Chapter 10
On the Use of Deep Learning for Geodata Enrichments

Alaeddine Moussa
Aix-Marseille University, France

Sébastien Fournier
Aix Marseille University, France

Bernard Espinasse
Aix Marseille University, France

ABSTRACT

Data is the central element of a geographic information system (GIS) and its cost is often high because of the substantial investment that allows its production. However, these data are often restricted to a service or a category of users. This has highlighted the need to propose and optimize the means of enriching spatial information relevant to a larger number of users. In this chapter, a data enrichment approach that integrates recent advances in machine learning; more precisely, the use of deep learning to optimize the enrichment of GDBs is proposed, specifically, during the topic identification phase. The evaluation of the approach was completed showing its performance.

INTRODUCTION

Data in a GIS (Faïz, 1999) are often collected for the specific needs of an institution or even a service. Faced with this reality, it becomes judicious to deploy new sources to meet the needs of a larger number of users. This is referred to as the enrichment of geographic databases (GDB). In this context, a new approach is proposed her

DOI: 10.4018/978-1-7998-1954-7.ch010

fitting (Mahmoudi & Faïz, 2006). The latter uses the technique of summarizing multiple documents (Barzilay & McKeown, 2005) to extract relevant information in an abbreviated form. In order to ensure extraction in a timely manner and in accordance with the multi-agent paradigm (Ferber, 1995).

The approach is based on the work of (Mahmoudi & Faiz, 2014). In fact, it is a modification of the latter to incorporate deep learning algorithms. The concept behind their approach is to have a condensed view relative to the geographic entity from a corpus of documents. Since the data is distributed between different documents, the best customized resolution approach is distributed processing. A multi-agent modeling is used to materialize this distribution. The authors adopt three classes of agents: interface agent, geographic agent and task agent. The interaction between the agents is completed by sending messages. Enrichment is carried out in three phases: identification of segments and themes, delegation and finally, text filtering. In addition to these basic steps, an on-demand approach is used to refine the process. In this work, we modified the implementation of the topic identification phase. Our idea stems from the fact that recent advances in deep learning and NLP can optimize the process of enriching GDB.

The rest of the paper is organized as follows: In the next section describes the fundamental concepts of semantic data enrichment of GDB. In the section after that, we outline the approach we propose to perform this enrichment. Section 5 reports the theme identification to label the detected segments. The implementation of our approach is reported in the next section. Finally, in the last section, we present the evaluation of our system.

BACKGROUND

The enrichment of the GIS allows the acquisition of additional information essential for good decision-making. We talk about spatial enrichment and semantic enrichment. Concerning the spatial aspect and within the framework of the generalization process (Plazanet, 1996), for example, enrichment provides the GDBs with information in terms of structure of forms, knowledge relating to the order of operations and the appropriate algorithms. Another stream of work relates to the semantic (also called factual or descriptive) aspect of GDBs. In this category, we can cite Metacarta (Kornai, 2005) and GeoNode (Hyland et al., 1999).

The Metacarta project accomplished the enrichment with its Geographic Text Search (GTS) product. GTS allows the linking of text documents to geographic features located on the map to enrich the GDB data. MetaCarta GTS is offered as an extension to the ArcGIS geographic information system.

Related Content