

СИСТЕМА ОПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ УСПЕШНОГО ПРОХОЖДЕНИЯ СОБЕСЕДОВАНИЯ СОИСКАТЕЛЕМ НА ОСНОВЕ ЭКСПЕРТНЫХ ОЦЕНОК

Нестеренков С. Н., Видничук В. Н., Шинкевич Н. Н

кафедра программного обеспечения информационных технологий, Белорусский государственный университет информатики и радиоэлектроники

Объединённый институт проблем информатики Национальной академии наук Беларуси

Минск, Республика Беларусь

E-mail: nsn@bsuir.by, {vadimxpzl, sn0wf1llin}@gmail.com,

Предложен алгоритм определения вероятности успешного прохождения собеседования соискателем на основе экспертных оценок его навыков, который в дальнейшем будет использован в интеллектуальной системе подбора команды специалистов для решения поставленной задачи.

ВВЕДЕНИЕ

В современном мире всё чаще используются нейронные сети для решения различных задач. Одним из приоритетных направлений является решение задачи классификации. Она заключается в отнесении определённого образца к не пересекающимся множествам. В данной статье предлагается рассмотреть задачу классификации для определения вероятности успешного прохождения собеседования в компании соискателем на основе экспертных оценок. На протяжении некоторого времени собирается статистика о компании, её работниках и в зависимости от этой статистики определяются пространства распределения.

I. СБОР СТАТИСТИКИ

Статистика о компании включает в себя данные о:

- сотрудниках, работающих в данной компании;
- сотрудниках, уволившихся по собственному желанию;
- соискателях, успешно прошедших собеседования в данной компании;
- соискателях, его не прошедших;
- навыках соискателей и сотрудников, требующихся для успешной работы в данной компании.

В статистику соискателей собираются данные с мест его работы, экспертные оценки навыков и времени работы. Экспертные оценки навыков собираются путём прохождения пользователем курсов, его обучения в различных заведениях. Они характеризуют знания соискателя в определённых областях. Каждому навыку ставится своя экспертная оценка. Далее рассчитывается достоверная экспертная оценка, при наличии нескольких факторов, по которым следует оценить каждый из объектов [1-2]. Средняя оценка каждого объекта может быть рассчитана по формуле:

$$m_i = \frac{\sum_{j=1}^m m_{ij}}{\sum_{j=1}^m \sum_{i=1}^n m_{ij}}$$

где m_{ij} – вес i -го объекта, подсчитанный по оценкам всех экспертов.

$$m_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}$$

где x_{ij} – оценка фактора i , данная экспертом j ; n – число факторов; m – число экспертов. Далее выполняется метод простой ранжировки, согласно которому каждый эксперт располагает признаки, основываясь на личных предпочтениях, а затем вычисляется среднее значение важности признака.

Таблица 1 – Таблица ранжировки признаков объекта

	1	2	...	j	m
1	a_{11}	a_{12}	...	a_{1j}	a_{1m}
2	a_{21}	a_{22}	...	a_{2j}	a_{2m}
..
i	a_{i1}	a_{i2}	...	a_{ij}	a_{im}
n	a_{n1}	a_{n2}	...	a_{nj}	a_{nm}

Где a_{ij} – оценка признака экспертом, n – количество признаков, m – количество экспертов.

Данная средняя экспертная оценка используется в различных алгоритмах системы подбора персонала.

Далее собранные данные применяются в качестве своеобразного описания соискателя для последующей обработки и анализа. Целевой переменной является вероятность успешного прохождения собеседования соискателем.

II. ЦЕЛИ И ЗАДАЧИ КЛАССИФИКАЦИИ

Основной целью данной статьи является разработка системы классификации соискателей по группам. Каждая группа представляет собой то, принят ли был на работу соискатель или нет. Так же обязательным является определение вероятности успешного прохождения собеседования соискателем с последующим наймом его на работу. В данном алгоритме каждый образец (соискатель) представлен вектором, включающим в

себя экспертные оценки, идентификатор соискателя, а так же некоторые специфические характеристики данного соискателя, причем последние зависят от конкретной компании и по её желанию могут изменяться. Вышеописанные параметры вектора влияют на принятие решения, а именно к какому из классов относится данный образец. Таким образом классификатор на основе некоторой информации относит объект к одному из классов в связи с определённым распределением пространства.

III. МОДЕЛЬ ДАННЫХ

Для решения поставленной задачи требуются следующие данные: ФИО соискателя и место его работы. Эти данные можно найти в свободном доступе, или же их может предоставить сама компания, например, данные системы подбора персонала. Также требуется информация об экспертных оценках навыков соискателя, которые можно получить по запросу в соответствующие компании. Тут стоит отметить, что не все компании изъявляют желание делиться подобными данными, поэтому с наиболее заинтересованными из них будут заключены соответствующие дополнительные соглашения. Полученные данные будут использованы для обучения нейронной сети с целью классификации соискателей и отнесение их к одной из двух групп: потенциальных работников данной компании или тех, у кого пока недостаточно соответствующих навыков для прохождения собеседования. Таким образом, это поможет ускорить время обработки заявки на подбор персонала в команду, а так же проинформирует соискателя о его сильных сторонах, о навыках, которые необходимо «подтянуть», чтобы получить место в соответствующей компании.

IV. КЛАССИФИКАТОР

Для решения поставленной задачи классификации используются нейронная сеть следующей структуры: один слой Input, на вход которому подаются необработанные данные, и который отвечает за образование вспомогательных параметров для анализа, очистку от строк, содержащих NaN или пустые данные, масштабирование. Далее следует несколько полносвязных слоев. На данном этапе работ структура сети достаточно тривиальна и является по сути многослойным перцептроном. Однако с учетом полученных результатов стоит отметить, что вышеописанную структуру можно усовершенствовать, добавив сверточный слой, который в свою очередь позволит перейти от N-мерного пространства, где расположены параметры каждого вектора характеристик соискателей, к M-мерному, посредством задания соответствующих параметров свертки [3-4]. Активационная функция выходного слоя

нейронной сети представлена на следующем рисунке:

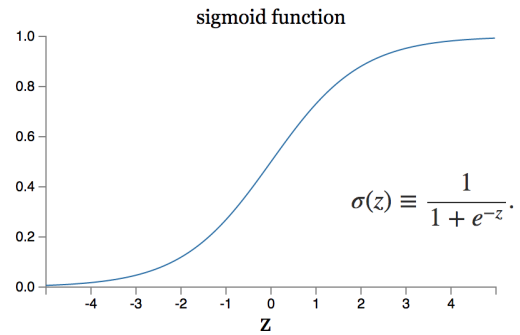


Рис. 1 – Активационная функция выходного слоя нейронной сети

Обновление весов модели происходит по следующей формуле:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t). \quad (1)$$

Где w_{ij} - параметр веса а Δw_{ij} считается по следующей формуле :

$$\Delta w_{ij}(t) = \eta(\delta e_p / \delta w_{ij} + \delta) + \epsilon \Delta w_{ij}(t-1). \quad (2)$$

В качестве примера были сгенерированы данные в размере 100.000 записей. В процессе работы с данными использовался язык программирования Python, пакет обработки данных Pandas, NumPy, пакеты sklearn, keras, matplotlib, mpl_toolkits, причем последние два использовались для построения наглядных изображений разбиения данных на классы [5-6].

1. Нестеренков, С. Н. Математическая модель процесса оценки и подбора персонала на основе многих параметров /С. Н. Нестеренков, В. Н. Видничук // Информационные технологии и системы 2017 (ИТС 2017) : материалы междунар. науч. конф., Минск, 25 окт. 2017 г. / Белорус. гос. ун-т информатики и радиозлектроники ; редкол.: Л. Ю. Шилин [и др.]. - Минск, 2017. - с. 200-201.
2. Нестеренков, С. Н. Генетический алгоритм как способ решения задач многомерной оптимизации / С. Н. Нестеренков, К. П. Белов // Информационные технологии и системы 2017 (ИТС 2017) : материалы междунар. науч. конф., Минск, 25 окт. 2017 г. / Белорус. гос. ун-т информатики и радиозлектроник ; редкол.: Л. Ю. Шилин [и др.]. - Минск, 2017. - с. 196-197.
3. Вандер, Д. П. Python для сложных задач. Наука о данных и машинное обучение /Д. П. Вандер // Питер, 2018. – 576 с.
4. Маккини, У. Python и анализ данных. /У. Маккини // Издательство: ДМК Пресс, 2015. – 482 с.
5. Силен, Д. Основы Data Science и Big Data. Python и наука о данных. /Д. Силен, А. Мейсман, А. Мохамед // Питер, 2017. – 336 с.
6. Machine Learning [Электронный ресурс] / Проект по публикации образовательных материалов в интернете в виде набора онлайн-курсов. – Режим доступа: <https://ru.coursera.org/learn/machine-learning>. – Дата доступа: 16.08.2018.