

## ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ В ПРОГНОЗИРОВАНИИ ПРЕДАВАРИЙНЫХ СИТУАЦИЙ В СИСТЕМАХ ТЕПЛОСНАБЖЕНИЯ

*А. А. Ахvaeв, В. Ф. Шуриев*

*Астраханский государственный технический университет,  
Астрахань, Российская Федерация*

Решение задачи прогнозирования в системах, характеризующихся проблемой подбора традиционного алгоритма для ее описания, сводится к технологии машинного обучения. В контексте прогнозирования аварий в системах теплоснабжения эта технология является наиболее эффективной. Осуществление прогноза сводится к задаче восстановления функции в общем контенте обучения с учителем. Из имеющихся инструментов машинного обучения следует использовать градиентный бустинг. Он работает по следующему принципу: на первых итерациях используются слабые алгоритмы, затем наращивается ансамбль постепенными улучшениями тех участков данных, где предыдущие модели «не доработали». Но при построении следующей простой модели она строится не просто на перевзвешенных наблюдениях, а так, чтобы лучшим образом приближать общий градиент целевой функции. Градиентный бустинг является одним из эффективных алгоритмов прогнозирования, и точность прогноза зависит от правильных входных данных (обучающая выборка). Исследуемая предметная область, а именно исследование аварийных ситуаций на тепловых сетях, имеет достаточные объемы накопленных данных для применения бустинга в качестве основного инструмента для прогноза.

**Ключевые слова:** машинное обучение, бустинг, прогнозирование, мониторинг, функция потерь.

**Для цитирования:** *Ахvaeв А. А., Шуриев В. Ф.* Применение машинного обучения в прогнозировании предаварийных ситуаций в системах теплоснабжения // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2020. № 3. С. 74–81. DOI: 10.24143/2072-9502-2020-3-74-81.

### **Введение**

Применение машинного обучения приобретает актуальность в решении задач прогнозирования. Это связано с тем, что в ряде некоторых задач, таких как предсказание экономических и социальных процессов или осуществление прогноза погоды, зачастую становится сложно, а иногда и невозможно применить явный алгоритм. Упомянутые ранее процессы представляют собой сложные системы, в которых невозможно учесть все существующие факторы, влияющие на состояние или поведение системы. Разрешить эту проблему можно путем выявления некоей функциональной зависимости, выстраивая функцию, которую необходимо приближать к реальной функции сложной системы. Приближение осуществляется обучением на имеющихся парах входных и выходных параметров. Одним из эффективных методов обучения является бустинг, основной принцип которого заключается в использовании «слабых» алгоритмов прогнозирования в совокупности. В этот принцип заложена идея совместного использования слабых алгоритмов, что может дать более качественный прогноз в сравнении с использованием их по отдельности.

### **Прогнозирование аварийных ситуаций**

**Постановка задачи.** В основе данного исследования лежит задача прогнозирования аварийных ситуаций в системах теплоснабжения. Как показывают исследования в этой области, чаще всего для этого применяют экспертные системы, которые основываются на опыте экспертов. Такие системы имеют преимущество, если применена эффективная методика формирования экспертного мнения и подобрана «правильная» группа экспертов. Под «правильными» экспертами понимаются те эксперты, опыт которых приемлем для формирования экспертного мне-

ния. Поэтому целесообразно применить машинное обучение в качестве инструмента прогнозирования аварийных ситуаций в системах теплоснабжения. В качестве инструмента машинного обучения рассматривается градиентный бустинг [1–4].

Проблема аварийных ситуаций в системах теплоснабжения остается актуальной и на данный момент. В первую очередь это связано с естественными процессами изнашивания трубопроводов или воздействия внешней и внутренней среды. На рис. 1 показаны основные характеристики трубопровода, которые следует принять во внимание в рамках задачи прогнозирования.

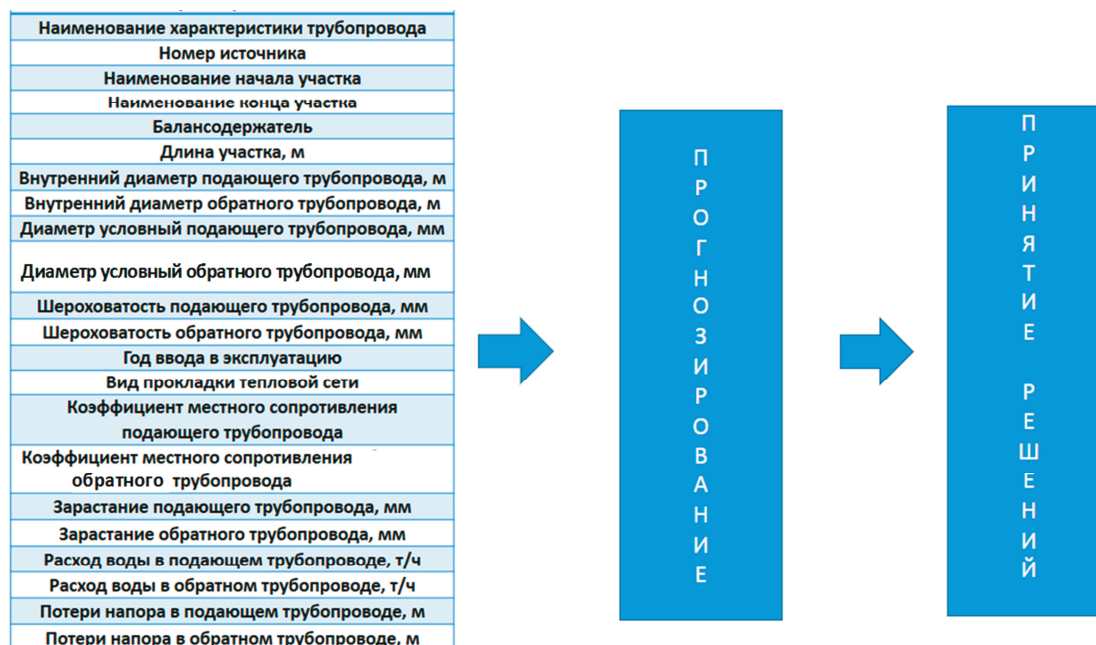


Рис. 1. Параметры трубопроводов, формирующие обучающую выборку

Данные параметры можно классифицировать по следующему принципу:

- группа характеристик источника теплоснабжения;
- группа характеристик трубопровода;
- группа характеристик химической очистки;
- группа характеристик факторов внешней среды.

Среди представленных ранее групп характеристик можно выделить третью группу, т. к. наличие химической очистки в системе теплоснабжения значительно уменьшает риск износа теплотрасс.

Осуществление прогноза сводится к задаче восстановления функции в общем контенте обучения с учителем. Следует использовать градиентный бустинг. Он работает по следующему принципу: на первых итерациях используются слабые алгоритмы, затем наращиваем ансамбль постепенными улучшениями тех участков данных, где предыдущие модели «не доработали». Но при построении следующей простой модели она строится не просто на перевзвешенных наблюдениях, а так, чтобы лучшим образом приближать общий градиент целевой функции [5, 6].

Что касается модели прогнозирования предаварийных ситуаций, то ее можно представить следующим образом:

$$M = \{MP, OB, MA, MFP, VI\},$$

где MP – входное множество параметров; MA – множество базовых алгоритмов, используемых в градиентном бустинге; OB – обучающая выборка; MFP – множество функций потерь; VI – выходная информация.

Общая концепция данной задачи прогнозирования, с учетом применения машинного обучения, включает в себя следующие процессы (рис. 2):

- инициализация и нормирование входных параметров;

- осуществление прогноза в заданный временной интервал;
- визуализация результата прогнозирования;
- поддержка принятия решений.

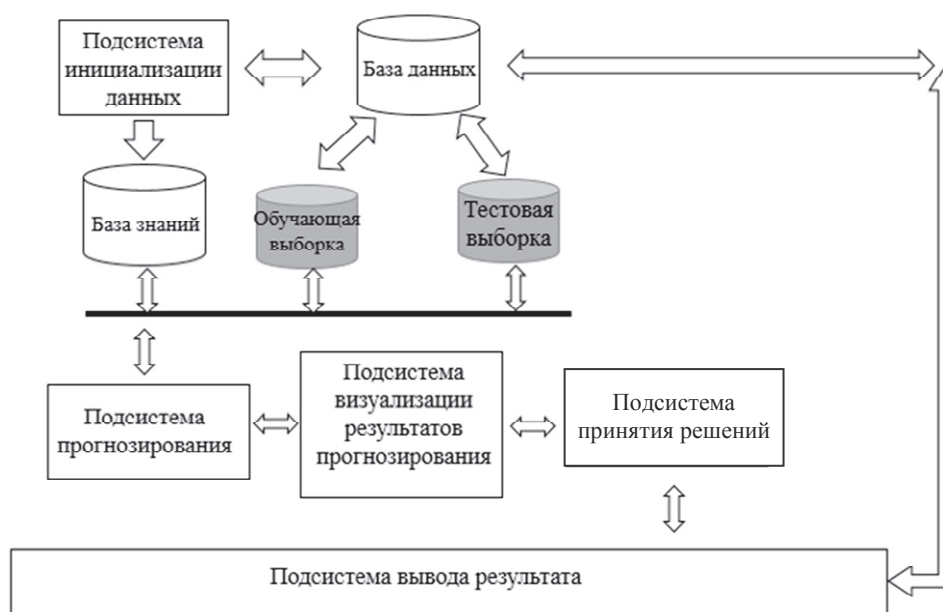


Рис. 2. Концептуальная схема прогнозирования и принятия решений

В данной концепции следует уделить должное внимание процессу формирования базы данных, которая включает в себя обучающую и тестовую выборку. Правильное формирование этих множеств определяет результат прогнозирования.

**Метод градиентного бустинга.** Решение проблемы прогнозирования осуществляется различными методами, в том числе и методом машинного обучения. Использование нейронных сетей или дерева решений позволяет на наборе пар входных и выходных параметров моделировать процессы. Поэтому решение проблемы прогнозирования сводится к использованию «хороших» входных и выходных параметров. Чем больше параметров можно использовать на входе, тем лучше будет происходить обучение модели, которое основывается на градиентном бустинге [7–13].

Градиентный бустинг может решать две задачи: задачу классификации и нахождения целевой функции. В нашем случае это второй вариант. Нам необходимо осуществить восстановление функции зависимости входных и выходных данных, указанных ранее. Имеется набор признаков  $x$  и результирующих переменных  $y$ , которые берутся из множества входных параметров МП. Эти наборы формируют пару множеств, на котором будет обучаться алгоритм (ОВ), восстанавливая функциональную зависимость  $y = f(x)$ :

$$ОВ = \{(x_i, y_i)\}_{i=1..n},$$

где  $x_i$  – набор признаков;  $y_i$  – результирующие переменные из МП.

Восстановление происходит путем приближения  $\hat{f}(x)$ , но для определения наиболее приемлемого приближения необходимо ввести функцию потерь  $L(y, f)$ . Задача сводится к минимизации функции потерь:

$$y \approx \hat{f}(x);$$

$$\hat{f}(x) = \arg \min L(y, f(x)).$$

Следует отметить, что прогнозирование происходит на основе имеющихся в распоряжении данных, и тогда

$$\hat{f}(x) = \arg \min_{f(x)} = E_{x,y} [L(y, f(x))].$$

Что касается подбора функции  $f(x)$  для решения задачи, следует ограничить поиск семейством функций  $f(x, \theta)$ ,  $\theta \in \mathbb{R}^d$  с параметрами. Такой подход значительно упрощает решение и сводится к решаемой минимизации параметров:

$$\hat{f}(x) = f(x, \hat{\theta});$$

$$\hat{\theta} = \arg \min_{\theta} E_{x,y} [L(y, f(x, \theta))].$$

Чтобы получить наиболее оптимальные параметры  $\hat{\theta}$ , их необходимо будет приближать итеративно. Приближение и функция потерь нашей модели с учетом выполнения  $M$  раз (числа итераций) выглядит следующим образом:

$$\hat{\theta} = \sum_{i=1}^M \hat{\theta}_i;$$

$$L_0(\hat{\theta}) = \sum_{i=1}^N L(y_i, f(x_i, \hat{\theta})).$$

В качестве подходящего рабочего алгоритма для решения данной задачи следует использовать градиентный спуск. Суть его заключается в том, чтобы к градиенту  $\nabla L_0(\hat{\theta})$  добавить итерационные оценки  $\hat{\theta}_i$ .

С учетом ранее обозначенных параметров алгоритм градиентного бустинга имеет следующий вид:

1. Инициализировать начальное приближение параметров  $\hat{\theta} = \hat{\theta}_0$ .
2. Для каждой итерации  $t = 1, \dots, M$  повторять:
  - а) посчитать градиент функции потерь  $\nabla L_0(\hat{\theta})$  при текущем приближении  $\hat{\theta}$

$$\nabla L_0(\hat{\theta}) = \left[ \frac{\partial L(y, f(x, \theta))}{\partial \theta} \right]_{\theta=\hat{\theta}} ;$$

б) задать текущее итеративное приближение  $\hat{\theta}_t$  на основе посчитанного градиента  $\hat{\theta}_t \leftarrow -\nabla L_0(\hat{\theta})$ ;

- в) обновить приближение параметров  $\hat{\theta}$ :

$$\hat{\theta} \leftarrow \hat{\theta} + \hat{\theta}_t \sum_{i=0}^t \theta_i.$$

3. Сохранить итоговое приближение  $\hat{\theta}$ :

$$\hat{\theta} = \sum_{i=0}^M \hat{\theta}_i.$$

**Описание параметров градиентного бустинга.** Чтобы определить, какие составляющие необходимы для решения задачи прогнозирования, необходимо перейти к оптимизации в функциональном пространстве. Тогда приближение  $\hat{f}(x)$  следует искать в виде самих функций. Чтобы решить задачу прогнозирования, необходимо ограничивать поиск группой функций  $\hat{f}(x) = h(x, \theta)$ .

Таким образом, для работы алгоритма необходимо иметь:

- набор данных  $OB = \{(x_i, y_i)\}_{i=1, \dots, n}$ ;
- число итераций  $M$ ;
- выбор функции потерь  $L(y, f)$  с выписанным градиентом;
- выбор семейства функций базовых алгоритмов  $MA = h(x, \theta)$ , с процедурой их обучения;
- дополнительные гиперпараметры  $h(x, \theta)$ , например глубина дерева у деревьев решений.

Набор данных обучающей выборки  $OB$  представляет собой набор параметров участка тепловых сетей, а также параметр, определяющий возникновение аварии. Что касается выбора семейства функций базовых алгоритмов, для прогнозирования могут быть применены следующие алгоритмы:

- линейная регрессия;
- логистическая регрессия;
- дерево решений.

Применение градиентного бустинга для прогнозирования аварийных ситуаций на участках тепловых сетей позволяет на ранних этапах исследования решить задачу классификатора (классификация: авария/не авария) и (при более подробных данных  $y_i$ ) задачу восстановления целевой функции.

### Заключение

Таким образом, в рамках поставленной задачи была определена общая концепция прогнозирования, которая включает в себя немаловажные для машинного обучения этапы. Разработана модель, учитывающая необходимые составляющие алгоритма прогнозирования популяционно-го процесса. В качестве подхода к прогнозированию был выбран бустинг, основанный на градиентном спуске. В данном случае градиентный бустинг рассматривается в контексте решения задачи регрессии, т. е. осуществления поиска функциональной зависимости между парой множеств входных и выходных данных.

### СПИСОК ЛИТЕРАТУРЫ

1. Квятковская И. Ю., Фам К. Х. Система показателей оценки качества телекоммуникационных услуг и метод их оценки // Вестн. Астрахан. гос. техн. ун-та. Сер.: Управление, вычислительная техника и информатика. 2013. № 2. С. 98–103.
2. Бялецкая Е. М., Квятковская И. Ю., Шуришев В. Ф. Формирование набора показателей для оценки качества управления жилыми домами // Вестн. Астрахан. гос. техн. ун-та. Сер.: Управление, вычислительная техника и информатика. 2011. № 2. С. 143–149.
3. Шуришев В. Ф., Демич Н. В. Алгоритм комплексного эволюционного метода поиска решений задачи синтеза состава смесей холодильных агентов // Юж.-Рос. вестн. геологии, географии и глобальной энергии. 2006. № 11. С. 65–68.
4. Лупян Е. А., Бурцев М. А., Балашов И. В., Барталев С. А., Ефремов В. Ю., Кашицкий А. В., Мазуров А. А., Матвеев А. М., Суднева О. А., Сычугов И. Г., Толпин В. А., Уваров И. А. Центр коллективного пользования системами архивации, обработки и анализа спутниковых данных ИКИ РАН для решения задач изучения и мониторинга окружающей среды // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 5. С. 263–284.
5. Ажогин Е. Ю., Квятковская И. Ю. Корпоративная база знаний как инструмент обеспечения бесперебойной работы информационных систем // Вестн. Астрахан. гос. техн. ун-та. Сер.: Управление, вычислительная техника и информатика. 2010. № 2. С. 7–14.
6. Шуришев В. Ф., Буй Л. В. Модель и алгоритм рационального выбора сканирующих приемников и трансиверов // Вестн. Сарат. гос. техн. ун-та. 2015. Т. 3. № 1 (80). С. 166–175.

7. Шуршев В. Ф., Кочкин Г. А., Кочкина В. Р. Модель системы поддержки принятия решений на основе рассуждений по прецедентам // Вестн. Астрахан. гос. техн. ун-та. Сер.: Управление, вычислительная техника и информатика. 2013. № 2. С. 175–183.

8. Квятковская И. Ю., Чертина Е. В., Белов С. В. Использование системы дополнительного профессионального образования в процессе повышения квалификации ИТ-специалистов: опыт Астраханского государственного технического университета // Информационно-коммуникационные технологии в науке, производстве и образовании ICIT-2017: сб. ст. Междунар. науч.-практ. конф. (Саратов, 21–22 сентября 2017 г.). Воронеж: Науч. книга, 2017. С. 411–419.

9. Шуршев В. Ф., Умеров А. Н. Моделирование процесса принятия решений при идентификации режимов течения смесей холодильных агентов // Вестн. Кузбас. гос. техн. ун-та. 2005. № 5 (50). С. 27–29.

10. Ахvaeв А. А., Шуршев В. Ф. Распознавание биообъектов в грунте с использованием радиофизических свойств // Передовые научно-технические и социально-гуманитарные проекты в современной науке: сб. ст. II Междунар. науч.-практ. конф. (Москва, 15 июня 2018 г.). М.: Науч.-издат. центр «Актуальность.РФ», 2018. С. 69–71.

11. Шуршев В. Ф., Бялецкая Е. М. Алгоритм экспертной оценки показателей контрольно-измерительной системы жилого объекта // Вестн. Астрахан. гос. техн. ун-та. Сер.: Управление, вычислительная техника и информатика. 2010. № 2. С. 117–121.

12. Исаев А. С., Киселев В. В., Калашиников Е. Н., Плешиков В. П. и др. Геоинформационные системы в прогнозировании и контроле массового размножения лесных насекомых // Лесоведение. 1999. № 5. С. 15–23.

13. Коэльо Л. П., Ричард В. Построение систем машинного обучения на языке Python. М.: ДМК Пресс, 2016. 302 с.

Статья поступила в редакцию 23.04.2020

### ИНФОРМАЦИЯ ОБ АВТОРАХ

**Ахvaeв Арман Амангельдыевич** – Россия, 414056, Астрахань; Астраханский государственный технический университет; аспирант кафедры автоматизированных систем обработки информации и управления; astu.arm@yandex.ru.

**Шуршев Валерий Фёдорович** – Россия, 414056, Астрахань; Астраханский государственный технический университет; д-р техн. наук, профессор; профессор кафедры прикладной информатики; v.shurshev@mail.ru.



## APPLICATION OF MACHINE TRAINING FOR FORECASTING EMERGENCIES IN HEAT SUPPLY SYSTEMS

*A. A. Akhvaev, V. F. Shurshev*

*Astrakhan State Technical University,  
Astrakhan, Russian Federation*

**Abstract.** The article touches upon the forecasting problem, the solution of which in systems characterized by selecting a traditional algorithm for its description is reduced to machine learning technology. In the context of predicting emergencies in heat supply systems this technology is the most effective. Carrying out the forecast is reduced to the problem of restoring the function in the general content of training by the teacher. Of the available machine learning tools, gradient boosting should be used. It works according to the following principle: at the first iterations the weak algorithms are used, then there increases the ensemble by gradual improvements of those data sections where the previous models have not been finalized. But when constructing the next simple model, it is built not just on reweighted observations, but in such a way as to better approximate the overall gradient of the objective function. Gradient boosting is one of the effective forecasting algorithms and the accuracy of the forecast

depends on the correct input data (training sample). The subject area under study, namely the study of emergency situations on heating networks, has sufficient accumulated data to use boosting as the main tool for forecasting.

**Key words:** machine learning, boosting, forecasting, monitoring, loss function.

**For citation:** Akhvaev A. A., Shurshev V. F. Application of machine training for forecasting emergencies in heat supply systems. *Vestnik of Astrakhan State Technical University. Series: Management, Computer Science and Informatics*. 2020;3:74-81. (In Russ.) DOI: 10.24143/2072-9502-2020-3-74-81.

#### REFERENCES

1. Kviatkovskaia I. Iu., Fam K. Kh. Sistema pokazatelei otsenki kachestva telekommunikatsionnykh uslug i metod ikh otsenki [System of indicators for assessing telecommunication service quality and method of their assessment]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naiia tekhnika i informatika*, 2013, no. 2, pp. 98-103.
2. Bialetskaia E. M., Kviatkovskaia I. Iu., Shurshev V. F. Formirovanie nabora pokazatelei dlia otsenki kachestva upravleniia zhilyimi domami [Formation of set of indicators for assessing quality of housing management]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naiia tekhnika i informatika*, 2011, no. 2, pp. 143-149.
3. Shurshev V. F., Demich N. V. Algoritm kompleksnogo evoliutsionnogo metoda poiska reshenii zadachi sinteza sostava smesei kholodil'nykh agentov [Algorithm of complex evolutionary method for finding solutions to problem of synthesis of refrigerant mixtures composition]. *Iuzhno-Rossiiskii vestnik geologii, geografii i global'noi energii*, 2006, no. 11, pp. 65-68.
4. Lupian E. A., Burtsev M. A., Balashov I. V., Bartalev S. A., Efremov V. Iu., Kashnitskii A. V., Mazurov A. A., Matveev A. M., Sudneva O. A., Sychugov I. G., Tolpin V. A., Uvarov I. A. Tsentr kollektivnogo pol'zovaniia sistemami arkhivatsii, obrabotki i analiza sputnikovnykh dannykh IKI RAN dlia resheniia zadach izuchenii i monitoringa okruzhaiushchei sredy [Center for collective use of systems for archiving, processing and analysis of satellite data of IKI RAS for solving problems of studying and monitoring environment]. *Sovremennye problemy distantsionnogo zondirovaniia Zemli iz kosmosa*, 2015, vol. 12, no. 5, pp. 263-284.
5. Azhogin E. Iu., Kviatkovskaia I. Iu. Korporativnaia baza znaniia kak instrument obespecheniia bespreboinoi raboty informatsionnykh sistem [Corporate knowledge base as tool for ensuring smooth operation of information systems]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naiia tekhnika i informatika*, 2010, no. 2, pp. 7-14.
6. Shurshev V. F., Bui L. V. Model' i algoritm ratsional'nogo vybora skaniruiushchikh priemnikov i transiverov [Model and algorithm for rational selection of scanning receivers and transceivers]. *Vestnik Saratovskogo gosudarstvennogo tekhnicheskogo universiteta*, 2015, vol. 3, no. 1 (80), pp. 166-175.
7. Shurshev V. F., Kochkin G. A., Kochkina V. R. Model' sistemy podderzhki priniatiia reshenii na osnove rassuzhdenii po pretseidentam [Decision support system model based on case studies]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naiia tekhnika i informatika*, 2013, no. 2, pp. 175-183.
8. Kviatkovskaia I. Iu., Chertina E. V., Belov S. V. Ispol'zovanie sistemy dopolnitel'nogo professional'nogo obrazovaniia v protsesse povysheniia kvalifikatsii IT-spetsialistov: opyt Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta [Using system of additional professional education in process of advanced training of IT specialists: experience of Astrakhan State Technical University]. *Informatsionno-kommunikatsionnye tekhnologii v nauke, proizvodstve i obrazovanii ICIT-2017: sbornik statei Mezhdunarodnoi nauchno-prakticheskoi konferentsii (Saratov, 21–22 sentiabria 2017 g.)*. Voronezh, Nauchnaia kniga Publ., 2017. Pp. 411-419.
9. Shurshev V. F., Umerov A. N. Modelirovanie protsessa priniatiia reshenii pri identifikatsii rezhimov techeniia smesei kholodil'nykh agentov [Modeling decision-making process in identifying flow patterns of mixtures of refrigerants]. *Vestnik Kuzbasskogo gosudarstvennogo tekhnicheskogo universiteta*, 2005, no. 5 (50), pp. 27-29.
10. Akhvaev A. A., Shurshev V. F. Raspoznavanie bioob"ektov v grunte s ispol'zovaniem radiofizicheskikh svoistv [Recognition of bioobjects in soil using radiophysical properties]. *Peredovye nauchno-tekhnicheskie i sotsial'no-gumanitarnye proekty v sovremennoi nauke: sbornik statei II Mezhdunarodnoi nauchno-prakticheskoi konferentsii (Moskva, 15 iunია 2018 g.)*. Moscow, Nauchno-izdatel'skii tsentr «Aktual'nost'.RF», 2018. Pp. 69-71.
11. Shurshev V. F., Bialetskaia E. M. Algoritm ekspertnoi otsenki pokazatelei kontrol'no-izmeritel'noi sistemy zhilogo ob"ekta [Algorithm for expert evaluation of indicators of control and measuring system of residential facility]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naiia tekhnika i informatika*, 2010, no. 2, pp. 117-121.

12. Isaev A. S., Kiselev V. V., Kalashnikov E. N., Pleshikov V. P. i dr. Geoinformatsionnye sistemy v prognozirovanii i kontrole massovogo razmnozheniia lesnykh nasekomykh [Geographic information systems in predicting and controlling mass reproduction of forest insects]. *Lesovedenie*, 1999, no. 5, pp. 15-23.

13. Koel'o L. P., Richard V. *Postroenie sistem mashinnogo obucheniia na iazyke Python* [Building machine learning systems in Python programming language]. Moscow, DMK Press, 2016. 302 p.

The article submitted to the editors 23.04.2020

### ***INFORMATION ABOUT THE AUTHORS***

***Akhvaev Arman Amangeldyevich*** – Russia, 414056, Astrakhan; Astrakhan State Technical University; Postgraduated Student of Department of Automated Informayion Processing and Control Systems; astu.arm@yandex.ru.

***Shurshev Valeriy Fyodorovich*** – Russia, 414056, Astrakhan; Astrakhan State Technical University; Doctor of Technical Sciences, Professor; Professor of the Department of Applied Informatics; v.shurshev@mail.ru.

