JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

NAMED ENTITY RECOGNITION

[1]Darsana V P, [2]Dr.Deepa A

[1]MCA Scholar, [2] Assistant Professor,
[1] Department of MCA,
[1]Nehru College of Engineering and Research Centre, Pampady, India

Abstract : Named Entity Recognition (NER) represents a critical task in natural language processing, particularly when applied to the domain of journal articles. This paper presents a comprehensive examination of supervised learningmethodologies employed for NER within journal articles, emphasizing the intricacies of evaluation methodologies andthe significance of key performance metrics. The study delves into the nuanced evaluation metrics of precision, recall,and F1-score, offering insights into their utilization for assessing model accuracy at both token and entity levels.Furthermore, the discussion extends to the implementation of cross-validation techniques, essential for ensuring therobustness and generalizability of NER models across diverse datasets. In addition, the paper underscores the importance of comparative analyses against baseline models to discern the efficacy and potential areas of improvement of supervised learning strategies. Error analysis emerges as a pivotal step, facilitating the identification of recurrenterror patterns and informing targeted model refinements. Moreover, the evaluation of model generalization to unseen data is highlighted as a crucial aspect, shedding light on the practical applicability of supervised learning approaches in real-world settings. By offering an in-depth exploration of evaluation methodologies and performance considerations in NER for journal articles, this review aims to equip researchers and practitioners with valuable insights, fosteringadvancements in the realm of natural language processing research and its practical applications.

IndexTerms - **Deep Learning,Evaluation Metrics, Multi-task Learning,Named Entity Recognition,Natural language Processing, Supervised Learning, Transfer Learning.**

I. INTRODUCTION

Named entity recognition (NER), it is a technique in natural language processing (NLP) that aims to identify and classify entities. The goal of NER is to extract data generated from unstructured text so that machines can understand and classify entities useful for a variety of applications such as text recognition, typing, creating knowledge maps, tests, and knowledge graphs. This article examines the principles, methods, and applications of the NER model. NER is also known as entity extraction, fragmentation and recognition. It is used in many areas of artificial intelligence (AI), including machine learning (ML), deep learning, and neural networks. NER is a key component of NLP systems such as chat bots, sentiment analysis tools, and search engines. It is used in healthcare, finance, human resources (HR), customer support, higher education, and social analytics. NER identifies, classifies and extracts the most important information from raw documents without the need for time-consuming manual searches. It is particularly useful for extracting important data from large files because it can speed up the extraction process. NER models help advance artificial intelligence as they improve their ability to analyze important data. These systems have improved their ability to understand AI language in areas such as content analysis and translation, as well as the ability of AI systems to analyze text. The NER grammar uses algorithms based on NLP models and prediction models. These algorithms are trained on data that people save with predefined names such as people, places, organizations, expressions, percentages, financial values, etc. Categories are defined by abbreviations; e.g. LOC for location, PER for people, ORG for organization.

II. LITERATURE SURVEY

The first study on NER was conducted by Nadeau and Sekine (2007) and included various supervised, semi- supervised and unsupervised NERs, revealed the characteristics of the NER systems used at that time and described the NER systems reported to be still in use today. measurement. Sharnagat (2014) presents new research on NER, including supervised, semi-supervised, and unsupervised NER, and some introductions to neural network NER systems. There are also studies of NER systems focusing on specific names and languages, including biomedical NER (Leaman and Gonzalez, 2008), Chinese medical NER (Lei et al., 2013), Arabic NER (Shaalan, 2014 ; Etaiwi et al., 2014).). , 2017) and NER for Indian languages (Patil et al., 2016). Most current work involves machine learning engineering models (such as supervised, semi-supervised, and unsupervised) and usually focuses on one language or one domain. To my knowledge, there has been no study of modern neural network NER systems or a comparative analysis between different languages (CoNLL 2002 and CoNLL 2003) and various (e.g., feature engineering and neural network systems). . , News and Medicine) site. I searched Google, Google Scholar, and Semantic Scholar to identify articles for this survey. My questions include name recognition, neural architecture for name recognition, neural network-based entity recognition models, deep learning models for name recognition, and other things. I sorted the data returned from each query by counting and reading at least the first three; I determined whether the data we analyzed revealed neural architectures for naming or represented the bestperforming model of the NER dataset. I added an article reporting only the neural architecture. If it was the first article reporting architecture; Otherwise, I follow the text until we find the first part of the architecture. I follow the same approach in our NER with process engineering. I also include articles that use these systems for different words or names. A total of 154 articles were reviewed and 83 were selected for research.

III. OBJECTIVES

The main purpose of NER is to scrape random text and identify specific blocks after pre-grouping them based on namespaces. Converting old documents into data structures makes data more efficient, thus simplifying tasks such as data analysis, data retrieval, and information retrieval. Name recognition (NER) acts as a bridge between unstructured and structured data, allowing machines to crawl large data sets and extract important information in a categorized format. NER transforms the way we process and use data by identifying specific sites in large volumes of data. Name recognition (NER) recognizes names in natural language processing (NLP) as personal names, appropriate entities, location, treatment numbers, teaching hours, dosage, financial cost, etc. It is a data extraction process that divides data into predefined categories such as. . Understanding these areas in the field of NLP is important for many applications because they often contain the most important information in the text.

IV. METHODOLOGY

SUPERVISED METHOD

Supervised algorithms are a class of algorithms that learn patterns by looking at training examples. In NER's supervised learning algorithm, the real work is done using Hidden Markov Model (HMM), Decision Trees, Maximum Mean Model (ME), Support Vector Machine (SVM) and Process Random Field (CRF). In general, thefocus is on examining conflicting rules based on discrimination or trying to examine conflicting distributions thatlead to better outcomes of the study material. We will examine these methods in detail in the next section.

3.1Hidden Markov Models

HMM is the first model developed by Bikel et al. It was applied to solve the NER problem. (1999) English version.Bikel introduced his IdentityFinder, a system for detecting NER.

person	doctor	**organization**	terrorist_organization
actor	engineer	airline	government_agency
architect	monarch	company	government
artist	musician	educational_institution	political_party
athlete	politician	fraternity_sorority	educational_department
author	religious_leader	sports_league	military
coach	soldier	sports_team	news_agency
director	terrorist		

location	body_of_water	**product**	camera	**art**	written_work
city	island	engine	mobile_phone	film	newspaper
country	mountain	airplane	computer	play	music
county	glacier	car	software	**event**	military_conflict
province	astral_body	ship	game	attack	natural_disaster
railway	cemetery	spacecraft	instrument	election	sports_event
road	park	train	weapon	protest	terrorist_attack
bridge					

building	time	chemical_thing	website
airport	color	biological_thing	broadcast_network
dam	award	medical_treatment	broadcast_program
hospital	educational_degree	disease	tv_channel
hotel	title	symptom	currency
library	law	drug	stock_exchange
power_station	ethnicity	body_part	algorithm
restaurant	language	living_thing	programming_language
sports_facility	religion	animal	transit_system
theater	god	food	transit_line

FIG.1.0

According to Bikel's formulation of the Identifier system problem, only one name can be assigned to a word in a context. Therefore, the model assigns each word a NOT-A-NAME label representing either one of the desired classes or "no desired class." The state diagram for this model is shown in Figure 2. When marking sentences, thetask is to find the most likely sequence of name classes (NC) given a sequence of words (W).

$$\max \Pr(NC|W)$$

HMM is a generative model. H. We try to generate data, word sequence W, and label NC from the distributionparameters.

$$\Pr(NC|W) = \Pr(W, NC)/\Pr(W)$$

Using the Viterbi algorithm Forney (1973), Pr(W, NC). -Maximize class allocation. Bikel modeled a generationon her three steps:

- Select name class nc depending on previous name class and word.

- Create first word in name class considering current name and previous name.

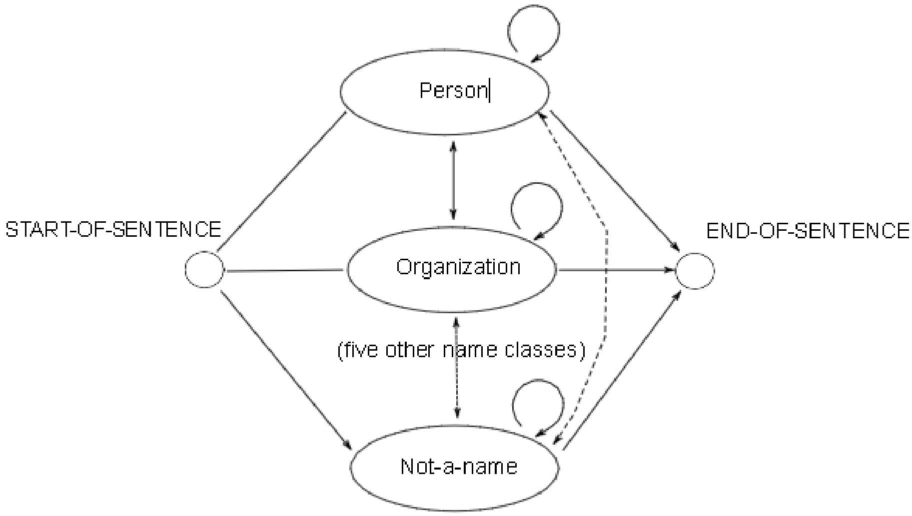

Fig.2.0

3.2 Maximum Entropy Based Models

Maximum entropy models, unlike HMMs, are discriminative models. Using a set of features and training data, the model directly learns to weight discriminative features for classification. The goal of maximum entropy models is to maximize the entropy of the data to generalize as much of the training data as possible. In the ME model, each feature is associated with a parameter λ_i . Therefore, the conditional probability is obtained as follows:

$$P(f|h) = \frac{\prod_i \lambda_i^{g_i(h,f)}}{Z_\lambda(h)}$$

$$Z_\lambda(h) = \sum_f \prod_i \lambda_i^{g_i(h,f)}$$

Maximizing the entropy ensures that for each feature gi, the expected value of gi according to the M.E. model is equal to the empirical expectation of gi in the training corpus. Finally, using the Viterbi algorithm, the most probable path is found from the probability line and thus the required action is created.

The MENE System

Borthwick's (1999) MENE system uses multiple types of information when deciding on a domain name, drawing on a comprehensive dictionary and one or more single-word dictionaries such as name, company name, company success. It uses many features such as binary properties, lexical properties, partial properties, external system output, compatibility and problem solving. MUC-7's tag 29 is the future location of the maximum entropy formula for NE discovery. The maximum entropy solution for this allows computing p (f |h) for any f from the future location F and for each h from the location of past results H. Section Event information that helps the maximum entropy model make future decisions.

Curran's ME Tagger

Curran and Clark (2003) used the maximum entropy model for the naming problem. They use the softmax method to generate the probability P(y|x). Tagger uses a pattern of the form:

$$P(y|x) = \frac{1}{\quad} \exp(\sum \lambda f(x, y))$$

where y is the tag, x is the context and fi(x, y) is the feature with associated weight λi.

Hence the overall probability for the complete sequence of y1 • • • yn and words sequence w1 • • • wn is approximated as:

$$P(y_1 \cdots y_n | w_1 \cdots w_n) \approx \prod_{i=1}^{n} Pr(y_i|x_i)$$

where x_i is a context vector for each word w_i. The tagger uses ray detection to find the most likely result for a sentence. Curran reported 84.89% accuracy for English test materials and 68.48% accuracy for German test materials for the CoNLL-2003 joint task.

3.3 SVM Based Models

Support vector machine was first proposed by Cortes and Vapnik (1995) as a linear hyperplane learning concept that can distinguish good and bad models. A large margin means that the plane has the greatest distance from the point through a sample. The points closest to the hyper plane on both sides are called support vectors. Section Figure 3 shows the geometric interpretation. Linear classifiers are based on two parameters: the weight vector W perpendicular to the elevation plane (used for separation) and the deviation b (used to determine the deviation of the hyper plane from the initial history). If f(x) = wx + b> 0, the sample x is classified as a positive sample, otherwise it is a negative sample. If the data points are not linearly distributed, a break is used to accommodate some error in the distribution. This prevents the classifier from over fitting the data.

When there are more than two groups, a set of criteria is used to divide the sample. Chapter McNamee and Mayfield (2002) solve the problem as a binary decision problem; if the word belongs to one of 8 groups, for example, it starts with B-, I- internal tags for people, organizations, places and other tags. For this purpose, 8 classifiers were trained. All functions used are binary. 258 characters and symbols and 1000 related words are used. The size of the display area is 7, which increases the number of features used to 8806. The S letter set is defined to create a label for each symbol. If S is empty, label O is given, otherwise, at most labels are given. If both open and internal are available, the open label is selected. For CoNLL 2002 data, Spanish and Dutch exposure data are 60.97 and 59.52 respectively.

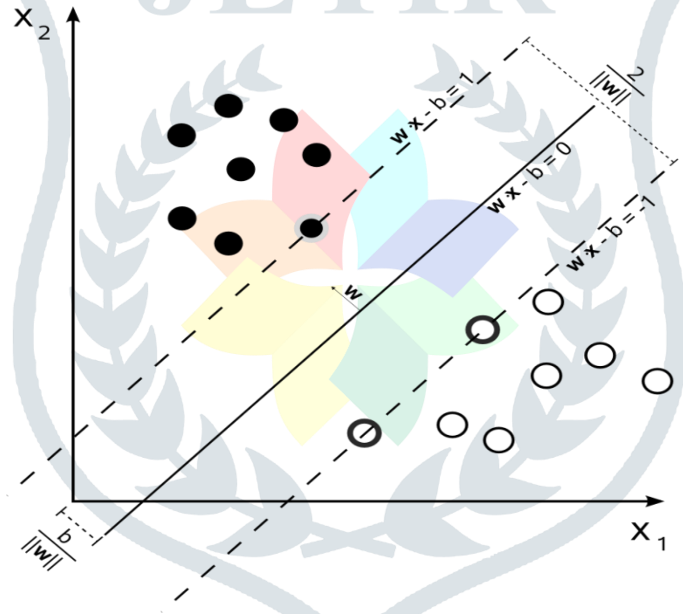

Fig 3.0

3.4 CRF Based Model

Random field conditions were introduced by Lafferty et al. (2001) as a benchmark for pattern recognition and machine learning using predictive models. McCallum and Li (2003) proposed a feature extraction method for CRF in NE. Let's make some visual data connections like o = < o_1, o_2, …..ˆ , o_t > connecting words in the text in the file (the value of the n input nodes of the graphical model). Let S be a set of FSM states; each state is associated with an L label (e.g. ORG). Let s = < s_1, s_2, ….. s_t > be a system state (value of output T). Accordingto Hammersley Clifford's theorem, CRF defines the probability of a set of states given an input:

$$P(s|o) = \frac{1}{Z} \exp \left(\sum_{t=1}^{T} \sum_k \lambda_k f_k(s_{t-1}, s_t, o, t) \right)$$

where Z is the normalization factor obtained by marginalizing over all state sequences, $f_k(s_{t-1}, s_t, o, t)$ is an arbitrary feature function and λ_k is the learned weight for each feature function.By using dynamic programming, state transition between two CRF states can be efficiently calculated. The modified forward values, αT (si), to bethe "unnormalized probability" of arriving state si given the observations < o_1, o_2, • • • , o_t >. α0(s) is set to probability of starting in each state s, and recursively calculated as :

$$\alpha_{t+1}(s) = \sum \alpha_t(s') \exp \sum \lambda_k f_k(s', s, o, t)$$

V. RESULT EVALUATION

The evaluation process and methods should be provided when evaluating the effectiveness of the so-called recognition (NER) system designed for newspapers. Precision, recall, and F1 scores are the main parameters that measure the performance of the system and take into account both correct and incorrect results. While conducting a comprehensive evaluation of the so-called recognition (NER) system designed for newspapers, the evaluation process and process should be used to better understand the performance.; Location evaluation provides an agreement by evaluating the ability to capture the finished product. Considering the different types of fields available in text, from common fields such as dates to specific fields such as search terms and proper names, it should be the same to examine the performance of the system well in different environments to evaluate the effectiveness of the system in different environments. fields. . Additionally, delving deeper into error analysis can reveal valuable insights by uncovering error types, which can lead to opportunities for remediation and improvement. The use of separate tests and cross-validation procedures ensures stability and generality by providing an unbiased assessment of the system's performance and ability to adapt to unseen data. Benchmarking baselines and identifying details contributes to the analysis process, allowing for a more comprehensive understanding of process performance in the affected area. Especially in systems designed to increase the value of products, it has become important to measure efficiency and performance to ensure that they can be used in real situations. Finally, human assessment works as a complementary, qualitative assessment that complements environmental assessment by capturing more nuance than mechanical assessment to provide a complete assessment full of quality and practicality.

VI. CONCLUSION

This report for Named Entity Recognition, covers supervised learning methodology. Named Entity Recognition (NER) is a technique in natural language processing (NLP) that focuses on identifying and classifying entities. In the realm of supervised learning for NER within journal articles, the evaluation process stands as a cornerstone for effectively gauging the prowess of the model. The goal of NER is to extract information generated from unstructured text so that machines can understand and classify entities useful for a variety of applications such as text recognition, typing, creating knowledge maps, tests, and knowledge graphs. This multifaceted endeavor encompasses a series of critical stages and metrics, all geared towards obtaining a comprehensive understanding of the model's capabilities and limitations. Initially, the dataset undergoesmeticulous partitioning into distinct training and test sets, with the former earmarked for model training andthe latter reserved for rigorous evaluation. Within this evaluation framework, precision, recall, and F1-score emerge as pivotal metrics, furnishing quantitative insights into the model's accuracy across both individual tokens and entire entity spans. Furthermore, the adoption of cross-validation techniques assumes paramount importance, ensuring the model's robustness and generalizability by subjecting it to scrutiny across multiple data subsets. Supervised methods is a class of algorithms that learn patterns by looking at training examples.

In NER's supervised learning algorithm, the real work is done using Hidden Markov Model (HMM), Decision Trees, Maximum Model (ME), Support Vector Machine (SVM) and Process Random Field (CRF). In general, the focus is on examining conflicting rules based on discrimination or trying to examine conflicting distribution that maximizes the outcomes of the study material. We will examine these methods in detail in the next section. A comparative analysis against baseline approaches serves to contextualize the model's performance, illuminating potential areas for enhancement and innovation. Concurrently, delving into error analysis plays a pivotal role in discerning prevalent sources of inaccuracies and guiding targeted refinements aimed at bolstering model efficacy. Moreover, the assessment of the model's ability to seamlessly generalize tounseen data via evaluation on separate test sets constitutes a critical litmus test for practical deployment and real-world applicability. Through meticulous consideration of these multifaceted factors and metrics, researchers are empowered to conduct thorough and insightful evaluations of supervised learning models tailored for NER in journal articles, thereby fostering notable advancements in the landscape of natural language processing research and applications.

VII. REFERENCES

[1] Yoshua Bengio, R´ejean Ducharme, Pascal Vincent, and Christian Janvin. A neural probabilistic language model. J. Mach. Learn. Res., 3:1137–1155, March 2003. ISSN 1532-4435.

[2] URL http://dl.acm.org/ citation.cfm?id=944919.944966.

[3] Daniel M. Bikel, Richard Schwartz, and Ralph M. Weischedel. An algorithm that learns whats in a name. Mach. Learn., 34(1-3):211–231, feb 1999. ISSN 0885-6125. doi: 10.1023/A:1007558221122.

[4] URL http://dx.doi.org/10.1023/A:1007558221122.

[5] Andrew Eliot Borthwick. A maximum entropy approach to named entity recognition. PhD thesis, New York, NY, USA, 1999. AAI9945252.

[6] Xavier Carreras, Llu´ıs M`arquez, and Llu´ıs Padr´o. Named entity extraction using adaboost. In proceedings of the 6th conference on Natural language learning - Volume 20, COLING-02, pages 1–4, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics. doi: 10.3115/1118853.1118857. URL http://dx.doi.org/10.3115/1118853.1118857.

[7] William W. Cohen, Robert E. Schapire, and Yoram Singer. Learning to order things. J. Artif. Int.Res.,10(1):243–270,May1999.ISSN1076-9757.

[8] URL http://dl.acm.org/citation.cfm?id=1622859.1622867.

[9] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning, ICML ’08, pages 160–167, New York, NY, USA, 2008. ACM. ISBN978-1-60558-205-4.doi:10.1145/1390156.1390177.

[10] URL http://doi.acm.org/10.1145/1390156.1390177.

[11] Ronan Collobert, Jason Weston, L´eon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. J. Mach. Learn. Res., 12:2493–2537,November2011.ISSN1532-4435.

[12] URL http://dl.acm.org/citation.cfm?id=1953048.2078186.

[13] Corinna Cortes and Vladimir Vapnik. Support-vector networks. In Machine Learning, pages 273–297, 1995.