

ПРИМЕНЕНИЕ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ УСПЕШНОСТИ ЛЕЧЕНИЯ ИНФАРКТА МИОКАРДА

Колмакова Елизавета Сергеевна

Студент

ФГАОУ ВО "Казанский (Приволжский) федеральный университет"

Аннотация: для задачи прогнозирования степени успешности лечения инфаркта миокарда в условиях стационарного наблюдения были разработаны: программный модуль векторизации посредством применения технологии Word2Vec и рекуррентная трехслойная нейронная сеть, один из слоев которых есть LSTM слой (слой кратковременной и долговременной памяти) для более качественной обработки естественного языка. Сеть дала процент корректности от 81% на разных тестовых наборах.

Ключевые слова: рекуррентная нейронная сеть, прогнозирование, инфаркт миокарда, LSTM, слой кратковременной и долговременной памяти, векторизация текста, Word2Vec.

THE USAGE OF RECURRENT NEURAL NETWORKS TO PREDICT SUCCESS OF MIOCARDIAL INFARCTION TREATMENT

Kolmakova Elizaveta Sergeevna

Abstract: For the purpose of predicting the degree of success of myocardial infarction treatment in stationary conditions, a software vectorization module has been developed through the application of Word2vec technology and a recurrent three-layer neural network, one layer of which is a LSTM layer (long short term memory layer) for better processing of natural language. The network gave a percentage of correctness of 81% on different test sets.

Key words: recurrent neural network, predict, myocardial infarction, LSTM, long short term memory, text digitalizing, Word2Vec.

Инфаркт – серьезное заболевание сердечно-сосудистой системы, в тяжелых случаях требующее хирургического вмешательства. Наше сердце – уникальный орган,

поддерживающий постоянный ежесекундный поток крови, жизненно необходимого для каждого органа и клетки тела. Жизнь не возможна без постоянного движения крови. По причине колоссальной важности сердца задача превентивного обнаружения инфаркта является более чем важной. Эта задача может быть решена несколькими способами: врачебный опыт и знания или компьютерное прогнозирование. Ни один из них не является абсолютom в вопросе обнаружения предрасположенности к инфаркту, но вкупе это максимально эффективный способ.

Существует множество методов компьютерного прогнозирования, один из которых - использование нейронных сетей для анализа данных пациентов, наблюдавшихся ранее, чтобы по этим данным научиться определять вероятность на успешное лечение для текущих пациентов. Для решения данной задачи потребуется затронуть множество других важных подзадач.

Во-первых, необходимо максимально структурировать входные данные, представленные в виде полной истории болезни пациента. Этот процесс заключается в анализе истории болезни и извлечении важной информации, а так же корректировки самих записей. В среднем, по нормам министерства здравоохранения Российской Федерации врач-кардиолог должен тратить на пациента 24 минут [1, с.3]. За это время необходимо не только расспросить пациента о его жалобах, но и занести эту информацию в систему. Ввиду того, что врач ограничен во времени, он заносит информацию о пациенте настолько быстро, насколько он может. Поэтому в таких записях много сокращений и ошибок. Ввиду этого одной из подзадач является исправление типичных и часто встречающихся ошибок.

Во-вторых, необходимо максимально эффективно перевести текст в понятный для нейронной сети вид – в числовой вектор. Объем информации о пациентах огромен: в рамках работы больше восьми тысяч историй болезни необходимо обработать, каждая из которых содержит от ста слов до нескольких тысяч. По сей причине недопустимо использование неэффективных алгоритмов, поскольку они потребуют колоссального запаса времени и памяти, что является ценным ресурсом.

Приступая к решению задачи, необходимо исследовать входные данные в истории болезни. Проанализировав их, можно выделить тот объем информации, который

определяет ответ и который необходимо использовать в качестве входных и выходных данных для нейронной сети:

1. Результат лечения. Эта информация необходима в качестве источника сверки выходных данных нейронной сети. Созданная нейронная сеть обучается с учителем, что подразумевает наличие ответа, с которым сеть должна сравнивать свои вычисленные выходные данные. Этим ответом и будет являться поле с названием «Результат лечения» из типичной истории больного.
2. Первичный осмотр. Отображает анамнез пациента, то есть все его симптомы, перенесенные заболевания, хронические заболевания, наследственность, вредные привычки и т.д.
3. Ежедневные осмотры. Отображают информацию о пациенте в стационаре: как меняется его состояния в ходе лечения.
4. Медицинские манипуляции. Результаты раздела, где описаны все процедуры, которые были назначены пациенту, такие как ЭКГ или электрокардиограмма, коронарография и другие, так же является ценным источником информации.
5. Операции. Отражает исходы всех хирургических вмешательств в организм пациента.

Следующим этапом будет являться создание векторной модели корпуса слов. Для данного этапа была применена технологии Word2Vec. Технология Word2Vec – это общее название для совокупности моделей на основе искусственных нейронных сетей, созданная для представления слов на естественном языке в векторном виде. Является инструментом компании Google, созданном в 2013 году, который завоевал популярность из-за удобства использования и скорости работы. Эта технологий является одной из самой эффективной. Модель Word2Vec создана таким образом, чтобы располагать похожие слова максимально близко друг к другу. Это вычисляется на основе контекстной близости: слова, встречающиеся чаще остальных рядом с каким-нибудь зафиксированным словом, будет иметь близкие векторы в пространстве [2]. Пример такой реализации изображен на Рис. 1.

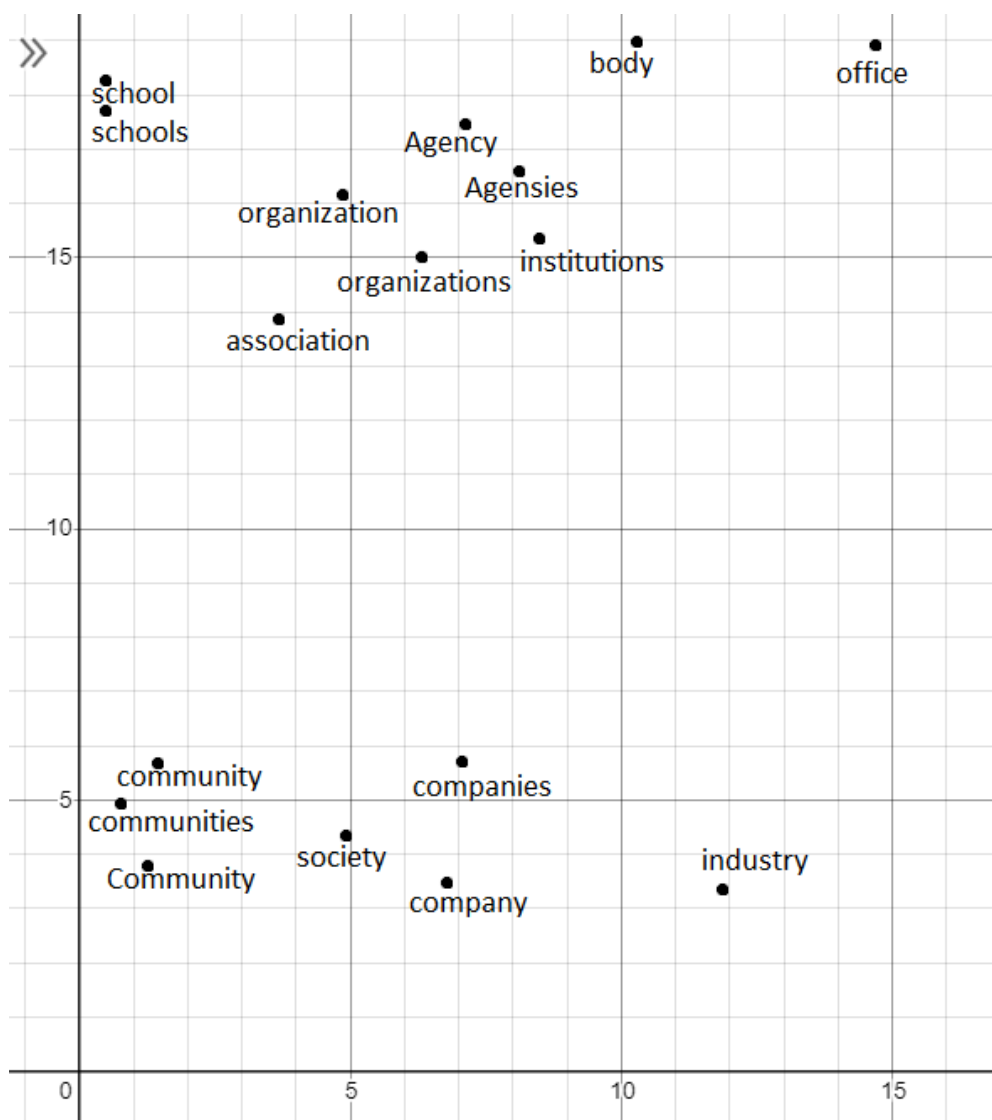


Рис.1. Изображение векторного представления похожих слов в модели Word2Vec.

Следующим шагом является этап выбора критерия разделения данных. Критерий деления – важный момент, на основании этого критерия нейронная сеть будет определять результаты. Вначале было выбрано два критерия деления – «жив/не жив пациент» и «окончено/не окончено лечение». Но такое разбиение дало крайне низкий процент корректного прогнозирования нейросети – среди умерших и выписанных без улучшения нейронная сеть давала корректность на уровне 20%, что является низким показателем. Такой результат был получен по той причине, что в исходных данных около 96% информации о пациентах относилось к одной категории, а оставшиеся 4% делились на другие категории. Поэтому нейронной сети было трудно определить закономерности, на основании которых можно было точно определить пациента из тех 4% пациентов. Ввиду данной причины критерий был сменен на «успешно/неуспешно лечение». Причем успешность лечения определялась двумя факторами – у пациента

должна была наблюдаться положительная динамика, то есть должно было наблюдаться улучшение или выздоровление, и стандарт лечения должен был быть выполненным.

В задаче обработки естественного языка конфигурация нейронной сети важна. Наибольшее применение в этой области нашли рекуррентные нейронные сети, главным отличием которых является наличие обратной связи. Реализаций у рекуррентных сетей немало, но именно LSTM (Long Short Term Memory – сети кратковременной и долговременной памяти) в настоящее время стала наиболее популярной в сравнении с остальными типами, продемонстрировав большой успех в решении многих задачах NLP (neuro-linguistic programming – нейролингвистическое программирование). LSTM сеть легко может разбирать смысл предложений, что является ключевой подзадачей при анализе медицинских документов с целью прогнозирования наступления инфаркта миокарда в определенных условиях [3, с.6-9].

Приступая непосредственно к реализации нейронной сети, необходимо загрузить в модель входные данные. Две группы пациентов, поделенные критерием «успешное/неуспешное лечение» на основании отмеченных результатов лечения в истории болезни, надо разделить на два набора – тестовый и тренировочный. На тренировочном наборе сеть будет учиться, методом проб и ошибок пытаться находить нужные веса, которые будут приводить к нужным результатам. На тестовом наборе нейронная сеть будет непосредственно проверять свои вычисления. Разбиение оптимально проводить в отношении 70 на 30%. Это наиболее эффективно, так как у сети будет возможность лучше поучить на большем количестве данных.

Следующим этапом идет конфигурирование нейронной сети. Для начала составляется архитектура сети, потом она компилируется. И только затем она обучается и проверяется корректность на тестовом наборе.

Нейронная сеть в рамках этой работы состоит из четырех слоев: слоя векторного представления, слоя типа Dropout, LSTM слоя и финального выходного слоя. Последний слой состоит всего из одного нейрона, главной задачей которого является формирование одного числа – ответа. Функцией активации этого слоя является сигмоида, которая переводит все поступающие сигналы в одно число от 0 до 1. Чем ближе к единице, тем вероятность успешного лечения выше. Третий слой – LSTM слой – непосредственно является основной частью нейронной сети. Этот слой и определяет все выходные

значения, которые потом станут одним финальным числом. Слой векторного представления - Embedding слой - создает многомерное пространство заданного размера для того, чтобы сопоставить каждому слову его точку в пространстве. И, наконец, второй слой - слой типа dropout – важный этап, созданный для придания нейронной сети большей устойчивости. Слой dropout обнуляет малую часть весов, чтобы уменьшить переоснащение информацией. Это позволяет делать узлы в сети более устойчивыми к входам, особенно если объем входной информации велик [4, с 11-13].

Следующим шагом является само обучение нейронной сети. На вход подается тестовый набор, тренировочный набор, количество эпох (это есть количество заходов обучения) и показатель batch_size, указывающий на размер «пакета», которыми итерирует нейронная сеть при обучении. После этого проверяется работа на тестовых наборах, на основе которых можно сделать вывод о корректности обученной модели.

Подав данные, разбитые по критерию «успешное/неуспешное лечение» нейронная сеть показала довольно хорошие результаты. Для проверки качества обучения нейронной сети было создано три тестовый набора:

- Набор №1 - это 13% успешных историй на 87% неуспешных
- Набор №2 - это 100% успешных историй на 0% неуспешных
- Набор №3 - это 45% успешных историй на 55% неуспешных

Результаты по наборам получились следующие (Рис.2): на первом наборе процент корректности составил 89,72%, на втором наборе это 81,17%, на третьем - 87,78%. Данное разбиение показало себя более эффективным, обучив модель с точностью в 80-90% определять успешность лечения инфаркта миокарда.

128/1722 [=>.....]	- ETA: 2:10		
192/1722 [==>.....]	- ETA: 2:06		
256/1722 [===>.....]	- ETA: 2:01		
320/1722 [====>.....]	- ETA: 1:56		
384/1722 [=====>.....]	- ETA: 1:51		
448/1722 [=====>.....]	- ETA: 1:45		
512/1722 [=====>.....]	- ETA: 1:40		
576/1722 [=====>.....]	- ETA: 1:34		
640/1722 [=====>.....]	- ETA: 1:29		
704/1722 [=====>.....]	- ETA: 1:24		
768/1722 [=====>.....]	- ETA: 1:18		
832/1722 [=====>.....]	- ETA: 1:13		
896/1722 [=====>.....]	- ETA: 1:08		
960/1722 [=====>.....]	- ETA: 1:03		
1024/1722 [=====>.....]	- ETA: 57s		
1088/1722 [=====>.....]	- ETA: 52s		
1152/1722 [=====>.....]	- ETA: 47s		
1216/1722 [=====>.....]	- ETA: 41s		
1280/1722 [=====>.....]	- ETA: 36s		
1344/1722 [=====>.....]	- ETA: 31s		
1408/1722 [=====>.....]	- ETA: 25s		
1472/1722 [=====>.....]	- ETA: 20s		
1536/1722 [=====>.....]	- ETA: 15s		
1600/1722 [=====>.....]	- ETA: 10s		
1664/1722 [=====>.....]	- ETA: 4s		
1722/1722 [=====>.....]	- 142s 83ms/step		
Точность на тестовых данных: <u>89.72%</u>			
64/223 [=====>.....]	- ETA: 7s		
128/223 [=====>.....]	- ETA: 4s		
192/223 [=====>.....]	- ETA: 1s		
223/223 [=====>.....]	- 10s 46ms/step		
Точность на тестовых данных: <u>81.17%</u>			
64/499 [=>.....]	- ETA: 36s		
128/499 [=====>.....]	- ETA: 30s		
192/499 [=====>.....]	- ETA: 25s		
256/499 [=====>.....]	- ETA: 20s		
320/499 [=====>.....]	- ETA: 15s		
384/499 [=====>.....]	- ETA: 9s		
448/499 [=====>.....]	- ETA: 4s		
499/499 [=====>.....]	- 42s 85ms/step		
Точность на тестовых данных: <u>87.78%</u>			



Тестовый набор #1 из 1722 случаев
223/1499
лечение успешно/не успешно



Тестовый набор #2 из 223 случаев
223/0
лечение успешно/не успешно



Тестовый набор #3 из 499 случаев
223/256
лечение успешно/не успешно

Рис.2. Результат работы нейронной сети

A heart attack or myocardial infarction is a serious cardiovascular disease that requires surgery in severe cases. Our heart is a unique organ that supports the constant flow of blood every second for every organ and cell of the body. Life is impossible without constant blood flow. Because of the great importance of the heart, the task of preventive detection of heart infarction is more than important. This problem can be solved in several ways: medical experience and knowledge or computer prediction. None of them is absolute about finding a predisposition to a heart attack, but together they are the most effective way.

There are many methods of computer prediction, one of which is the usage of neural networks to analyze the data of patients observed in the past, in order to learn from these data the probability of successful treatment for current patients. Many other important sub-tasks would need to be addressed in order to achieve this.

First, the input data presented in the patient's medical history should be as structured as possible. This process consists of analyzing the medical history and extracting important information, as well as adjusting the records themselves. On average, the Ministry of Health of the Russian Federation requires a cardiologist to spend 24 minutes per patient [1, p.3]. During this time, it is necessary not only to ask the patient about his complaints, but also to enter this information into the system. Because of the limited time available, the doctor records the

patient as quickly as he can. So there are a lot of acronyms and mistakes in these records. Therefore, one of the sub-tasks is the correction of typical and frequent errors.

Second, it is necessary to translate the text into a form that is intelligible to the neural network - into a numerical vector - as efficiently as possible. The amount of information about patients is enormous: more than 8,000 case files must be processed, each of which contains between 100 words and several thousand. For this reason, inefficient algorithms are unacceptable, as they will require a huge amount of time and memory, which is a valuable resource.

When starting to solve the problem, it is necessary to examine the input data in the medical history. After analysing them, one can distinguish the amount of information that determines the answer, which should be used as input and output data for the neural network:

1. Result of treatment. This information is needed as a source of reconciliation of neural network outputs. The created neural network is taught by a teacher, which implies that there is an answer to which the network must compare its computed output. This answer will be the field with the name «Result of treatment» from the typical history of the patient.
2. Primary examination. It displays the patient's history, i.e., all his symptoms, diseases, chronic diseases, heredity, bad habits, etc.
3. Daily check-ups. Information about the patient in hospital is displayed: how his condition changes during treatment.
4. Medical manipulation. The results of the section, which describes all procedures that have been prescribed to the patient, such as ECG or electrocardiogram, coronary and other, are also a valuable source of information.
5. Operations. It reflects the outcome of all surgical interventions in the patient's system.

The next step is to create a vector model of the word corpus. Word2vec technology was used for this phase. Word2vec technology is a generic name for a collection of models based on artificial neural networks, created to represent words in a natural language in vector form. It is a tool of Google, established in 2013, which has gained popularity due to its usability and speed. This technology is one of the most effective. Word2vec is designed to have similar words as

close as possible to each other. This is computed on the basis of contextual proximity: words that occur more often than others next to any fixed word will have close vectors in space [2]. An example of such an implementation is shown in Fig. 1.

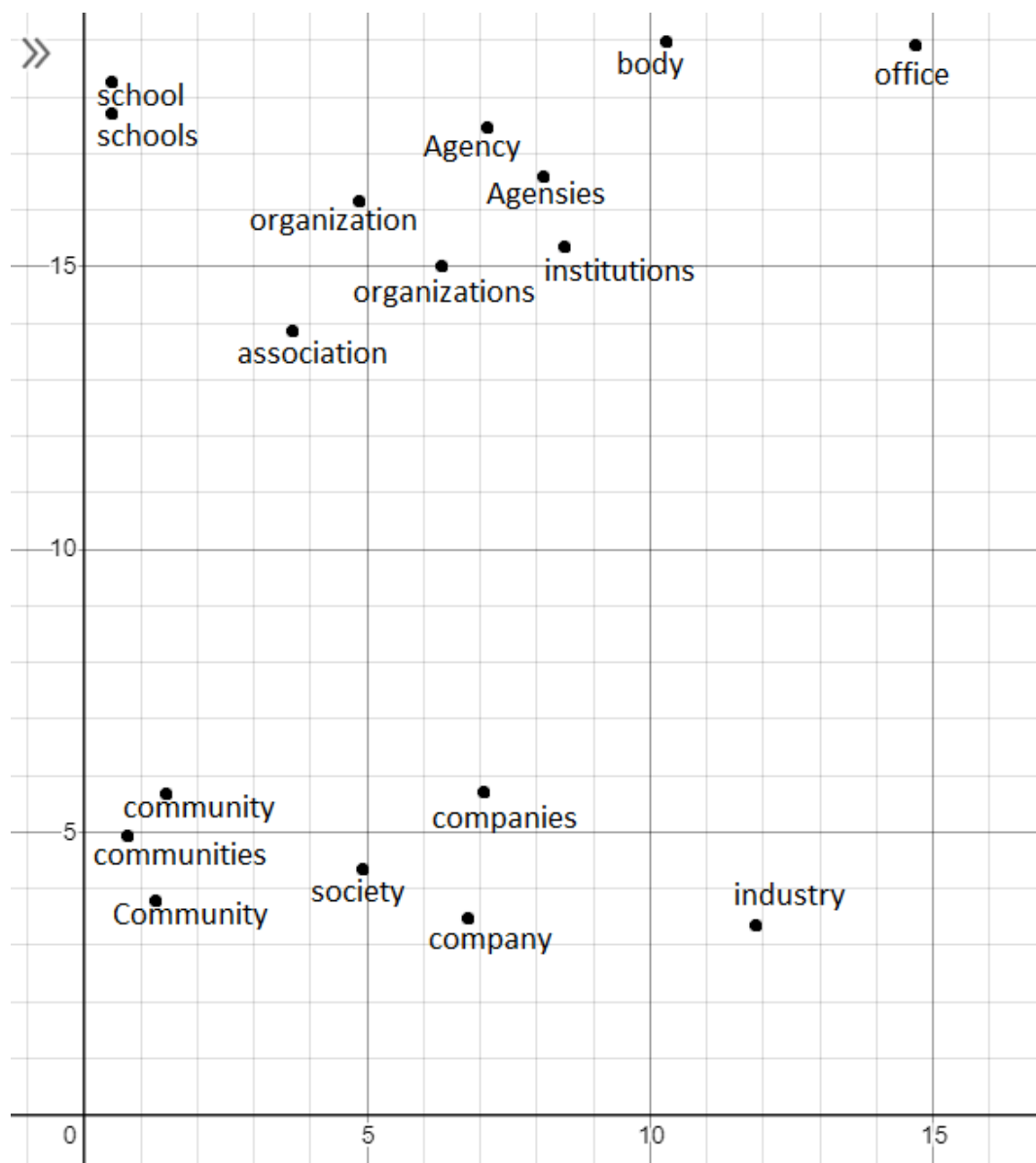


Figure 1. Image of a vector representation of similar words in the Word2vec model.

The next step is the selection of the data separation criterion. The fission criterion is an important point, so the neural network will determine the results. At first two criteria of division were chosen - «alive/not alive patient» and «finished/not finished treatment». But this partition produced an extremely low percentage of correct prediction of the neural network - among the dead and discharged without improvement neural network gave correctness of 20%, which is a low rate. This result was achieved because, in the original data, about 96 per cent of patient information was in one category and the remaining 4 per cent was divided into other categories. So it was difficult for the neural network to determine the patterns from which it was possible

to pinpoint the patient from the four percent of patients. Due to this reason the criterion was changed to «successful/unsuccessful treatment». The success of the treatment was determined by two factors - the patient had to have a positive dynamic, that is, there had to be improvement or recovery, and the standard of treatment had to be met.

In natural language processing, neural network configuration is important. Most of the applications in this area are in recurrent neural networks, the main difference of which is having the availability of feedback. There are a number of implementations of recurrence networks, but it is LSTM (Long Short Term Memory) that has now become the most popular network compared to other types, demonstrating great success in solving many NLP problems (neuro-linguistic programming). The LSTM network can easily disassemble sentences, which is a key sub-task in the analysis of medical documents to predict myocardial infarction under certain conditions [3, p.6-9].

When starting to implement a neural network directly, input data must be loaded into the model. Two groups of patients divided by the criterion «successful/unsuccessful treatment» on the basis of the observed results of treatment in the history of the disease should be divided into two sets - test and training. In a training set, the network will learn by trial and error to try to find the right weights that will lead to the desired results. On the test set, the neural network will directly check its calculations. Optimally split 70 by 30 per cent. This is most effective because the network will have a better opportunity to learn from more data.

The next step is to configure the neural network. First, the architecture of the network is constructed, then it is compiled. It is only then that it is trained and validated on the test set.

The neural network within this work consists of four layers: the vector representation layer, the Dropout layer, the LSTM layer, and the final output layer. The last layer consists of just one neuron whose main task is to form one number - the answer. The function of activating this layer is a sigmoid, which translates all incoming signals into a single number from 0 to 1. The closer to one, the more likely it is to be successful. The third layer, the LSTM layer, is directly a fundamental part of the neural network. This layer determines all output values, which then become one final number. The Embedding layer of a vector representation creates a multi-dimensional space of a given size to compare each word its point in space. And finally, the second layer, the dropout layer, is an important stage created to make the neural network more stable. The dropout layer resets a small portion of the weights to reduce the retrofitting of

information. This makes network nodes more resistant to inputs, especially if the amount of input is large [4, p.11-13].

The next step is learning the neural network itself. A test set, a training set, the number of epochs (this is the number of study visits) and a batch_size indicator indicating the size of the «package» by which the neural network iterates during training are submitted to the input. The work on the test sets from which the correctness of the trained model can be inferred is then checked.

Having collected data broken down according to the criterion «successful/unsuccessful treatment» neural network has shown quite good results. Three test sets have been developed to test the quality of neural network learning:

- o Set 1 is 13% successful stories 87% unsuccessful
- o Set 2 is 100% successful stories 0% unsuccessful
- o Set 3 is 45% successful stories by 55% unsuccessful

The results for the sets were as follows (Figure 2): the first set of correctness was 89.72%, the second set 81.17%, and the third set 87.78%. This splitting proved to be more effective by training the model to determine with 80-90% accuracy the success of myocardial infarction treatment.

```
128/1722 [=>.....] - ETA: 2:10
192/1722 [==>.....] - ETA: 2:06
256/1722 [===>.....] - ETA: 2:01
320/1722 [====>.....] - ETA: 1:56
384/1722 [=====>.....] - ETA: 1:51
448/1722 [=====>.....] - ETA: 1:45
512/1722 [=====>.....] - ETA: 1:40
576/1722 [=====>.....] - ETA: 1:34
640/1722 [=====>.....] - ETA: 1:29
704/1722 [=====>.....] - ETA: 1:24
768/1722 [=====>.....] - ETA: 1:18
832/1722 [=====>.....] - ETA: 1:13
896/1722 [=====>.....] - ETA: 1:08
960/1722 [=====>.....] - ETA: 1:03
1024/1722 [=====>.....] - ETA: 57s
1088/1722 [=====>.....] - ETA: 52s
1152/1722 [=====>.....] - ETA: 47s
1216/1722 [=====>.....] - ETA: 41s
1280/1722 [=====>.....] - ETA: 36s
1344/1722 [=====>.....] - ETA: 31s
1408/1722 [=====>.....] - ETA: 25s
1472/1722 [=====>.....] - ETA: 20s
1536/1722 [=====>.....] - ETA: 15s
1600/1722 [=====>.....] - ETA: 10s
1664/1722 [=====>.....] - ETA: 4s
1722/1722 [=====>.....] - 142s 83ms/step
Точность на тестовых данных: 89.72%

64/223 [=====>.....] - ETA: 7s
128/223 [=====>.....] - ETA: 4s
192/223 [=====>.....] - ETA: 1s
223/223 [=====>.....] - 10s 46ms/step
Точность на тестовых данных: 81.17%

64/499 [=>.....] - ETA: 36s
128/499 [==>.....] - ETA: 30s
192/499 [===>.....] - ETA: 25s
256/499 [====>.....] - ETA: 20s
320/499 [=====>.....] - ETA: 15s
384/499 [=====>.....] - ETA: 9s
448/499 [=====>.....] - ETA: 4s
499/499 [=====>.....] - 42s 85ms/step
Точность на тестовых данных: 87.78%
```

Figure 2. Neural network output

Список литературы.

1. Приказ Минздрава РФ От 19.12.2016 N 973н
"Об Утверждении Типовых Отраслевых Норм Времени На Выполнение Работ,
Связанных С Посещением Одним Пациентом Врача - Кардиолога, Врача -
Эндокринолога, Врача - Стоматолога - Терапевта" (Зарегистрировано В Минюсте
РФ 13.01.2017 N 45216)
2. Jay Alammar, The illustrated Word2Vec [Электронный источник]/ URL:
<https://jalammar.github.io/illustrated-word2vec/>
3. Sepp Hochreiter, Jurgen Schmidhuber, "Long Short Term Memory"/ Neural
Computation 9(8): 1735-1780/ 1997 – URL:
<https://www.bioinf.jku.at/publications/older/2604.pdf>
4. Yarin Gal, Zoubin Ghahramani /_A Theoretically Grounded Application of Dropout
in Recurrent Neural Networks/ [Электронный ресурс] URL:
<https://arxiv.org/pdf/1512.05287.pdf>
5. Filippo Amato, Alberto López, Eladia María Peña-Méndez, Petr Vaňhara, Aleš
Hampl, Josef Havel /_Artificial neural networks in medical diagnosis / Journal of
APPLIED BIOMEDICINE/ ISSN 1214-0287/ стр.48-56