

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IJCSMC, Vol. 5, Issue. 1, January 2016, pg.74 – 82

REVIEW on ENHANCE APPROACH for AUTO CAPTION GENERATION on DIFFERENT NEWS IMAGES DATASET USING FUZZY RULES

Priyanka V. Thakre, A.S.Sambare, N.S.Khade

Research Scholar, Department of Computer Science and Engineering, Priyadarshini Institute of Engineering and Technology Nagpur (MS), India

Professor, Department of Computer Science and Engineering, Priyadarshini Institute of Engineering and Technology Nagpur (MS), India

Abstract: These times, whenever retrieving images from the search Engines that retrieves images without analyzing their include restrain, simply by matching user inquires against the image's file name and format, user comment the tags, captions, and, generally, text surrounding the image. Also the retrieved image contains any textual data along with the images. Our announced the task of automatic caption generation for news images. The task fuses insights from computer vision and natural language processing and holds promise for various multimedia applications, such as image retrieval, development of tools supporting news media management, and for individuals with visual impairment. It is possible to learn a caption generation model from weakly labeled data without costly manual involvement. Instead of manually creating annotations, image captions are treated as labels for the image. Although the caption words are admittedly noisy compared to traditional human-created keywords, we show that they can be used to learn the correspondences between visual and textual modalities, and also serve as a gold standard for the caption generation task. 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. We proposed fuzzy rules for generating caption of the new image.

Keywords: Caption generation, Stop word removal, Stemming, News Images, Fuzzy logic.

1. Introduction

Lots of search engines available on the web retrieve images without analyzing their content, simply by matching user inquiries against collocated textual set of data. For example metadata, user-comment tags, captions, and, generally, text surrounding the image. But this has major disadvantages; they challenge the applicability of search engines. So there is a require of appear to proceed on the development of methods that produce description words for a picture automatically. In our plan, we tackle the related problem of produce captions for news images. Our focus on captioned images embedded in news dataset, and makes use of both models of content selection and surface realization from data and thus avoid expensive manual comment. For example, an image commented with the words blue, sky, car could represent a blue car or a blue sky, whereas the caption “car slang to go away opposition the bright clear blue” would make the relations between the words outspoken. Also, image descriptions need to be abrupt, focusing on the most important depicted objects or events. A method that generates such descriptions automatically could also assist journalists in creating being describe for the images associated with their articles or in finding images that appropriately illustrate their text. Also the linking of images with textual descriptions would facilitate the retrieval and management of multimedia data .It could also assist journalists in creating descriptions for the images associated with their articles or in finding images that appropriately illustrate their text.

The most recent quaternary periods have witnessed an unprecedented growth in the multiple of digital lots of data available on the Internet. Flickr, one of the best known photos sharing websites, hosts more than three very large numbers of images, with approximately 2.5 million images being uploaded every one day. Many on-line news sites like CNN, Yahoo!, and BBC publish images with their stories and even provide photo feeds related to current events. Browsing and finding pictures in large-scale and heterogeneous collections is an important problem that has attracted much interest within information retrieval. Many of the search engines deployed on the web retrieve images without analyzing their content, simply by matching user queries against collocated textual information. Examples include meta-data (e.g., the image’s file name and format), user-comment tags, captions, and generally text surrounding the image. As this limits the applicability of search engines (images that do not coincide with textual data cannot be retrieved), a great deal of work has focused on the development of methods that generate description words for a picture automatically. With the exponential growth in the

quantity and complexity of information sources on the internet, it has become increasingly important to provide improved mechanisms to user to find exact information from available New Images. Caption generation has become an important and timely tool for helping and interpreting the large volumes of text available in documents. Automatic Caption generation is the summary of the source text created by machine to present the most important information in a shorter version of the original text while keeping its main content and helps the user to quickly understand large volumes of information. Caption generation can handle the problem of selecting the most important portions of text as well as the problem of generating coherent summaries. Automatic Caption generation is significantly different from that of human based text summarization since humans can capture and relate deep meaning and themes of News images while automation of such a skill is very difficult to implement. Automatic Caption generation can be carried out mainly by two methods: extraction and abstraction. Extraction summary method is the selection of sentences or phrases having highest score from the original text and put it together to a new shorter text without changing the source text. Abstraction summary method uses linguistic methods to examine and interpret the text. Automatic Caption generation works best on well-structured documents such as news, reports, articles and scientific papers. The Extraction method for summarization involves identifying the features such as sentence length, sentence location, term frequency, number of words occurring in title, number of proper nouns, number of numerical data and thematic word. Our approaches uses feature fusion technique in order to decide which features are actually useful out of the available ones.

2. Existing System

All previous methods attempt to learn the correlation between image features and words from examples of images manually annotated with keywords. They are typically developed and valued on the Corel database, a collection of stock photographs, divided into themes (e.g., tigers, sunsets) each of which are associated with keywords (e.g., sun, sea) that are in turn considered appropriate descriptors for all images belonging to the same theme.

1. Automatic Image Description Generation

This application follows two stage architecture. The images first analyzed using image processing techniques into an abstract representation, which is then rendered into a natural

language description with a text generation engine. A common theme across different models is domain specificity, the use of hand-labeled data indexed by image signature(e.g. Color and texture), and reliance on background ontological information.

2. Automatic Description for human activities in video

The idea is to extract features of human motion from video key frames and interleave them with a concept hierarchy of actions to create a case frame from which a natural language sentence is generated.

Author in this paper focus on [1] concerned with the task of automatically generating captions for images, which is important for many image related applications. Examples include video and image retrieval as well as the development of tools that aid visually impaired individuals to access pictorial information. Our approach leverages the vast resource of pictures available on the web and the fact that many of them are captioned and collocated with thematically related documents. Our model learns to create captions from a database of news articles, the pictures embedded in them, and their captions, and consists of two stages. Content selection identifies what the image and accompanying article are about, whereas surface realization determines how to verbalize the chosen content. We approximate content selection with a probabilistic image annotation model that suggests keywords for an image. The model postulates that images and their textual descriptions are generated by a shared set of latent variables (topics) and is trained on a weekly labeled dataset (which treats the captions and associated news articles as image labels). Inspired by recent work in summarization, we propose extractive and abstractive surface realization models. Experimental results show that it is viable to generate captions that are pertinent to the specific content of an image and its associated article, while permitting creativity in the description. Indeed, the output of our abstractive model compares favorably to handwritten captions and is often superior to extractive methods.

Author in this paper [2] concerned with the task of automatically generating captions for images, which is important for many image related applications. Our model learns to create captions from publicly available dataset that has not been explicitly labeled for our task. A dataset Consists of news articles; the pictures embedded in them, and their captions, and consist of two stages. First stage consists of content selection which identifies what the image and

accompanying article are about, whereas second stage surface realization determines how to put the chosen content in a proper grammatical caption. For content selection, we are using probabilistic image annotation model that suggests keywords for an image. This model postulates that images and their textual descriptions are generated by a shared set of latent variables (topics) and is trained on a weekly labeled dataset (which treats the captions and associated news articles as image labels). The abstractive surface realization model generates captions that are favorable to human generated captions.

Author in this paper [3] tackle the problem of automatic caption generation for news images. Our approach leverages the vast resource of pictures available on the web and the fact that many of them are captioned. Inspired by recent work in summarization, we propose extractive and abstractive caption generation models. They both operate over the output of a probabilistic image annotation model that preprocesses the pictures and suggests keywords to describe their content. Experimental results show that an abstractive model defined over phrases is superior to extractive methods.

Author in this paper [4] focus on inclination to tend to introduce the novel task of automatic caption generation for news footage. The task fuses insights from laptop computer vision and tongue technique and holds promise for varied multimedia system applications, like image retrieval, development of tools supporting medium management, and for people with incapacity. It's potential to be told a caption generation model from sapless labeled knowledge whereas not costly physical involvement. Rather than physically making annotations, image captions area unit treated as labels for the image. Although the caption words area unit confessedly shouting compared to ancient human-created keywords, we've an inclination to tend to suggests that they are attending to be accustomed learn the correspondences between visual and matter modalities, and in addition perform a gold customary for the caption generation task. We've got given extractive and speculative caption generation models. A key facet of our approach is to permit every the visual and matter modalities to influence the generation task.

Author in this paper [5] try to do provides a model for automatically generating captions for news images, which is used to support development of news media management and many

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 involvement and hence a time consuming process.

3. Proposed Work

The proposed system uses a two-stage framework for automatically generating captions for news images: 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. The images are analyzed using the image annotation technique. It uses a multimodal vocabulary consisting of textual words and visual terms. The textual words obtained from annotation are clustered with the words in the news document. Here the extractive and abstractive models for generating short, meaningful and precise captions for the news image are used.

The advantages of this model are:

- (1) It does not need automatically annotation of images
- (2) It reduces the need for human supervision.



File Photo: Prime Minister Narendra Modi

NEW DELHI: Prime Minister Narendra Modi today greeted the people of Vietnam on their National Day.

"My best wishes to the people of Vietnam on their National Day," PM Modi tweeted.

My best wishes to the people of Vietnam on their National Day. Our Government attaches great importance to strong ties with Vietnam.

- Narendra Modi (@narendramodi) September 2, 2015

"Our government attaches great importance to strong ties with Vietnam," he added.

In early classic summarization system the important summaries were created according to the most frequent words in the text. Luhn created the first summarization system in 1961 proposed empirical evidences for difficulties inherent in the notion of ideal summary. Both studies used one thematic feature called term frequency, thus they are characterized by surface level approaches. In the early 1960's new approaches called entity level approaches appeared; the first approach of this kind used syntactic analysis. The location features were used, where

key phrases that are used dealt with three additional components: pragmatic words; title and heading words and structural indicators.

In our proposed method uses fuzzy rules and fuzzy set for selecting sentences based on their features. Fuzzy logic technique in the form of approximate reasoning provides decision support and expert system with powerful reasoning capabilities. The permissiveness of fuzziness in human thought processes suggests that much of the logic behind human reasoning is not only a traditional two-values or multi-valued logic, but also logic with fuzzy truths, fuzzy connectives, and fuzzy rules of inference. Fuzzy set proposed by Zadeh is a mathematical tool for dealing with uncertainty, imprecision, vagueness and ambiguity. Fuzzy logic in caption generation needs more investigation. A few studies were done in this area, Witte and Berger presented a fuzzy-theory based approach to co-reference resolution and its application to text summarization. Automatic determination of co-reference between noun phrases is fraught with uncertainty.

The main target is to find the exact or similar documents by querying a printed or scanned document image over a large document corpus. Previous works typically rely on Optical Character Recognition (OCR) techniques. More recently, visual matching is becoming a promising alternative to solve the limitation of poor OCR performance in scanned documents retrieval.

The feature extraction:

The feature extraction techniques are used to obtain the important sentences in the text. In feature extraction technique some of the features have more importance and some have less so they should have balance weight in computations and we use fuzzy logic to solve this problem by defining membership function for each feature.

4. Conclusion

The proposed methods infer that news images are find the captions on dataset and retrieved the text of that particular image. An efficient method of retrieving news images from content based information retrieval and searching methods that search the text with respect to

content of images. Also getting the optimized output at the user end in an efficient manner. The any image compression is used to low complexity is introduced to reduce the query while maintaining comparable search accuracy.

References

- [1] Yansong Feng, Mirella Lapata “Automatic Caption Generation for News Images” *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, VOL. 35, NO. 4, APRIL 2013.
- [2] Priyanka Jadhav, Sayali Joag, Rohini Chaurse, Sarika Koli “Automatic Caption Generation for News Images” *Priyanka Jadhav et al, / (IJCSIT) International Journal of Computer Science and Information Technologies*, Vol. 5 (2) , 2014, 1675-1677
- [3] Yansong Feng and Mirella Lapata “How Many Words is a Picture Worth? Automatic Caption Generation for News Images” *Association for Computational Linguistics Uppsala*, Sweden, 11-16 July 2010
- [4] Dr. M Nagaratna, Nafees Fathima “Automatic Caption Generation for Pictures Available on Web” *International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Volume 3 Issue 8*, August 2014
- [5] N.P. Kiruthika, V. Lakshmi Devi “EXTRACTIVE AND ABSTRACTIVE CAPTION GENERATION MODEL FOR NEWS IMAGES” *International Journal of Innovative Research in Technology & Science (IJIRTS) 2014*
- [6] C. Wang, D. Blei, and L. Fei-Fei, “Simultaneous Image Classification and Annotation,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1903-1910, 2009.
- [7] V. Lavrenko, R. Manmatha, and J. Jeon, “A Model for Learning the Semantics of Pictures,” *Proc. 16th Conf. Advances in Neural Information Processing Systems*, 2003.
- [8] S. Feng, V. Lavrenko, and R. Manmatha, “Multiple Bernoulli Relevance Models for Image and Video Annotation,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1002-1009, 2004.
- [9] L. Ferres, A. Parush, S. Roberts, and G. Lindgaard, “Helping People with Visual Impairments Gain Access to Graphical Information through Natural Language: The iGraph System,” *Proc. 11th Int'l Conf. Computers Helping People with Special Needs*, pp. 1122-1130, 2006.
- [10] A. Abella, J.R. Kender, and J. Starren, “Description Generation of Abnormal Densities Found in Radiographs,” *Proc. Symp. Computer Applications in Medical Care*, Am. Medical Informatics Assoc., pp. 542-546, 1995.
- [11] A. Kojima, T. Tamura, and K. Fukunaga, “Natural Language Description of Human Activities from Video Images Based on Concept Hierarchy of Actions,” *Int'l J. Computer Vision*, vol. 50, no. 2, pp. 171-184, 2002.
- [12] A. Kojima, M. Takaya, S. Aoki, T. Miyamoto, and K. Fukunaga, “Recognition and Textual Description of Human Activities by Mobile Robot,” *Proc. Third Int'l Conf. Innovative Computing Information and Control*, pp. 53-56, 2008.
- [13] P. He'de, P.A. Moe'llic, J. Bourgeois, M. Joint, and C. Thomas, “Automatic Generation of Natural Language Descriptions for Images,” *Proc. Recherched'Information Assistee par Ordinateur*, 2004.
- [14] B. Yao, X. Yang, L. Lin, M.W. Lee, and S. Chun Zhu, “I2T: Image Parsing to Text Description,” *Proc. IEEE*, vol. 98, no. 8, pp. 1485-1508, 2009.
- [15] G. Kulkarni, V. Premraj, S. Dhar, S. Li, Y. Choi, A.C. Berg, and T.L. Berg, “Baby Talk: Understanding and Generating Image Descriptions,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1601-1608, 2011.
- [16] A. Farhadi, M. Hejrati, A. Sadeghi, P. Yong, C. Rashtchian, J. Hockenmaier, and D. Forsyth, “Every Picture Tells a Story: Generating Sentences from Images,” *Proc. 11th European Conf. Computer Vision*, pp. 15-29, 2010.
- [17] V. Ordonez, G. Kulkarni, and T.L. Berg, “Im2Text: Describing Images Using 1 Million Captioned Photographs,” *Advances in Neural Information Processing Systems*, vol. 24, pp. 1143-1151, 2011.
- [18] A. Makadia, V. Pavlovic, and S. Kumar, “Baselines for Image Annotation,” *Int'l J. Computer Vision*, vol. 90, no. 1, pp. 88-105, 2010.

- [19] C.-F. Chai and C. Hung, "Automatically Annotating Images with Keywords: A Review of Image Annotation Systems," *Recent Patents on Computer Science*, vol. 1, pp. 55-68, 2008.
- [20] J.-Y. Pan, H.-J. Yang, and C. Faloutsos, "MMSS: Multi-Modal Story-Oriented Video Summarization," *Proc. Fourth IEEE Conf. Data Mining*, pp. 491-494, 2004.
- [21] T. Hofmann, "Unsupervised Learning by Probabilistic Latent Semantic Analysis," *Machine Learning*, vol. 41, no. 2, pp. 177-196, 2001.
- [22] F. Monay and D. Gatica-Perez, "Modeling Semantic Aspects for Cross-Media Image Indexing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1802-1817, Oct. 2007.
- [23] T.L. Berg, A.C. Berg, J. Edwards, and D. Forsyth, "Who's in the Picture," *Advances in Neural Information Processing Systems*, vol. 17, pp. 137-144, 2005.