PURIFY NOISY DATA FROM ANNOTATED IMAGES USING MONTYLINGUA AND CONTROL REDUNDANT TERM

Rooh Ullah, Jafreez Jaafar and Abas B Md Said
Universiti Teknologi PETRONAS, Tronoh, Perak, Malaysia
E-Mail: Roohullah.orc@mail.com

ABSTRACT

Dynamic growths in the field of digital data and new techniques (manual and automatic) are introduced to tag images. Tagging of an object within the image is labeled in different terms base on the user perception. LabelMe is the image datasets that give a user online access to labeled object through by webtool. However, there are a number of noisy terms and errors found in the annotated list. Nevertheless, sometime a user tags the same objects with repeated terms. It requires to pruning the dataset from errors, noisy keywords and reduces to one instance term. This paper uses Montylingua for two purposes. First, it converts the tag term into base form. Second it purifies the irrelevant terms from the list. Next reduce the repeated terms into one instance and display their total count of occurrence. An experiment work, it shows that the purified list of the tagging has successfully removed from the annotated images. The result depicts through tagging ratio as well as degree of retrieval for effective achieved.

Keywords: purify, noisy terms, labelme, montylingua, annotation.

INTRODUCTION

From the last decades, multimedia content has been the largest dataset. It organizes contents in a rich and complex way through label tag activities. Current existing annotation techniques [3, 4] considered the labels associated with the images to be devoid of errors and connection to a small fixed vocabulary. This is used directly for annotation schemes designing. Image annotation is vital issue that has proficiency of retrieving relevant images from the huge digital datasets depend on semantic concepts or keyword annotations. On the other hand, the labels collected by collaborative tagging websites are noisy such as misspelled, redundant, irrelevant to content, and/or unlimited in numbers. Since each document often has only a few annotation words that describe the contents of images (incomplete). From time to time, progress in the image classification problem has been achieved by using more powerful classifiers and building or learning better image representations. In automatic image annotation, computers are able to learn low level features correspond to high-level concepts. Several developed system can automatically tags the image object [5] based on bag-of-words (BoW). However, there is a problem that some of the objects tagged insufficient words, which in turn leads to the problem of retrieval with the low accuracy result. A particular example is a LabelMe which gives free web based annotation and has no proper checked to handle the user mistake or unusual label terms. In other words, user can tag irrelevant terms in the image object. Thus, an interesting problem to address is on how to purify the noisy tag terms from the label images.

Digital data-sharing web services such as social website pages, news image archives all offer collections of images that have been manually tagged by keywords [6]. It often tags mark physical things inside the image (name of object, landmark, location), which grants the user to retrieve the essential image within the massive collections using a simple text-based search. However, images naturally multi-label, which mean each image, should assign more than single keywords. A times a user label word to an image with irrelevant, which cannot be handled properly [2]. There is no such method that can stop this type of irrelevant label terms. This paper, introduce the purified method that can help to stop these types of irrelevant terms on image. It can be possible for the user to search the relevant information from the massive dataset.

The aim of this paper is to purify the irrelevant terms from the annotated list and reduces the same label terms into one instance. First, it verifies the tag terms with the help of Montylingua, to lemmatize the labeled terms into base form. Subsequently, the base form of label terms makes a sequence as a unique. Then, it filters the terms through tagging the part-of-Speech (noun, verb and adverb). The decision of qualifying / disqualifying of any object or object name is based on the tagging. If the object name got the tag (noun, verb, adverb), it declares as qualifying else declares as a disqualifying and stops to allow for further process. After that, redundancy control is used to reduce the occurrence of the same term into one instance. It selects the object name from the annotation set (purified: stopping-words) of the image and then count their occurrences within the purified annotation list and record their count along with the object name. The purified image terms will store in the new created XML file for future uses.

LITERATURE REVIEW

Digital dataset (LabelMe, Flickr) websites believed that most of the image labels are accurate, even though there are many irrelevant and redundant label's tags. It can be observed that around 40-50% of the label's tag [7] is inappropriate, and the image representation is out of context. The degree of freedom while using the LabelMe online Annotation tool makes the users comfortable on one side, but it gains complexity in terms of usability of datasets for research. It creates a problem such as redundancy, irrelevant and inapplicable keywords.

The existing object recognition techniques usually apply in the human labeled training images. It then classifies and attempts to conclude the correlation or join probabilities between the query and Annotation keywords for images [8]. Semi-supervised learning methods have been widely used for image Annotation [9]. It gives limited training dataset to label the image objects. Its learning methods can influence both labeled and unlabeled data. Its performance still quite depends on the amount of labeled data. Therefore, it is usually rather time-consuming and costly to assemble enough well labeled training data to accomplish functioning in large-scale scenarios.

Image is one of the significance information of digital data. It represents information in many words. The searcher who retrieves images depends on Annotation. It believes that the quality of the Annotation should be taken into narratives. Manually label annotations are mostly more reliable than automatically assigned ones. However, it is time-consuming and costly. Most of the researchers have focused to automatically assigned Annotation [5]. Two types of automatic label assign methods used for image annotations. First method is based on information extraction techniques. The image annotates through by World Wide Web (WWW) extraction on surrounding texts or anchor texts linked to the images [1]. The Annotation on this technique depends on the textual information corresponding that used for annotate the images. Nevertheless, there are no rules between WWW extraction and Annotation, which images can correct to annotate. The second method is of automatically assign label which is based on the classification techniques. This procedure for assigning keywords to images is an active research topic [11].

Annotation can be regarded as a category of multi-class image classification. The number of classes is large, and the amount of data for each class is small. It is the crucial problem for the automatic assigning labels that represents correlations between the visual features and linguistic concept [10]. The user can assign different keywords for the same object on different environments, which make noisy words, and it reduces the accuracy of the information retrieval system.

Purify of these noisy words from the tagged image, researchers have developed methods to reduce it. Unsupervised label refinement (ULR) technique has developed by [12] to reduce the unusual words from the tagged list. Their technique was based on Matrix to describe the weak label information on the face detection to stop the unusual terms. Furthermore, Correlated Multi-label refinement technique has developed by Tie et al [13]. It has refined the semantic noisy retrieval result based on a clustering algorithm Double-Circles to remove a label. In a different work Chandrashekar et al [7] defined K-nearest neighbor method to tag the images with the sample of neighbor images used for Annotation of unseen image. However, the noisy and unusual terms are still not controlled by these methods. To that effect, the purification method to filter the label terms from images based on Montylingua [14] is removed in order to the stop words and control redundant instances of keywords attached to one instance.

METHODOLOGY

The purpose of this paper is to purify each contributor (object) in the image toward image semantics. For experimental work, choose LabelMe images dataset. The problem with the LabelMe dataset is their open nature of Annotation on the web. It allows everyone to draw a sketch on the image and tag with user-define words, and hence generate a lot of noisy data. It is needed to prune the dataset from such kinds of noises. Purification model define as a filtration process to purify the image annotation list and save as a new created XML file. The proposed model for the Purification is shown in Figure-1.

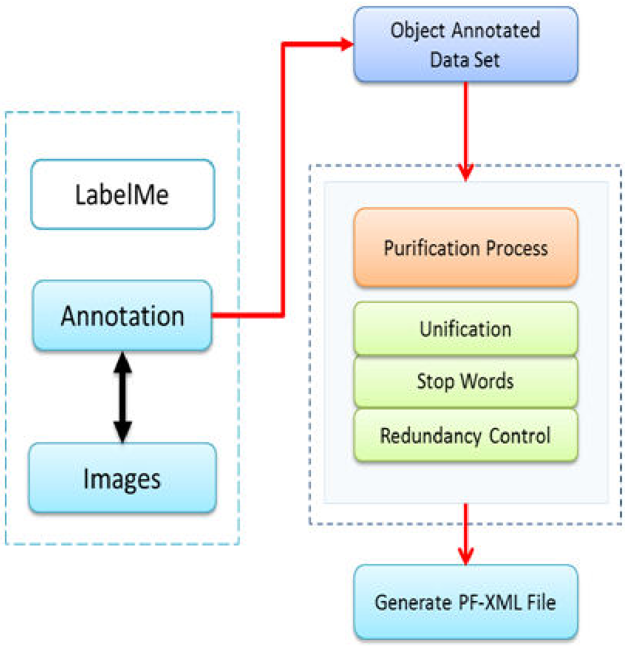

Figure-1. Purification of Annotated images.

Figure-1 illustrates the model of Purification process. LabelMe image datasets use as an input, which provide object-annotated data to the Purification. The Purification process is used to purify and verify the input by applying 3-level of the filtration process. After filtration, the output is stored in a new created structure of PF-XML file with the purified object-names.

LabelMe with Object Annotated Dataset

For the experimental purpose, this work uses an open source which is freely available object-annotated image data sources called LabelMe. It contains 187,240 images, 62,197 annotated images, and 658,992 labelled objects as of October, 2010 [15]. The flexible nature of the LabelMe web-tool provides an opportunity to the user to annotate an object in the image by sketching the border of the object and tagged with user-defined words. As a result,

VOL. 10, NO. 23, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

a noisy and unnecessary data is generated. The structure of the XML file used by the LabelMe webtool is shown in Figure-2.


```
<?xml version="1.0"?>
- <annotation>
    <filename>p1010755.jpg</filename>
    <folder>05june05_static_street_boston</folder>
  - <source>
      <sourceImage>The MIT-CSAIL database of objects and scenes</sourceImage>
      <sourceAnnotation>LabelMe Webtool</sourceAnnotation>
    </source>
    <scenedescription>street urban city outdoor</scenedescription>
  - <object>
      <name>building</name>
      <deleted>0</deleted>
      <verified>0</verified>
      <date>24-Oct-2005 20:38:25</date>
    - <polygon>
        <username>lhr</username>
      - <pt>
          <x>323</x>
          <y>341</y>
        </pt>
      - <pt>
          <x>309</x>
          <y>460</y>
        </pt>
      - <pt>
          <x>305</x>
          <y>628</y>
        </pt>
      - <pt>
          <x>286</x>
          <y>924</y>
        </pt>
      - <pt>
          <x>282</x>
          <y>1334</y>
        </pt>
      - <pt>
          <x>318</x>
          <y>1398</y>
        </pt>
      - <pt>
```


Figure-2. Structure of XML file used by LabelMe webtool.

Figure-2, XML includes source folder, image/file name, object name and their polygon values. It also describes sourceimage, sourceAnnotation, date and time of creating and annotation perform on the specified image. The XML file used by the LabelMe webtool is the main source of input for the Purification Model.

Purification Module

The flexible nature of the webtool for the LabelMe online Annotation tool makes the users comfortable in one side of image annotation. However, it increases the difficulty in terms of usability of such data for research. It creates problems like unusual, irrelevant and redundancy keywords are incessantly produced during the Annotation. The best approach to minimize the risk during the Purification Model (PM), elongate the PM to further sub-three-modules, i.e. unification, stop words and redundancy control. The result produced by the PM is the purified form of data for the source image. The following is the details of each sub-module.

Unification

Unification is the procedure of adapting the multifaceted terms into the simple term. The aim of this module is two folds. First convert terms into base form. For example, words like “fished”, “fishing”, “fisher” while their base form is “fish” which will convert with the help of Lemmatization in Montylingua. Second it makes unique form regardless terms in sequence. The entire unified object name list will pass through the ascending sorting process to avoid any confusion during data reading.

Stopping Words

The stopping words sub-module is used to stop noisy object tagged with the images from being further processing. For example, the irrelevant words like “aahrdls”, “dasf”, “ddddd”, “zaxwxc” need to be discarded straight away from the Annotation list. A Part-of-Speech Tagger of the Montylingua is used once again for stopping words. During the processing, the name of each object is tagged with Part-of-Speech tagger. The decision of qualifying/ disqualifying of any object or object name is based on the tagging, if the object name got the tag (noun, verb, adverb), it is declared as qualifying else declares as a disqualifying. It is restricted to allow for further processing.

Redundancy Control

Redundancy is the crowning communal problem subsists in the LabelMe object annotated datasets. It is due to the presence of too many similar objects in the image. Redundancy control has two aims (1) to decrease the processing load and (2) preventing redundancy in outcome.

It selects the object name from the Annotation set (purified: stopping-words) of the image and then count their occurrences within the purified Annotation set. It records their count along with the object name. The process modifies the existing Annotation to set by reducing all the redundant object name to one instance. It does not change their other related properties and put their polygon values under one object name.

Generate XML for PF-Model

The next task is to store the data in the same format and structure as extracted from the original XML file. It creates a separate new file called PF-XML and stores these files in a separate folder within the Annotation folder. The PF-XML file contains only the purified data. The structure of the PF-XML file is shown in Figure-3.

Figure-3 demonstrates the PF-XML file structure. The structure of the PF-XML file is almost the same as presented in Figure-2. However, the exception is the updates of Purified and control of total occurrence of the terms with <*count*>. Moreover, the PF-XML file includes only the purified data with their total count terms.

VOL. 10, NO. 23, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com


```
<?xml version="1.0"?>
- <annotation>
    <filename>p1010755.jpg</filename>
    <folder>05june05_static_street_boston</folder>
  - <source>
      <sourceImage>The MIT-CSAIL database of objects and scenes</sourceImage>
      <sourceAnnotation>LabelMe Webtool</sourceAnnotation>
    </source>
    <scenedescription>street urban city outdoor</scenedescription>
  - <object>
      <name>air conditioning</name>
      <deleted>0</deleted>
      <verified>0</verified>
      <count>1</count>
      <date>28-Apr-2008 15:13:33</date>
      <id>24</id>
    - <polygon>
        <username>admin</username>
      - <pt>
          <x>225</x>
          <y>220</y>
        </pt>
      - <pt>
          <x>247</x>
          <y>238</y>
        </pt>
      - <pt>
          <x>245</x>
          <y>265</y>
        </pt>
      - <pt>
          <x>228</x>
          <y>250</y>
        </pt>
      </polygon>
    </object>
  - <object>
      <name>armchair</name>
      <deleted>0</deleted>
      <verified>0</verified>
      <count>2</count>
      <date>28-Apr-2008 15:13:33</date>
      <id>46</id>
```


Figure-3. PF-XML-file structure.

Purification Module

All the annotations were firstly passed through the purification module to get the purify data. After passing through the purification module, a lot of unnecessary data were removed from the Annotation. Figure-4 represents the annotated tagged with the image before the purification process.

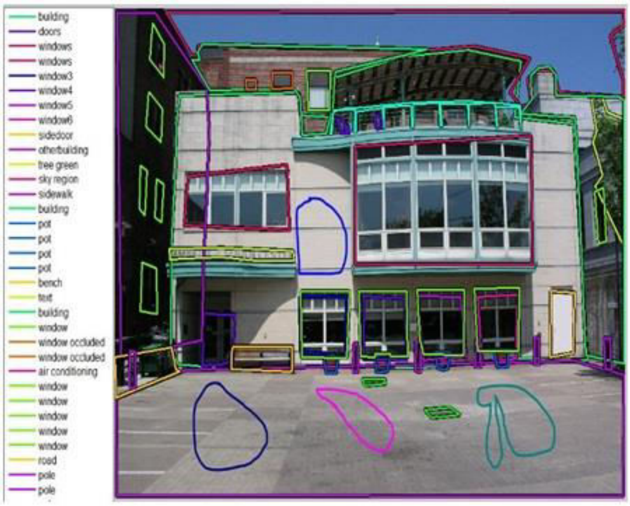

Figure-4. LabelMe image with noisy and unusual terms.

Figure-4 represents the LabelMe image along with the Annotation set. It includes a lot of unusual and noisy objects that are needed to be removed from the annotation list before processing. For example, the colour lines on the road and on the building are drawn by the users without any Annotation or the Annotation (object) that is meaningless. All of these kinds of words/ object need to be removed from the Annotation set before processing. The three steps purification process is performed on the Annotation set like this to purify their data. Figure-5, shows the first step of the purification model.

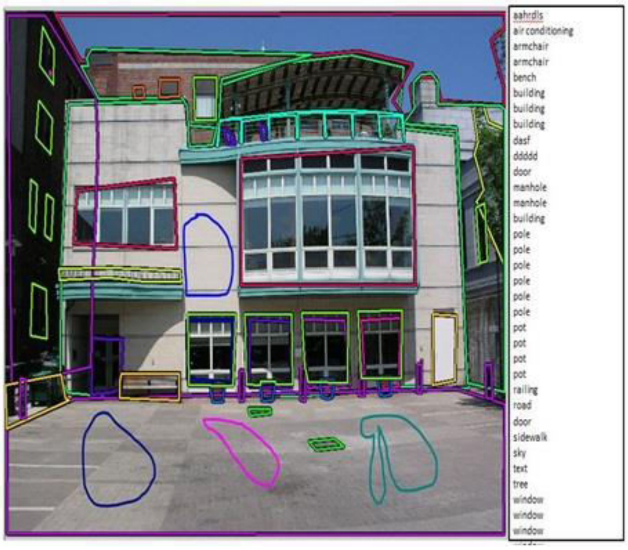

Figure-5. Unification form of the LabelMe image.

Figure-5 represents the output of the unification, where each of the composite words is removed and changes the keywords to its base form. For Unification use Montylingua (an open sources NLP software freely available) the limitation process that first removed fabricated word and change it into keywords form and then convert into root form. After it, the annotation list is done alphabetically sequence from ‘A/a’ to ‘Z/z’. For this use the Matlab building unique function that sequence overall the annotation list. However, still in the annotation list have unusual terms such as ‘asdwd’, ‘hjkhkhkghk’ available. The subsequent figure shows the filtration process of the stopping words performed over the Annotation sets of the image. Figure-6 represents the second step of the purification module which is stopping words and their output is shown.

Figure-6 illustrates the view of the objects without noisy data. All the noisy data were removed during the stopping words process of the Purification model. The stopping words removed the unusual terms such as ‘asdwd’, ‘hjkhkhkghk’ from the annotation list. For this process use in Montylingua the Part-of-Speech (POS) tagging process. Part-of-Speech tagging only work on the root form of terms. It filters only the POS terms (noun, verb, and adverb) from the annotation list. In the figure shows the terms that match with POS still remain in the annotation list and the rest were removed. The stopping word method was exercised on different selected images. Next it will check the tagging ratio of the

VOL. 10, NO. 23, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

annotated images before and after the stop words of the purification module.

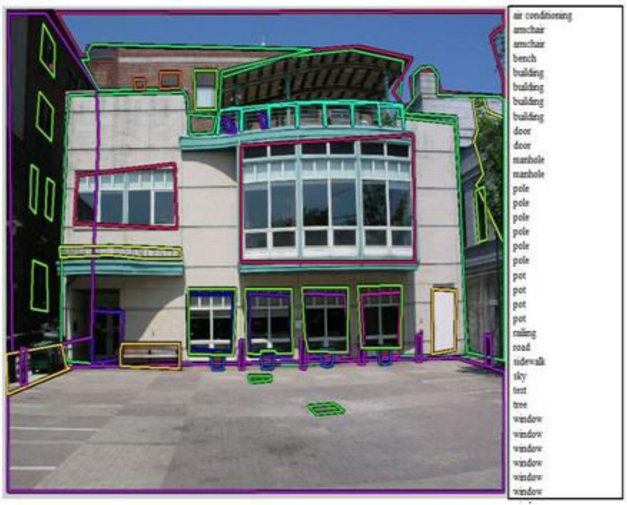

Figure-6. Stop words of the unified image.

Tagging Ratio

Tagging is the average number of labels tag per image. It represents the tagging ratio within the image. Tagging formulas are clarified in the following equations,

$$T_1 = \frac{\sum_{i=1}^{n}(C_i)}{N} \quad (1)$$

$$T_2 = \frac{\sum_{i=1}^{n}(C_i)}{N} \quad (2)$$

An equation (1), T_1 is the tagging ratio before data filtration process, and equation (2), T_2 is the tagging ratio after data filtration process, while C_i represents the number of concepts tag with in the image, N is the total number of images in the corpus and 'n' is the total number of tags of the image respectively. Figure-7 shows the tagging ratio of the purification model.

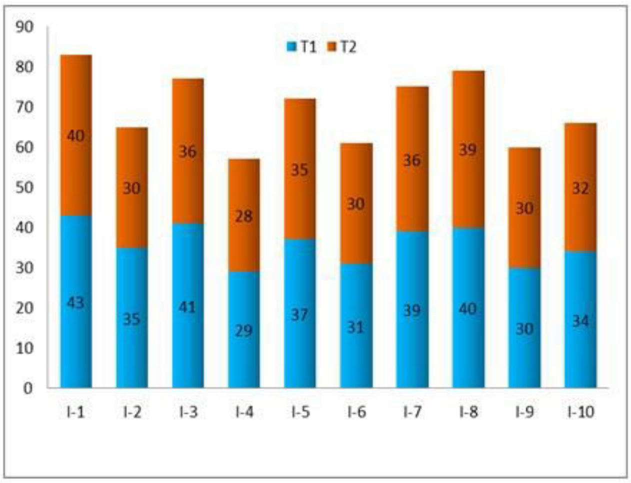

Figure-7. Shows the number of tags per image of the 10 sample images.

In Figure-7, images taken from the LabelMe dataset, where T1 and T2 represents the number of tags before and after data filtration process. It depicts the tagging ratio of the randomly selected 10 sample images. The image I-1 in Figure-7 was initially tagged with 〖T〗_1←43, these tags are then refined to 〖T〗_2←40. This decreases the number of tags as there were three unusual terms removed in the purification process. Similarly, the same processes have done for I-2 to I-10 respectively. For I-2 T1⌈ was 35 and then 5 irrelevant terms removed in the purification process and remain T2 ⌉30. Maximum number of irrelevant terms removed from the I-2 and I-3. Some o the sample have less numbers of irreleant terms. Such as I-4, I-6 and I-8 have only one term removed. I-9 has no irrelevant term and their T1 and T2 has same numer of tagging ratio. It is observed variations among the tagging ratio for different images, which is because of some of the images are simple while some of them are semantically enriched. The concepts in the simple images are limited and as result there has limited objects tag. While the semantically enriched images consist of a large number of concepts and constitute a large number of tagging. As a result, their a large number of irrlevant terms exist in the tagging ratio. Figure-8 shows the redundancy control of the annotated images.

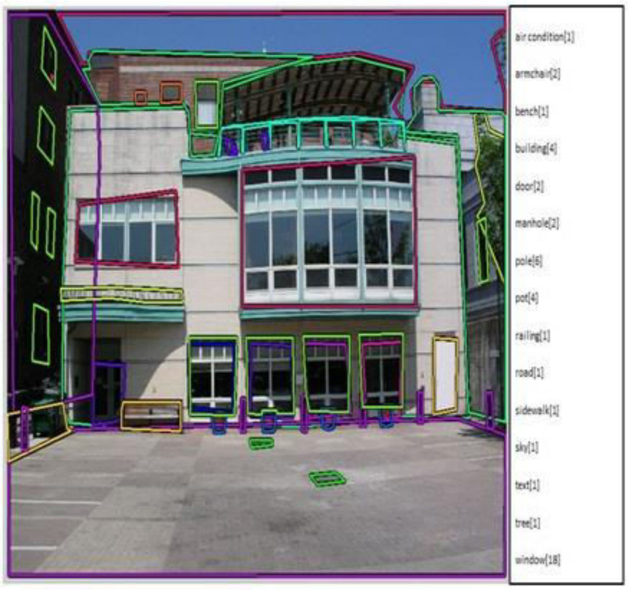

Figure-8. Redundancy control term.

Figure-8 illustrates the image of the Annotation information. It represents the effectiveness of the Redundancy control. The Redundancy control has designed to collect the occurrences of all the objects in the Annotation set of the images and store the resultant/ output in the form as specified. The annotation list already sequence by the Unification process, while irrelevant and unusual terms removed by the stop words process is suitable for Redundancy control. The Redundancy control reduce the repeated terms into one and their total occurrence write in bracket. For example, the keyword 'window' Eighteen (18) times occurrence in the annotation list. The redundancy control method reduce to only one time occurrence and in bracket write total numbers of occurrence in the overall image annotation list. Same as used for other terms in the annotation list, such as 'pole' six (6) times repeated and so on for others repeated objects. From the experimental work, it depicts that the method designed for the redundancy control perform efficiently in the process. The Figure-9 shows the object name with their total occurrence in the annotation list.

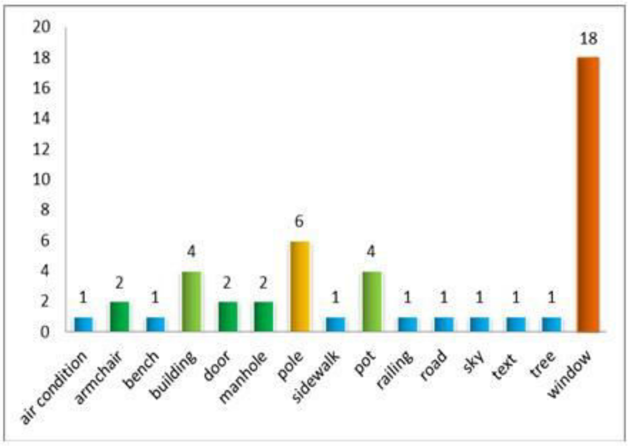

Figure-9. Chart representation of same term occurred in Figure-8.

Figure-9 illustrates the occurrences of an object of the image that shows after performing the redundancy module in the chart form. The colour of each of the object name is represented in the Bar-Chart having the number of occurrences on the top of each of the bar. It showed the object name with their number of tags in the annotation list. It represents number of same type object exist in the image. The highest occurrence is the "Window" which is Eighteen (18) times repeated in the image. The next highest object is "pole" which is Six (6) times repeated, "Building" and "pot" which are repeated 4 times in the Annotation set of the image.

Degree of Retrieval

Retrieval degree is the number of correct images retrieved with a simple single word query. For the purpose of experiments, the LabelMe search engine was used, which is freely available with the LabelMe toolset and is working through the string matching mechanism. The outputs produced by the LabelMe search engine with the original data as a baseline and compare the result with the proposed work. It achieved a great improvement in terms of degree of retrieval. The Figure-10 shown the retrieval degree of the different words queries performed on LabelMe dataset.

Figure-10 describes the retrieval degree of the randomly selected terms from the LabelMe corpus. The selected terms are either single concept words or multi-concept words. For instance, the term like 'car' is a single concept word, while the term like 'building' is a combination of several other concepts such as 'window', 'door', 'wall', and 'floor' are existed. Figure-10 shows a significant improvement of the proposed model over the baseline in terms degree of retrieval. It is due to the fact that original dataset consists of a limited number, irrelevant and redundant of tags attached to the images. Due to query searching, the same type of an image retrieved, which is low the accuracy. However, the same approach attempts on the purified dataset and retrieved the high accuracy result. The baseline approach on original dataset for the 'car' query retrieved 40% out of hundred, while the propose model dataset retrieved 67% relevant. Same their result for the others search query. All these results lead to the considerable improvement in term of degree of retrieval. These results exhibit that searching and retrieval for images of the purified model dataset is highly achieved as compared with the original dataset.

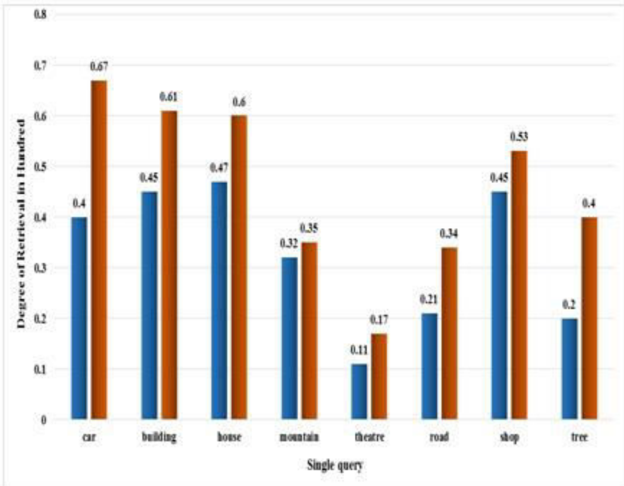

Figure-10. Degree of retrieval.

CONCLUSIONS

In this paper purification model was presented to purify the label tag list of images. There are numbers of images annotation list found with the irrelevant and unusual object tag on it. The purification model was used to purify the Annotation set from the noisy/unnecessary data and get the purified list of the object name. It purified the dataset by used Montylingua. It first used the unification process that converted the label terms into base from and then unique all the terms in ascending. Subsequently, it used stopping words process that removed the irrelevant terms with the help Part-of-Speech tag. Next it has used redundancy control to reduced repeated object name into one instance and showed their total occurrence in a bracket. All the experiments have done on the LabelMe dataset. Results showed that successfully removed the noisy terms from the Annotation list and object occurrences reduced into one term in the list. The Purified model dataset achieved a high accuracy result in term of degree of retrieval as compared with the original dataset.

REFERENCES

[1] D. Zhang, M. Md. Islam and G. Lu. 2013. Structural image retrieval using automatic image annotation and region based inverted file. Journal of Visual Communication and Image Representation. Vol. 24, No. 7, pp. 1087-1098.

[2] A. Makadia, V. Pavlovic and S. Kumar. 2010. Baselines for image annotation. International Journal of Computer Vision. Vol. 90, No. 1, pp. 88–105.

[3] D. Zhang, M. Md. Islam and G. Lu. 2011. A review on automatic image annotation techniques. Journal of Pattern Recognition. Vol. 45, pp. 346–362.

[4] D. P. Tian, X. F. Zhao and Z. Z. Shi. 2014. An Efficient Refining Image Annotation Technique by Combining Probabilistic Latent Semantic Analysis and Random Walk Model. Journal of Intelligent Automation and Soft Computing. Vol. 20, No. 3, pp. 335-345.

[5] C. F. Tsai. 2012. Bag-of-Words Representation in Image Annotation: A Review. Journal of International Scholarly Research Network Artifical intiligence. Vol. 2012, pp. 1-19.

[6] S. J. Hwang and K. Grauman. 2012. Reading between the Lines: Object Localization Using Implicit Cues from Image Tags. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 34, No. 6, pp. 1145-1158.

[7] V. Chandrashekar, S. Kumar and C. V. Jawahar. 2013. Image Annotation in Presence of Noisy Labels. PReMI 2013, LNCS 8251, Springer-Verlag Berlin Heidelberg. pp. 381–389.

[8] D. Wang, S. C. H. Hoi, Y. He, J. Zhu, T. Mei and J. Luo. 2014. Retrieval-Based Face Annotation by Weak Label Regularized Local Coordinate Coding. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 36, No. 3, pp. 550-563.

[9] P. Ji, N. Zhao, S. J. Hao and J. G. Jiang. 2014. Automatic image annotation by semi-supervised manifold kernel density estimation. Journal of Information Sciences. Vol. 281, pp. 648-660.

[10] J. Jeon, V. Lavrenko and R. Manmatha. 2003. Automatic image annotation and retrieval using cross-media relevance models. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 119–126.

[11] J. W. Jeong and D. H. Lee. 2014. Automatic image annotation using affective vocabularies: Attribute-based learning Approach. Journal of Information Science. Vol. 40, No. 4, pp. 426-445.

[12] W. Dayong, C.H. H. Steven and H. Ying. 2011. Mining Weakly Labeled Web Facial Images for Search-based Face Annotation. International conference SIGIR’11, ACM. pp. 1-10.

[13] H. Z. Tie, W. Ling, S. S. Ho, K. L. Yang and H. R. Keun. 2010. Correlated Multi-label Refinement for Semantic Noise Removal. ICIC 2010, LNAI 6216, Springer-Verlag Berlin Heidelberg. pp. 309–316.

[14] http://web.media.mit.edu/~hugo/montylingua/version 2.1. Last visited 12-10-2015.

[15] I. Ullah, N. Aslam, J. Loo, R. Ullah and M. Loomes. 2011. Adding semantics to the reliable object annotated image databases. Journal of Procedia Computer Science. Vol. 3, pp. 414–419.