created after the calculation of true-positive, true-negative, false-positive and false-negative. Then, the overall Precision, Recall and F1 score is calculated for the occurrence of an activity. F1 score refers to the harmonic mean of Precision and Recall. The graph for the accuracy containing Precision, Recall and F1 score is plotted and compared with the results derived using the existing method.

4. Conclusion and Future Scope

A new approach is being implemented here to Web-based activity mining, which is based on image content on web pages rather than textual data present on the web pages. The technique presented here extracts appropriate images of an activity from the web using the search engines. It identifies the crucial objects using the computer vision API. Further, computing the relevant weights for each of the object.

This work has a lot of potential for future enhancements. The methods to obtain relevant images can be more refined and improved by also analysing the textual content present on the web pages on which the

images appear. Instead of using a general purpose computer vision API, the recognition for the objects present in the images can be improved by adopting particular methods to recognize the interaction between the human and the objects.

References

- [1] N. Davies, D. P. Siewiorek, and R. Sukthankar, "Activity-based computing," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 20–21, 2008.
- [2] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Proceedings of Pervasive Conference*. Springer, 2004, pp. 1–17.
- [3] A. Calatroni, D. Roggen, and G. Tröster, "Collection and curation of a large reference dataset for activity recognition," in *Proceedings of Systems, Man, and Cybernetics Conference*. IEEE, 2011, pp. 30–35.
- [4] D. Riboni and C. Bettini, "OWL 2 modeling and reasoning with complex human activities," *Pervasive and Mobile Computing*, vol. 7, no. 3, pp. 379–395, 2011.
- [5] M. Perkowitz, M. Philipose, K. P. Fishkin, and D. J. Patterson, "Mining models of human activities from the web," in *Proceedings of WWW Conference*. ACM, 2004, pp. 573–582.
- [6] P. P. Palmes, H. K. Pung, T. Gu, W. Xue, and S. Chen, "Object relevance weight pattern mining for activity recognition and segmentation," *Pervasive and Mobile Computing*, vol. 6, no. 1, pp. 43–57, 2010.
- [7] K. Ihianle, U. Naeem, A. H. Tawil, and M. A. Azam, "Recognizing activities of daily living from patterns and extraction of web knowledge," in *UbiComp Adjunct Proceedings*. ACM, 2016, pp. 1255–1262.
- [8] L. Chen and C. Nugent, "Ontology-based activity recognition in intelligent pervasive environments," *International Journal of Web Information Systems*, vol. 5, no. 4, pp. 410–430, 2009.
- [9] D. Riboni, L. Pareschi, L. Radaelli, and C. Bettini, "Is ontology based activity recognition really effective?" in *Proceedings of the Ninth Annual IEEE International Conference on Pervasive Computing and Communications Workshops*. IEEE, 2011, pp. 427–431.
- [10] R. Helaoui, D. Riboni, and H. Stuckenschmidt, "A probabilistic ontological framework for the recognition of

multilevel human activities,” in Proceedings of ACM UbiComp. ACM, 2013, pp. 345–354.

[11] prints,” Data Knowl. Eng., vol. 69, no. 6, pp. 533–544, 2010.

[12] M. Rajesh, Manikanthan, “GET-UP-AND-GO EFFICIENT MEMETIC ALGORITHM BASED AMALGAM ROUTING PROTOCOL”, International Journal of Pure and Applied Mathematics, ISSN NO:1314-3395, Vol-116, No. 21, Oct 2017.

[13] T. Padmapriya, V.Saminadan, “Performance Improvement in long term Evolution-advanced network using multiple input multiple output technique”, Journal of Advanced Research in Dynamical and Control Systems, Vol. 9, Sp-6, pp: 990-1010, 2017.

