Improving Talent Management with Automated Competence Assessment: Research Summary

N.S. Nikitinsky

Abstract Improvement and automatization of Human Resource Management processes is an aspiring trend in modern business. In this paper, we summarize the results of the research we conducted in order to create a formal Text Mining-based competence assessment model and develop a decision support system for talent management (DSSTM) based on this model. We also conducted several experiments in order to improve the performance of the competence assessment model. The resulting prototype of DSSTM currently undergoes validation in a software development company and shows promising results.

Keywords Competence assessment • Text Mining • Decision support system
Word embeddings • Classification

Introduction

Currently, Human Resource Management (HRM) becomes a strategic trend for business. Managers from science-based industries pay special attention to Talent Management, becoming a popular tool in effective organization. As certain parts of HRM can be supported by technology, many enterprises use special software such as HR Information Systems or HR Management Systems.

Competence Management is an old, but still promising area in Talent Management field, consisting of all of a company's formal, organized approaches to ensuring that it has the human talents needed to meet its business goals. Now, there are several Competency Management Systems (CompMS) offering functionality to make the process of competency assessment more formal.

However, many contemporary CompMS propose only limited set of tools, based on traditional competence evaluation methods such as 360 degree-feedback or ordinal professional skills tests. The extensive use of Data Mining and, particularly,

N.S. Nikitinsky (✉)
NAUMEN, Moscow, Russia
e-mail: nnikitinskij@naumen.ru

Text Mining can greatly improve speed and quality of competence assessment, making it less human-biased at the same time. This approach should improve business processes and significantly decrease HR expenses. Several researchers and developers work in the field of applying modern Data Science approaches to the field of Competence Management.

Related Work

In our short survey, we will focus on several aspects. First, we will briefly describe attempts to introduce Data Mining into Decision Support Systems (DSS) field. Second, we will turn to contemporary full-cycle HRM systems and outline modern attempts to improve the Talent and Competence Management with Data Mining, including both theoretical models and practical applications.

Currently, several researchers offer promising Data Mining solutions for Decision Support Systems. For example, Dai et al. presented the Mining Environment for Decisions (MinEDec) model that introduced Text Mining technologies into Competitive Intelligence [4]. Bara and Lungu made a further effort to justify the importance of using Data Mining in DSS. In the paper, they described the process of developing DSS using Data Mining techniques and finished with the case study of applying data mining for forecasting the wind power plants' energy production [1].

To support decision making in Talent Management, there exist several multi-functional HRM software products, such as SAP SuccessFactors [17], Oracle Taleo [12], Cornerstone OnDemand [3], and People Fluent, Inc. [14]. These solutions cover the full cycle of employment and all HR processes, including hiring, developing, and retaining the best staff. Nevertheless, their approach to assessment of competences and qualifications is still mainly based on traditional methods.

In the meantime, various approaches for competence assessment have been discussed in academic literature. Some researches try to formalize the process of competence and qualification assessment and create theoretical models for competence assessment.

Thus, Berio and Harzallah in their earlier work discuss the model to manage four essential processes: identification, assessment, acquisition, and usage of the competences [2]. García-Barriocanal et al. developed a generic ontological model, which should facilitate the quantitative competences analysis [5]. Work of Rauffet et al. propose a methodology for assessment of organizational capabilities [16]. Haimovich proposed a mathematical model for assessing qualification efficiency of professionals [6]. However, it suits better for evaluation of qualification of university graduates, not professionals. Hassan et al. proposed theoretical model for an automated competency management system [7].

Researchers also conduct a study on applying standalone Data Mining modules to assist decision-making process in HRM. Ranjan presented the Data Mining approach to discover and extract useful patterns from this large data set to find

observable patterns in HR and thus improve the quality of the decision-making process in HRM [15]. Jantan et al. demonstrated how clustering, classification and association rules can be applied to solve modern Talent Management challenges such as detecting employees with similar characteristics, predicting the employee performance and associating the employee's profile to the most appropriate job [8].

DSSTM Concept

The idea to conduct research in the area of Competence Assessment originated from HR managers trying to cope with challenging Talent Management tasks such as searching for the most skillful employees or group of employees or improving speed and quality of employees' competences evaluation. As the result of research, we created a formal model for automating competence assessment and developed a prototype of Decision Support System for Talent Management (DSSTM) based on the model. The main functionality of DSSTM is competence assessment and services built on the top of it.

Currently for competence assessment, the system combines the information from three large sources: employee HR profile, all the text documents produced by employees (e.g., scientific publications, work reports) and results of professional skill tests and other traditional competence assessments methods. To obtain this data, DSSTM uses connectors to the most popular HR software systems, databases, and directory services.

As DSSTM focuses on text document analysis, it obtains texts of documents produced by employees. The system takes into account all the metadata derived from the documents, including document type (report, specification, etc.), co-authors, and results of document evaluation provided by colleagues (such as likes or comments), whereas some data (e.g., information from the employee HR profile) demands relatively simple transformation (e.g., ranking or averaging), preprocessing of documents is a more complex task. Apart from tokenization, stop words removal, and morphological analysis of the contents (i.e., part-of-speech tagging, lemmatization, etc.), we employ word2vec and latent semantic analysis (LSA) for text classification and key terms extraction.

Word2vec is a tool implementing two neural network architectures, which is used for word embeddings analysis. The assumption behind the model is that words located in the similar contexts tend to have semantic closeness (i.e., similar meanings) [11].

LSA is a natural language processing technique, which analyzes relationships between a set of documents and the terms they contain [10]. The assumption behind the algorithm is very similar to that of word2vec. LSA constructs a weighted term-document matrix, where rows represent unique words, and columns represent documents.

Text classification allows DSSTM to define document subject areas such as scientific areas (physics, chemistry, etc.) and then compare vectorized document of

certain subject area to a vector of a benchmark document in order to define its quality.

The algorithm for text classification in DSSTM consists of five steps. First, the word2vec model is trained with given text corpus where each text is tagged with certain topic name. Second, the text from each topic in corpus is projected into the word2vec model and transformed into the sum of word vectors, thus we obtain a so-called “topic vector.” Third, the input text undergoes the same transformation. Fourth, the system then compares vectorized input text to each topic vector. Fifth, the input text is then tagged with topic names with semantic similarity above the threshold.

For key term extraction, we apply a combination of the LSA approach with a rule-based approach. We first select candidates for key terms from the document with predefined part-of-speech-based rules. Then, we estimate cosine similarity between each candidate-term vector and document vector in the LSA space (thus, we obtain a list of so-called local key terms). After that, we estimate similarity between the document vector and all lemmas in the semantic space (and thus we obtain list of so-called global key terms). Finally, we select top-n key terms (both for global and local key terms) and obtain a list of the most appropriate human-interpretable key terms (including n-grams). The extracted key terms are later used for competence assessment and in other additional modules.

Competence Assessment Model

By a competence, we mean a combination of a skill applied to a certain domain. For instance, the skill “Analytics” applied to a domain of Biology is a competence “Analytics in Biology.”

Competence assessment consists of two steps: First step is the identification of competence, when the system checks if an employee has certain competence. Second step is the evaluation of competence, when DSSTM evaluates the competence with help of set of modifiers.

To identify competence, DSSTM uses a rule-based approach. Based on possible features, obtained from profile, text documents or professional skills tests, a user may create rules to identify the presence of competence for an employee. For instance, there may be a rule similar to this: *check for presence of at least 3 documents of type “RnD report” with topic “Physics” AND average semantic similarity of that three documents to the benchmark RnD Physics report must be at least 0.7 AND key terms from that three documents must match benchmark key terms for this competence at least for 70%.*

The result of the competence identification must be a number in order to be used in the process of competence evaluation.

The general formula for competence evaluation thus may be expressed as

$$rate_{comp} = B_{comp} * \left(\frac{1}{3} (B + HR + TXT) \right), \quad (1)$$

where $rate_{comp}$ —competence score, B_{comp} —basic score, B —basic parameters modifier score, HR —HR modifier score, and TXT —text modifier score.

We scale the resulting competence score to the conventional scale (say, 1–5, where 1 means junior level of competence and 5—expert compared to other employees with the same competence). Therefore, the results from the formula (1) need to be adjusted accordingly:

$$rate_{comp} = \begin{cases} MaxScale & \text{if } rate_{comp} > MaxScale \\ rate_{comp} & \text{if } rate_{comp} \in Scale \\ MinScale & \text{if } rate_{comp} < MinScale \end{cases}, \quad (2)$$

where $rate_{comp}$ —competence score, $MaxScale$ —the maximum value for the scale, $MinScale$ —the minimum value for the scale, and $Scale$ —the scale value within the range.

As the result, an employee gets his competence level evaluated and scaled to a conventional scale. For example, “Analytics in Biology—3.”

The formula of competence evaluation (1) is composite and consists of basic score and several modifier scores—Basic parameters modifier, HR modifier, and Text modifier. To calculate every modifier, currently we use more than 10 parameters.

To make sure that all the modifier parameters produce values between 0 and 1, we apply scaling all the modifiers’ parameters using well-known Min-Max scaling technique.

Basic score indicates the evidence of the identified competence. It is calculated as ratio of the result of competence identification for certain employee to maximum result of competence identification in a department or whole organization:

$$B_{comp} = \frac{CI_{res}}{CI_{max}}, \quad (3)$$

where B_{comp} —basic score, CI_{res} —the result of competence identification for certain employee, CI_{max} —maximum result of competence identification in a department or whole organization.

Basic parameters modifier takes into account the numerical or categorical parameters from the employee HR profile, such as overall work experience, work experience in the company, amount of KPI achieved, etc.

HR modifier is based on traditional evaluation methods such as 360 degree-feedback, various professional tests, and different surveys. The modifier utilizes scores earned by employee in the tests.

Text modifier evaluates the indirect quality of text documents by estimating text parameters such as readability (calculated via different methods such as Flesh-Kinkaid and SMOG readability indices), uniqueness of lexis, etc.

The general formula to calculate any modifier is the following:

$$MOD = gl * \left(\sum_{i=1}^n (D(mod_i) * imp) \right), \quad (4)$$

where $D(mod_i)$ —adjusted modifier element from 0 to 1 for computational convenience, imp —the weight for each element (by default equal 1), gl —global weight for the element (by default equal $1/n$).

The competence assessment algorithm is customizable and allows adding more rules and parameters into the formulas.

Evaluation of separate competences cannot tell CEO or HR manager, how qualified each employee is in general and in comparison to others with same competences and same position in a company. Thus, we also provide evaluation of the *overall qualification level*. Average qualification evaluates the current qualification level of employee. The index represents the average of present competences:

$$Q = \frac{1}{n} \sum_{i=1}^n rate_i, \quad (5)$$

where Q —the level of employee qualification and $rate_i$ —the current level of certain employee competence.

We also calculate a *Qualification with reference to job requirements*, which evaluates the employee's qualification required for promotion to the next position within the company structure. This parameter compares the current level of employee's qualification with the required level for specific position, which has to be defined in advance.

Professional interest discovery at the same time is intended to define the subject areas an employee is interested in. DSSTM detects two types of interests: global and local. Global interests are determined as *top-n* most frequent subject areas from the employee's text documents. Local interests are determined as *lists of key terms* from clustered personal employee's semantic LSA subspace. For LSA subspace clustering, we employ Clustering by Committee (CBC) algorithm [13]. Lists of key terms are extracted via the above described key terms extraction algorithm.

As a result, we obtain a list of key terms for every local professional interest, which is human-interpretable. Either, we sort the discovered local interests according to their significance (based on occurrence in texts) and relate them to global ones to let CEO or HR managers find out that, for example, certain employee is interested in specific branch of biology.

Application of Assessed Competences

The assessed competences, qualification, and professional interests are used to create a personal DSSTM profile for every employee, perform search for employees and let the employees with similar competences communicate and share knowledge. Furthermore, DSSTM computes competence statistics for a company, which can be used for analysis purposes—for example, to find out strategic competences or the competences that are weakly represented in the company.

DSSTM has several modules built over the competence assessment model, which help the managers in their daily Talent Management tasks:

Information Retrieval System module provides search for employees and task teams based on specified or unknown parameters of competence, qualification, and/or professional interests.

Content recommendation system (CRS) is designed to share knowledge between employees by recommending content created by one employee to other employees. The system is based on two approaches: content filtration and rule-based method to fit both employees who produced sufficient amount of documents and newcomers.

Employee recommendation system (ERS) is designed to recommend employees to each other, assuming that the employees with close profiles (including competences and qualification) and semantically close documents tend to have common tasks and interests and may help each other in their daily work. The system is based on the same approach as CRS.

Automatic employee catalog was created to group employees with close professional characteristics (including competences) based on all the data about the employees in the system. This catalog is intended to help employers with organization of task team in a bit different way than task-team search.

Additional Experiments

After we created DSSTM prototype, we started to research possibilities of applying some sociological and psychological aspects to the system as a possible parameters or modifiers in order to improve the quality of the competence assessment, since current version of DSSTM lacked such parameters. As first step, we decided to try to work in the Generation Theory framework. As part of the research, we conducted several experiments in order to create a model, which could detect a generation of a person based only on the texts, produced by the person.

Generation Theory is an approach used in management, which is based on specificity of various generations. The key notion of Generation Theory is a generation—a group of people born in a certain period of time who acquired common behavior scenarios, habits, beliefs, and language peculiarities [18].

For the study, we randomly selected in popular social networks (namely, Facebook and Vk.com) 600 Russian-speaking people from one of the five largest

Russian cities, who have been born between 1968 and 1981 (for generation X, excluding borderline years) or between 1989 and 1999 (for generation Millennium, excluding borderline years) and posted on social network more than five large texts created by themselves (no reposts).

We preprocessed all the texts with tokenization, stop words removal and morphological analysis. We extracted key terms from every text, using algorithm for local key terms extraction, which was described previously in the paper. Then, every person's keywords were transformed into vector in word2vec space. We used the pretrained word2vec model, namely, Web corpus from RusVectors [9].

We randomly selected 25% of all person's vectors for every generation. Then, we averaged the vectors, thus obtaining the so-called "generation vector." Therefore, we got Generation X vector and generation Millennium vector. The other 75% of persons from each generation were selected as validation sample. We iteratively compared each vector from validation sample to the Generation X vector and generation Millennium vectors. We related each vector from validation sample to certain generation if it was closer to this "generation vector" than to another.

We conducted 20 iterations, each time randomly selecting the generation samples in order to verify the stability of the approach and obtained average accuracy of classification equal to 0.951 and 0.953 for generation X and Millennium, respectively. Average F1-measure on 20 iterations was at the same time 0.797, which is a decent result.

Results and Discussion

The prototype version of DSSTM has been functioning for 3 months in the software developing company in Russia. The practical implementation of the system demonstrated promising results such as average time to find an employee with required competencies within the company decreased by almost 35% from 2–3 to 0.7–1.5 h. However, we also revealed certain limitations of the model—for instance, the model can detect and estimate mostly technical competencies and it works well only for those employees, who produce at least some text (e.g., programmers, analysts, marketing managers etc.). In addition, it may be a challenging task to create detection conditions for some competences, which are difficult to formalize.

Additional experiments showed us that text-based generation classifier might be utilized as an additional parameter for the DSSTM or other similar system. As a standalone module, the classifier may help HR managers better understand the life values of certain persons who were born in borderline years for two types of generations (e.g., X and Millennium) and may belong to either one or the other generation. Certainly, this classifier model also has limitations, e.g., it should be retrained on newly acquired data often enough, as persons write texts in social networks under influence of various social or political issues, which tend to change

in the course of time. Thus, the model built on the texts written recently, will tend to produce worse classification results in future.

Conclusion

Nowadays, automation and formalization of HRM and Talent Management activities such as competence assessment are in high demand for large companies. Several researchers focus on this topic and apply Data Mining techniques to assist decision-making process in HRM.

In this paper, we briefly described the results of our research, which was focused on designing a theoretical framework for competence assessment, its practical implementation, developing additional modules for that implementation, which are useful for business, and conducting experiments in order to find new features for the model, which will be able to improve its quality and business-value.

We created a prototype version of Decision Support System for Talent Management, which was based on a text mining-based theoretical framework for competence assessment. This prototype version has been undergoing validation in a software development company and showing promising results. For a future study on this, we would suggest focusing on improving current limitations of a competence assessment model.

Additional experiments showed that the word2vec-based classifier can be applied to pure text-based detection of generation of a person.

Acknowledgements Research was conducted with the financial support of the state represented by the Ministry of Education and Science of the Russian Federation. Agreement no. 14.579.21.0091 28. Nov 2014. Unique project Identifier: RFMEFI57914X0091.

We would like to acknowledge hard work and commitment from Alexey Korobeynikov and Alexey Nesterenko throughout this study.

References

1. Bara, A., Lungu, I.: Improving decision support systems with data mining. *Advances in data mining knowledge discovery and applications* (2012)
2. Berio, G., Harzallah, M.: Towards an integrating architecture for competence management. *Comput. Ind.* **58**(2), 199–209 (2007)
3. Cornerstone. Cornerstone OnDemand official website [Electronic source] URL: <https://www.cornerstoneondemand.com/> (date accessed 20.11.2016)
4. Dai, Y., Kakkonen, T.: MinEDec: a decision-support model that combines text-mining technologies with two competitive intelligence analysis methods. *Int. J. Comput. Inf. Syst.* **3** (1), 165–173 (2011)
5. García-Barriocanal, E., Sicilia, M.A., Sánchez-Alonso, S.: Computing with competencies: modelling organizational capacities. *Expert Syst. Appl.* **39**(16), 12310–12318 (2012)

6. Haimovitch, I.N.: Metodika integral'noi otsenki effektivnosti kvalifikatsii spetsialista [Methods of integrated performance assessment training]. *Sovremennye problemy nauki i obrazovaniya* **6**, 1–8 (2013)
7. Hassan, W., et al.: Automated competency management system: an advanced approach to competence management efficiency. *Eur. J. Business Manage.* **5**(16), 64–74 (2013)
8. Jantan, H., Hamdan, A.R., Othman, Z.A.: Towards applying data mining techniques for talent management. *International conference on computer engineering and applications (IACSIT)* **2**, 476–481 (2011)
9. Kutuzov, A., Andreev, I.: Texts in, meaning out: neural language models in semantic similarity task for Russian. In: *Proceedings of the Dialog 2015 Conference*, vol. 2, pp. 133–145 (2015)
10. Landauer, T.K., McNamara, D.S., Dennis, S., Kintsch, W. (eds.): *Handbook of latent semantic analysis*. Lawrence Erlbaum Associates, (2007)
11. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. In: *Proceedings of workshop at ICLR*, (2013)
12. Oracle.: Oracle Taleo Enterprise Cloud Service [Electronic source] URL: <http://www.taleo.net/> (date accessed 20.11.2016)
13. Pantel, P.A.: *Clustering by committee*. University of Alberta, Diss (2003)
14. PeopleFluent.: PeopleFluent HCM Software official website [Electronic source] URL: www.peoplefluent.com/ (date accessed 20.11.2016)
15. Ranjan, J.: Data mining techniques for better decisions in human resource management systems. *Int. J. Business Inf. Sys.* **3**(5), 464–481 (2008)
16. Rauffet, P., Da Cunha, C., Bernard, A.: A dynamic methodology and associated tools to assess organizational capabilities. *Comput. Ind.* **65**(1), 158–174 (2014)
17. SAP SuccessFactors official website [Electronic source] URL: <https://www.successfactors.com/> (date accessed 20.11.2016)
18. Strauss, W., Howe, N.: *Generations. Morrow*, New York (1991)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.