

Towards Image and Text Integration – Understanding Image Caption

E. Apostolova, D. Demner-Fushman , S. Antani, G. Thoma

Abstract:

Biomedical images are one of the main tools in establishing diagnosis, facilitating training and research. A vast amount of the “image data” available is embedded in biomedical articles and thus not directly accessible. Our goal is to automatically annotate images extracted from scientific publications with respect to their usefulness for clinical decision support and instructional purposes. In this study, text analysis was applied to the captions of 67,115 figures from biomedical articles with the goal of automatic image annotation. An algorithm was developed to segment multiple figure image captions (92% accuracy) and identify in the text three types of concepts – 1) the object of the image, e.g. body parts 2) findings demonstrated in the image, and 3) the medical device/diagnostic procedure used to produce the image. The concepts were mapped to the National Library of Medicine’s Unified Medical Language System terms and were correctly identified with an F-measure of 0.78. In addition, it was recognized that pointers (arrows, lines, letters, etc) are commonly used as an overlay in images and described in captions to emphasize the main image finding. An algorithm was developed to extract from the text such “image pointers” (F-measure of 0.81) together with their referents.