International Journal of Web Services Research
Volume 17 • Issue 2 • April-June 2020

A Tool to Extract Name Entity Recognition From Big Data in Banking Sectors

C. Janarish Saju, Anna University, Tamil Nadu, India
S. Ravimaran, M.A.M. College of Engineering, Tamil Nadu, India

ABSTRACT

Generally, the Internet is the global system of interconnected computer networks, connecting millions of computers as well as people, and thus generates a massive quantity of information on a daily basis. This leads to extracting the necessary information using information filtering (IF) in several domains. In our implementation, the named entity recognition (NER) technique is employed to automatically extract valuable data from the unstructured natural language texts. As several works has been outlined in detecting named entities, plenty of very different NER tools exist for several domains. However, NER remains a giant challenge so to solve this problem we proposed a novel framework by combining three efficient classifiers. This article proposes a three-layered neural network approach with conditional random field (CRF), the Pachinko allocation model (PAM), and the Adaptive Neuro-Fuzzy Inference System (ANFIS) for detecting named entities in three steps. First, a classifier based on CRF is employed to train the input file. Second, PAM is employed to boost the previous output created by CRF to enhance the label annotation. Third, the ANFIS captures the deep features of the information by itself from the pre-trained information to attain accurate predictions. Experimental results show that the learned model yields a banking domain with a recall rate of 92%, a precision rate of 95% and F-measure of 92% by implementing it in the R Platform.

KEYWORD

Adaptive Neuro-Fuzzy Inference System, Conditional Random Field, Information Filtering, Named Entity Recognition, Natural Language Processing, Pachinko Allocation Model

1. INTRODUCTION

The Internet is a revolution in information technology, which also plays an important role in our daily life, because of its services and features. In today's networked international world, people, goods, data, and information move very widely, quickly and independently within the speed of light across various boundaries. This new environment has brought tremendous opportunities in the form of challenges for businesses, governments and scientific communities (Lee, 2015) as well as for the people, politics and other knowledge (Lee & Olson, 2010). The advancement in Information and communication technology (ICT) have brought an incredible increase within the quantity of data created and shared (Big Data). Knowledge analytics are also used for a spread of purposes (business, security and safety, scientific discovery, etc.), domains (biology, medicine, education, etc.), and stakeholders (businesses, governments, scientists, and consumers) to understand the impact of knowledge management (KM), since we handle a large amount of data on a daily basis (Kim, Trimi, & Chung, 2014). Therefore, data extraction is very essential for several sectors like academia, the trade, banks and governments.

DOI: 10.4018/IJWSR.2020040102

Copyright © 2020, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Numerous organizations are continuously assembling knowledge (grabbed from different sources and languages) by some procedure, and analyzing many data in textual format. The assembling and analyzing of such type of data are turned out to be more and harder, because of the blemished, fragmented, and unstructured nature of data at all time. The developing territories for text analytics are (1) Information extraction (IE) consequently extracting organized information from archives; (2) Topic model (TM) finding the most topics in an exceptionally monstrous and unstructured accumulation of records by utilizing calculations; (3) Opinion mining get to, separate, group, and see the conclusions communicated in a few sources together with interpersonal organizations; assessment investigation is moreover utilized for sentiment mining; and (4) Question answering (Q&A)- noting accurate inquiries (e.g., IBM's Watson, Apple's Siri, Amazon's Alexa, and so forth) in the view of methods from statistical Natural Language Processing (NLP), Information (IR), and Human-Computer Interaction (HCI) (Chen H, Chiang RH, Storey VC 2012).

Named Entity Recognition (NER) is the task of recognizing named elements like a person, area, association, time, amount and so on in the content. NER frameworks are regularly utilized as the initial phase being referred to replying, information retrieval, co-reference goals, and topic modeling, and so on (Abraham, Liu, Lin, & Sun 2012). The web is a noteworthy wellspring of data in this cutting edge world and so as to handle a lot of data stream we need Information Filtering (IF) (Mamat, Mansouri, & Suriani, 2008).The clarified extraction technique improves the extraction nature of the framework by increasing the precision of entity extraction (McIntosh, Murphy, & Curran, 2006), because in our approach the first separated elements will be utilized as a seed for preparing other AI (ML) frameworks (Kanimozhi & Manjula, 2016), across domains and languages. Along these lines, disposing of requirement for manual training and seed extending (through learning) will keep on expanding the accuracy and utilization of element extraction across domains and languages (Ahmed & Sathyaraj, 2015). Energy investment for the preparation of framework (as is finished by the seed elements made by us, and used by Machine Learning and deep learning) (Powley & Dale, 2007) will be reduced by the specialists' involvement and use. Furthermore this framework will takes less time for fixing faulty entities and connection removal to make better data-driven decision (Yoshioka & Thaer, 2015).

Natural Language Processing (NLP) can be taken as a procedure of making a framework whose primary center is to play out specific task like named entity recognition, question answering, sentiment analysis, speech recognition, machine translation (Ekbal & Bandyopadhyay, 2009) and so on. NLP is principally used to investigate the content, remove the importance, and produce common language from it (Chopra & Morwal, 2013). NLP is mostly instituted as an entangled procedure, since computers are typically great about preparing numbers however managing a lot of content will prompt trouble (Bao, Dai, Deng, Kong, & Ren, 2016). In the translation process in NLP, a computer ought to comprehend the contribution and produce the yield in more than one language and translate it (Cao, Long, Sun, Wang, & Yu, 2016). NLP is commonly utilized for handling human-generated language, it does not include preparing things like mathematical equations or programming structures (Hinton, Krizhevsky, Salakhutdinov, Srivastava, & Sutskever, 2014). The traditional method of NLP for the most part needs a great deal of domain knowledge in etymology. The NLP framework performs well, when the input is in the terms of simple text, However, so as to see all the possible combinations and meanings of complex text, it needs a talented etymological. The principle objective of NLP takes it in to another dimension by making less demand for human-machine correspondence (Cai, Qu, Ren, Wang, Wen, Yu, & Zhang, 2016). In our day by day life, we experience numerous things which include NLP like, when an person experiences the remarks segment on Facebook, he may experience a circumstance in which there is a remark in Russian language and it inquires as, whether we need to decipher it or not (Dong, Guan, Huang, Qian, Yang, & Yu, 2016). This is a straightforward ongoing case of NLP. NLP has appeared as a potential business esteems in some ongoing applications, for example, language modeling (Mori, Tomori, & Ninomiya, 2016), website translations, sentiment

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/a-tool-to-extract-name-entity-recognition-from-big-data-in-banking-sectors/250238

Related Content

Big Data Analytics for Childhood Pneumonia Monitoring
Suresh Kumar Peddoju, Kavitha K.and Sharma S. C. (2019). *Web Services: Concepts, Methodologies, Tools, and Applications (pp. 1129-1145).*
www.irma-international.org/chapter/big-data-analytics-for-childhood-pneumonia-monitoring/217880

Business Process Driven Trust-Based Task Scheduling
Jian Shu, Hemant Jainand Changyong Liang (2019). *International Journal of Web Services Research (pp. 1-28).*
www.irma-international.org/article/business-process-driven-trust-based-task-scheduling/231447

Selective Service Provenance in the VRESCo Runtime
Anton Michlmayr, Florian Rosenberg, Philipp Leitnerand Schahram Dustdar (2012). *Web Service Composition and New Frameworks in Designing Semantics: Innovations (pp. 372-394).*
www.irma-international.org/chapter/selective-service-provenance-vresco-runtime/66968

A Similarity Measure Across Ontologies for Web Services Discovery
Aissa Fellah, Mimoun Malkiand Atilla Elci (2019). *Web Services: Concepts, Methodologies, Tools, and Applications (pp. 859-881).*
www.irma-international.org/chapter/a-similarity-measure-across-ontologies-for-web-services-discovery/217867

Affordances of Data Science in Agriculture, Manufacturing, and Education
Krishnan Umachandranand Debra Sharon Ferdinand-James (2019). *Web Services: Concepts, Methodologies, Tools, and Applications (pp. 953-978).*
www.irma-international.org/chapter/affordances-of-data-science-in-agriculture-manufacturing-and-education/217872