Automatic Caption Generation for News Images

[1]RINKU BALAN K, [2]LISSY ANTO P, [3]GREESHMA SUNNY
[1] *MSc .COMPUTER SCIENCE*, [2]*ASSOCIATE PROFESSOR*, [3]ADHOC FACULTY
[1]*DEPT. OF COMPUTER SCIENCE*,
[1]*ST.JOSEPH'S COLLEGE*, [1] *IRINJALAKUDA*,[1]*INDIA*

Abstract— **Captions are essential components associated with images to make search engines to respond easily with user queries. Making appropriate captions for images is a difficult task. By making the appropriate caption will help the user to search images with long queries. But most of the images are associated with user annotated tags, captions and text surrounding the images. This paper is concerned with the task of automatic caption generation for news images in association with the related news article. In this method we will input one image and news article to the system. The system will generate most important keywords which are associated with the image. And using these features we will compare the image with the images which are stored in the database. After finding the best matched image we will extract the keywords associated with that image. After applying grammatical rules to the keywords an appropriate caption is generated. Here we are combing the textual modalities with the visual one. In the existing method the captions are not efficiently generated and there is no mapping between the image and the text associated with it. But we are introducing a best method for news image caption generation without costly manual involvement.**

Specifically, we exploit data resources where images and their textual descriptions co-occur naturally. We present a new dataset consisting of news articles, images, and their captions that we required from the BBC News website. Rather than laboriously annotating images with keywords, we simply treat the captions as the labels. We show that it is possible to learn the visual and textual correspondence under such noisy conditions by extending an existing generative annotation mode. We also find that the accompanying news documents substantially complements the extraction of the image content. In order to provide a better modelling and representation of image content, We propose a probabilistic image annotation model that exploits the synergy between visual and textual modalities under the assumption that images and their textual descriptions are generated by a shared set of latent variables. Using Latent Dirichlet Allocation, we represent visual and textual modalities jointly as a probability distribution over a set of topics. Our model takes these topic distributions into account while finding the most likely keywords for an image and its associated document. The availability of news documents in our dataset allows us to perform the caption generation task in a fashion akin to text summarization; save one important difference that our model is not solely based on text but uses the image in order to select content from the document that should be present in the caption. We propose both extractive and abstractive caption generation models to render the extracted image content in natural language without relying on rich knowledge resources, sentence-templates or grammars. The backbone for both approaches is our topic-based image annotation model. Our extractive models examine how to best select sentences that overlap in content with our image annotation model. We modify an existing abstractive headline generation model to our scenario by incorporating visual information. Our own model operates over image description keywords and document phrases by taking dependency and word order constraints into account. Experimental results show that both approaches can generate human-readable captions for news images. Our phrase-based abstractive model manages to yield as informative captions as those written by the BBC journalists.

***Index Terms*—Captiongeneration,imageannotation,summarization**

I. INTRODUCTION

A caption is an important factor for non textual media objects like image, video etc. A good caption will describe the complete content in a short manner. Without more time loss we can understand a context in a glance with an efficient caption. Generally caption generation is a difficult task. An expertise journalist can also feel difficulty in this generation task. Here we are introducing an automatic caption generation method for news images in an efficient way. This method is helpful in text summarization with the help of visual modality. Many people do summarization based on text alone, which will not have an relation with the associated image. Captions are essential components associated with images to make search engines to respond easily with user queries. Making appropriate captions for images is a difficult task. By making the appropriate caption will help the user to search images with long queries. But most of the images are associated with user annotated tags, captions and text surrounding the images. This paper is concerned with the task of automatic caption generation for news images in association with the related news article.

II. AUTOMATIC CAPTION GENERATION FOR NEWS IMAGES

Good caption should be informative. It must clearly identify the objects of the picture. And follow a good summarization method to summarize the textual content. And provide a relation between the textual and visual contents. Moreover lead the reader in to the article. Worthless words are avoided from a good caption. While a short caption is appropriative, but it should be informative. Following correct grammatical rule will generate a good caption. In the existing method the captions are not efficiently generated and there is no mapping between the image and the text associated with it. But we are introducing a best method for news image caption generation without costly manual involvement. We are not using any dictionaries to provide text to image relation. By using the extractive and abstractive caption generation process a good caption is generated. Amount of images which are available on internet are huge today therefore for the better image retrieval process caption generation is very important task. Making his process automatic means it will reduce the costly manual involvement in caption generation process. There are many problems in image captioning. Because content based image retrieval uses visual similarities of images to find out the matched image. But it is suffering with these mantics information loss. Manual annotated words provides solution to this problem but it is time consuming and costly. In our system this problem is avoided with using the image to text correspondence. A good caption for image is generated in association with the associated article.

Our model consists of two stages, content selection and surface realization. Content selection identifies the important keywords from the text and image. And surface realizations verbalize the keywords which are extracted from the image and text.

Image Content Selection

Content selection is the procedure in which the keywords for the caption generation process are extracted from the image and the associated article. We will find out the important keywords using a text frequency calculator. Text frequency is measure used to figure out how much important a word in a document. This method reads one word at time through the document. And a hash table is build using each word. The hash table has two entries, one the word as the key and other is number of times that word appears in the document. And the word with higher count is considered as an important word. In order to generate efficient keywords for the particular image with respect to article made easily with the stop-word removal and stemming procedure. Using those techniques we have to remove the unnecessary letters in the document. Stemming is a process of linguistic normalisation, in which the variant forms of a word are reduced to a common form. To describe features of an object it is necessary to take out required points from the object available in an image. In order to ensure the scale invariance internal representation of the original image is created. Next we will find out the interesting points of the object in the image using Laplacian of Guassian approximation. After that find out the keypoints that is maxima and minima among the points of interest. Edges and low contrast regions are bad key points. Eliminating these makes the algorithm efficient and robust. An orientation is calculated for each key point. Further calculations are done relative to this orientation assigned to the key points. This effectively cancels out the effect of orientation, making it rotation invariant. Finally, with scale and rotation invariance in place, one more representation is generated.

This helps to uniquely identify features of the image. The comparison between input image and image stored in the database are done with the Euclidian distance of their feature vectors. After that associated keywords of images are extracted. All the keywords from image and document are classified under pronoun, noun, verb etc. By applying stop word removal and stemming algorithm unwanted letters of word are removed. Then by applying extractive and abstractive caption generation process automatic caption is generated. Caption generation is done automatically without costly manual involvement. The abstractive and extractive caption generation procedures are explained below.

Extractive Caption Generation

Extractive caption generation is concerned with extracting a single sentence which contains maximum predicted keywords. It is a text summarization technique in which a sentence from the article itself is retrieved based on maximum occurrence of extracted words with that sentence.

Abstractive Caption Generation

The extractive caption generation method generates a caption which is naturally grammatical but it is not possible to express a subject with a single sentence. And also the selected sentence may be long making a caption inefficient. For these reasons we will go for abstractive caption generation procedure which is more efficient than the extractive one.

III. CONCLUSION

In this paper, we introduced the novel task of automatic caption generation for news images. This becomes useful for various multimedia applications, such as image and video retrieval and development of tools supporting news media management. We have presented extractive and abstractive caption generation models. A key aspect of our approach is to allow both the visual and textual modalities to influence the generation task. This is achieved through an image annotation model that characterizes pictures in terms of description keywords that are subsequently used to guide the caption generation process. Here using Content selection and surface realization stages to generate data without requiring expensive manual annotation. We can also use the following additional algorithms to improve the efficiency of caption, they are: stemmer and text frequency calculator. In information retrieval, stemming is the process for reducing inflected (or sometimes derived) words to their stem, base or root form generally a written word form. Text frequency is a numerical statistic which reflects how important a word is to a document in a collection or corpus. And the not only generates caption according to the image but also the related article is taken into account. By making the appropriate caption will help the user to search images with long queries. It makes an efficient image retrieval process.

REFERENCES

[1] Yansong Feng, Member, IEEE, and Mirella Lapata, Member, IEEE "Automatic Caption Generation for News Images,"IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 35, NO. 4, APRIL 2013

[2] D. Lowe, "Object Recognition from Local Scale-Invariant Features," Proc. IEEE Int'l Conf. Computer Vision, pp. 1150-1157, 1999.

[3] P. He´de, P.A. Moe¨llic, J. Bourgeoys, M. Joint, and C. Thomas, "Automatic Generation of Natural Language Descriptions for Images," Proc. Recherche d'Information Assiste´e par Ordinateur, 2004.