|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Министерство науки и высшего образования Российской Федерации | | | | | | | | | | |
|  | | | |  | | |  | | | |
|  | | Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования  Пермский государственный национальный исследовательский университет  (ПГНИУ) | | | | | | |  | |
|  | | | |  | | |  | | | |
|  | | | отчет  Определение вкусов в кино при помощи нейросетевой модели | | | | |  | | |
|  | | | |  | | |  | | | |
|  | Работу выполнил  студент гр. ПМИ-3,4  Волобоев С.В.  \_\_\_\_\_\_\_\_\_\_  (подпись)  «\_\_\_» \_\_\_\_\_\_\_\_ 2019 | | | |  | Проверил  Профессор кафедры ПМИ  Ясницкий Л.Н.  \_\_\_\_\_\_\_\_\_  (подпись)  «\_\_\_» \_\_\_\_\_\_\_\_ 2019 | | | |  |
|  | | | |  | | |  | | | |
|  | | | | Пермь 2019 | | |  | | | |

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ В ПРЕДМЕТНУЮ ОБЛАСТЬ 3](#_Toc11890767)

[1 Описание модели 4](#_Toc11890768)

[2 Структура нейронной сети 5](#_Toc11890769)

[3 Проверка на тестовом множестве 14](#_Toc11890770)

[4 Определение значимости параметров 15](#_Toc11890771)

[Выводы 16](#_Toc11890772)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 17](#_Toc11890773)

# ВВЕДЕНИЕ В ПРЕДМЕТНУЮ ОБЛАСТЬ

В эпоху стремительного развития интернета и медиа у каждого человека есть свободный доступ почти ко всем кинокартинам, но выбор сугубо индивидуален. У всех есть предпочтения по жанрам кино, и хотелось бы узнать, что любит конкретный человек, чтобы, например, иметь хорошую тему для разговора.

Конечно, можно всегда создать опрос вида «Какие фильмы из предложенных вам нравятся?», но в процессе устного диалога такого точно не сделаешь, да и заполнять подобные анкеты крайне утомительно.

Однако почти все любят отвечать на нестандартные вопросы, поданные в шуточной манере. Но можно ли из них извлечь какую-то закономерность в увлечениях человека?

Ответ на этот вопрос будет интересно узнать людям, которым не всегда удаётся найти точки соприкосновения в диалоге, потому что случае создания рабочей нейросетевой модели откроются новые возможности для социальных взаимодействий.

# Описание модели

Нейронная сеть имеет 22 входных параметра, которые описаны в таблице 1.1 ниже.

|  |  |
| --- | --- |
| **Вход** | **Описание** |
| X1 | Пол: 0 – женский, 1 – мужской. |
| X2 | Как часто били в детстве: не били – 0, разве что по делу – 0.3, часто – 1 |
| X3 | Что такое двач: не знаю : 0, паблик в вк : 0.2, лучшая борда рунета : 0.7, ненавижу Абу : 1 |
| X4 | Отклонение от стандартной ориентации: гетеро – 0, би – 0,5, альтернативная - 1 |
| X5 | Предпочтение активного отдыха: 0 – нет, 1 – да |
| X6 | «Вы таки антисемит?»(шуточный вопрос): 0 – нет, 0,5 – я просто боюсь евреев, 1 – да |
| X7 | Как вы относитесь к аннексии Крыма: 0 – мне всё равно, 0.5 – отрицательно, 1 - положительно |
| X8 | Знание английского: оригатоё гайзаймас : 0, чуть лучше, чем ничего : 0.2, нормально : 0.4, хорошо : 0.6, отлично : 0.8, НЭЙТИВ ИНГЛИШ СПИКЕР : 1 |
| X9 | Было ли раньше лучше: 0 – нет, 1 – да. |
| X10 | Любимый предмет – литература: 0 – нет, 1 – да |
| X11 | Любимый предмет – история: 0 – нет, 1 – да |
| X12 | Любимый предмет – иностранные языки:  0 – нет, 1 – да |
| X13 | Любимый предмет – физика: 0 – нет, 1 – да |
| X14 | Любимый предмет – биология: 0 – нет, 1 – да |
| X15 | Любимый предмет – химия: 0 – нет, 1 – да |
| X16 | Любимый предмет – математика: 0 – нет, 1 – да |
| X17 | Любимый предмет – информатика: 0 – нет, 1 – да |
| X18 | Насколько вы вегетарианец: от 0 до 1 |
| X19 | Насколько любите игры: от 0 до 1 |
| X20 | Насколько любите сериалы: от 0 до 1 |
| X21 | Насколько любите книги: от 0 до 1 |
| X22 | Оригинальность ответа в вопросе про отношение к президенту(были даны 2 варианта и поле для своего ответа): 0 – ответ стандартный, 1 – оригинальный |

Таблица 1 – Описание входных параметров

Выходных параметров 10 штук, отвечающих за предрасположенность к жанрам фильмов: боевик, детектив, аниме, драма, комедия, мелодрама, исторический фильм, трагедия, фильм ужасов, фантастика. Все они могут принять 2 значения: -1 – жанр неинтересен, 1 – жанр интересен (выбор значений выходов обусловлен биполярной функцией активации)

# Структура нейронной сети

Для получения изначального множества был использован сервис «Google Формы» [1], в котором был создан социологический опрос, содержащий вопросы из таблицы 1.

Всего было опрошено 190 респондентов, предпочтительно студентов с различных факультетов. Результаты приведены в диаграммах 1, 2, 3, 4, 5, 6, 7, 8,9, 10, 11, 12, 13, 14, 15

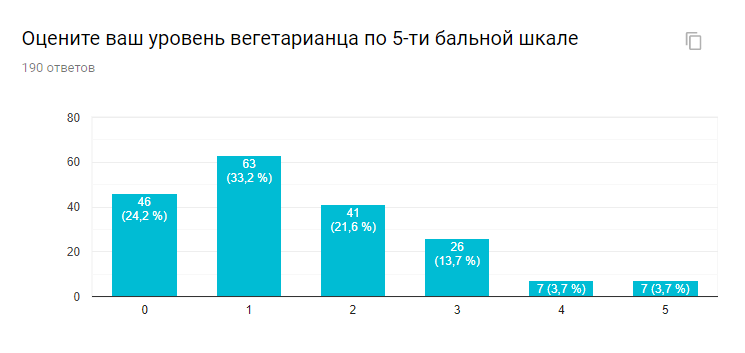
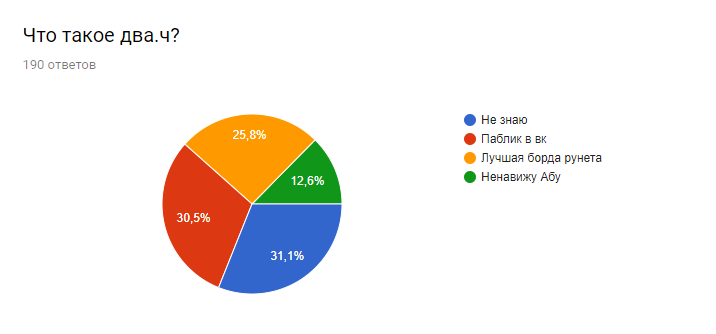
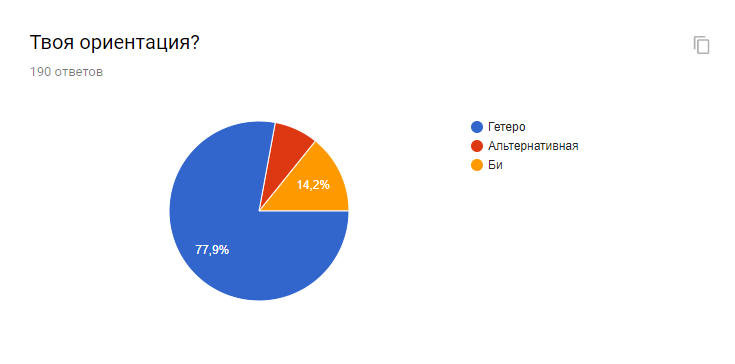


Диаграмма 1

Диаграмма 2



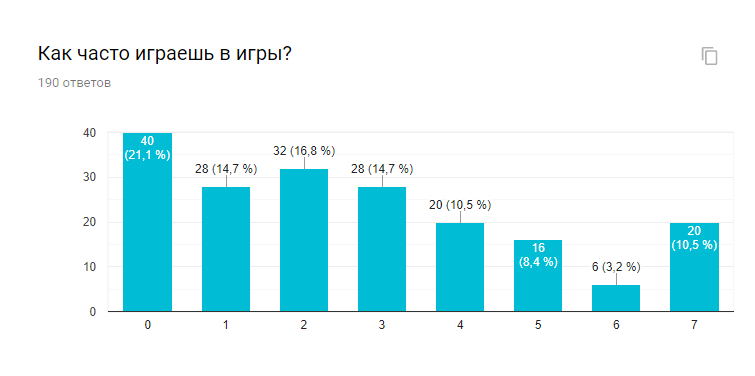
Диаграмма



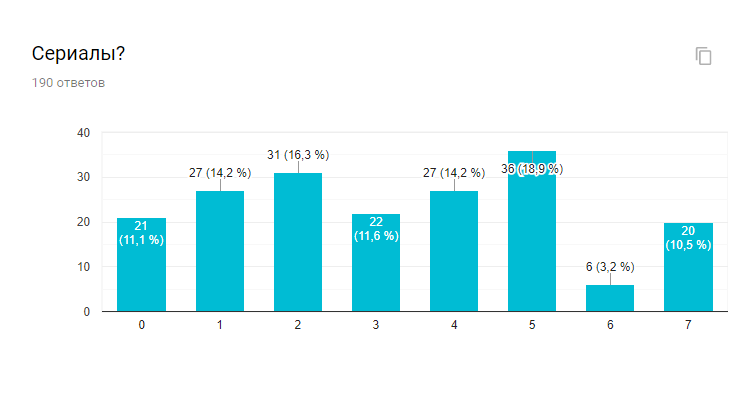
Диаграмма



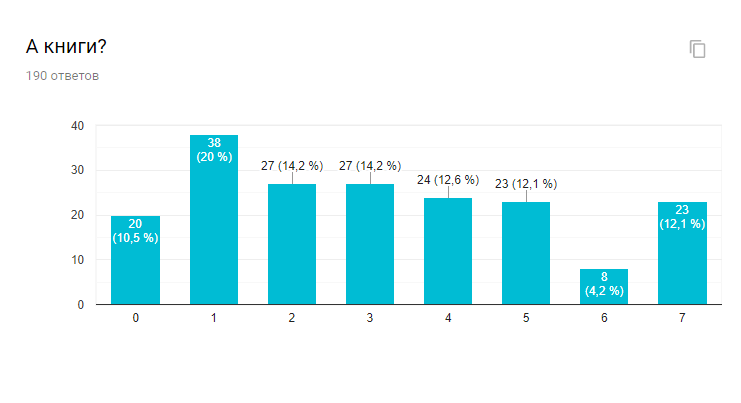
Диаграмма



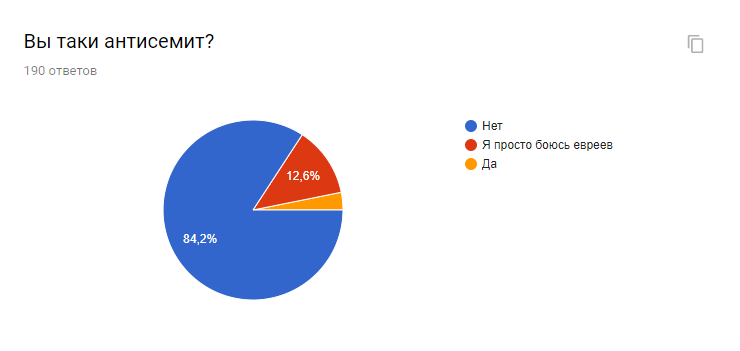
Диаграмма



Диаграмма



Диаграмма



Диаграмма



Диаграмма



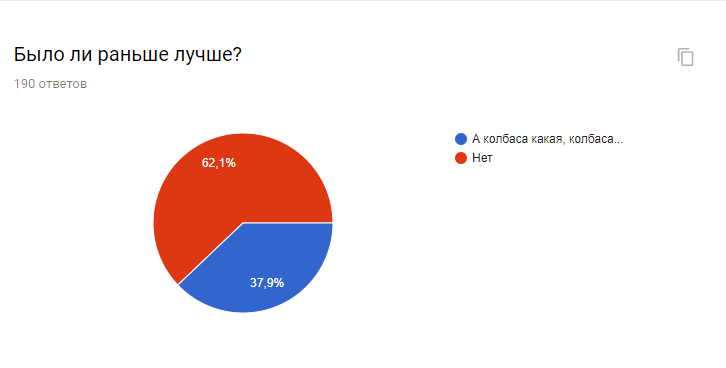
Диаграмма



Диаграмма



Диаграмма



Диаграмма

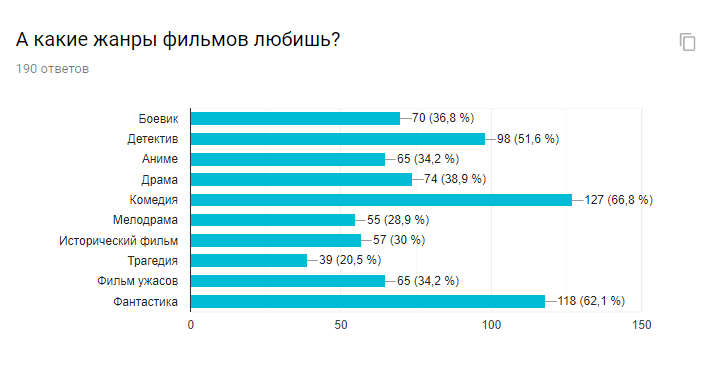


Диаграмма 15

Результаты опроса обработаны и приведены к удобному для использования виду с помощью Python.

Для создания самой нейронной сети было использовано программное обеспечение «Нейросимулятор 5.0». [2]

Сеть состоит из 22 входных нейронов, скрытого слоя и 10 выходных нейронов. Для определения числа нейронов в скрытом слое персептрона была

использована формула, являющаяся следствием теорем Арнольда – Колмогорова – Хехт-Нильсена, которую можно найти, например, в уч. пособии [3] на стр. 106. Полученная оценка лежит в пределах от 5 до 75 нейронов.

Результат анализа квадратичной ошибки обобщения приведён на рисунке 1.



Рисунок 1

Дальнейший анализ показал, что чаще всего наименьшая ошибка выходит при 17 нейронах.

Таким образом, оптимальной оказалась конструкция нейросети, в которой на скрытом слое находится 17 нейронов, функция активации скрытого слоя – сигмоида. Структура показана на рисунке 2.

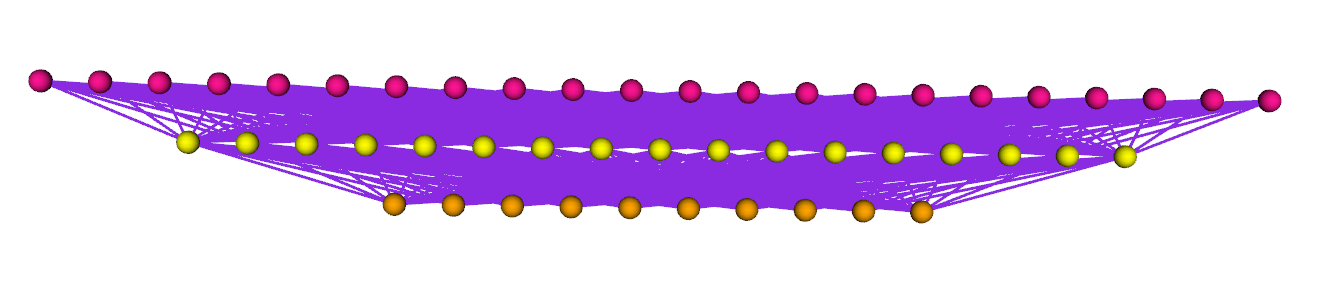


Рисунок 2

# Проверка на тестовом множестве

Для тестового множества было отобрано 20% случайных ответов, реализация разделения на тестовое и обучающее множество также было решено сделать в Python.

График обучения представлен на рисунке 3

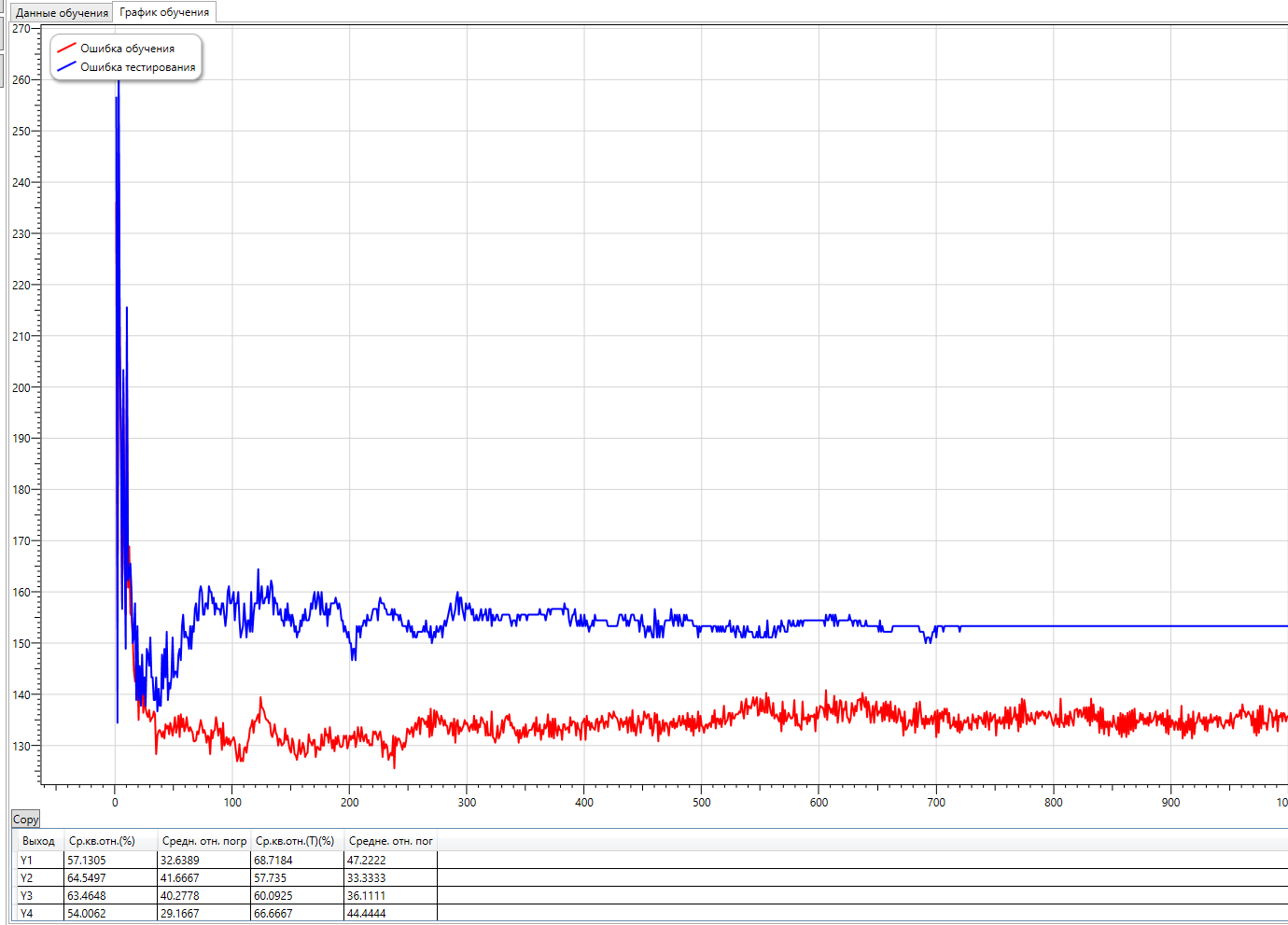


Рисунок 3

Видно, что нейронная сеть усвоила какие-то закономерности, но насчёт её объективности судить сложно, так что составим гистограмму точности принятия/отрицания каждого жанра фильма на тестовом множестве. Она показана на рисунке 4.

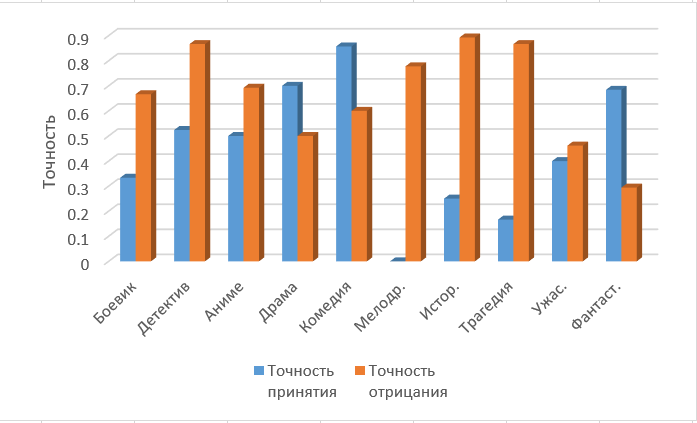
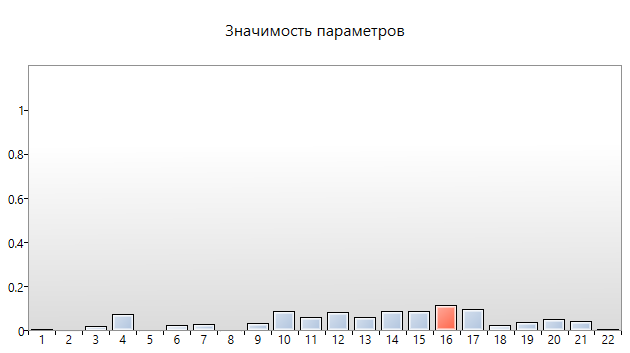


Рисунок 4

# Определение значимости параметров

Существует несколько способов определения важности входных параметров. [4]. Воспользуемся встроенным функционалом в инструментарии ПО «Нейросимулятор 5» [2]. Результаты работы программы изображены на рисунке 5.



Как можно заметить, больше всего влияют вопросы про любимый предмет(10-17) и ориентацию(4). Вопросы про насилие в детстве, активный отдых, знание английского и оригинальное политическое мнение вообще не влияют на результат.

# Выводы

Нейросеть определяет общие предпочтения в кино крайне грубо, и всё же она показывает интересные результаты. Она научилась весьма точно говорить о том, что человек не любит популярные жанры, такие как боевик, детектив и комедия. Также стоит обратить на любовь к драме, ведь в исходной выборке её выбрало 38,9% людей, а точность принятия 70%, что говорит о полученных знаниях. На высокую точность принятия у комедии и фантастики можно не обращать особого внимания, так как они являются популярным выбором. Аналогично с точностью отрицания у непопулярных жанров.

Как можно убедиться, шуточные вопросы(«Вы таки антисемит?», оригинальность в политике) или вопросы, на которые человеку сложно ответить честно (насилие в детстве, знание английского) не привносят ничего в результат.

Увеличение выборки, причём горизонтальное(среди разных слоёв населения, а не только студентов), точно улучшит результаты, но насколько хорошо – сложно сказать. Отредактировать состав вопросов в анкете также непросто, так как при расширении выборки могут начать влиять новые аргументы. Но её определённо стоит расширить, так как текущие параметры достаточно малозначимы.

Также можно предположить и психологическую тенденцию к потере внимания, т.к. точность оценки резко начинает падать после первых пяти жанров.

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 <https://forms.gle/upUa7WqfpJEjJ6Sb7>.

2 Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 5.0 // Свидетельство о государственной регистрации программы для ЭВМ № 2014618208. Заявка Роспатент № 2014614649. Зарегистрировано в Реестре программ для ЭВМ 12 августа 2014г.

3 Ясницкий Л.Н. Интеллектуальные информационные технологии и системы: учеб.-методпособие / Л. Н. Ясницкий; Перм. Ун-т. – Пермь, 2007. – 271 с.

4 Ясницкий Л.Н. Нейронные сети – инструмент для получения новых знаний: успехи, проблемы, перспективы // Нейрокомпьютеры: разработка, применение. – 2015. – № 5. – С. 48-56.

5 Ясницкий Л.Н. Интеллектуальные информационные технологии и системы: учеб.-метод пособие / Л. Н. Ясницкий; Перм. Ун-т. – Пермь, 2007. – 271 с