THINK INDIA JOURNAL
ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

A Deep Insight of Automatic Resume Classifiers For Skill Mapping By Recruiters

Sakshi Takkar[1], Mohit Arora[2], Shivali Chopra[3]

School of Computer Science and Engineering

Lovely Professional University, Phagwara, Punjab, India

[1]sakshi.22258@lpu.co.in, [2]mohit.15980@lpu.co.in, [3]shivali.19259@lpu.co.in

Abstract

IT industries are currently facing a major challenge of mapping candidate's skill set from the available pool to fill various job positions.As the data is increasing at enormous speed, the issue is escalating to next higher levels wherein expertise repositories are piled up with profile records (say CVs), but mining becomes a cumbersome task. In this paper, we made an attempt to sensitize stakeholders of domains of interest to get acquainted with the present State-Of-The-Art of automatic classification of resumes. Researchers classified resumes into various categories and traversed it end to end to extract the relevant features using text mining, Natural Language Processing, crawlers, Deep learning, glove-word embedding, convolution neural network, K-Means Clustering, support vector machine and decision tree classifier etc. Different Machine Learning Algorithms are applied on various Resume Analysis Frameworks and compared on the basis of various metrics like Root Mean Square Error (RMSE), accuracy, recall, relative absolute error and precision etc.

Keywords: Machine learning, Text Mining, Resume Analytics, CV Parser, Natural Language Processing

1. Introduction

Recruitment in the current IT scenario has taken up the pace, as market is flooded with various cross cutting technologies. Software organizations are on the chase to enroll raw talents directly from the colleges through employment fairs. Usually the Human Resource (HR) department of the company works on the recruitment and hires the employees or candidates on the basis of their skills and educational qualifications. Manually, to recruit the candidates with the matching job profiles for the HR department and for the candidates to search the job according to their interest or skills is a very tough job[1]. In this paper, we have reviewed the work of various contextual authors who presented different frameworks toget the useful information from the CV and to identify the most eligible aspirants with their matching job profiles. Eachcandidate assembles

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

his/her CV in his/her own particular manner. This exploratory style of creating CVs impede the talent acquisition division of IT industries for hiring suitable candidates to cater the need of immediate environment. Researchers took this issue categorically by arranging structured and unstructured resume repositories. A lot of tailored CASE tools have been proposed to handle dynamic specificities of resumes[2].

Interviewers make unreasonable choices in view of their abstract perspectives and individual feelings while reviewing candidate's profile during or before interviews. Subsequently, the chance to employ skilled people can be missed. Since contracting a magnificent staff is basic for the accomplishment of a firm. HR staff in the company tries to find the candidates with extraordinary talent and skills. Moreover, when candidates contact any (HR) office or business, they can analyze whether the organization is worried about treating every candidate equally. The impartial behavior of the HR cell of the company can make a good impression in the psyche of a candidate. Such an impression will help the candidates to make a decision whether they have to accept or reject an offer by the company and this can also affect the chance for a company to hire the eligible or potential candidates[3].

Analysis of Resume is one of the most costly and time takingtask during the time spent choosing another expert. So as to enlist another equipped representative, you should look through numerous resumes and lead an individual meeting, which takes from 0.5 to 1.5 hours, and some of the time more[4]. In this way, before welcoming a potential workerfor an interview, it is important to examine the resume. It helps the company to identify the most eligible candidate for the matching job profile. Researchers have introduced automatic resume analysis systems or frameworks with the help of latest machine learning techniques. This paper will assist you to analyze the current resume analysis systems proposed by different researchers which helps to extract the useful information from the resumes andhelps to sort the resumesby finding eligible candidates on the basis of their job profiles[5].

2. Background and Related Work

Yi-Chi Chou et al. in their paper made an interview robot using techniques like NLP, Text Mining etc. It analyzed various resumes and provides recommendations based on skills acquired by an aspirant accordingly. The words in the resumes are graded as per the need of skills required by the immediate environment. The developed system targets the Chinese market. They

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

have performed data preprocessing, transformation, operational analysis and generating real time reports. The different types of reports generated at the end can be used as a ready reckoner by various corporates looking for particular skill sets. This AI-inspired resume vetting and interviewing system helps in handling "individual bias" and "subjective assessment" by panel members[6].

Maria V. Belova et al. in their paper have discussed various state-of-the-art methods used by industries for automatic processing of resumes. The factors they filtered for analysis are content and personal scores. Various systems they reviewed includes Applicant Tracking System, AlphaRecruiter, HRP, ZohoRecruit, Experium, Rezjumaks, BOSS-Kadrovik, Lumesseetc[7].

GopalaKrishnaet al.[8] in their paper have created an automated resume classifier using ensemble learning so that the manual process of fitting an individual in a particular job profile can be avoided. A large-scale recruitment processes at all times are running across the globe, keeping this in context authors have clustered various domain of expertise/skill sets. An exploration of new set of domains has also been done using dynamic and automated resume classifier. Voting mechanism has been employed for classification. A two-step simple yet robust process of depicting candidate mapping to various job profiles includes NLP and classification modules.

Zhang Chen et al.[9]has proposed a novel system, Resume Vis to handle semi-structured resume data. Resumes has been classified into three perspectives viz. Individual, interpersonal, and collective. Deep mining and crawling have been done for extraction, parsing, and rank quantification. Various graphs and charts like career trajectory chart, mobility map of organization has been demonstrated for visual analytics. The system has also been verified by various domain experts.

Papiya Das et al.[10]has considered CV parser model under consideration and presented review of 9 survey papers. They deduced that Big Data is coming so there is a need to powerful CASE tools to process it. CV parsers provide us the ability to hire efficient recruits in different domains of interest. As data is increasing at rapid rate, techniques to handle structured, semi, un-structured and meta data are also booming at same pace.

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

Nasser et al. [11]have created different resume classes using CNN and Glove word embedding. The proposed approach with various levels has been applied on wide variety of data sets and has presented promising results.

AmalaDeshpande et al. [1]described thatrecruiters of companies, organizations and universities receive number of resumes of candidates after the job postings. It is very difficult for the recruiters or HR department to go through each resume and select the appropriate profile of a candidate. Every resume is of different format. The system is proposed where the resume will be downloaded and analysis will be done on the resume by using machine learning techniques. First, the resume will be downloaded using web crawler. The resume is unstructured and in pdf format. After then, the information will be extracted and converted to json format. The converted information is normalized and scaled to make the resumes in same format. Then the resumes will be clustered or collected on the basis of parameters like education qualifications, skills or certifications using the k-means clustering algorithm and stored into the database. After the clustering, each resume is scored using the unit scoring method on the basis of job specifications provided by recruiters. Further parameters can be decided to make other possible clusters. In this way, company can shortlist the resumes of eligible candidates and can find top N candidates.

Jacob Bollinger et al. [12]described that bright has made an Automated Resume Analysis System which finds the list of top candidates who got shortlisted for interview on the basis of job description. However, this paper is giving more complete system by matching candidate's resume with his/her social media profile. This augmented or complete system of aspirant's CV and social meansoutline are analyzed to find the overlapping features by using a very popular machine learning algorithm "BRIGHT SCORE". This Bright Score algorithm helps to check the candidate's capability and helps recruiters or employers to find the most eligible candidates for the job opening. Bright Score algorithm's main utilization is to elaborate the limited information in the resume by using the various social media resources like Linkedin, Facebook, and Google. This helps the recruiters to differentiate the qualified candidates from the non-qualified candidates using public data resources.

Jiaze Chen et al.[2] explained the way to extract the information from Resume in PDF format. As in previous studies, we were extracting the information and converting them in JSON format. The recruiters or HR department of big companies, organizations or universities have to manage

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

large number of resumes for one job profile. The resumes are in different formats and in different layouts. To extract the information from resume documents with high accuracy is a very difficult task. PDF format is the one which gives the more detailed information as compare to plain text format or HTML format. This paper is proposing an extraction method in hierarchical manner. Firstly, a resume document can be converted into different high level blocks. Further, high level blocks can be converted into low level blocks. The classification of each block is done by Conditional Random Field (CRF) model. The detailed information from each block is extracted on the basis of layout and content of resume document. The hierarchical extraction method to extract the information is showing more Fi-Score value as compare to Flat method.

Kun Yu et al.[5] have proposed a Cascaded Information Extraction Framework to extract the information from resume. First, the resume is divided into different high-level blocks with the labels of information types. Each high-level block is further divided into low level blocks. HMM model is applied in the first step where each resume document is divided into number of different blocks with the type of information. Moreover, Classification based model SVM (Support Vector Machine) is applied in the second step to extract the personal information. This cascaded information extraction model helps to extract the detailed information from the complete text of the resume document. No hierarchical extraction method is used in these two models. The hybrid model is giving more F-Score value as compare to the flat models. It is concluded that cascaded hybrid model is used to extract the information with high precision.

JunejaAfzalAyubZubeda et al.[3]has developed a resume ranking system for the two entities. One entity is the company who wants to hire the best candidates and the other entity is the students who wants best job according to his/her own skills. After graduation, student wants to get placed in company according to his/her skills. If they are not getting any job then they are ready to do work in a company on any post. However, this paper provides some machine learning algorithms which to help the candidates in searching job according to their skills. In this approach, first the resume of candidates stored in the database and parsing of resume was done. Then skills information was obtained from social profiles like Github, linkedIn etc. According to the skill set of stored resumes, Github and linkedin, candidate resumes were ranked using the Natural Language Processing of Machine Learning.

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

ManopPhankokkruad et al. [13]proposed a recommender system for the recruiters or HR of the company to find the suitable candidates for their company. This recommender system helps them in making decisions and managing recruitments. Matching and Collaborative filtering are the two used techniques in the recommender system. In the matching technique, it matches the profile data of the candidate and generates the score to rank the resume of the candidate. This technique has one problem of scoring because of less variation in scores. Then collaborative technique came into existence which solves the problem of scoring that was in matching technique. Collaborative technique increases the variation in the scores and helps to find the most eligible candidates on the basis of their qualification. This recommender system reduces the error rate of 28.65 percent with the help of collaborative technique used in system.

Sunil Kumar Kopparapu[14]has found that the resume of candidates is in unstructured or in free format. It is very difficult to identify useful information from the unstructured resumes. This paper proposes an automated system to get the usable information from the free format resumes. Natural Language Processing (NLP) technique is used to make an automated system to extract various fields of information. Six main information fields are extracting like total experience, passport number, date of birth, email id, and qualification. Numbers of NLP algorithms are used to extract the information and performance of the system totally depends on the span of knowledge base and accuracy. Metrics like accuracy and recall are used to identify the performance of an automate resume system. This proposed system can deal with large number of unstructured or free format resumes with the precision of 91% and recall of 88%.

Mohamed El Mohadab et al.[15] has proposed a framework that is used to hire the scientific researchers by university managers. In this paper, researcher has used Natural Language Processing (NLP) technique with the decision tree classifier of data mining. One Rule and Naïve Bayes are the two classifiers of Decision tree. The evaluation of both the classifiers are done by comparing some performance metrics like recall, precision, root mean square error, relative absolute error, correctly classify instance and incorrectly classify instance.

Askin Karakas et al. [16]has described in this paper that main objective is to extract useful information such as experience, business and education related information. Ontology based extraction system is proposed to analyze various unstructured or free format resumes and to convert them to structured form.

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

FaizanJaved et al.[17] have proposed an automatic text document classification system which uses machine learning algorithms. Researchers have used Carotene, a machine learning based semi-supervised job title classification. Carotene is a collection of various classification and clustering tools which is used to classify various types of job categories for online recruitment. Cascade classifier architecture has been developed to encompass various techniques and various experimental results have been represented.

WenHua et al. [18]described that understanding short texts is very important task. To understand the short texts in applications, semantic knowledge is required. This paper has proposed a dummy system to understand the short text which makes use of semantic information gave by a notable knowledgebase and naturally gathered from a web corpus. Our insight concentrated methodologies upset customary strategies for assignments, for example, content division, grammatical feature labeling, and idea marking, as in we center around semantics in every one of these errands.

Zhao Jianqiang et al. [19]described the sentiment analysis on twitter to oversee the feeling of public towards events of the real world. Preprocessing of twitter data is the first step of sentiment analysis. Six preprocessing methods have been used for sentiment analysis. Five datasets have been used on which four classifiers and two feature models have been applied for twitter sentiment classification. Models like prior polarity model, Nave Bayes Classifier models and N-grams model have been proposed in this paper.

BichitraMandal et al.[20]have proposed MapReduce framework for identifying gigantic or large amount of data by using collection of computers. To keep analyzing the large volume of datasets by using MapReduce, we work in two phases. On phase is mapping and the other is reducing. To improve its adaptability and effectiveness of word processor in huge information condition, redundancy of sequential words check with rearranging is executed on Hadoop.

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

The consolidated literature review has been tabulated in Table 1 given beneath.

Table1. Literature Review Summary

Author Name	**Paper title**	**Techniques used**	**Pros**	**Research Directions**
Yi-Chi Chou et al.[6]	A Résumé Evaluation System Based on Text Mining	Natural Language Processing, Text Mining	Classification of personality traits a) Dominance b) Influence c) Steadiness d) Compliance e) Education f) Experience g) Skills	Developed system lacked in satisfying requirements of a particular firm
Gopala Krishna et al. [8]	Automated tool for resume classification using semantic analysis	Semantic Analysis, NLPP, Classifiers based on ensemble learning	a) Dynamic and automated resume classification using NLPP and Classifiers. b) Major domain mapping with profiles	Ensemble deep learning using GANs
Zhang et al.[9]	ResumeVis: A Visual Analytics System to Discover Semantic Information in Semi-structured Resume Data	Text Mining, Human-centered computing, visual Analytics	a) Handles semi-structured resume data well b) Over 2500 gazette officers resumes process and evaluated	Authors aim to extend the work tovarious geographical dimensions and finding hidden patterns in trajectories

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

Papaya das et al.[10]	A Review on Text Analytics Process with a CV Parser Model	CV Parsers	Review of 9 survey papers	Real time data analytics with historical models
Nasser et al.[11]	Convolutional Neural Network with Word Embedding Based Approach for Resume Classification	Convolutional Neural Network(CNN), Glove-word Embedding(GWE)	Increased accuracy of resume categorization	As an extension, custom deep neural network model with more sample data can be used for categorization
AmalaDeshpande et al.[1]	Proposed System for Resume Analytics	K-Means Clustering and Hierarchical Clustering	Represents the resume information in standardized format and Scoring of resumes in decreasing order	Authors aim to do the scoring on the basis of weights assigned to the different parameters of resume.
Jacob Bollinger et al.[12]	Using Social Data for Resume Job Matching	Bright Score Machine Learning Algorithm	Giving more complete system by matching candidate's resume with his/her social media profile.	Explore other methods to optimize the effects. Authors will explore the use of analysis of variance (ANOVA) main effects.
Jiaze Chen et al.[2]	Information Extraction from Resume Documents in PDF Format	Conditional Random Field model, Hierarchical Extraction Method	Extraction of resume information in PDF format which gives more detailed information	Authors will plan to update the methods by applying other page segmentation algorithms to increase the accuracy.
WenHua et al. [18]	Understand Short Texts by Harvesting and Analyzing Semantic	Monte Carlo method, Pair wise model and chain model.	To maintain accuracy and efficiency in short texts to	Authors will try to incorporate spatial temporal features in the framework to

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

	Knowledge		get useful information from resumes.	understand short texts easily.
Kun Yu et al.[5]	Resume information extraction with cascaded hybrid model	Hidden Markov model (HMM) and Support Vector Machine (SVM)	Give more accurate results in extracting information from resumes.	Finite state automation will be applied to improve precision and recall.
JunejaAfzalAyubZubeda et al.[3]	Resume ranking using NLP and machine learning	Natural Language Processing (NLP)	Optimized and effective results for the client company and hired candidate	This application can be further extended to other domains like telecomm, healthcare and other public sector jobs
FaizanJaved et al. [17]	Carotene: A Job Title Classification System for the Online Recruitment Domain	Carotene Classification System.	Accurate job classification by SVM-KNN method.	Authors will conduct research consolidate advancement vectors for all the more semantically adjusted characterizations.
ManopPhankokkruad et al.[13]	An improvement of recommender system to find appropriate candidate for recruitment with collaborative filtering.	Matching and Collaborative Filtering methods	Improve scoring of resumes and helps recruiter in decision and manage recruitments.	Authors will use other filtering methods to analyze the performance of the system.
Sunil Kumar Kopparapu[14]	Automatic extraction of usable information from unstructured resumes to aid search.	Natural Language Processing (NLP) technique	Handles unstructured resume data well and increases the precision and recall	Aims to use other machine learning techniques.

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

			value.	
BichitraMandal et al. [20]	Architecture of efficient word processing using HadoopMapReduce for big Data applications	Hadoop Distributed File System (HDFS) and Hadoop Map Reduce	To tally the number of back to back words and rehashing lines	It tends to be adaptable and proficiently utilized for better uses of word handling like; number of copy words, mix of copy words.
Mohamed EIMohadab et al.[15]	Automatic CV Processing for scientific research using data mining algorithm.	Natural Language Processing (NLP) with data mining classifier decision tree	Mainly focus on analyzing CVs of researchers of different disciplines.	A new method will be introduced which will be the improvement of existing method.
Askin Karaka et al.[16]	Towards an information extraction system based on ontology to match resumes and jobs	Ontology based resume parser	Converts plain text resume into ontology based using ontology knowledge base.	Further details of the system properties will be discussed.

3. Conclusion and Future Scope

Resume Analysis with automated systems is a very potential area of research. A lot of research contributions are in this research field as number of IT companies are increasing and hiring number of candidates. To recognize an eligible candidate with the matching job profile is a very significant task. Automated Resume Analysis Systems helps the companies to find the candidates for the job post on the basis of his/her skill or talent. These systems are made with the help of machine learning techniques or algorithms.Different algorithms have different features which are compared on the basis of various metrics like Root Mean Square Error (RMSE), accuracy, recall, relative absolute error and precision etc.. Analysis of different Automated Resume Analysis Systems has been done to find the future possibilities of improving the system.

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

In the future scope, one can aim for the creation of an end to end continuous expert system to handle the complete recruitment process of an organization.

REFERENCES

[1] Amala Deshpande, Deepika Khatri, Divya Deshpande, Prarthita Das, and Sujata Khedkar, "Proposed System for Resume Analytics," *Int. J. Eng. Res.*, vol. V5, no. 11, pp. 468–471, 2016.

[2] J. Chen, L. Gao, and Z. Tang, "Information extraction from resume documents in PDF format," *IS T Int. Symp. Electron. Imaging Sci. Technol.*, pp. 1–8, 2016.

[3] T. Khan, "Resume Ranking using NLP and Machine Learning," no. 2, pp. 2015–2016, 2016.

[4] S. Sanyal, S. Hazra, S. Myashree Adhikary, and N. Ghosh, "Resume Parser with Natural Language Processing," *Int. J. Eng. Sci. Comput.*, vol. 7, no. 2, p. 4484, 2017.

[5] K. Yu, G. Guan, and M. Zhou, "Resume information extraction with cascaded hybrid model," *ACL-05 - 43rd Annu. Meet. Assoc. Comput. Linguist. Proc. Conf.*, no. June, pp. 499–506, 2005.

[6] Y. C. Chou, C. Y. Chao, and H. Y. Yu, "A Résumé Evaluation System Based on Text Mining," *1st Int. Conf. Artif. Intell. Inf. Commun. ICAIIC 2019*, pp. 52–57, 2019.

[7] M. V. Belova and A. B. Zhernakov, "Modern Methods of Resume Processing in Recruiting Information Systems," *Proc. 2018 17th Russ. Sci. Pract. Conf. Plan. Teach. Eng. Staff Ind. Econ. Complex Reg. PTES 2018*, pp. 19–22, 2019.

[8] S. T. Gopalakrishna and V. Varadharajan, "Automated Tool for Resume Classification Using Sementic Analysis," *Int. J. Artif. Intell. Appl.*, vol. 10, no. 01, pp. 11–23, 2019.

[9] C. Zhang and H. Wang, "Resumevis: A visual analytics system to discover semantic information in semi-structured resume data," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 1, pp. 1–25, 2018.

[10] P. Das, B. Sahoo, and M. Pandey, "Parser Model," *2018 3rd Int. Conf. Converg. Technol.*, pp. 1–7, 2018.

ISSN: 0971-1260
Vol-22-NO-30 Special Issue-December-2019

[11] S. Nasser, C. Sreejith, and M. Irshad, "Convolutional neural network with word embedding based approach for resume classification," *2018 Int. Conf. Emerg. Trends Innov. Eng. Technol. Res. ICETIETR 2018*, pp. 1–6, 2018.

[12] J. Bollinger, D. Hardtke, and B. Martin, "Using social data for resume job matching," *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 27–30, 2012.

[13] P. Ri, V. Wr, D. I. R. U. Zlwk, and R. Lowhulqj, "$Q ,Psuryhphqw Ri 5Hfrpphqghu 6\Vwhp Wr)Lqg $Ssursuldwh &Dqglgdwh Iru 5Hfuxlwphqw Zlwk &Roorerudwlyh)Lowhulqj," pp. 72–77.

[14] S. K. Kopparapu, "Automatic extraction of usable information from unstructured resumes to aid search," *Proc. 2010 IEEE Int. Conf. Prog. Informatics Comput. PIC 2010*, vol. 1, pp. 99–103, 2010.

[15] S. S. Mohamed EI Mohadab, Belaid Bouikhalene, "Automatic CV Processing for scientific research using data mining algorithm." 2018.

[16] D. Çelik, "Towards a semantic-based information extraction system for matching résumés to job openings," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 24, no. 1, pp. 141–159, 2016.

[17] F. Javed, Q. Luo, M. McNair, F. Jacob, M. Zhao, and T. S. Kang, "Carotene: A job title classification system for the online recruitment domain," *Proc. - 2015 IEEE 1st Int. Conf. Big Data Comput. Serv. Appl. BigDataService 2015*, pp. 286–293, 2015.

[18] W. Hua, Z. Wang, H. Wang, K. Zheng, and X. Zhou, "Understand Short Texts by Harvesting and Analyzing Semantic Knowledge," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 3, pp. 499–512, 2017.

[19] Z. Jianqiang and G. Xiaolin, "Comparison research on text pre-processing methods on twitter sentiment analysis," *IEEE Access*, vol. 5, pp. 2870–2879, 2017.

[20] B. Mandal, S. Sethi, and R. K. Sahoo, "Architecture of efficient word processing using Hadoop MapReduce for big data applications," *Proc. - 2015 Int. Conf. Man Mach. Interfacing, MAMI 2015*, pp. 1–6, 2016.