Application of GIS technology and natural language processing technology in automatic generation of marine weather bulletin

Xinping Bai, Zhongliang Lv, and Hui Wang
National Meteorological Center of China Meteorological Administration, Beijing, China

Marine Weather Bulletin is the main weather service product of China Central Meteorological Observatory. Based on five-kilometer grid forecast data, it comprehensively describes the forecast information of wind force, wind direction, sea fog level and visibility in eighteen offshore areas of China, issued three times a day. Its traditional production process is that the forecaster manually interprets the massive information from grid data, then manually describes in natural language, including the combined descriptions to highlight the overall trend, finally edits manually including inserting graphics and formatting, which causes low writing efficiency and quality deviation that cannot meet the timeliness, refinement and diversity. The automatic generation of marine weather bulletins has become an urgent business need.

This paper proposes a method of using GIS technology and natural language processing technology to develop a text feature extraction model for sea gales and sea fog, and finally using Aspose technology to automatically generate marine weather bulletins based on custom templates.

First, GIS technology is used to extract the spatiotemporal characteristics of meteorological information, which includes converting grid data into vector area data, performing GIS spatial overlay analysis and fusion analysis on the multi-level marine meteorological areas and Chinese sea areas to dig inside Information on the scale, Influence area, and time frequency of gale and fog in different geographic areas.

Next, natural language processing, as an important method of artificial intelligence, is performed on the spatiotemporal information of marine weather elements. Here, it is mainly based on statistical machine learning. By data mining from more than 1000 historical bulletins, Content planning focuses on putting large numbers of marine weather element words and cohesive words into automatic word segmentation, part-of-speech statistics and word extraction, then creating preliminarily classified text description templates of different elements. Through long machine learning processes, sentence planning refines sea area filtering and merging rules, wind force and wind direction merging rules, sea fog visibility describing rules, merging rules of different areas of the same sea area, merging rules of multiple forecast texts, etc. Based on these rules, omitting, referencing and merging methods are used to make the descriptions more smooth, natural and refined.

Finally, based on Aspose technology, a custom template is used to automatically generate marine weather bulletins. Through file conversion, data mining, data filtering and noise removal of historical bulletins, a document template is established in which the constant domains and variable domains are divided and general formats are customized. Then use the Aspose tool to call the template, fill in its variable fields with actual information, and finally export it as an actual document.

Results show that the automatically generated text has a precise spatial description, accurate merge and no scales missed, the text sentence is smooth, semantically and grammatically correct, and conforms to forecaster's writing habits. The automatically generated bulletin effectively avoids common mistakes in manual editing and reduces many tedious manual labor. This study has been put into operation in China Central Meteorological Observatory, which has greatly mproved the efficiency of marine weather services.