

УДК 004.89
DOI 10.17513/snt.39757

МЕТОДИКА ПРОВЕДЕНИЯ КЛАССИФИКАЦИИ РЕЗУЛЬТАТОВ ТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ РЕЗЮМЕ КАНДИДАТОВ ПО КОМАНДНЫМ РОЛЯМ

¹Безруких А.Д., ¹Горлушкина Н.Н., ¹Черепанов М.Д., ²Безруких Ю.А.

¹ФГАОУ ВО «Национальный исследовательский университет ИТМО», Санкт-Петербург,
e-mail: anastasiyabzr@gmail.com;

²ФГБОУ ВО «Сибирский государственный университет науки и технологий
имени академика М.Ф. Решетнева», Красноярск, e-mail: expert-sib@yandex.ru

На сегодняшний день информационные технологии являются востребованной сферой деятельности, и количество специалистов в этой области постоянно растет, как и частота сменяемости кадров. В связи с этим возрастает и нагрузка на рекрутеров, из-за чего крупные компании вынуждены расширять штаты специалистов по подбору кадров и тратить на это дополнительные средства. Большую часть работы рекрутера занимает выполнение рутинных задач, основной из которых является скрининг резюме. При этом специалист может просмотреть, упустить или забыть важную информацию из-за влияния человеческого фактора. Использование нейронных сетей для решения узкоспециализированных, рутинных задач в управлении человеческими ресурсами – целесообразно. Благодаря такому внедрению, время на выполнение стандартных, повторяющихся задач, таких как скрининг резюме, у рекрутера сокращается. Автоматизация позволит исключить ошибки, которые совершает человек, при этом увеличить количество рассматриваемых резюме. В данной статье описывается разработка методики классификации резюме с использованием ролей по Белбину, которая позволит автоматизировать процесс подбора специалистов. В рамках статьи рассматриваются современные методы тематического моделирования, а также проводится сравнительный анализ моделей определения командных ролей. Разработана и описана методика классификации результатов тематического моделирования резюме с использованием модели командных ролей Белбина.

Ключевые слова: тематическое моделирование, командные роли, классификация, методика, формирование команды

CLASSIFICATION METHODOLOGY FOR THEMATIC MODELING RESULTS OF CANDIDATES BY TEAM ROLE

¹Bezrukikh A.D., ¹Gorlushkina N.N., ¹Cherepanov M.D., ²Bezrukikh Yu.A.

¹ITMO University, Saint Petersburg, e-mail: anastasiyabzr@icloud.com;

²Reshetnev Siberian State University of Science and Technology, Krasnoyarsk,
e-mail: expert-sib@yandex.ru

Today, information technology is a sought-after field of activity, and the number of specialists in this field is constantly growing, as is the turnover rate. In this regard, the burden on recruiters is increasing, which is why large companies are forced to expand the staff of recruiters and spend additional funds on this. Most of the work of a recruiter is occupied by performing routine tasks, the main of which is resume screening. At the same time, a specialist can overlook, miss or forget important information due to the influence of the human factor. The use of neural networks for solving highly specialized, routine tasks in human resource management is advisable. Thanks to this implementation, the recruiter's time to perform standard, repetitive tasks, such as resume screening, is reduced. Automation will eliminate the mistakes that a person makes, while increasing the number of resumes considered. This article describes the development of a methodology for classifying resumes using roles according to Belbin, which will automate the process of selecting specialists. Within the framework of the article, modern methods of thematic modeling are considered, as well as a comparative analysis of models for determining team roles. As a result, a methodology for classifying the results of resume thematic modeling using the Belbin team role model was developed and described.

Keywords: thematic modeling, team roles, classification, methodology, team building

В сфере информационных технологий рынок труда становится все более конкурентным. И крупные компании сталкиваются с рядом проблем, которые связаны с поиском и наймом квалифицированных ИТ-специалистов. В условиях постоянной текучести кадров и быстрого технологического развития бизнесу необходимо быстро находить новых сотрудников, чтобы не допустить простоя и сбоев в работе.

Сегодня рекрутеры больших фирм физически не успевают просмотреть все поступающие резюме. Согласно статистике за март 2022 г. с сайта HeadHunter, крупнейшей российской компании интернет-рекрутмента, более 92% резюме кандидатов остались без внимания. В среднем специалист тратит около 2 мин на первичный анализ одного резюме [1]. Человеческий фактор при такой рутинной работе может привести к ошиб-

кам при выборе кандидатов, основанным на субъективных предпочтениях и предвзятости рекрутеров, и, как следствие, к потере ресурсов и времени компании.

В условиях ухода с рынка иностранного программного обеспечения компании сталкиваются и с необходимостью импортозамещения, внедрения новейших технологий для управления кадровыми ресурсами. Для решения данных проблем активно разрабатываются методы автоматической обработки и анализа резюме с использованием современных информационных технологий. Автоматизация процесса отбора и классификации резюме позволяет значительно увеличить производительность труда рекрутеров, сократить временные и финансовые затраты компании на подбор персонала и повысить объективность в выборе кандидатов [2].

Актуальность статьи заключается в возрастающих трудозатратах при ручной обработке резюме и недостатке объективности у рекрутеров при формировании команд. Цель – разработка методики классификации результатов тематического моделирования резюме кандидатов по ролям.

Материалы и методы исследования

В современном мире коллективная работа и командное взаимодействие имеют большое значение в достижении успеха в различных сферах, поэтому понимание ролей, которые есть в команде, влияет на ее успех. Командная роль описывает функции и ответственности, которые индивиды выполняют внутри команды, в то время как оценка успеха команды помогает определить степень достижения поставленных целей и выполнения задач.

Существует множество моделей, разработанных для определения командных ролей, такие как модель Белбина, Маргерисона, Мансона и др. Каждая из этих моделей предлагает свой подход к классификации и описанию ролей, основываясь на различных теоретических концепциях и исследованиях.

На сегодняшний день определение командных ролей осуществляется через прохождение специального теста. Он представляет собой серию вопросов или утверждений, на которые респонденты должны ответить, выбирая наиболее подходящий вариант или оценивая себя по определенным параметрам.

В ходе сравнительного анализа моделей командных ролей выбрана модель Белбина, так как она является одной из наиболее широко известных и исследованных. Ее эф-

фективность и достоверность подтверждены рядом научных работ [3].

Технология отбора кадров и формирования команд в каждой организации зависит от специфики деятельности. Использование методов психологического тестирования рационально при комплексном подходе к выбору кандидатов.

Для выделения тем и ключевых слов из резюме кандидатов можно использовать тематическое моделирование. Это статистический подход к анализу текстов, который позволяет автоматически выявлять темы в больших коллекциях текстовых документов. В основе этого подхода лежит предположение о том, что каждый документ представляет собой смесь нескольких тем, а каждая тема определяется своим уникальным распределением слов. Таким образом, тематическое моделирование позволяет найти скрытые темы в текстах и определить, какие слова наиболее характерны для каждой темы.

Несмотря на то, что тематическое моделирование традиционно применялось для анализа больших коллекций текстов, есть исследования [4–6], в которых подтверждается эффективность метода на коротких текстах. Так как короткие тексты содержат ограниченное количество слов, это затрудняет точное определение и выявление скрытых тем. Однако существуют модели, которые показывают хорошие результаты на коротких текстах. Среди них: Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), BERT (Bidirectional Encoder Representations from Transformers).

Для более глубокого понимания и анализа текстовых данных необходимо корректно интерпретировать полученные результаты тематического моделирования. Для этого используется метод классификации текстов. Он позволяет автоматически присвоить текстовому документу определенную категорию или метку на основе его содержания. Учитывая особенности данных, доступные вычислительные ресурсы, а также требования к точности и интерпретируемости классификации, для классификации результатов тематического моделирования выбраны следующие подходы: Наивный Байесовский классификатор, методы опорных векторов и глубокого обучения, Случайный лес.

Классификация результатов тематического моделирования позволяет организовать полученные темы и интерпретировать их содержание. Кроме того, классификация позволяет автоматически классифицировать тексты на основе имеющихся тематических моделей [5].

Сравнительная оценка эффективности методики классификации результатов тематического моделирования с ручной классификацией, выполняемой рекрутером, позволит определить, насколько успешно методика может заменить или дополнить человека в процессе формирования команд на основе резюме кандидатов.

Результаты исследования и их обсуждение

Методика классификации результатов тематического моделирования по ролям Белбина представляет собой разработку гибридного подхода, комбинирующего преимущества различных методов, а также адаптацию и оптимизацию существующих решений для совершенствования системы управления кадровыми ресурсами в ор-

ганизационных системах. Подход состоит из пяти этапов.

На первом производится необходимая предобработка данных для последующей классификации результатов тематического моделирования по ролям Белбина. В табл. 1 описаны основные действия по каждому шагу, предполагаемый результат и используемые средства.

На втором этапе используется ранее описанная методика тематического моделирования для выделения тем и ключевых слов из резюме кандидатов. Результат этапа – список тем, связанных с каждым резюме, и набор ключевых слов, характеризующих каждую тему. В табл. 2 описаны основные шаги второго этапа процесса тематического моделирования резюме.

Таблица 1

Основные шаги первого этапа

№	Действие	Результат	Используемые средства
1	Сбор реальных данных для классификации резюме	Набор данных, состоящий из резюме и уже известной информации о том, какая роль Белбина присвоена методом прохождения теста	Запустить в социальных сетях опросники
3	Сортировка по профессиям	Отсортированный набор данных	Python, библиотеки: numpy, pandas
4	Выбор процентного соотношения деления данных на обучающую и тестовую выборки	Обучающая и тестовая выборки	Научная литература, подобные исследования на русском и английском языках
5	Очистка и нормализация данных	Очищенный и нормализованный набор данных	Python, библиотеки: nltk, pymorphy2

Таблица 2

Основные шаги второго этапа

№	Действие	Результат	Используемые средства
1	Создание векторного представления резюме	Корпус биграмм и/или триграмм корпус в виде «мешка слов» (bag of words)	Python, библиотеки: pymorphy2, nltk [7]
2	Модель LDA, NMF, BERT для 5, 7, 10 тем	Интерактивная визуализация расстояний между темами для каждой модели	Python, библиотеки: LatentDirichletAllocation, pyLDAvis, keras, и др.
3	Сравнение корпусов данных с помощью расстояния Левенштейна	Сравнительная таблица расстояний Левенштейна по каждой модели	Python, библиотеки: python-Levenshtein, textdistance; формула расстояния Левенштейна
4	Оценка качества моделей	Сравнительная таблица по когерентности для каждой модели на 3, 5, 10 темах	Когерентность темы PMI-Score [8, 9]
5	Назначение каждому резюме в корпусе подходящей темы	Дополнительный столбец в наборе данных с результатами по теме	Python + библиотеки моделей
6	Создание таблицы тем и топа слов для каждой модели	Таблица тем и топа слов для моделей	Python + библиотеки matplotlib, numpy, pandas
7	Результирующая таблица	Сводная таблица из исходных данных, списка тем и ключевых слов	Python + библиотеки matplotlib, numpy, pandas и др.

Таблица 3

Основные шаги третьего этапа

№	Действие	Результат	Используемые средства
1	Создание обучающей и тестовой выборок	Данные для обучения и оценки модели	Результирующая таблица второго этапа; Python, библиотека: scikit-learn
2	Обучение Наивного Байесовского классификатора, метода опорных векторов, Случайного леса, глубокое обучение	Обученные модели, время обучения классификатора	Python, библиотеки: scikit-learn, nltk и др.
3	Классификация тестовых данных	Классифицированные результаты	Обученные модели классификации

Таблица 4

Основные шаги четвертого этапа

№	Действие	Результат	Используемые средства
1	Оценка качества классификации	Сравнительная таблица	Метрики Accuracy, F-score и др. [10]
2	Выбор лучшей модели на основании метрик	Обученные модели классификации	Ручное сравнение
3	Проверка калибровки модели	Диаграмма надежности	Диаграмма надежности
4	Предсказание роли по резюме	Предсказанная роль	Метод predict
5	Подсчёт вероятности предсказанной роли	Вероятность	Метод predict_proba

Таким образом, на этом этапе выделяются темы и ключевые слова из предварительно обработанных данных. Для этого используются различные методы, включая алгоритмы тематического моделирования. Эти методы помогают выявить основные темы и ключевые слова, которые отражают навыки, опыт и интересы кандидатов.

Третий этап – этап обучения классификатора. Данный этап включает использование различных методов машинного обучения. Классификация основана на вероятностях, предсказываемых классификатором, на пороговых значениях, установленных на основе тренировочного набора данных. В табл. 3 описаны основные шаги третьего этапа.

Четвертый этап – интерпретация, оценка результатов классификации. Анализ соответствия выделенных ролей Белбина и оценка их значимости для каждого кандидата. В табл. 4 описаны основные шаги этапа.

Результатом четвертого этапа является оценка результатов классификации, включая вычисление метрик качества классификации и анализ полученных результатов.

Пятый этап – оценка эффективности методики в сравнении с ручной классификацией. Этап имеет решающее значение, так как он позволяет определить, насколько успешно методика может заменить или дополнить человека в процессе формирования команд на основе резюме кандидатов.

Перед тем как оценить эффективность описанной на предыдущих шагах методики, необходимо определить соответствующие метрики и критерии. Для этой цели определены следующие показатели эффективности:

- Сокращение трудозатрат: экономия рабочего времени рекрутера на скрининг резюме.

- Снижение затрат по сравнению с ручной обработкой: экономия ресурсов компании.

- Минимизация ошибок: снижение вероятности ошибок и улучшение качества подбора персонала.

- Скорость и объем обработки резюме.

- Исключение субъективности в процессе подбора персонала.

- Качество состава рабочих команд: насколько хорошо методика позволяет формировать сбалансированные и оптимальные команды по сравнению с человеком.

- Увеличение производительности: ускоряет процессы закрытия вакансий и уменьшает вероятность ошибок, производительность рекрутеров повышается.

На пятом этапе проводится сравнительная оценка эффективности методики классификации результатов тематического моделирования с ручной классификацией, выполняемой человеком-рекрутером. Ключевые шаги и аспекты на этом этапе описаны в табл. 5.

Таблица 5

Основные шаги пятого этапа

№	Действие	Результат	Используемые средства
1	Обучение рекрутера модели командных ролей Белбина	Рекрутер, способный классифицировать резюме; время, за которое обучается среднестатистический специалист	Литература, модель командных ролей Белбина, тест Белбина
2	Классификация резюме рекрутером	Время, за которое человек классифицирует 1 резюме, 20 резюме	Тестовый набор данных/резюме в количестве 20 штук из табл. 1 пункт 1
3	Классификация резюме обученным классификатором	Время, за которое классифицирует 1 резюме с помощью описанной методики, 20 резюме	Тестовый набор данных из предыдущего пункта; методика классификации резюме
4	Сравнение результатов классификации рекрутером и с помощью методики	Сравнительная таблица по критериям эффективности	Python + библиотеки matplotlib, numpy, pandas и др.

Важно отметить, что каждый рекрутер имеет различную квалификацию и опыт. Такая оценка и классификация может быть непостоянной и субъективной. Поэтому пятый этап позволит не только оценить эффективность описанной методики, но и определить ее применимость в конкретных условиях и задачах организации по сравнению с ручной обработкой.

Заключение

В данном исследовании проведен сравнительный анализ моделей командных ролей, выделены методы тематического моделирования, которые можно успешно применять на коротких текстах, а также определен основной подход классификации результатов тематического моделирования.

В результате разработана методика классификации результатов тематического моделирования резюме с использованием ролей Белбина. Она включает пять основных этапов. Важно отметить, что методика классификации резюме по ролям Белбина требует достаточного количества тренировочных данных с уже известными ролями кандидатов. Кроме того, она может быть уточнена и дополнена с учетом специфических требований и особенностей конкретной организации или команды.

Список литературы

1. Dixit V.V., Trisha Patel, Nidhi Deshpande, Kamini Sonawane Resume Sorting using Artificial Intelligence // International Journal of Research in Engineering, Science and Management. 2019. Vol. 2, Is. 4. P. 423–425. DOI:10.33545/26648776.
2. Блиникова А.В., Йинг Д.К. Использование искусственного интеллекта в процессах управления челове-

скими ресурсами // Вестник университета. 2020. № 7. URL: <https://vestnik.guu.ru/jour/article/view/2305> (дата обращения: 20.08.2023). DOI: 10.26425/1816-4277-2020-7-14-21.

3. Ruolan Bao. Research on the Relationship between Team Roles Theory and Team Effectiveness – Taking the Sales Team of Company A as an Example // 2019 3rd International Conference on Economics, Management Engineering and Education Technology (ICEMEET 2019). 2019. P. 2121–2125. DOI: 10.25236/icemeet.2019.424.

4. Jipeng Qiang, Zhenyu Qian, Yun Li, Yunhao Yuan, Xindong Wu. Short Text Topic Modeling Techniques, Applications, and Performance: A Survey // IEEE Transactions on Knowledge and Data Engineering. 2020. Vol. 34, Is. 3. P. 1427–1445. DOI: 10.1109/TKDE.2020.2992485.

5. Wen Hua, Zhongyuan Wang, Haixun Wang, Kai Zheng, Xiaofang Zhou. Short text understanding through lexical-semantic analysis // 2015 IEEE 31st International Conference on Data Engineering. 2015. P. 495–506. DOI: 10.1109/ICDE.2015.7113309.

6. Xiaobao Wu, Chunping Li, Yan Zhu, Yishu Miao. Short Text Topic Modeling with Topic Distribution Quantization and Negative Sampling Decoder // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020. P. 1772–1782. DOI: 10.18653/v1/2020.emnlp-main.138.

7. Jey Han Lau, Timothy Baldwin. An empirical evaluation of doc2vec with practical insights into document embedding generation // Proceedings of the 1st Workshop on Representation Learning for NLP. 2016. P. 78–86. DOI: 10.18653/v1/W16-1609.

8. Краснов Ф.В., Баскакова Е.Н., Смазневич И.С. Оценка прикладного качества тематических моделей для задач кластеризации // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2021. № 56. С. 100–111.

9. Булатов В.Г. Методы оценивания качества и многокритериальной оптимизации тематических моделей в библиотеке TopicNet: Специальность 05.13.18 «Математическое моделирование, численные методы и комплексы программ»: автореф. дис. ... канд. техн. наук. Москва, 2020. 147 с.

10. Jey Han Lau, David Newman, Timothy Baldwin. Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality // In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics. 2014. P. 530–539. DOI: 10.3115/v1/E14-1056.