Условия существования петель скрытой обратной связи в рекомендательных системах с учётом шума

А. А. Пилькевич[1]***, А. С. Хританков***[2]

anton39reg@mail.ru; anton.khritankov@phystech.edu

В работе исследуются петли скрытой обратной связи в рекомендательных системах. Под положительной обратной связью подразумевается неограниченный рост интереса пользователя к предлагаемым объектам. Решается задача поиска условий возникновения положительной обратной связи. Учитывается наличие шума в выборе пользователя. Рекомендательная система использует алгоритм Thomson Sampling Multi-armed Bandit. В задачах без шума известно, что существуют условия неограниченного роста. Но отсутствие шума не реализуется в реальных системах. Экспериментально проверяются полученные условия в имитационной модели.

The paper discovers hidden feedback loops in recommender systems. A positive feedback is an unlimeted growth of user interest in proposed objects. The paper looks for conditions for positive feedback loops. The paper takes into account a noise in user responses. The recommender system uses Thompson Sampling Multi-armed Bandit algorithm. In noise-free problems other works improved that conditions for unlimited growth exist. But noise-free is not true in a practice. The paper carries out simulation experiments to check found conditions.

Ключевые слова: *machine learning; hidden feedback loops*

DOI:

1 Введение

Рекомендательные системы являются важной составляющей социальных сетей, веб-поиска и других сфер [...]. Рассматриваются петли скрытой обратной связи, которые подразумевает рост качества предсказаний, как результат учёта принятых решений. Эффект петель скрытой обратной связи в реальных и модельных задачах в публикациях [...] опи-сыается как нежелательное явление. Частные и часто рассматриваемые случаи скрытой обратной являются echo chamber и filter bubles [1]. До сих пор нет строгой формализации условий возникновения этих эффектов при условиях приближенных к реальности [...].

Целью данной работы является нахождение условий существования петель обратной связи в рекоммендательной системе с алгоритмом Thomson Sampling в условиях зашумлённости выбора пользователя. Зашумлённость выбора рассматривается, как смещение первоначального интереса к исходному объект или категории. Предлагается способ отыскание условий модели исходя из теоретических свойств алгоритма TS. Под условиями подразумеваются параметры шума и параметры рекомендательной системы. Для описания условий предлагается выражение для математического ожидания интереса. Также рассмаривается вариант нахождения этих условий чисто из экспериментов. Целью является матетическое описание искомых условий с дальнейшим экспериментальным подтверждением полученных условий. Для проверки результатов используется имитационная модель, использующая синтетические данные.

Существует ранее описанання модель [1] петель в случае отсутствия шума в действиях пользователя. Подобное исследование проводилось в статье [1] на примере различных моделей (Oracle, Optimal Oracle, UCB, TS) в задаче многорукого бандита. Удалось показать условия существования неограниченного роста интереса пользователя. В работе [2] изучалась схожая постановка задачи и были получены условия возникновения, но рассматривалась линейная модель и градиентный бустинг. Но отсутвие шума в ответах пользователей в работах [1,2] не реализуется на практике. Важным отличием данной работы является факт рассмотрения более сложных условий модели, таких как шум в выборе пользователя и другой алгоритм рекомендательной системы.

В работе предлагается анализ роста интереса пользователя. Рассматривается математическое ожидание изменения интереса. Полученные условия проверяются в вычислительном эксперименте.

2 Петли скрытой обратной связи

Целью работы является теоретический анализ условий сходимости TS для различных параметров шума и экспериментальное подтверждение полученых соотношений. Также делается уточнений условий из [1].

2.1 Модель рекомендательной системы

Обозначим за t очередной момент выдачи рекомендаций. Рекомендательная система на шаге t выбирает элементы $(a_t^1, \ldots, a_t^l)$ из конечного набора M. Истинный *интерес* пользователя к элементу $a \in M$ описывается неизвестным отображением $\mu_t : M \to \mathbb{R}$. При этом считается, что чем больше значение $\mu_t(a)$, тем заинтересованнее пользователь в рекомендции a.

После очередного набора рекомендаций $a_t = (a_t^1, \ldots, a_t^l)$ пользователь возвращает *отклик* $c_t = (c_t^1, \ldots, c_t^l), c_t^i \in \{0, 1\}$. Предполагается, он выбирает элементы c_t^i случайно и независимо, пропорционально $\mu_t(a)$. Значит отклик имеет распределение Бернулли :

$$c_t^i \sim Bern(\sigma(\mu_t(a_t^i))), \text{ где } \sigma(x) = \frac{1}{1+e^x} \text{ — сигмоида }.$$

Предполагаем, что интерес пользователя во времени описывается как

$$\begin{cases} \mu_{t+1} \geqslant \mu_t, \text{ если } c_t = 1, \\ \mu_{t+1} < \mu_t, \text{ если } c_t = 0, \\ \mu_{t+1} = \mu_t, \text{ если элемент не попал в рекомендацию.} \end{cases}$$

Тогда петля обратной связи выражается как

$$\lim_{t\to\infty} \|\mu_t - \mu_0\|_2 = \infty.$$

Обновление интереса для элементов очередной рекомендации происходит по правилу:

$$\mu_{t+1} - \mu_t = \delta_t c_t - \delta_t(1 - c_t), \text{ где } \delta_t \sim U[0, 0.01].$$

Оптимизационной задачей рекомендательной системы является задача минимизации потерь. Максимальная сумма наград :

$$\max_{c_t^i} \sum_{t=1}^{T} \sum_{i=1}^{l} c_t^i = T \cdot l.$$

Тогда задача ставится так :

$$T \cdot l - \sum_{t=1}^{T} \sum_{i=1}^{l} c_t^i \to \min_b,$$

где b — используемый алгоритм в рекомендательной системе.

2.2 Алгоритм рекомендательной системы

Задача многорукого бандита состоит из k бандитов и системы взаимодействующей с ними. Каждый бандит имеет собственное распределение неизвестное для системы. Система "дёргает"за ручки бандита и получает награду из соответствующего распределения бандита. Задачей системы является максимизации суммы наград или же минимизации потерь.

В данной задаче рекомендательная система использует алгоритм Thompson Sampling [3] для задачи бернуллиевского бандита. Бандитами являются отклики пользователя c_t^i на очередую рекомендацию. Средняя награда равна: $\sigma(\mu_t(a_t^i))$.

В начальный момент времени определены вероятности бернуллиевских случайных величин c_t^i для элементов M равные $\pi_0(\theta_1), \ldots, \pi_0(\theta_m)$. Задаётся априорное распределение для θ_i равное бэта-распределению $Beta(1,1) = U[0,1]$. Апостериорное распределение для элемента $a^i \in M$ описывается бэта-распределением: $Beta(\alpha_t^i, \beta_t^i)$. Параметры после очередной рекомендации обновляются по закону :

$$\alpha_{t+1} = \alpha_t + c_t, \beta_{t+1} = \beta_t + 1 - c_t.$$

2.3 Учёт шума в поведении пользователя

Шум откликов описывается следующим образом:

$$\begin{gathered}
c_t^i \sim Bern\left(\sigma(s_t^i \cdot \mu_t(a_t^i) + q_t^i)\right), \\
P(s_t^i = 1) = p, \\
P(s_t^i = -1) = 1 - p, \\
q_t^i \sim U[-w, w].
\end{gathered}$$

Наличие q_t^i позволяет описать несмещённый аддитивный шум, то есть отклонение от истинного интереса пользователь. А s_t^i описывает кардинальное изменение интереса на противоположный.

3 Теоретическое обоснование

Назовём *особым режимом* работы TS поведение алгоритма, в котором TS не меняются элементы рекомендаций.

Утверждение 1. Пусть $p = 1$ и TS работает в особом режиме начиная с какого-то момента времени τ. Тогда при $w \geqslant 0 : \lim\limits_{t\to\infty} \|\mu_t - \mu_0\|_2 = \infty$.

Или же при фиксированном $p = 1$ и любых параметрах шума w возникает петля скрытой обратной связи.

Доказательство. Так как алгоритм работает в особом режиме, то при $t \geqslant \tau$ точно извествно, что он будет рекомендовать. Для случая нормы интересов:

$$\|\mu_t - \mu_0\|_2^2 = \sum_{i=1}^{M} (\mu_t^i - \mu_0^i)^2,$$

с ростом t сновной вклад будут давать только $l < M$ объектов попавших в рекомендацию. Причём эти объекты известны и не меняются для очередного шага.

Рассмотрим изменение интереса для произвольного $a \in M$. Обновление интереса происходит согласно: $\mu_t - \mu_{t-1} = \delta_t c_t - \delta(1 - c_t)$. Случайные величины δ_t, c_t независимы, поэтому:

$$\mathsf{E}\delta_t c_t = \mathsf{E}\delta_t \mathsf{E} c_t.$$

Для удобства будем считать, что у нас $c_t \sim \mathrm{Bern}_{\pm}(\sigma(s_t \cdot \mu_t(a_t) + q_t))$ Тогда:

$$\mathsf{E}(c_t | s_t = x, q_t = y) = 2\sigma(x \cdot \mathsf{E}\mu_{t-1} + y) - 1,$$
$$\mathsf{E}(\mathsf{E}(c_t | s_t, q_t = y)) = p \cdot (2\sigma(\mathsf{E}\mu_{t-1} + y)) - 1) + (1 - p) \cdot (2\sigma(-\mathsf{E}\mu_{t-1} + y)) - 1).$$

В случае $\mathsf{E}(\mathsf{E}(\mathsf{E}(c_t | s_t, q_t))) > 0$ петля будет возникать, так как рост интереса в среднем положителен.

Далее для простоты считается, что $\sigma(x) \approx \left(\frac{x}{4} + \frac{1}{2}\right) \cdot I[-2, 2] + I[2, \infty]$ и $p = 1$. Задача в этом случае записывается так:

$$\mathsf{E}(\mathsf{E}(c_t | s_t, q_t = y)) \approx 2\left(\frac{\mathsf{E}\mu_{t-1} + y}{4} + \frac{1}{2}\right) - 1.$$

Теперь петля возникает при условии: $\mathsf{E}\sigma(x) > \frac{1}{2}$.

Тогда остаётся посчитать:

$$\mathsf{E}\sigma(\mu_t) \approx \int_{-\infty}^{\infty} \left(\frac{\mathsf{E}\mu_t + y}{4} + \frac{1}{2}\right) I\{-2 < \mathsf{E}\mu_t + y < 2\} f(y) dy$$
$$+ \int_{-\infty}^{\infty} I\{2 < \mathsf{E}\mu_t + y\} f(y) dy =$$
$$\int_{-2}^{2} \left(\frac{z}{4} + \frac{1}{2}\right) f_s(z) dz + \int_{2}^{\infty} f_s(z) dz,$$

где $f_s(z)$ плотность $U[\mathsf{E}\mu_t - w, \mathsf{E}\mu_t + w]$. Таким образом у нас возникает 6 случаев.

1. $\mathsf{E}\mu_t + w < -2$. Тогда, очевидно:

$$\mathsf{E}\sigma(\mu_t) = 0 \to \mathsf{E}(\mathsf{E}(\mathsf{E}(c_t | s_t, q_t))) = -1.$$

   В этом случае интерес бесконечно убывает. Так как рассмытривается норма интересов, то всё равно $(\mu_t - \mu_0)^2 \to \infty$ при $t \to \infty$.

2. $\mathsf{E}\mu_t - w < -2 < \mathsf{E}\mu_t + w < 2$. Тогда:

$$\mathsf{E}\sigma(\mu_t) = \frac{1}{16w}(y + 2)^2 \Big|_{-2}^{\mathsf{E}\mu_t + w} = \frac{1}{16w}(\mathsf{E}\mu_t + w + 2)^2 > \frac{1}{2},$$
$$(\mathsf{E}\mu_t + w + 2)^2 > 8w,$$
$$\begin{cases} \mathsf{E}\mu_t > -w - 2 + \sqrt{8w}, \\ \mathsf{E}\mu_t < -w - 2 - \sqrt{8w}, \end{cases} \to \text{рост.}$$

3. $\mathsf{E}\mu_t - w < -2, \mathsf{E}\mu_t + w > 2$.

$$\mathsf{E}\sigma(\mu_t) = \frac{1}{16w}(y + 2)^2 \Big|_{-2}^{2} + \frac{1}{2w}(\mathsf{E}\mu_t + w - 2) =$$
$$\frac{1}{w} + \frac{\mathsf{E}\mu_t + w}{2w} - \frac{1}{w} = \frac{\mathsf{E}\mu_t + w}{2w} > \frac{1}{2} \Rightarrow$$
$$\mathsf{E}\mu_t > 0, w > 2 \to \text{ рост.}$$

4. $\mathsf{E}\mu_t - w > -2, \mathsf{E}\mu_t + w < 2$. Тогда:

$$\mathsf{E}\sigma(\mu_t) = \frac{1}{16w}(y+2)^2\Big|_{\mathsf{E}\mu_t - w}^{\mathsf{E}\mu_t + w} > \frac{1}{2},$$
$$(\mathsf{E}\mu_t + w + 2)^2 - (\mathsf{E}\mu_t - w + 2)^2 > 8w,$$
$$(2\mathsf{E}\mu_t + 4) \cdot 2w > 8w,$$
$$\mathsf{E}\mu_t > 0 \to \text{рост.}$$

5. $\mathsf{E}\mu_t - w > -2, \mathsf{E}\mu_t + w > 2$. Тогда:

$$\mathsf{E}\sigma(\mu_t) = \frac{1}{16w}(y+2)^2\Big|_{\mathsf{E}\mu_t - w}^{2} + \frac{1}{2w}\Big|_{2}^{\mathsf{E}\mu_t + w} =$$
$$\frac{1}{16w}\left(16 - (\mathsf{E}\mu_t - w + 2)^2\right) + \frac{1}{2w}(\mathsf{E}\mu_t + w - 2) =$$
$$\frac{1}{w} - \frac{(\mathsf{E}\mu_t - w + 2)^2}{16w} + \frac{\mathsf{E}\mu_t + w}{2w} - \frac{1}{w} =$$
$$-\frac{1}{16w}(\mathsf{E}^2\mu_t - 2(w-2)\mathsf{E}\mu_t + (w-2)^2) + \frac{\mathsf{E}\mu_t + w}{2w} > \frac{1}{2} \Rightarrow$$
$$\mathsf{E}^2\mu_t - 2(w-2)\mathsf{E}\mu_t + (w-2)^2 - 8(\mathsf{E}\mu + w) + 8w < 0,$$
$$\mathsf{E}^2\mu_t - 2(w+2)\mathsf{E}\mu_t + (w-2)^2 < 0,$$
$$(\mathsf{E}\mu_t - (w+2))^2 - (w+2)^2 + (w-2)^2 < 0,$$
$$(\mathsf{E}\mu_t - (w+2))^2 - 8w < 0,$$
$$\begin{cases} \mathsf{E}\mu_t < w + 2 + \sqrt{8w}, \\ \mathsf{E}\mu_t > w + 2 - \sqrt{8w}, \end{cases} \to \text{рост.}$$

6. $\mathsf{E}\mu_t - w > 2$. Тогда:

$$\mathsf{E}\sigma(\mu_t) = 1 > \frac{1}{2}.$$

4 Вычислительный эксперимент

Целью эксперимента является подтверждедние существования петель скрытой обратной связи для произвольных параметров шума w. Важной частью эксперимента является сравнения поведений рекомендательной системы с шумом в ответах пользователя и без.

4.1 Описание данных и работы модели

Перед началом эксперимента фиксируются следующие параметры: T — число итераций рекомендательной системы, $|M|$ — число рассматриваемых объектов для рекомендации, l — число элементов в одной выдачи. Также фиксируются параметры шума p, w, u. Далее случайным образом сэмплируются начальные значения интереса $\{\mu_0^i\}_{i=1}^{|M|}$. Параметры априорного распепеделения $\{\alpha_0^i, \beta_0^i\}_{i=1}^{|M|}$ также семплируются случайно.

Генерация элементов очередной рекомендации производится на основе текущего апостериорного распределения. Выбираются элементы с наибольшим значением. Получение отклика от пользователя заключается в генерации случайных величин на основе рекомендации. Обновление параметров апостериорного распределения происходит по правилу $\alpha_{t+1} = \alpha_t + c_t, \beta_{t+1} = \beta_t + 1 - c_t$. Интерес обновляется согласно $\mu_{t+1} - \mu_t = \delta_t c_t - \delta_t(1 - c_t)$.

Также рассматривается вариант эксперимента, когда используется случайная модель генерации рекомендации. В этом случае l элементов для очередной рекомендации выбираются случайным образом.

В каждый момент выдачи t фиксируются значения интереса μ_t^i, сумма откликов c_t^i и параметры апостериорного распределения. По полученным данным строятся графики для определения наличия петель скрытой обратной связи (см. рис. 1). Как определялось раньше, петля скрытой обратной связи выражается так: $\lim\limits_{t\to\infty} \|\mu_t - \mu_0\|_2 = \infty$.

4.2 Псевдокод проведения эксперимента

Вход: M, l, T, w, p
  BanditLoopExperiment.prepare()
  для t от 1 до T
    $r_t \leftarrow$ TSBandit.predict()
    $c_t \leftarrow$ make_response_noise(r_t, w, p)
    TSBandit.update(c_t)
    Model.interest_update(c_t)
    save_iter(t, c_t, μ_t)

5 Результаты

Из рис. 1 видно, что наблюдается эффект неограниченного роста интереса. Причём величина шума никак не ограничивает рост интереса, а лишь замедляет его, что согласуется с определением петли и утверждением 1.

Из графика суммы наград видно, что с определённого момента кривые начинают идти параллельно максимально возможной сумме. Это свидетельствует о наличии петель обратной связи.

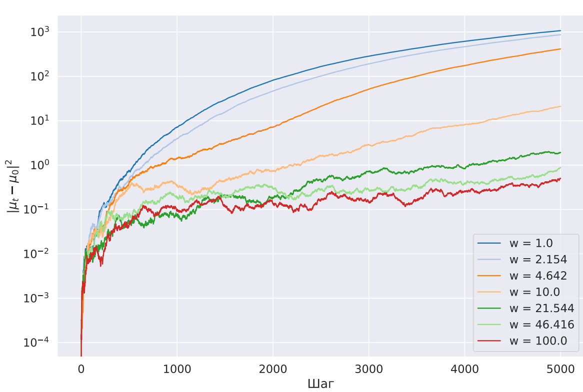

Рис. 1 Логарифм нормы интереса на очередном шаге рекомендации.

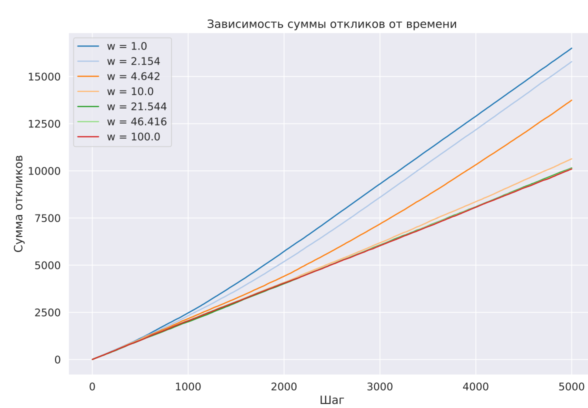

Рис. 2 Суммы наград на очередном шаге рекомендации.

Для случайной модели тоже наблюдается образование петли (см. рис. 4). Она более хаотична, но тренд неограниченного роста интереса всё равно присутствует.

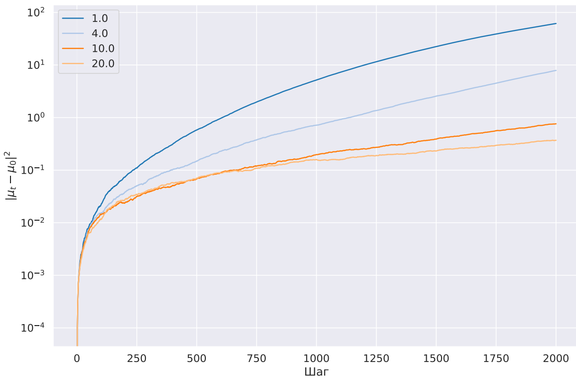

Рис. 3 Разброс логарифма нормы интереса от шага.

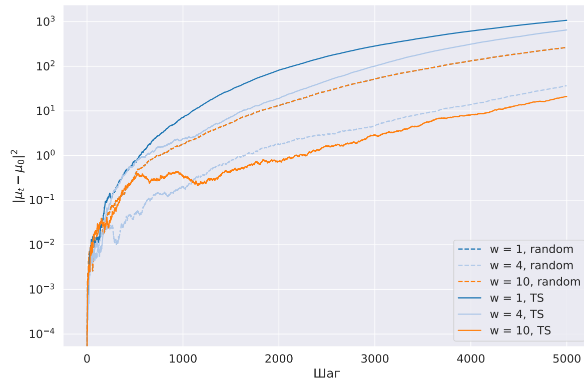

Рис. 4 Сравнение случайного алгоритма и TS.

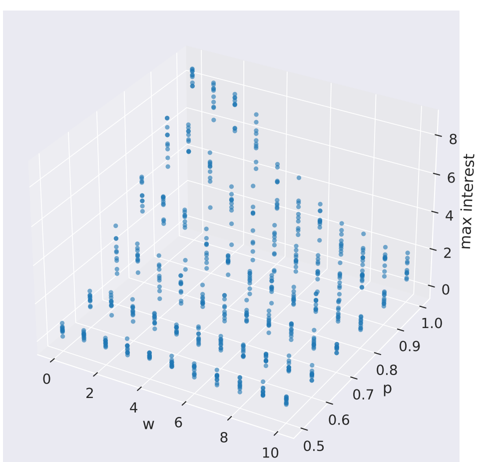

Рис. 5

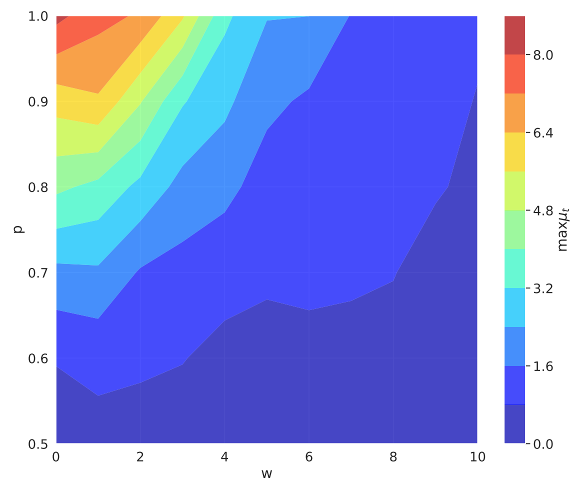

Рис. 6

Весь экперимент и исходники расположены в гите:

`https://github.com/Intelligent-Systems-Phystech/2021-Project-74`.

6 Заключение

Поставлена задача существования петель скрытой обратной связи при наличии шума в ответах пользователя. Для текущей модели шума было получено, что при любых параметрах возникают петли. Это также подверждается в эксперименте.

В дальнейшем требуется проверить гипотезу о возникновении петель при любом несмещённом аддитивном шума. Также стоит расмотреть другие модели шума. Так как после определённого значения интереса из-за сигмоиды любое влияние шума сводилось на нет.

Литература

[1] *Ray Jiang, Silvia Chiappa, Tor Lattimore,András György, Pushmeet Kohli* Degenerate Feedback Loops in Recommender Systems// CoRR, 2019, Vol. abs/1902.10730, URL: `https://arxiv.org/abs/1902.10730`.

[2] *Khritankov, Anton* Hidden Feedback Loops in Machine Learning Systems: A simulation Model and Preliminary Results// Springer, 2021, P. 54–65.

[3] *Daniel Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband* A Tutorial on Thompson Sampling// CoRR, 2017, Vol. abs/1707.02038, URL: `https://arxiv.org/abs/1707.02038`.

[4] *Shipra Agrawal, Navin Goyal* Analysis of Thompson Sampling for the multi-armed// CoRR, 2011, Vol. abs/1111.1797, URL: `https://arxiv.org/abs/1111.1797`.

[5] *Giuseppe Burtini, Jason L. Loeppky, Ramon Lawrence* Improving Online Marketing Experiments with Drifting Multi-armed Bandits// SciTePress, 2018, P. 630–636.

[6] *David Krueger and Tegan Maharaj and Jan Leike* Hidden Incentives for Auto-Induced Distributional Shift// CoRR, 2020, Vol. abs/2009.09153.

[7] *Wilbert Samuel Rossi, Jan Willem Polderman, Paolo Frasca* The closed loop between opinion formation and personalised recommendations// CoRR, 2018, Vol. abs/1809.04644.

[8] *Pedreschi, D. and Miliou, I. and European Parliament. Directorate-General for Internal Policies of the Union* Artificial Intelligence (AI): new developments and innovations applied to e-commerce// European Parliament, 2020.

Поступила в редакцию