

УДК 51–77

DOI: [10.26102/2310-6018/2021.33.2.010](https://doi.org/10.26102/2310-6018/2021.33.2.010)

Модель оценки удовлетворенности потребителей на основе онлайн-отзывов с помощью метода главных компонент

Е.Б. Грибанова, В.В. Саулин

*Томский государственный университет систем управления и радиоэлектроники,
Томск, Российская Федерация*

Резюме. Актуальность исследования обусловлена высокой популярностью интернет-сервисов для публикации отзывов о товарах / услугах, их влиянием на поведение покупателей, а также необходимостью автоматизации обработки данных таких сервисов из-за большого количества представленной информации. Разработана модель оценки удовлетворенности потребителей на основе их отзывов, учитывающая оценку товара потребителем, определяющую характер отзыва (отрицательный, положительный, нейтральный), и оценку отзыва другими участниками. Интегральный показатель оценки удовлетворенности формируется на основе нормированных значений средней оценки потребителей и суммарных оценок положительных, нейтральных и отрицательных отзывов. При этом определение весовых коэффициентов, используемых в расчете интегрального показателя, выполнено с помощью метода главных компонент. В статье представлены результаты расчета интегрального показателя для шести моделей видеокарт. На основе разработанной модели реализована программа, позволяющая автоматизировать сбор данных с интернет-сайта и выполнить расчет интегрального показателя. Реализация программы выполнена с помощью языка программирования Java и среды разработки IntelliJ IDEA. Разработанная модель и программа могут быть использованы как потенциальными покупателями, принимающими решение о покупке товаров, так и предприятиями, осуществляющими продажу товаров и стремящимися получить обратную связь, улучшить ассортимент и качество обслуживания.

Ключевые слова: отзывы потребителей, метод главных компонент, рейтинг, сбор данных, линейная модель

Для цитирования: Грибанова Е.Б., Саулин В.В. Модель оценки удовлетворенности потребителей на основе онлайн-отзывов с помощью метода главных компонент. *Моделирование, оптимизация и информационные технологии.* 2021;9(2). Доступно по: <https://moitvvt.ru/ru/journal/pdf?id=937> DOI: 10.26102/2310-6018/2021.33.2.010

Customer satisfaction assessment model based on online reviews using principal component analysis

E.B. Gribanova, V.V. Saulin

*Tomsk state university of control systems and radioelectronics,
Tomsk, Russian Federation*

Abstract: The relevance of the study is due to the high popularity of Internet services for publishing reviews of goods or services, their impact on consumer behavior, as well as the need to automate the processing of data from such services due to a large amount of information provided. A model for assessing customer satisfaction based on their feedback has been developed, taking into account the assessment of the product by the consumer, which determines the nature of the response (negative, positive, neutral) and the evaluation of the feedback by other participants. The integral indicator of satisfaction assessment is formed based on normalized values of the average merit of consumers and the validation of positive, neutral, and negative reviews. In this case, the weight determination of the coefficients used in the

integral indicator calculation was carried out using the principal components method. The article presents the results of calculating the integrated indicator for six models of video cards. Based on the developed model, a program has been implemented. That allows automating data collection from the website and calculating the integral indicator. The program is implemented using the Java programming language and the IntelliJ IDEA development environment. The developed model and program can both be used by potential buyers who decide to purchase goods, as well as enterprises selling goods, and seeking to get feedback, identify weak and strong points, improve the range and quality of service.

Keywords: consumer reviews, principal component method, rating, data collection, linear model

For citation: Griбанова E.B., Saulin V.V. Customer satisfaction assessment model based on online reviews using principal component analysis. *Modeling, Optimization and Information Technology*. 2021;9(2). Available from: <https://moitvvt.ru/ru/journal/pdf?id=937> DOI: 10.26102/2310-6018/2021.33.2.010 (In Russ).

Введение

В настоящее время стремительное развитие интернет-технологий и их высокая востребованность пользователями обусловили создание различных интернет-ресурсов для общения и обмена мнениями участников [1]. Такие сайты могут рассматриваться организациями как площадки для проведения маркетинговых мероприятий и получения обратной связи от потребителей. Одним из таких видов ресурсов являются интернет-сайты, предназначенные для публикации отзывов товаре или услуге. Отзыв представляет собой форму обратной связи, при которой потребитель делится своим опытом взаимодействия с организацией и/или использования товара или услуги. Существует множество сервисов, которые предоставляют возможность покупателю опубликовать свое мнение о товаре/услуге и оценить его в соответствии с определенной шкалой. Такая обратная связь позволяет выявить слабые и сильные места, улучшить ассортимент и качество обслуживания. Большая популярность ресурсов, посвященных отзывам, обусловлена стремлением потенциальных покупателей сэкономить денежные ресурсы и получить качественный товар или услугу. Однако такие сервисы могут быть использованы для публикации негативной информации о конкурентах, заказных положительных отзывов, кроме того, вследствие субъективности оценки потребителей, информация может быть противоречивой. Эти факторы приводят к тому, что анализ отзывов и идентификация реального качества продукта или услуги является трудоемкой задачей, требующей работы с большим количеством информации. В связи с этим разрабатываются модели и программные продукты, позволяющие выполнить оценку отзывов и автоматизировать процесс её получения.

Среди существующих работ в этой области можно отметить следующие.

В работе [2] авторами предложено два подхода для оценки отзывов о товарах с сайта Amazon: на основе ожидаемой полезности отзыва и ожидаемого объема продаж. Авторами работы [3] в рамках исследования влияния положительных и отрицательных отзывов на продажи товаров (DVD-, Blu-Ray-плееры) с сайта Amazon была построена регрессионная модель, описывающая зависимость рейтинга продаж от бренда товара, числа позитивных негативных отзывов, цены товара, расходов на рекламу, общего количества представленных моделей за рассматриваемый период. В статье [4] выполнено исследование влияния поведения организаций в отношении негативных отзывов на степень доверия потребителей. Также в литературе приводятся работы, посвященные оценке влияния на аудиторию анонимных [5] и негативных [6] отзывов, активности участников и полезности из отзывов [7], предсказанию оценок на основе анализа текста [8] и т.д.

Представленная работа посвящена анализу товаров на основе онлайн-отзывов. Цель работы заключается в разработке системы оценки удовлетворенности потребителей на основе их отзывов. Для достижения поставленной цели необходимо решить следующие задачи: разработать модель оценки удовлетворенности потребителей на основе отзывов, учитывающую оценку товара потребителем и оценку отзыва другими участниками; реализовать систему, позволяющую автоматизировать сбор исходных данных и вычисление интегрального показателя оценки удовлетворенности потребителей; провести вычислительные эксперименты по расчёту оценок удовлетворенности потребителей для выбранной группы товаров.

Для сбора данных был использован сервис Яндекс Маркет, который позволяет пользователям оценивать товар по пятибалльной шкале, оставлять отзывы о товаре и оценивать отзывы других пользователей. На данный момент Яндекс Маркет является популярным сервисом в русскоязычном сегменте интернета и интегрирован в некоторые интернет-магазины.

Модель оценки удовлетворенности потребителей на основе онлайн отзывов

Входными данными модели являются:

- оценка V , которую поставил товару отдельный потребитель (от 1 до 5);
- средняя оценка товара для всех потребителей (L);
- число отметок «мне нравится» U , которые поставили отзыву другие участники;
- число отметок «мне не нравится» H , которые поставили отзыву другие участники.

В зависимости от оценки V , которую поставил товару автор отзыва, отзывы группируются на положительные (оценка «5» или «4»), нейтральные (оценка «3»), отрицательные (оценки «2» или «1»).

Оценивание отзыва другими участниками позволяет судить о его качестве. В работе оценка отзыва R_l (l – номер отзыва) формируется как разница между числом отметок «мне нравится» и числом отметок «мне не нравится»:

$$R_l = U_l - H_l.$$

Высокая положительная оценка говорит о том, что большое число участников согласны с ним и находят его полезным. Высокая отрицательная оценка отражает большое расхождение с мнением большинства, что может свидетельствовать о его высоком уровне субъективности или заказном характере. При этом чем больше отзывов было оставлено участниками, тем меньше неопределенность относительно мнения об этом товаре.

Для формирования интегрального показателя необходимо выполнить обработку исходных данных, которая заключается в их нормировании и определении весовых коэффициентов с помощью метода главных компонент.

Оценки отзывов R нормируются по формуле:

$$R_l^* = \frac{R_l - R_{\min}}{R_{\max} - R_{\min}}, \quad (1)$$

где R_l^* – нормированное значение оценки отдельного отзыва;

R_l – оценка отдельного отзыва;

R_{\max} , R_{\min} – величины максимальной и минимальной оценки отзыва, определяемые на основе исходного набора данных оценок отзывов.

Далее необходимо выполнить преобразование оценок каждого отзыва в одну характеристику для положительных, отрицательных и нейтральных отзывов. Таким образом будут сформировать показатели, которые будут отражать величины оценок отзывов, а также количество оценок отзывов. Для этого определим суммы оценок отзывов:

$$\begin{aligned} Sp &= \sum_{j=1}^{n_1} R_j^{*(p)}, \\ Sn &= \sum_{j=1}^{n_2} R_j^{*(n)}, \\ Si &= \sum_{j=1}^{n_3} R_j^{*(i)}, \end{aligned} \quad (2)$$

где Sp – суммарная оценка положительных отзывов;

Sn – суммарная оценка отрицательных отзывов;

Si – суммарная оценка нейтральных отзывов;

n_1 – число положительных отзывов;

n_2 – число отрицательных отзывов;

n_3 – число нейтральных отзывов.

В этой формуле j представляет собой индекс отзыва, индексы p, n, i говорят о том, что для вычисления показателя используются положительные (p), нейтральные (i) и отрицательные (n) отзывы.

Далее необходимо выполнить линейную свертку, используемую при решении задач многокритериального выбора и заключающуюся в нормировании показателей и их сложении с учетом коэффициентов важности. В итоге получим значение интегральной оценки степени удовлетворенности. Выполним нормирование значений (2) для рассматриваемых товаров (k – номер товара):

$$\begin{aligned} L_k^* &= \frac{L_k - L_{\min}}{L_{\max} - L_{\min}}, \\ Sp_k^* &= \frac{Sp_k - Sp_{\min}}{Sp_{\max} - Sp_{\min}}, \end{aligned} \quad (3)$$

где L_k – средняя оценка k -го товара,

L_k^* – нормированное значение оценки k -го товара.

Нормирование отрицательных и нейтральных отзывов осуществляется аналогично.

Далее для формирования общей оценки необходимо определить весовые коэффициенты для каждого из показателей. Для этого был использован подход, рассмотренный в работах [9,10] и основанный на применении метода главных компонент. В работе [9] данный подход был использован для оценки конкурентоспособности предприятий, в статье [10] этот метод использован для оценки поставщиков. В результате применения метода происходит аппроксимация данных линейными многообразиями меньшей размерности. В работе [10] было отобрано три компонента, а весовой коэффициент, соответствующий определенной характеристике поставщика, вычисляется как произведение значения каждой факторной нагрузки, умноженной на процент общей дисперсии, которую соответствующая компонента объясняет (при одной компоненте умножения на дисперсию может не осуществляться).

Для реализации метода главных компонент в работе используется алгоритм NIPALS [11], в результате применения которого формируется матрица счетов и матрица нагрузок:

$$X = T \cdot P^T,$$

где X – матрица исходных переменных;

T – матрица счетов;

P – матрица факторных нагрузок.

Формула расчета весовых коэффициентов w будет иметь вид:

$$w_t = \sum_{r=1}^m P_{t,r} \cdot d_r, \quad (4)$$

где $P_{t,r}$ – факторная нагрузка, соответствующая переменной-характеристике t и компоненте r ;

d_r – процент дисперсии, объясняемый r -й компонентой;

m – число главных компонент.

Формула расчета интегрального показателя M оценки удовлетворенности потребителей на основе отзывов для продукта k формируется на основе нормированных значений показателей (3) и вычисленных коэффициентов важности (4):

$$M_k = L_k^* \cdot w_1 + Sp_k^* \cdot w_2 + Sn_k^* \cdot w_3 + St_k^* \cdot w_4. \quad (5)$$

Для исследования были взяты 6 моделей видеокарт. На Рисунке 1 представлен график, иллюстрирующий объем собранных данных по каждой видеокарте.

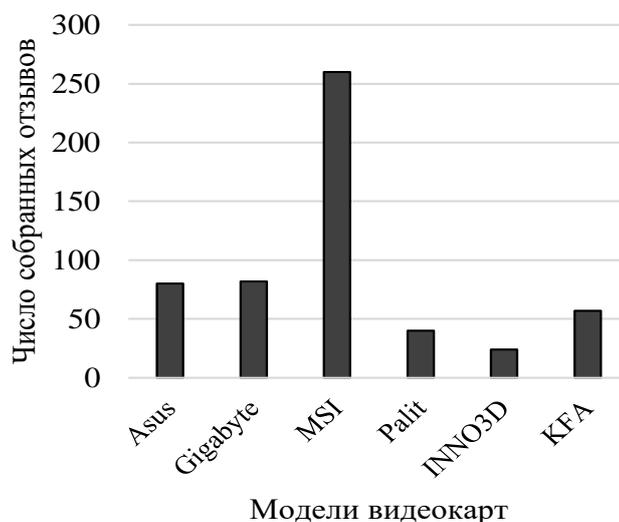


Рисунок 1 – Объем собранных данных
 Figure 1 – Amount of data collected

На Рисунке 2 представлены оценки отзывов модели видеокарты MSI до выполнения нормирования по формуле (1) (число отметок «мне нравится» минус число отметок «мне не нравится»).

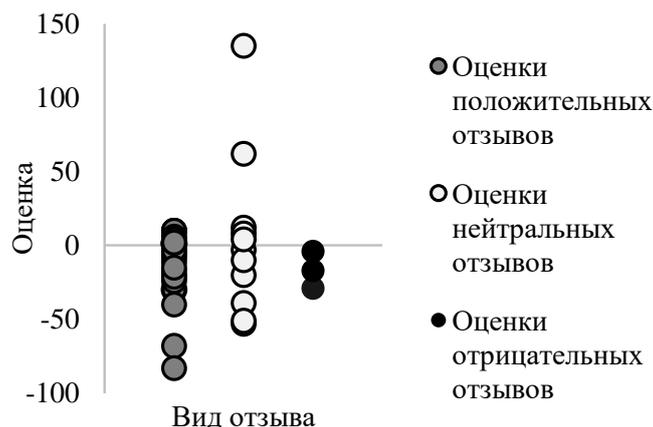


Рисунок 2 – Оценки отзывов R видеокарты модели MSI
Figure 2 – Review scores R for MSI graphics card

В Таблице 1 представлены исходные данные по шести видеокартам (P – цена товара в руб.).

Таблица 1 – Исходные данные по шести видеокартам
Table 1 – Source data for six video cards

Видеокарта	Характеристики видеокарт				
	L	Sp	Si	Sn	P
KFA2	4,5	15,03	1,01	0,92	10999
Palit	4,4	13,63	2,25	1,9	11299
INNO3D	4,3	9,48	0,5	1,45	12299
MSI	4,8	14,67	6,04	1,45	13299
Gigabyte	4,8	15,07	2,5	0,51	14799
Asus	4,8	14,75	2,4	0	15999

Полученная матрица нагрузок для первой компоненты представлена в Таблице 2.

Таблица 2 – Значения матрицы нагрузок
Table 1 – Load matrix values

Переменная	Первая компонента
L^*	0,977
Sp^*	0,861
Si^*	0,657
Sn^*	-0,623

Вычислив значение интегрального показателя M , можно выполнять сравнение между собой товаров с точки зрения их оценки пользователями.

На Рисунке 3 представлены значения интегрального показателя (5) и цены для рассматриваемых моделей видеокарт. Можно отметить, что эффективное по Парето множество образуют видеокарты моделей: KFA2, MSI, Asus.

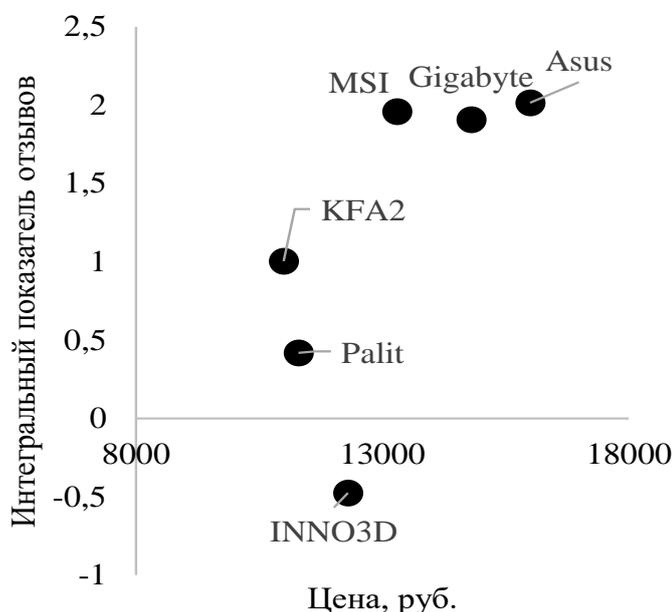


Рисунок 3 – Значения цены и интегральных показателей
 Figure 3 – Values of the price and integral indicators

Разработка программы оценки отзывов

Для автоматизации процесса вычисления интегрального показателя отзывов было разработано программное приложение. Реализация программы выполнена с помощью языка программирования Java и среды разработки IntelliJ IDEA. Для получения данных с сайта была использована библиотека Jsoup, работа с таблицами Excel осуществляется с помощью библиотеки Apache POI.

Программа состоит из следующих компонентов:

- модуль сбора данных, реализующий сбор и сохранение всей необходимой информации о товаре и отзывах;
- модуль группировки данных, выполняющий группировку информации о нескольких товарах и их нормирование;
- модуль анализа данных, осуществляющий определение интегрального показателя с помощью метода главных компонент.

На Рисунке 4 представлен фрагмент формы сбора данных о товаре с сайта. В данном окне пользователю необходимо указать ссылку на страницу товара в Яндекс Маркет, после чего, нажимая на кнопки «получить количество» и «получить оценки», производится сбор общей информации о товаре и информации об оценках отдельных отзывов. После сбора всех необходимых данных создается файл электронных таблиц.

При необходимости вычисления весовых коэффициентов для собранных данных выбирается соответствующее действие. Также можно воспользоваться вычисленными ранее весовыми коэффициентами для различных групп товаров (например, в случае, если для рассматриваемых товаров имеется небольшое количество отзывов).

После получения значений весовых коэффициентов может быть вычислено значение интегрального показателя и составление рейтинга товаров на основе отзывов.

Всего:

положительных:

нейтральных:

отрицательных:

общая оценка:

Положительные

Ссылки на страницы

Результат в виде массива

Величина с учетом оценки

Рисунок 4 – Фрагмент интерфейса
Figure 4 – Fragment of the interface

Заключение

В статье приводится описание модели оценки удовлетворённости потребителей на основе отзывов. В отличие от существующих работ модель учитывает не только характер отзыва, но и его оценку другими участниками. Описано формирование интегрального показателя с помощью нормирования исходных данных и определения весовых коэффициентов методом главных компонент. В качестве примера представлен расчет для шести моделей видеокарт одной ценовой категории. Также в статье представлено описание программы, позволяющей выполнять сбор данных и их обработку для автоматизированного вычисления интегрального показателя отзывов.

Данная система, а также представленная модель могут быть использованы как потребителями для принятия решения о покупке, так и организациями для формирования ассортимента и ценовой политики.

ЛИТЕРАТУРА

1. Грибанова Е.Б., Савицкий А.С. Разработка программы оценки времени размещения сообщения в онлайн-социальной сети ВКонтакте. *Моделирование, оптимизация и информационные технологии*. 2020. Доступно по: <https://moitvvt.ru/ru/journal/article?id=724>. DOI: 10.26102/2310-6018/2020.28.1.012 (дата обращения: 25.02.2021).
2. Ghose A., Ipeirotis P. Designing Ranking Systems for Consumer Re-views: The Impact of Review Subjectivity on Product Sales and Review Quality. *Proceedings of the 16th annual workshop on information technology and systems*. 2006:303–310.
3. Ho-Dac N.N., Carson S.J., Moore W.L. The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *Journal of Marketing*. 2013:37–53.
4. Zhao H., Jiang L., Su C. To Defend or Not to Defend? How Re-sponses to Negative Customer Review Affect Prospective customers' Distrust and Purchase Intention. *Journal of Interactive Marketing*. 2020:45–64.
5. Zhao K., Stylianou A.C., Zheng Y. Sources and Impacts of Social Influence from Online Anonymous User Reviews. *Information and Management*. 2018:16–30.
6. Gafni R., Gola O.T. The Influence of Negative Consumer Reviews in Social Networks. *Online Journal of Applied Knowledge Management*. 2016:44–58.
7. Mohammadiani R.P., Mohammadi S., Malik Z. Understanding the relationship strengths in users' activities, review helpfulness and influence. *Computers in Human Behavior*. 2017:117–129.
8. Xing S., Liu F., Wang Q., Zhao X., Li T. A hierarchical attention model for rating prediction by leveraging user and product reviews. *Neurocomputing*. 2019:417–427.
9. Кендюхов А.В., Толкачев Д.О. Использование метода главных компонент для оценки конкурентоспособности машиностроительных предприятий. *Маркетинг и менеджмент инноваций*. 2013:219–227.
10. Petroni A., Braglia M. Vendor Selection Using Principal Component Analysis. *The Journal of Supply Chain Management*. 2000:63–69.
11. Sarkodie S.A., Adom P.K. Determinants of energy consumption in Kenya: A NIPALS approach. *Energy*. 2018:696–705.

REFERENCES

1. Gribanova E.B., Savitsky A.S. Development of a program to estimate the time of posting a message in the online social network VKontakte. *Modeling, Optimization and Information Technology*. 2020. Available at: <https://moitvvt.ru/ru/journal/article?id=724>. DOI: 10.26102/2310-6018/2020.28.1.012 (In Russ) (accessed: 25.02.2021).
2. Ghose A., Ipeirotis P. Designing Ranking Systems for Consumer Re-views: The Impact of Review Subjectivity on Product Sales and Review Quality. *Proceedings of the 16th annual workshop on information technology and systems*. 2006:303–310.
3. Ho-Dac N.N., Carson S.J., Moore W.L. The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *Journal of Marketing*. 2013:37–53.
4. Zhao H., Jiang L., Su C. To Defend or Not to Defend? How Re-sponses to Negative Customer Review Affect Prospective customers' Distrust and Purchase Intention. *Journal of Interactive Marketing*. 2020:45–64.
5. Zhao K., Stylianou A.C., Zheng Y. Sources and Impacts of Social Influence from Online Anonymous User Reviews. *Information and Management*. 2018:16–30.

6. Gafni R., Gola O.T. The Influence of Negative Consumer Reviews in Social Networks. *Online Journal of Applied Knowledge Management*. 2016:44–58.
7. Mohammadiani R.P., Mohammadi S., Malik Z. Understanding the relationship strengths in users' activities, review helpfulness and influence. *Computers in Human Behavior*. 2017:117–129.
8. Xing S., Liu F., Wang Q., Zhao X., Li T. A hierarchical attention model for rating prediction by leveraging user and product reviews. *Neurocomputing*. 2019:417–427.
9. Kendyukhov A.V., Tolkachev D.O. Ispol'zovanie metoda glavnykh komponent dlya otsenki konkurentosposobnosti mashinostroitel'nykh predpriyatii. *Marketing i menedzhment innovatsii*. 2013:219–227.
10. Petroni A., Braglia M. Vendor Selection Using Principal Component Analysis. *The Journal of Supply Chain Management*. 2000:63–69.
11. Sarkodie S.A., Adom P.K. Determinants of energy consumption in Kenya: A NIPALS approach. *Energy*. 2018:696–705.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Грибанова Екатерина Борисовна, кандидат технических наук, доцент Томского государственного университета систем управления и радиоэлектроники, Томск, Российская Федерация.
e-mail: geb@asu.tusur.ru
ORCID: [0000-0001-6499-5893](https://orcid.org/0000-0001-6499-5893)

Ekaterina Borisovna Gribanova, Candidate of Sciences in Engineering, Associate Professor, Department of Automated Control Systems, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation

Саулин Вячеслав Валерьевич, магистрант, кафедра автоматизированных систем управления, Томский государственный университет систем управления и радиоэлектроники, Томск, Российская Федерация.
e-mail: sv4774@vtomske.ru

Vjacheslav V. Saulin, Master student, Department of Automated Control Systems, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation