Auto-Caption Generation for News Images

Kishor Prajapati
Computer Engineering
Atharva College of Engineering
Mumbai, Malad(E)

Shardul Wadekar
Computer Engineering
Atharva College of Engineering
Mumbai, Malad(E)

Bhushan Bobhate
Computer Engineering
Atharva College of Engineering
Mumbai, Malad(E)

Amruta Mhatre
Computer Engineering
Atharva College of Engineering
Mumbai, Malad(E)

***Abstract* - News forms an essential part of modern civilization. It plays a significant role in keeping everyone updated about the various events around the world. News consists of article, headlines, captions and generating captions is the most difficult task. So our system is concerned with the idea of automatically generating captions for images, which is vital for many image related applications. It helps in the improvement of the news media and multimedia applications. In the existing method, the captions for the news images are given manually by reading the text content. Thus the caption generation task requires human assistance and hence a time consuming process. Our project will help to generate automatic and appropriate caption for the desired news. The system learns to create captions from the accessible dataset that has not been explicitly defined for our task. A dataset consists of news article and the picture embedded in them. From the dataset the system will identify what the image is and suggest appropriate keywords and with the accompanying article it will generate a caption.**

Keywords:- Caption generation, SIFT, LDA, K-means

I. INTRODUCTION

Captions for images, also known as cutline are a few lines of text used to explain or elaborate photographs. Along with the headlines, captions are the most commonly read words in an article. Furthermore, image descriptions tend to be precise and concise focusing on the most important depicted objects or events. Our innovation is to exploit this implicit information and treat the surrounding document and caption words as labels for the image, thus reducing the need for human involvement.[1]

The system learns to create captions from the available dataset that has not been explicitly labeled. A dataset consists of news article and the image embedded in them. The dataset consists of extensive range of topics including technology, sports, education, military, politics and so on. [2] The standard technique to generate caption is based on two stage framework consisting of content selection and surface realization. Content Selection identifies what the image and accompanying article are about, whereas surface realization determines how to verbalize the chosen content.[2][3]

The backbone for surface realization is a probabilistic image annotation model which uses SIFT algorithm that suggests keyword for an image and it highlights important objects or events depicted in the image. Then by using extractive approach we identify the keywords in the article with the help of LDA algorithm which consist of different mechanism. After generating appropriate keywords from image as well as article we perform the operation of overlapping the keywords by using N-gram model. The system selects maximum overlap keywords and generates a caption for an image.

The dataset we employed contains real-world images and it exhibits a large vocabulary including both object names and keywords, so instead of manually generating annotations, image captions are treated as labels for the image. Simply extracting a sentence from the document often yields an inferior caption. It produces captions that are more grammatical than a closely related word-based system. [1]

II. RELATED STUDIES

An image is a work of art that depicts or records visual perception and should be able to describe the environment related to it. To generate description from the image two steps must be followed. The initial step is to examine the image with the help of image processing tool and extract leading factors from the images by means of some extraction methods which is then record into natural language text by taking into consideration the text generation engine. The work describe by Patrick Hede, Pierre-Alain Moellic, Joel Bourgeoys, Magali Joint, Corinne Thomas [4], has describe to represent images of objects in some natural language or in a human readable form.

The next step is to process the document or the article and generate the keywords from them. M. Banko, V. Mittal, and M. Witbrock [5], can help to generate headline and image description by means of a statistical machine conversion. From a source documents this technique produce summary in a succinct manner.

Automatic image annotation is the process by which a computer system automatically assigns metadata in the form of captioning or keywords to a digital image. Several steps must be followed in order to initiate such a description of images. The initial step is to perform segmentation technique over the images in accordance to the available objects in the image.

Next step is to retrieve the attributes from the resultant database, and finally description is generated using templates. B. Yao, X. Yang, L. Lin, M.W. Lee and S. Chun

Zhu [6], signify most effectual methods which are image parsing and text generation.

Database of images and its related documents is necessary for development of all the obligatory steps to perform annotation of images. But formation and collection of such database is challenging and time consuming. To overcome such a drawback Y. Feng and M. Lapata [1] create a database of images that are naturally associated with the news articles without overhead of manual annotation because documents or news articles associated with images can heighten the image annotation process. Such document contains foremost information that is used to create image description.

Therefore we will be creating a caption that not only summarize the document but is also trustworthy to the image's content.

III. NEED OF THE PROJECT

In the existing method, the captions generation process is done manually by reading the text information. Thus the caption generation task requires human involvement and hence it is a time consuming process. Our project will help to generate automatic and appropriate caption for the desired news.

The other important needs are:-

- In computer vision, there is an increasing demand for describing images or video frames more linguistically [7], i.e. with description sentences rather than isolated keywords list.
- Many of the search engines deployed on the web retrieve images without analyzing their content, simply by matching user input against collocated textual article. [1]
- In social media huge amount of images are uploaded on the web daily. So generating appropriate caption and within short interval of time has become a need.

IV. IMPLEMENTATION PROCESS

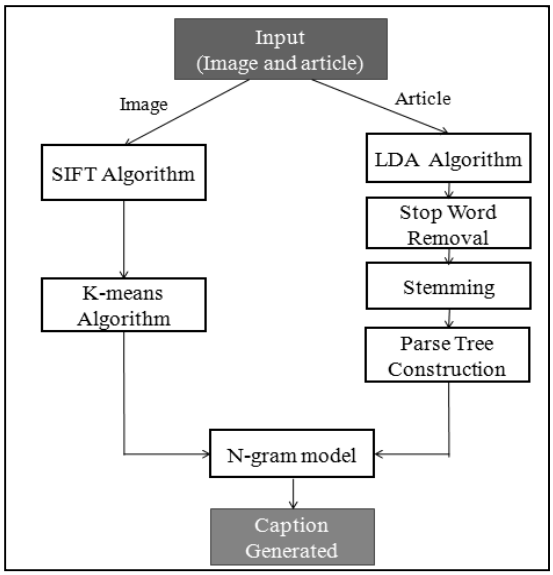

Fig. 1 System Block Diagram

The implementation is done in two stages:

- **Content Selection**: In content selection, from the image and article keywords will be generated by using two different algorithms. It includes two models-image annotation model and topic model.
- **Surface Realization**: From the keywords generated in content selection, keywords are arranged using language model and caption is generated.

A. System Design:

The input to the system is image and article. First the image will be processed using SIFT algorithm. SIFT stands for Scale Invariant Feature Transform algorithm. It is used to describe and detect the local features in the image. It identifies the multiple object with SIFT descriptor, which is histogram of direction at different location in detected region. [8]

Once object are identified the next process is Clustering. Clustering means grouping of data or dividing a large datasets into smaller data set of same similarity. For these k-means algorithm [9] is used. ' k' stands for number of different objects observed. From the observed objects keywords are generated for that object.

Further, the article will be given and processed using LDA algorithm. [10] LDA stands Latent Dirichlet Allocation. The process starts with stopword removal. Stopwords are nothing but words having less efficiency or less meaning than keywords. Stopwords include is, of, the, a, an, etc. The next process is stemming in which the derived words are reduced. For example searching, searched becomes search. Since the caption generated is always in the present tense. Finally, the parse tree is constructed in which the keywords obtained are arranged in sequential form.

Here keywords are generated from both image and article. They are put in N-gram Mode. [10] It is a language model. Here from the keywords similar keywords are overlapped and using language model phrase is constructed which then become appropriate caption.

V. CONCLUSION

In this paper, we introduced the task of automatic caption generation for news images. The key aspect of our approach is to allow both the textual and visual information to support the caption generation task. This is achieved through an image annotation model that characterizes pictures in terms of description keywords that are subsequently used to make the caption generation process easier. Then, the topic model helps in the abstraction of keywords from the article. Furthermore, the language model combines result of both the image annotation and topic model which gives appropriate caption for the news images.

VI. ACKNOWLEDGEMENT

We would also like to express our appreciation and gratitude to Atharva College of Engineering authority for helping us to develop this paper.

Furthermore, we would wish to give our sincere thanks to all staff members who have directly or indirectly influenced us.

VII. REFERENCES

[1] Yansong Feng,Mirella Lapata"**Automatic Caption Generation for NewsImages**", 2013
[2] Priyanka M. Kadhav,Pritam Nikam **"Automatic Caption Generation for News Images using Phrase Based Model",** 2015
[3] N.P.Kiruthika, V.Lakshimi Devi, Ms.N.D.Thamarai selvi**"Extractive And Abstractive Caption Generation Model For News Images**", 2014.
[4] P. He´de, P.A. Moe¨llic, J. Bourgeoys, M. Joint, and C. Thomas, "**Automatic Generation of Natural Language Descriptions for Images**", 2004.
[5] M. Banko, V. Mittal, and M. Witbrock, **"Headline Generation Based on Statistical Translation"**, 2000.
[6] B. Yao, X. Yang, L. Lin, M.W. Lee, and S. Chun Zhu, "**I2T: Image Parsing to Text Description**", 2009.
[7] Priyanka Jadhav, Sayali Joag, Rohini Chaure, Sarika Koli"**Automatic Caption Generation for News Images**", 2014
[8] David G. Lowe "**Object Recognition from Local Scale-Invariant Features**", Sept. 1999.
[9] Tapas Kanungo**,** Nathan S. Netanyahu, Angela Y. Wu **"An Efficient k-Means Clustering Algorithm: Analysis and Implementation",** 2002.
[10] D. Blei, A. Ng, and M. Jordan "**Latent Dirichlet Allocation**", January 2003.