An e-Recruitment System Exploiting Candidates' Social Presence

Evantheia Faliagka[1(✉)], Maria Rigou[2,3], and Spiros Sirmakessis[1]

[1] Department of Computer and Informatics Engineering, Technological Institution of Western Greece, National Road Antirrio-Ioannina 30020, Antirrio, Greece
faliagka@ceid.upatras.gr

[2] Department of Computer Engineering and Informatics, University of Patras, Rion Campus, Patras 26500, Greece

[3] Hellenic Open University, Parodos Aristotelous 18 26335, Patras, Greece

Abstract. Applicant personality is a crucial criterion in many job positions. Choosing applicants whose personality traits are compatible with job positions has been shown to increase their satisfaction levels, as well as the rate of employee retention. However, the task of assessing candidates' personality is not addressed in today's online recruitment systems, but is typically handled during the interview process. The rapid deployment of social web services has made candidates' social activity much more transparent, giving us the opportunity to infer features of candidate personality with web mining techniques. In this work, a novel approach is proposed and evaluated for automatically extracting candidates' personality traits based on their social media use.

Keywords: e-recruitment systems · Personality mining · Social web mining

1 Introduction

The amount of increased information at all levels of people's electronic social environment has rapidly in the last years due to the expansion of the social web user base and frequency of recorded user activity in the so called social media [1]. Among the reasons that bring people to the web is knowledge and skills improvement [2], as well as career development [3]. Lately, people seeking for a job are increasingly using web 2.0 services like LinkedIn and job search sites [4], while a lot of companies also use web-based knowledge management systems to hire employees. These are termed e-recruitment systems and automate the process of publishing open job positions and receiving candidates' CVs. Online recruitment can be seeker-oriented or company-oriented. In the first case the e-recruitment system recommends to the candidate a list of job positions that better fit his profile. In the latter, recruiters publish the specifications of available job positions and the candidates can apply.

Typically in on-line recruitment systems, candidates upload their CVs in the form of a document with a loose structure, which must be considered by an expert recruiter. However this incorporates a great asymmetry of resources required from candidates and recruiters and potentially increases the number of unqualified applicants. This

F. Daniel and O. Diaz (Eds.): ICWE 2015 Workshops, LNCS 9396, pp. 153–162, 2015.
DOI: 10.1007/978-3-319-24800-4_13

situation might be overwhelming to HR agencies that need to allocate human resources for manually assessing the candidate resumes and evaluating their suitability for the positions at hand. Several e-recruitment systems have been pro-posed with an objective to automate and speed-up the recruitment process, leading to a better overall user experience and increasing efficiency. For example, after deploying an e-recruitment system SAT telecom reported 44% cost savings and a drop in the average time needed to fill a vacancy from 70 to 37 days [5].

In practice though, the required skills of the applicant are not the only factor that determines the final decision about hiring him or not, as personality is also considered crucial in many job positions. Candidates' personality is assessed by human recruiters at the interview stage, and is usually limited to candidates that passed the screening phase. However, it is known that the vast majority of recruiters perform a preliminary background check on applicants, based on their (social) web presence. This approach has limitations, as it is still a time consuming process and does not work well when the applicants use pseudonymous accounts or have a very common name and surname. A better approach would be to automate this background check, so that this background check could be performed by the e-recruitment system.

To this direction in [6] an integrated company oriented e-recruitment system is proposed. The system automates the candidate pre screening process providing an overall candidate ranking based on a combination of supervised learning and semantic skills matching. Applicant evaluation is based on a predefined set of objective criteria evaluated on the basis of applicant's skills that are directly extracted from his LinkedIn profile, as well as his personality traits extracted by textually analyzing blog posts. One drawback of the aforementioned approach is that the popularity of the blog as medium of self-expression has been steadily declining. Instead, web users have turned to micro-blogging platforms that function with "status updates" instead of free text, such as Facebook, twitter, and Instagram.

In this work, we propose a different approach to infer a candidate's personality traits from his social web activity. More specifically, the personality characteristics that are assessed by the new system regard the candidate's interests as expressed by his social presence, how collaborative he is, whether his interests are related to his line of work and if he demonstrates leadership skills and is an influential individual for his electronic social environment.

The proposed company oriented e-recruitment system provides automated applicant ranking and personality mining with the purpose of restricting interviewing and background investigation of applicants to the top candidates identified by the system. This can lead to a major positive impact on the efficiency of the recruitment process and to significant cost savings. To showcase the effectiveness of the proposed scheme, a prototype has been implemented. The produced candidates' listing combining character features automatically extracted from each candidate's social media activity has been assessed against actual personality traits manually evaluated by human recruiters. The next sections provide details regarding the architecture of the proposed system comprising three main modules, followed by the description of the new personality module. The discussion moves to the pilot recruitment scenario that was used to evaluate the effectiveness of the proposed approach and the paper ends with the main conclusions reached.

2 Architecture

Applicant personality is considered crucial in many job positions. It is assessed by human recruiters at the interview stage and is usually limited to candidates that passed the screening phase. However, it is a common practice among the majority of recruiters to perform a preliminary background check on applicants, based on their (social) web presence. This approach has limitations, as it is still a time consuming process and does not work well when the applicants use pseudonymous accounts or have a very common name and surname. A better approach would be to automate this background check, so that it is performed by the e-recruitment system. In [7] the authors proposed a first approach to such a system that performs automated extraction of candidate personality traits based on linguistic analysis of blog posts and automates the candidate evaluation and pre-screening process. One drawback of this approach is that it relies on the LIWC system to perform a textual analysis of candidates' blog posts to infer the "Big-5" personality directions. However the popularity of the blog as a medium of self-expression has been steadily declining. Instead, web users have turned to micro-blogging platforms that use "status updates" instead of free text, such as Facebook, Twitter and Instagram. Short "status" messages are typically accompanied with images and / or URLs to share information. Although a positive / negative tag can still be attributed to each status update [8] [9] [10], these lack the linguistic markers required by LIWC.

In this work, we propose a different architecture for extracting candidates' personality traits based on their social media use. An overview of the proposed system architecture is depicted in Fig. 1, and consists of the following components:

- *Job Application* Module: It implements the input forms that allow the candidates to apply for a job position. The candidate is given the option to log into the system with the account credentials of his micro-blogging platform of choice (i.e., either Twitter of Facebook).
- *Personality Mining* Module: If the candidate opts to log into the system with his Facebook or Twitter credentials, the system gains access to his "timeline" (i.e., stream of status updates) and interactions with other members of the community.
- *Applicant Grading* Module: The applicant's score in a set of predefined criteria is calculated. Web mining techniques are employed and applied to the candidate's social activity.

Although this is out of scope of this work, as can be seen in Fig. 1 a fully functional system is expected to also implement a ranking function that will assess the candidates' overall relevance to a specific job position, based on the scores of individual selection criteria. Numerous ranking functions based on AHP [11] or Machine Learning techniques [12] can be found in the literature.

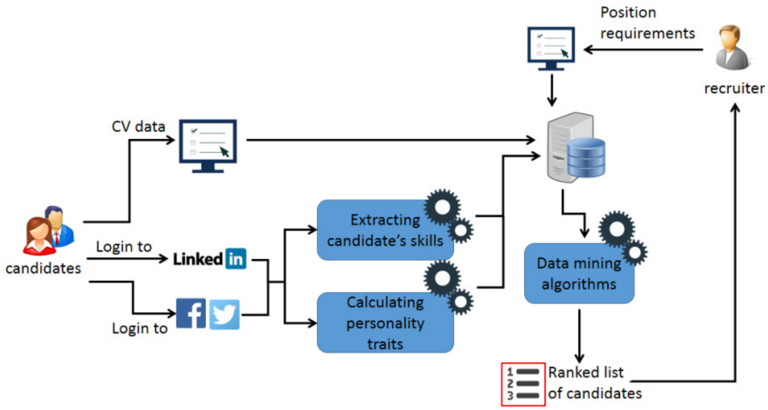

Fig. 1. System's architecture

3 The New Personality Module

An inherent limitation of automated recruitment systems is their over-reliance on formal qualifications. Hiring managers are able to see beyond skill-sets, and assess candidates' personality and how they would fit in the corporate culture. Judging applicants' creativity, inspiration or their ability to work with people is a hard problem to solve for automated systems (assuming that it can be solved at all). A new approach is proposed in this section. The proposed system is able to make inferences on the applicants' suitability for a specific position, by inferring his personality traits based on his social media use. Specifically, asking the applicant to log-in to the system with his Facebook or twitter credentials, we get access to a large amount of information, such as status updates, interaction with other users and interests ("likes") or side projects. These can demonstrate his strengths and passions in a way that the employment history can't always do [13]. For example, an individual involved in the organization of community events could boost an applicants' suitability for positions where organizational skills are critical. But even activities that do not directly relate to a specific position can translate to a well-rounded individual who takes time to "refuel and recharge" after work. In the following table (Table 1) we summarize the main personality traits that are of interest to the hiring managers, and how they are related to candidates' hobbies and leisure activities.

Table 1. Personality traits and the according activities

The candidate is a team player	Participation in a team sport or collaborative activity could translate to a person that is able to function in an activity that requires group interaction.
The candidate has leadership skills	Candidates that possess a leadership role in their social interactions are favored in management positions.
Actively working to improve his skills	Candidates that work on learning a new language or participate in seminars about public speaking are viewed positively.
The candidate is passionate	Good candidates are passionate about their activities, whether inside or outside of the office. Leisure activities relative to the position are always a plus
He is a well-rounded individual	Well-rounded personalities have an array of interests and are not merely focused on work.

3.1 Method

To assess personality we define the following criteria that quantify aspects of candidates' personality and can be mapped to real numbers in the interval [0, 1]. Their values can be calculated with web mining techniques, exploiting the candidates' social presence with respect to the following four criteria:

- *well-rounded*: quantifies in what degree a candidates' personality is fully developed, which is indicated by a number of non-work related leisure activities. Well-rounded candidates are persons with good social skills and are expected to function adequately in a team.
- *cooperative*: indicates well-rounded individuals with a demonstrated capacity to cooperate, as indicated by their participation in social after-work activities.
- *passionate*: candidates who are involved in activities related to the job position, or actively working in sharpening their skills after work.
- *influential*: candidates that others tend to use as a source of information or arguments. They tend to receive a high number of social interactions (e.g., re-shares, comments, etc.).

It is evident that these four criteria are not independent from each other. Rather, cooperative individuals are actually a subset of well-rounded personalities, while influential individuals are typically passionate persons that have the tendency to reveal information about their passions and promote them to other people.

The first step required to calculate the abovementioned selection criteria is to define a set of leisure activities that are supported by the system. These were selected from the categories of the Open Directory Project (ODP20) [14] and are shown in the following table (Table 2). Activities shown in bold are the collaborative activities that can indicate social and pleasant candidates.

Table 2. Activities that show a candidates' personality

1. Science	**2. Sports**
5. Travel	4. Computers
7. News	**6. Computer-Board Games**
9. Arts	8. Business

The second step is to create a corpus of words which are representative for each category. To create the corpus we used the Open Directory Project predetermined hierarchy of categories so as to access websites of a specific topic that respond to specific information needs. Then, we parsed these websites to make a corpus of words for each topic. Specifically, we used features extracted from the HTML code of the webpage including the pure text contained in it and the specific tag values as headings (<title>, <h1>) and meta information (<meta>). By discarding terms that appear less than 100 times (process only adjectives, verbs and nouns) we form a vocabulary that represents each category.

To find which categories are mentioned by the candidate often we analyze not only the hashtags that can directly give the information we need, but also the raw text of tweets and Facebook posts. We process the text in the tweets and Facebook posts and compute daily unigram frequencies. By discarding terms that appear less than 100 times, we form a vocabulary of size $|V| = 71,555$. We then form a user term-frequency matrix with the mean term frequencies per user during the time interval Δt. All term frequencies are normalized with the total number of tweets and Facebook posts posted by the user. The final step is to compute a topic score for each user-topic pair to assign the interest categories to the candidates. For that reason, we used the Jaccard index, also known as the Jaccard similarity coefficient [15]. The Jaccard coefficient measures similarity between finite sample sets and is defined as the size of the intersection divided by the size of the union of the sample sets:

$$\mathrm{J(A,B)} = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

Clearly,

$$0 \leq J(A,B) \leq 1 \tag{2}$$

To calculate the "well-rounded" metric we calculate the Jaccard coefficient for every pair of {activity words, candidate words} and the score is the greatest of the calculated values. In the same way we calculate the "cooperative score" and the "passionate score" using the corresponding activity words.

3.2 Identification of User Impact

As mentioned in previous sections, candidates that demonstrate passion for their leisure activities are also regarded more likely to be passionate in the workplace. One special category of passionate individuals are the ones that tend to influence others, attracting interest for their hobbies and leisure activities. In the social web, this is translated in a high degree of interactions with others. The most significant form of endorsement in the social media is the "share" or "retweet" function, in Facebook and Twitter platforms respectively. Influential users receive a high number of shares (i.e. other users repeat influential users' status updates on their own timelines). It must be noted that a high number of friends or followers is not a reliable metric of user impact, as shown in [16]. On the other hand, interactions (and especially shares) are indicative of a person that attracts interest and is viewed by others as a source of interesting information.

In order to quantify the user impact, we only take into account "status updates" (or "tweets") that fall into one of the categories illustrated in Table 2. The methodology detailed in Section 3.1 is employed to filter-out status updates or tweets that are not relevant to candidate interests. The rationale is that we are not interested in the candidate's impact as a whole in the social web, but rather on how successful he is in disseminating information about his leisure activities. This will generally imply that they are influential individuals. Thus, candidates' influence score is defined as the percentage of relevant status updates or tweets that their friends or followers interacted with (i.e., shared or retweeted). Notice that in the proposed method there is no dependency with the overall number of friends or followers, as would be the case if we counted the aggregate number of shares or retweets.

4 Pilot Scenario

In this section we evaluate the effectiveness of the proposed personality mining approach as detailed in Section 3, in a pilot recruitment scenario. The system's performance is assessed based on how effective it is in assigning consistent personality scores to the candidates, compared to the ones assigned by human recruiters. In our pilot scenario we compiled a corpus of 100 random twitter users, that reported to be working in the financial sector in their twitter bio. We subsequently derived their personality scores using the methodology detailed in Section 3. These scores were cross-compared with scores assigned by an expert recruiter, who had access to the same tweets.

The performance of the proposed system is evaluated based on how effective it is in discriminating the top candidates of each category, and providing a rank that is consistent with the one provided by the human recruiters. Three metrics were used for comparing rankings; the simplest one is the overlap size of the top-k list selected by the system and the human recruiter for each job position, where k=25 corresponds to 25% of overall applicants. The second metric is the correlation coefficient (Spearman's rho) of the top-k candidates per category. The third metric is the mean absolute difference (ranking error) of top-k candidate's ranks. The performance metrics for all three positions can be seen in Table 3.

Table 3. Performance evaluation metrics per job position

	Top-k	Correlation	Ranking error
Well-rounded	16 (64%)	0.66	4,8
Social	18 (72%)	0.68	4,1
Passionate	19 (76%)	0.71	3,9
Influential	22 (88%)	0.79	3,1

The calculation of the influential score was based on more objective criteria and thus this metric performed better than the others, with a correlation coefficient of 0.79. On the other hand the metrics that are based on the candidate's social presence had the lowest performance. Nevertheless, our method was able to output a top-25 list that overlapped at least 64% (for the well-rounded metric) and the correlation reached 0,71 (for the passionate metric).

Finally, in Fig. 2 we represent the candidates with circles positioned in a 2D plane based on the two most important personality scores (social and influential), while the circle radius is proportional to the candidate overall score, assigned by the recruiter. It is evident that most highly ranked candidates are clustered in the top right quadrant (i.e. with high social and influential scores), which attests that our tool assigned high ranks to candidates with the desired personality.

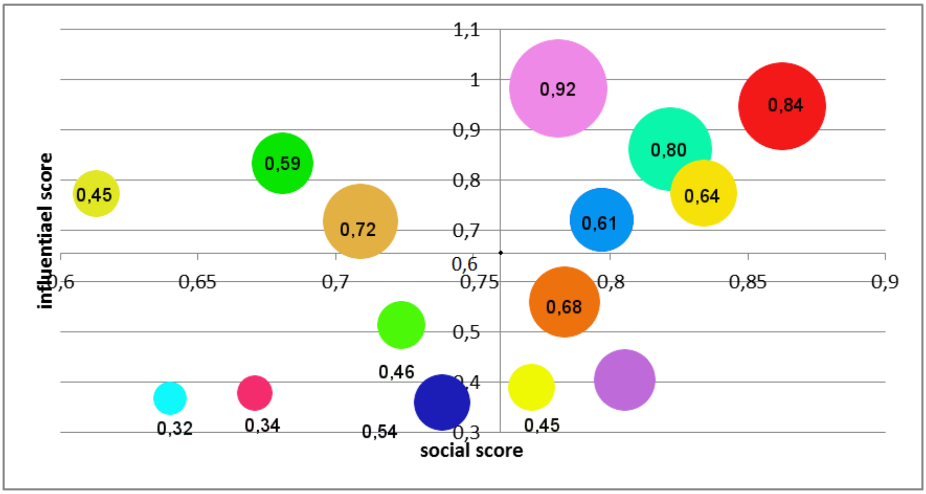

Fig. 2. The candidates' personality and overall scores

5 Conclusions

An inherent limitation of automated recruitment systems is their over-reliance on formal qualifications. Hiring managers are able to see beyond skill-sets, and assess candidates' personality and how they would fit in the corporate culture. Judging applicants' creativity, inspiration or their ability to work with people is a hard problem to solve for automated systems (assuming that it can be solved at all). This paper proposed a new approach in company-oriented recruitment by means of a system that provides automated candidates ranking and personality mining with the purpose of restricting interviewing and background investigation of applicants to the top candidates identified by the system. The proposed system is able to assess the applicants' suitability for a specific position, by inferring his personality traits based on his social media use. More specifically each candidate is evaluated based on how well-rounded, social, passionate and influential he is according to information collected by his social media activity. The performance of the system has been evaluated on the basis of how effective it is in discriminating the top candidates of each category, and outputting a ranking that is consistent with the one provided by the human recruiters. Among the four personality scores, the influential score obtained the highest accuracy, while the more subjective social score had the lowest. Regardless though of the partial scoring, it is quite positive that the system managed to output a top-25 list that overlapped with the corresponding ranking of the human recruiters at least 64% (for the well-rounded metric) and the correlation reached 0,71 (for the passionate metric). We plan to experiment with the system in full scale in order to investigate potentially better fine-tuning in the algorithm that assesses the four personality scores and set up recruitment scenarios in various domains.

References

1. Neuman, C.: Prospero: a tool for organizing Internet resources. Internet Research **20**, 408–419 (2010)
2. Ho, L., Kuo, T., Lin, B.: Influence of online learning skills in cyberspace. Internet Research **20**, 55–71 (2010)
3. Jansen, B., Jansen, K., Spink, A.: Using the web to look for work: Implications for online job seeking and recruiting. Internet Research **15**, 49–66 (2005)
4. Bizer, R.H., Rainer, E.: Impact of Semantic web on the job recruitment Process. Wirtschaftsinformatik, Physica-Verlag HD (2005)
5. Pande, S.: E-recruitment creates order out of chaos at SAT Telecom: System cuts costs and improves efficiency. Human Resource Management International Digest **19**, 21–23 (2011)
6. Faliagka, E., Iliadis, L., Karydis, I., Rigou, M., Sioutas, S., Tsakalidis, A., Tzimas, G.: On-line Consistent Ranking on e-Recruitment: Seeking the Truth Behind a Well-formed CV. Artif. Intell. Rev. **42**, 515–528 (2014)
7. Faliagka, E., Tsakalidis, A., Tzimas, G.: An integrated e-recruitment system for automated personality mining and applicant ranking. Internet Research **22**(5), 551–568 (2012)
8. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.: Sentiment analysis of twitter data. In: Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics (2011)

9. Kouloumpis, E., Wilson, T., Moore, J.: Twitter sentiment analysis: The good the bad and the omg! ICWSM **11**, 538–541 (2011)
10. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: LREC, vol. 10 (2010)
11. Saaty, T.L.: How to make a decision: The analytic hierarchy process. European Journal of Operational Research (1990). Elsevier Science B.V.
12. Basak, D., Srimanta, P., Dipak, C.P.: Support vector regression. Neural Information Processing-Letters and Reviews **10**, 203–224 (2007)
13. http://www.businessinsider.com/why-job-interviewers-ask-about-hobbies-2014-8#ixzz3VCFcHCja
14. http://www.dmoz.org/
15. Suphakit, N., Jatsada, S., Ekkachai, N., Supachanun, W.: Using of jaccard coefficient for keywords similarity. In: Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1 (2013)
16. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K.: Measuring User Influence in Twitter: The Million Follower Fallacy. ICWSM **10**, 10–17 (2010)