

Методи прогнозування рівня російського контенту на українському телебаченні

Виконала: Онанко Анастасія Максимівна

Науковий керівник:

д.т.н. проф. Зайченко Олена Юріївна

Постановка задачі

- Провести дослідження задач, пов'язаний з визначенням рівня російського контенту на українському телебаченні, виділення факторів, які впливають на його кількість
- Проаналізувати вхідні дані, привести їх попередню обробку, а також продумати архітектуру моделі для моніторингу російського телевізійного продукту
- Провести прогнозування рівня російського телепродукту на українському телебаченні за допомогою МГУА та нейронної мережа і порівняти їх результати

Актуальність роботи

Телебачення і в XXI сторіччі залишається головним джерелом інформації. У контексті активного розвитку телебачення та його використання як інструменту пропаганди та зомбування людей, завданням сучасного українського телебачення є формування незалежного інформаційного простору від держави-агресора, адже “Морські дияволи”, “Менти” та інші серіали про силові структури Росії шкодили формуванню українських поглядів та світогляду на нашу армію, поліцію. Після Революції Гідності і початку військового конфлікту на Донбасі, анексії Криму, російські телепередачі, що транслюються на українських телеканалах, використовувалися як елемент пропаганди. Глядачі в Україні спостерігали і, можливо, пишалися російськими спецпризначенцями, шпигунами, оперативниками ФСБ та поліцейськими. У зв'язку з цим важливо розробляти методи прогнозування, які будуть працювати в умовах постійних змін і покажуть, що незважаючи на всі складнощі, українське телебачення починає активно розвиватись, а вплив російського телепродукту зменшується.

Об'єкт дослідження

Об'єктом дослідження було обрано ключові показники, які впливають на перегляд людьми саме російського телепродукту

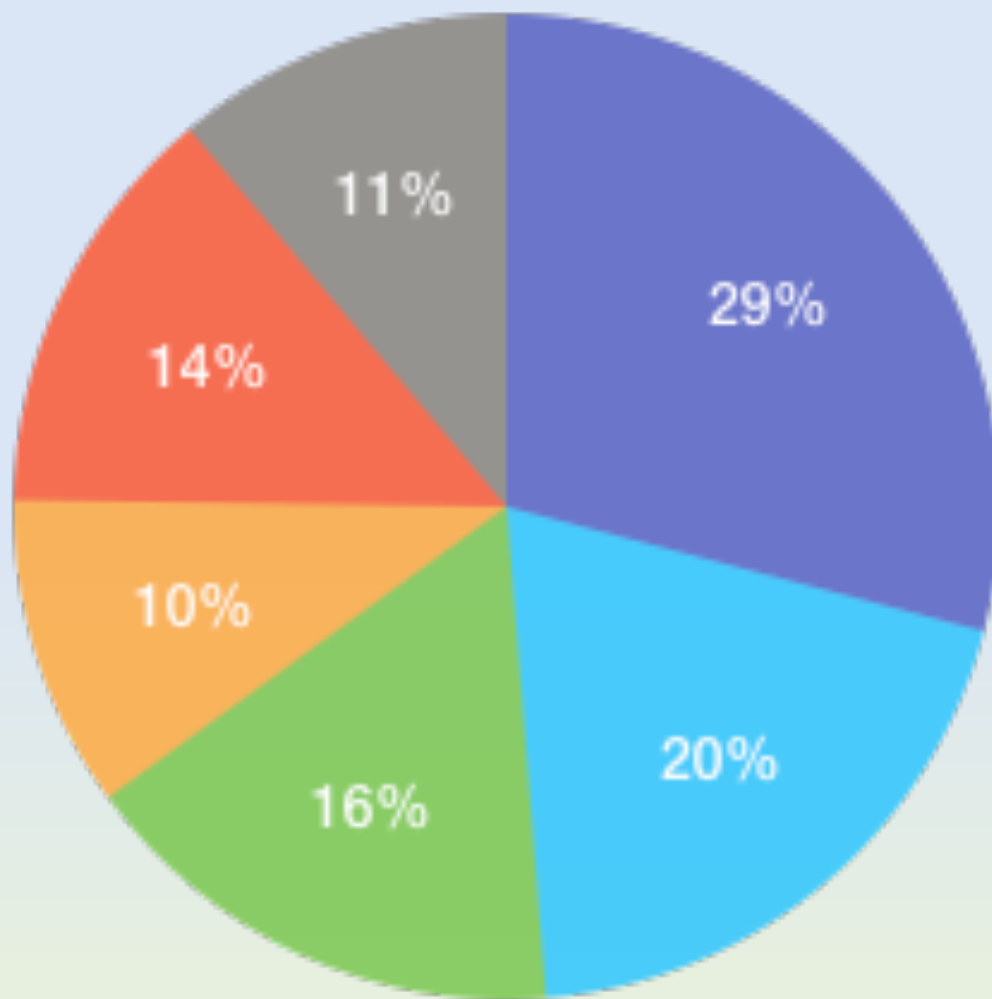
Предмет дослідження

Предметом дослідження є інтелектуальні методи прогнозування на основі метода групового урахування аргументів і штучних нейронних мереж прямого розповсюдження.

Мета роботи

Побудова прогностичної моделі для здійснення короткострокового прогнозу відсотку російського контенту

Обсяг закупленого російського контенту станом на 2014 рік



● Україна ● Інтер ● СТБ ● 1+1 ● ICTV ● Новий канал

Вхідні параметри для прогнозування

У якості вхідних параметрів для прогнозування було обрано наступні помісячні показники починаючи з січня 2014 року і до травня 2018:

1. Середня заробітна плата (в доларовому еквіваленті)
2. Відсоток українців, що ставляться позитивно до Російської Федерації (за даними соціологічного опитування КМІС та “Левада-центр”)
3. Вартість показу російського контенту у категорії
4. Відсоток глядачів, що дивляться російськомовний контент цієї категорії
5. Відсоток російського контенту у цій категорії.

*під категорією тут і надалі будемо розуміти: анімація, документальні фільми, концерти, серіали, фільми.

Метод групового урахування аргументів

Дано: N вхідних числових $X = (x_1, x_2, \dots, x_n)$, і вихідний числовий ряд $Y = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ijk} x_i x_j x_k + \dots$, де усі коефіцієнти a_0, a_i, a_{ij} невідомі.

Пошук найкращої моделі відбувається таким чином:

- уся вибірка ділиться на навчальну та перевірочну: $N_{\text{виб}} = N_{\text{навч}} + N_{\text{перев}}$;
- на навчальній вибірці $N_{\text{навч}}$ визначаються оцінки коефіцієнтів $\bar{a}_0, \bar{a}_i, \bar{a}_{ij}$;
- на перевірочній вибірці $N_{\text{перев}}$ обираються найкращі моделі.

У роботі проводився прогноз на основі даних за останні n місяців, де n - від 1 до 5.

Метод групового урахування аргументів

Принцип свободи вибору (незакінченості проміжного рішення):

1. Для кожної пари x_i та x_j будуюмо часткові описи (усього C_n^2) вигляду:

- або $y_s = \varphi(x_i, x_j) = a_0 + a_i x_i + a_j x_j$, $s = 1 \dots C_n^2$ (лінійні);
- або $y_s = \varphi(x_i, x_j) = a_0 + a_i x_i + a_j x_j + a_{ii} x_i^2 + a_{ij} x_i x_j + a_{jj} x_j^2$, $s = 1 \dots C_n^2$ (квадратичні).

2. Визначаємо коефіцієнти цих моделей за МНК, використовуючи навчальну вибірку, тобто знаходимо оцінки $\bar{a}_0, \bar{a}_1, \dots, \bar{a}_j, \dots, \bar{a}_N, \bar{a}_{11}, \dots, \bar{a}_{ij}, \dots, \bar{a}_{NN}$.

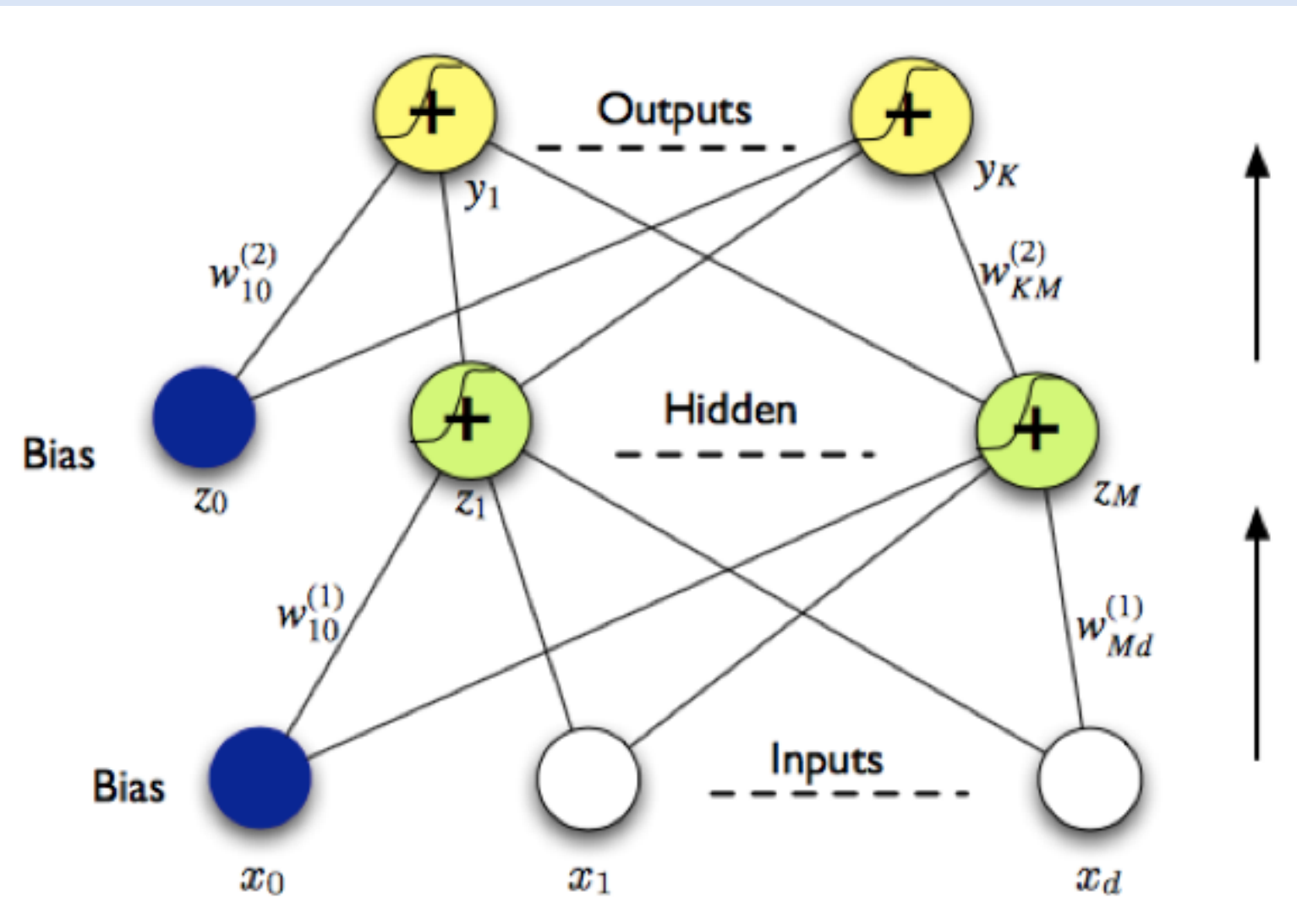
3. Далі на перевіірочній вибірці для кожної з цих моделей шукаємо оцінку за критерієм регулярності:

$$\bar{\delta}_s^2 = \frac{1}{N_{\text{перев}}} \sum_{i=1}^{N_{\text{перев}}} [Y(k) - \bar{Y}_s(k)]^2,$$

Де $Y(k)$ – дійсне значення вихідної змінної в k -тій точці перевіірочної вибірки; $\bar{Y}_s(k)$ – вихідне значення в

k -тій точці перевіірочної вибірки у відповідності з s -ою моделлю; $N_{\text{перев}}$ – число точок перевіірочної вибірки.

Багатошаровий персеptron

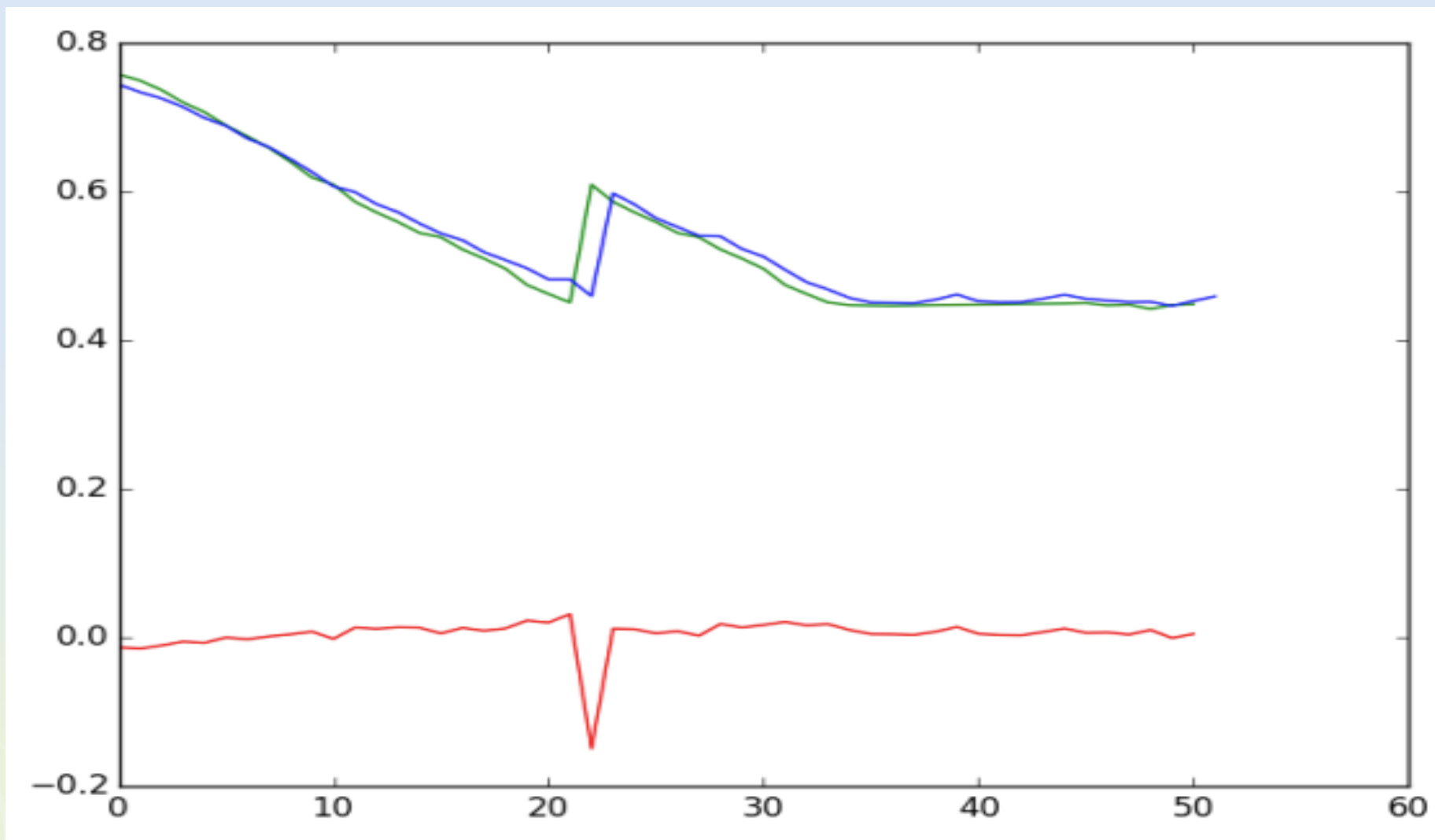


Досліджувались нейронні мережі з такими параметрами:

- від одного до трьох слоїв
- 20, 50, 100, 200 або 500 нейронів у слої
- активаційна функція: лінійна, гіперболічний тангенс, сигмоїд

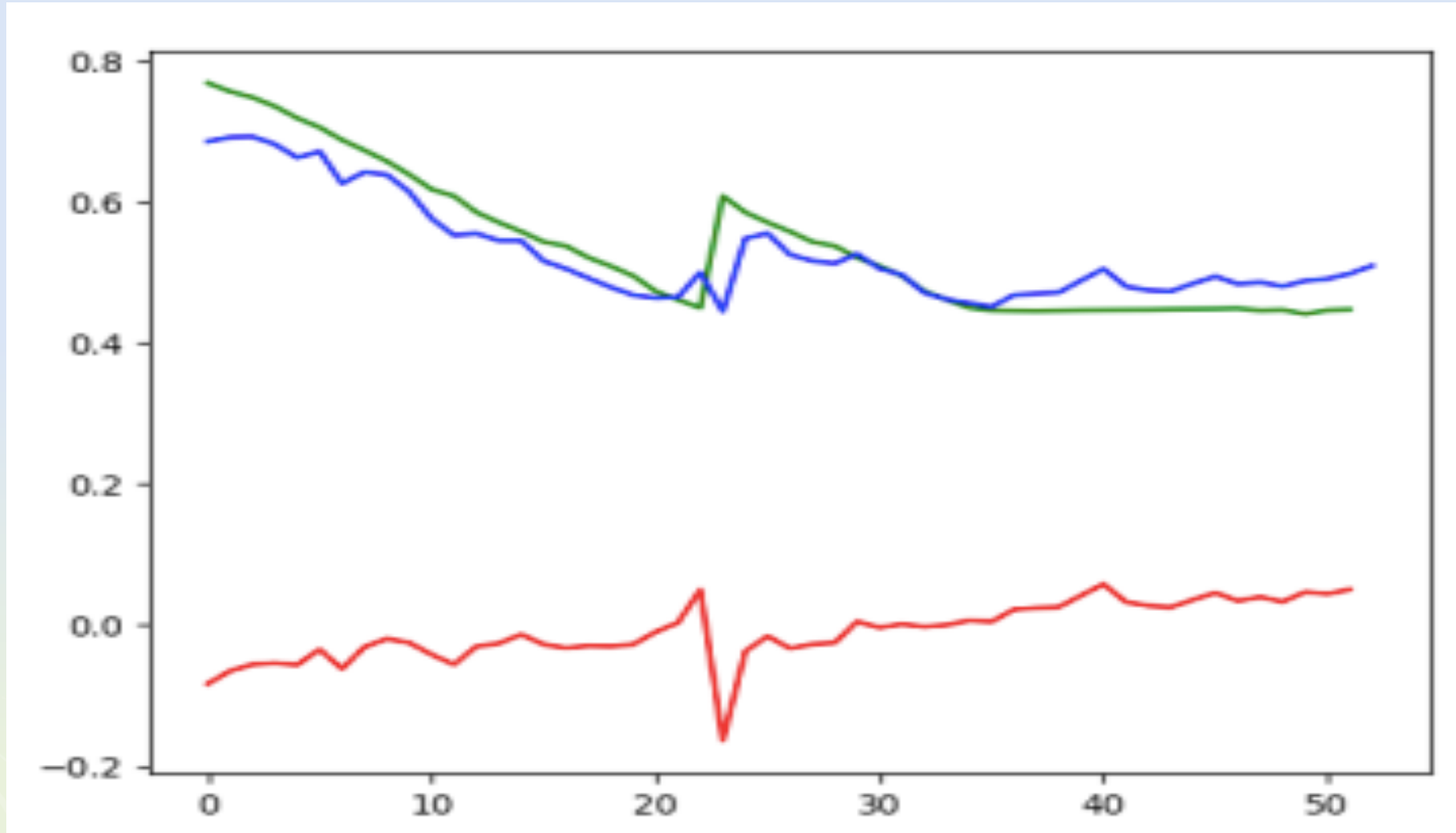
Кількість нейронів і активаційна функція різних шарів, а також активаційна функція виходу можуть відрізнитись.

Найкращий результат прогнозування методом МГУА в категорії “серіали”



Зелений – дані, які подаються моделі на вхід.
Синій – вихід моделі.
Червоний – різниця між вхідними даними і вихідними

Найкращий результат прогнозування щільними нейронними мережами в категорії “серіали”



Зелений – дані, які подаються моделі на вхід.
Синій – вихід моделі.
Червоний – різниця між вхідними даними і вихідними

Порівняння моделей

	Нейронна мережа	МГУА	НМ	МГУА	НМ	МГУА	НМ	МГУА
Категорія контенту	Середньоквадратичне відхилення (standard deviation)		Середній квадрат похибки (mean squared error)		Найбільша похибка		Прогноз на червень 2018	
Анімація	0,01794	0,008618	0,0003219	7,428e-05	0,0491	0,04498	0,0367	0,01750
Концерти	0,08454	0,03395	0,007148	0,001152	0,2140	0,1701	0,2515	0,09680
Фільми	0,03919	0,01412	0,001536	0,0001994	0,0788	0,08494	0,3349	0,4161
Документальні фільми	0,03671	0,01052	0,001348	0,0001107	0,08655	0,06321	0,7780	0,8210
Серіали	0,04279	0,02244	0,001831	0,0005034	0,1636	0,1523	0,5101	0,4430

Висновки

- Проведено детальний аналіз предметної області та визначено основні параметри, які вплинули на кількість російського телевізійного продукту
- Сформована вибірка, на основі якої було проведено прогнозування відсотку російського контенту у категорії на наступний місяць за даними за попередні періоди
- На основі проведених моделювань було проведено порівняльний аналіз результатів МГУА і щільної нейронної мережі, який показав, що МГУА в цілому дає кращі результати при прогнозуванні відсотку російського контенту в певній категорії телевізійних передач, аніж щільні нейронні мережі.
- Розроблено програмний продукт для прогнозування рівня російського контенту.
- Тези даної роботи були опубліковані у матеріалах міжнародної SAIT

Подальші дослідження

У ході дослідження було встановлено, що МГУА дає хороші результати на досліджуваних даних. Планується розвивати роботу у напрямку дослідження застосування модифікацій МГУА з метою подальшого зменшення похибки прогнозування у різних задачах, пов'язаних з телебаченням.

Дякую за увагу!

КОНЕЦ

ФИЛЬМА

