

Improving Biomedical Article Retrieval Using Figure Content Analysis

D. You, S. Antani, D. Demner-Fushman, G. Thoma

Abstract:

Authors often use text annotations and pointers on figures and illustrations in biomedical articles to indicate regions of interest (ROI). These annotations are often referenced in the caption and discussion of the figure in the text. Automatically localizing and extracting symbols and ROIs and correlating them with biomedical concepts extracted from the article text could significantly improve indexing and retrieval of biomedical articles. Most difficulties in automatic localization of symbols in biomedical images are due to the large variety of symbol shapes, arbitrary sizes and locations, low image quality, and interference due to image background. In this research, we have developed an automatic symbol localization and ROI extraction algorithm for biomedical images. We have used an edge-based segmentation method and have detected symbol contours from binarized edge images using 8-direction chain code. We have developed a boundary matching algorithm based on dynamic time warping (DTW) method. Several boundary preprocessing steps such as line segment approximation, rotation, size normalization, and boundary point rearrangement are applied before the DTW step. Evaluation results on three test sets (each consisting of 100 images) show an average precision of 92.3% and 75.3% recall.