









МРНТИ 70.17.55

<https://doi.org/10.32523/2616-7263-2024-149-4-202-214>

Научная статья

Микроклимат параметрлерін кластеризациялау: әдістер мен математикалық сипаттамалар

Дауренбаева Н.А.¹, Нұрланұлы А.², Атымтаева Л.Б.³, Быков А.А.¹,
Ергалиев Д.С.², Әбдірашев Ө.К.*⁴

¹Халықаралық ақпараттық технологиялар университеті

²Азаматтық авиация академиясы

³Сүлеймен Демирел атындағы университеті

⁴Л.Н.Гумилев атындағы Еуразия ұлттық университеті

(E-mail: ¹omeke_92@mail.ru)

Аңдатпа. Қазіргі уақытта микроклимат параметрлерін бақылау және талдау өндіріс, экологиялық зерттеулер және ғимараттарды басқаруды қоса алғанда, әртүрлі салаларда маңызды рөл атқарады. Микроклимат параметрлері – температура, ылғалдылық, ауа қысымы және басқа физикалық көрсеткіштер-өндіріс процестерінің тиімділігі мен өнім сапасын анықтауда үлкен маңызға ие. Үлкен көлемдегі деректермен жұмыс істеу кезінде, әсіресе көптеген өлшемдер мен параметрлері бар микроклиматтық деректерді талдау кезінде кластерлеу әдістері ерекше маңызға ие болады. Кластерлеу - бұл олардың арасындағы ұқсастықтар мен айырмашылықтарға негізделген деректерді топтарға немесе кластерлерге бөлу процесі. K-Means және DBSCAN (Density-Based Spatial Clustering of Applications with Noise) сияқты әдістер деректердегі ауытқуларды анықтауға және жүйенің бұзылуын немесе бұзылуын анықтауға көмектеседі.

Бұл мақалада микроклимат параметрлерін кластеризациялау үшін қолданылатын әдістер мен олардың математикалық модельдері қарастырылады. DBSCAN әдісінің ерекшеліктері мен артықшылықтары, сондай-ақ оның микроклиматтық деректердегі кластерлер мен аномалияларды анықтаудағы тиімділігі талданады. Мақалада DBSCAN әдісін қолданудың нақты мысалдары келтіріледі, оның көмегімен микроклимат параметрлерін тиімді талдау мен басқару үшін математикалық формулалар мен есептеулер ұсынылады.

Кластеризация әдістерінің математикалық негіздері мен олардың микроклимат параметрлерін талдаудағы рөлі туралы түсініктерді тереңдету арқылы, біз микроклиматтық жүйелерді басқару мен оңтайландыруда жаңа тәсілдер мен шешімдерді ұсынуға ұмтыламыз.

Түйін сөздер: Микроклимат, кластерлеу, DBSCAN, VAE, K-means, аномалияларды анықтау, машиналық оқыту, микроклиматтық параметрлер, тығыздыққа негізделген кластерлеу, деректерді талдау, ақауларды анықтау.

Түсті 06.11.2024. Жөнделді 06.11.2024. Мақұлданды 13.11.2024. Онлайн қолжетімді 31.12.2024

*хат хабар үшін авторы

Кіріспе

Микроклимат параметрлерін бақылау мен талдау қазіргі таңда өндіріс, экология және тұрмыстық жағдайлар сияқты көптеген салаларда маңызды рөл атқарады. Микроклиматтың негізгі параметрлері, мысалы, температура, ылғалдылық, ауа қысымы және т.б., қоршаған ортаның жай-күйін және жүйелердің жұмысын тиімді басқаруға мүмкіндік береді. Бұл параметрлерді дұрыс талдау үшін машиналық оқыту әдістері, соның ішінде кластеризация алгоритмдері кеңінен қолданылады.

Мақсат: Бұл жұмыстың мақсаты – микроклимат параметрлерін талдау үшін K-means және DBSCAN кластеризация әдістерін қолдану арқылы жүйедегі аномалияларды анықтау және алынған нәтижелер негізінде микроклиматты тиімді басқару жолдарын ұсыну.

Талдаудың басында микроклимат деректері өңдеу кезеңінен өтеді. Өткізіп алған мәндерді толтыру үшін деректердің үздіксіздігін қамтамасыз ете отырып, өткізіп алған мәндерді соңғы белгілі мәнмен толтыратын тікелей сүзу әдісі қолданылады. Z-бағалау орташадан үш стандартты ауытқудан асатын шығарындыларды жою үшін қолданылады:

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

мұндағы X – белгінің мәні, μ – белгінің орташа мәні, σ – стандартты ауытқу.

Сондай-ақ, деректер RobustScaler (квартильаралық масштабқа сәйкес масштабтау) және StandardScaler (орташа мәнді алып тастау және бірлік дисперсияға келтіру) [2] көмегімен масштабталады, бұл шығарындылардың әсерін азайтады және деректерді стандартты қалыпты үлестіруге әкеледі.

Болашақта деректердің өлшемін азайту және олардың жасырын көріністерін алу үшін вариациялық автоэнкодер қолданылады. Кодер ReLU белсендіру функциясы бар бірнеше жасырын қабаттардан және жасырын кеңістіктегі қалыпты үлестіру параметрлерін білдіретін Z_{mean} және $Z_{\text{log var}}$ шығыс қабаттарынан тұрады. Декодер кодердің архитектурасын бейнелейді және жасырын көріністерді бастапқы белгілер кеңістігіне айналдырады. Оқудағы дифференциалдылықты қамтамасыз ету үшін градиенттерді есептеу мүмкіндігін сақтай отырып, кездейсоқ айнымалыларды модельдеуге мүмкіндік беретін репараметризация қолданылады:

$$Z = Z_{\text{mean}} + e^{0,5 \cdot Z_{\text{log-var}}} \cdot \epsilon \quad (2)$$

мұндағы ϵ – стандартты қалыпты үлестірімнің кездейсоқ шамасы.

Шығын функциясы қайта құру қатесі мен KL дивергенциясын қамтиды.

Оқу процесінде VAE Моделі тұрақты конвергенцияны қамтамасыз ету үшін Adam төмен оқу жылдамдығының оптимизаторын ($\alpha=10^{-4}$) қолдана отырып құрастырылады. Оқыту валидация үшін деректердің бір бөлігін пайдалана отырып, қалыпқа келтірілген

деректерде жүргізілді. VAE жаттығуынан кейін кластерлеу үшін қолданылатын деректердің жасырын көріністері (z_{mean}) алынады. Жасырын көріністерде кластерлеу үшін K-means $k=2$ алгоритмі қолданылып, деректерді екі негізгі кластерге бөлді.

Деректердегі ауытқуларды анықтау үшін бастапқы деректер X реконструкциясын алу үшін оқытылған VAE арқылы өткізілді, содан кейін әрбір үлгі үшін орташа квадраттық MSE қатесі есептелді. Шекті мәннен жоғары қате үлгілері (99,95-ші процентиль) ауытқулар ретінде анықталды және бөлек кластерге тағайындалды.

Кідіріс кеңістігінің жоғары өлшемді деректерін визуализациялау мақсатында t-SNE (t-distributed Stochastic Neighbor Embedding) көмегімен сызықтық емес өлшемді азайту пайдаланылды, бұл деректерді екі өлшемді кеңістікте көрсетуге мүмкіндік береді [3].

K-Means алгоритмімен біріктірілген вариациялық автоэнкодерді пайдалану микроклимат деректерін тиімді кластерлеуге және сенсорлардың дұрыс жұмыс істемеуін немесе дұрыс емес өлшемдерді көрсетуі мүмкін ауытқуларды анықтауға мүмкіндік береді. Бұл тәсіл DBSCAN сияқты дәстүрлі әдістерді толықтырады және күрделі және жоғары өлшемді деректерді талдауға икемді мүмкіндіктер береді.

Қосымша зерттеулер VAE моделінің гиперпараметрлерін оңтайландыруға, нейрондық желілердің басқа архитектураларын және мүмкіндіктерін зерттеуге мүмкіндік береді. Нәтижелерді жақсарту алатын Isolation Forest (оқшауланған орман) немесе One-Class SVM (Support Vector Machine) сияқты аномалияларды анықтаудың балама әдістерін қолдану. Сондай-ақ олардың себептерін анықтау және деректердің сапасы мен микроклиматты бақылау жүйесінің сенімділігін жақсарту стратегияларын әзірлеу үшін анықталған ауытқуларға егжей-тегжейлі талдау жүргізу маңызды.

Ұсынылған нәтижелер микроклимат параметрлерін кластерлеу үшін терең оқыту әдістерін қолдану деректерді дәлірек және тиімді талдауға, жасырын құрылымдарды анықтауға және ауытқуларды анықтауға мүмкіндік беретінін көрсетеді. Бұл микроклиматты бақылау және басқару жүйелерінің сенімділігін арттыруға, сондай-ақ өнім сапасы мен қоршаған орта жағдайларын жақсартуға ықпал етеді.

Әдіснама

DBSCAN әдісі деректерді кластерлеудің қуатты құралы болып табылады, әсіресе кластерлердің пішіні алдын ала белгісіз және ерікті болуы мүмкін кезде пайдалы. Оның математикалық негізі кластерлеу кезінде нүктелердің тығыздығына сүйенеді, бұл әдісті жоғары өлшемді деректерді талдау үшін тиімді етеді.

Автоматтандырылған кластерлеу: DBSCAN кластерлердің санын алдын ала анықтауды қажет етпейді, бұл оның икемділігін және қолдану аясын кеңейтеді. Бұл әдіс деректердің ішкі құрылымын анықтауға және кластерлерді автоматты түрде айқындауға мүмкіндік береді.

Шу мен шектен тыс деректерді өңдеу: DBSCAN кластерлердің сыртында орналасқан деректер нүктелерін шу немесе шектен тыс деректер ретінде анықтайды. Бұл мүмкіндік

деректер жинау жүйелеріндегі немесе микроклимат датчиктеріндегі ақауларды анықтау үшін аса маңызды болып табылады.

Кластердің пішіні мен тығыздығына бейімделу: DBSCAN әртүрлі пішіндегі және тығыздығы өзгеше кластерлерді анықтауға қабілетті. Бұл микроклиматтық деректерде кездесетін күрделі немесе сызықтық емес кластерлердің шекараларын анықтауда маңызды рөл атқарады.

DBSCAN әдісін қолдану процесі

Аномалияларды автоматты түрде анықтау: DBSCAN микроклимат параметрлерінің кеңістігіндегі деректер нүктелерінің тығыздығына негізделген кластерлерді анықтайды. Нүктелердің басқа нүктелерден қашықтығына немесе олардың төмен тығыздығына байланысты ешбір кластерге жатпайтын нүктелер анықталады.

Шығарушыларды анықтау: DBSCAN негізгі нүктелерді (кластердің ішіндегі нүктелер) шу нүктелерінен (кластерге жатпайтын нүктелер) ажыратады. Бұл деректерді жинау жүйесіндегі ақауларды немесе микроклиматтық сенсорлардың істен шығуын көрсетуі мүмкін ауытқуларды анықтау үшін пайдалы.

Кластердің пішіні мен тығыздығына бейімделу: DBSCAN әртүрлі пішіндегі және тығыздығы әртүрлі кластерлерді анықтауға қабілетті. Бұл микроклиматтық деректерде күрделі немесе сызықтық емес кластерлердің шекараларын тиімді анықтауға мүмкіндік береді.

Аномалияларды анықтау және нәтижелерді талдау: DBSCAN әдісін қолдану арқылы микроклимат деректеріндегі кластерлер мен шу нүктелерін айқындау арқылы жүйедегі ықтимал ақауларды анықтауға болады. Нәтижелерді талдау арқылы сенсорлардың істен шығуы, параметрлердің қалыптан тыс өзгеруі немесе дұрыс өлшенбеуі сияқты мәселелер анықталады.

Ақауларды түзету және жою: Анықталған ауытқулар негізінде жабдыққа техникалық қызмет көрсету, сенсорларды калибрлеу немесе жүйенің қалыпты жұмысын қалпына келтіру сияқты түзету шаралары қабылданады.

DBSCAN алгоритмі кластерлерді нүктелердің тығыздығына қарай анықтайды. Негізгі параметрлері:

'*eps*' (*epsilon* – ϵ): Әрбір нүкте үшін радиус, егер жақын орналасқан нүктелер осы радиус ішінде болса, олар бір кластерге жатқызылады.

'*min_{samples}*': Нүктенің негізгі нүкте ретінде классификациялануы үшін ерс радиусында кем дегенде болуға тиіс нүктелердің саны.

DBSCAN әдісінің математикалық тұжырымдамасы

Бізде Қысым – (P), Шық нүктесі – (D), Көмірқышқыл газ – CO₂ (C), TVOC(T) – ұшпа органикалық қосылыстар және т.б. сияқты параметрлері бар деректер жиынтығы бар делік. Біз кластерлеу үшін DBSCAN алгоритмін қолданамыз.

Қашықтықты есептеу:

Бізде (p_p, p_D, p_C, p_T) параметрлері бар және сәйкесінше екі $p \wedge q$ деректер нүктесі бар делік. Осы нүктелер арасындағы қашықтықты, мысалы, Евклидтік қашықтықты пайдаланып есептеуге болады:

$$\text{distance } (p, q) = \sqrt{(p_P - q_P)^2 + (p_D - q_D)^2 + (p_C - q_C)^2 + (p_T - q_T)^2} \quad (3)$$

Негізгі нүктелер және кластерлеу:

Қашықтықты есептеу: деректер жиынындағы нүктелердің әрбір жұбы арасындағы қашықтықты есептеңіз. Бұл барлық бірегей жұп нүктелердің тіркесімі бар массив құруды қажет етеді.

Қашықтықты есептеу және матрицаны бейнелеу

а) Жұптық қашықтықты есептеңіз

Деректер жиынындағы нүктелердің әрбір жұбы арасындағы қашықтықты есептеу үшін барлық бірегей жұптар үшін Евклидтік қашықтықты есептеу керек. Бұл қашықтық матрицасына әкеледі, мұнда әрбір жазба нүктелер мен арасындағы қашықтықты білдіреді.

Нәтижелер мен талқылау

Сізде n деректер нүктесі бар делік, қашықтық матрицасы матрица болып табылады, мұндағы:

$$D_{ij} = \sqrt{(p_iP - p_jP)^2 + (p_iD - p_jD)^2 + (p_iC - p_jC)^2 + (p_iT - p_jT)^2} \quad (4)$$

Мұндағы p_iP, p_iD, p_iC, p_iT i -ші нүктенің ерекшеліктері, p_jP, p_jD, p_jC, p_jT j -ші нүктенің ерекшеліктері. Мысалы, егер сіз келесідей үш ерекшелікті алсаңыз:

$$\begin{aligned} p_1 &= (p_1P, p_1D, p_1C, p_1T) \\ p_2 &= (p_2P, p_2D, p_2C, p_2T) \\ p_3 &= (p_3P, p_3D, p_3C, p_3T) \end{aligned} \quad (5)$$

D қашықтық матрицасы келесідей болады:

$$D = \begin{matrix} d(p_2, p_1) & d(p_1, p_2) & d(p_1, p_3) \\ d(p_3, p_1) & d(p_3, p_2) & d(p_2, p_3) \end{matrix} \quad (6)$$

Негізгі нүктелерді анықтау:

p әрбір нүкте үшін '*eps*' радиуста көршілерінің санын N_n есептеңіз. Егер бұл саннан үлкен немесе оған тең болса '*min samples*', негізгі нүкте ретінде белгілеңіз. яғни, негізгі нүкте шарттарды қанағаттандыруы керек.

$X = \{x_1, x_2, \dots, x_n\}$ d – өлшемді кеңістіктегі деректер нүктелерінің жиынтығы болсын және k жасалатын кластерлер саны болсын.

1. Центроидты инициализациялау: $\{c_1, c_2, \dots, c_k\}$. бастапқы центроидтар ретінде k нүктелерді X кездейсоқ таңдау.

2. Тапсырма Қадамы:

x_i әрбір деректер нүктесін ең жақын c_j центроидқа тағайындаңыз, мұндағы

$$j = \arg \min_j \sum_{x_i \in S_j} \|x_i - c_j\|^2 \quad (7)$$

3. Жаңарту Қадамы:

c_j әрбір центроидты оған тағайындалған барлық деректер нүктелерінің орташа мәні ретінде жаңартыңыз:

$$c_j = \frac{1}{S_j} \sum_{x_i \in S_j} x_i \quad c_j = \frac{1}{S_j} \sum_{x_i \in S_j} x_i \quad (8)$$

S_j – центроидқа берілген нүктелер жиыны

4. Итерация: тапсырманы қайталаңыз және конвергенцияға дейін қадамдарды жаңартыңыз.

Бізде екі ерекшелігі бар мәліметтер жиынтығы бар делік: Қысым (P) және Температура (T).

Деректерді үш кластерге $k = 3$ бөлгіміз келетін мағынаны алайық.

Баптандыру (Инициализация):

Бастапқы центроидтар ретінде кездейсоқ үш нүктені таңдаңыз.

Тапсырма қадамы:

Әрбір нүктені Евклид қашықтығына қарай ең жақын центроидқа тағайындаңыз.

Жаңарту қадамы:

Ағымдағы тапсырмалар негізінде центроидтарды қайта есептеңіз.

Итерация:

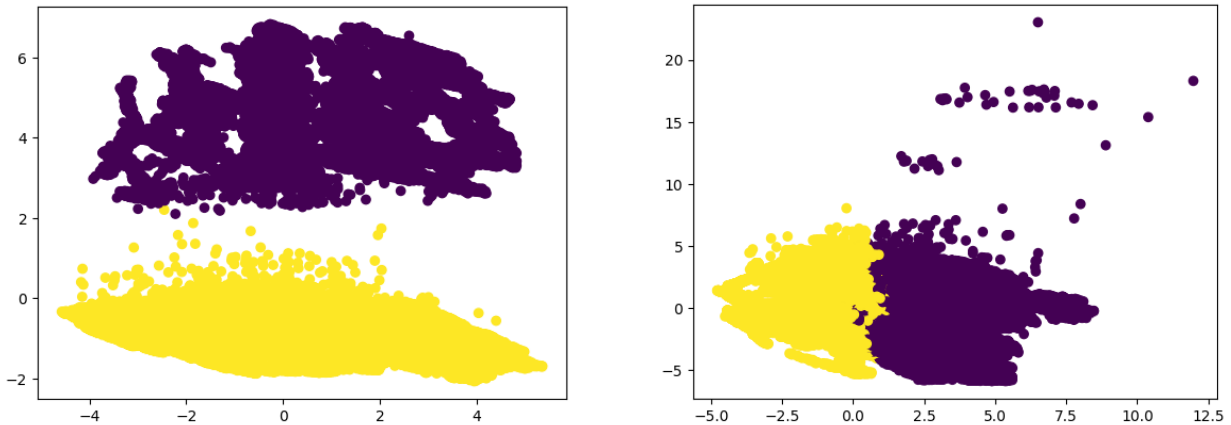
Тапсырманы қайталаңыз және конвергенция (центроидтар тұрақтанғанша) немесе итерациялардың максималды санына жеткенше қадамдарды жаңартыңыз.

Әдетте, DBSCAN-ға артықшылық беріледі: K-Means саны белгісіз немесе сызықтық емес шекаралары мен тығыздығы әр түрлі күрделі деректер құрылымдарымен жұмыс жасайтын мәліметтер жиынтығына арналған құралдар. Ол кластерлеуге икемді және автоматтандырылған тәсілді ұсынады, бұл оны нақты әлемдегі деректерді талдаудың қуатты құралына айналдырады.

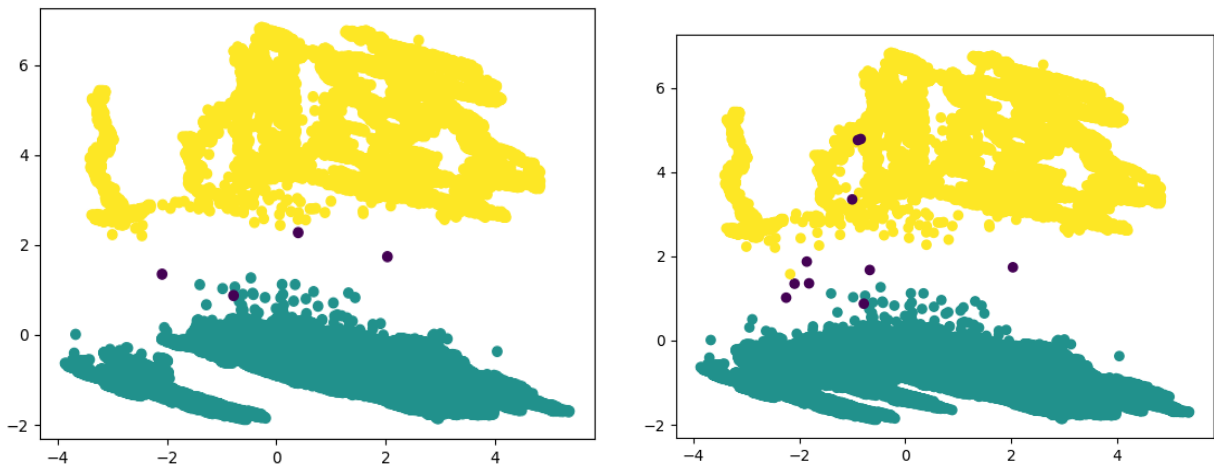
Бұл мақалада біз кластеризация әдістерінің математикалық формулаларын нақты мысалдар арқылы қарастырдық. Әрбір әдіс деректерді тиімді топтастыру үшін белгілі бір алгоритмдерді пайдаланады және нәтижелердің дәлдігі алынған кластерлердің сапасына байланысты болады. Жұмыста қолданылған әдістердің көмегімен талдау жасалып, деректердің ішкі құрылымын анықтауға және маңызды үлгілерді табуға мүмкіндік беретін нәтижелер алынды. Сонымен қатар, ұсынылған формулалар мен алгоритмдер мысалдың нәтижелілігін айқын көрсетті, бұл әдістерді болашақ зерттеулер мен нақты қолданбалы міндеттерде қолдануға болады.

Біріктірілген температура, ылғалдылық және шық нүктесі датчиктері бар note MCU негізіндегі әзірленген аппараттық кешен әрбір 10 секунд сайын тұрақты деректер жинауды қамтамасыз етеді, бұл үй-жайлар мен көшедегі климаттық жағдайларды нақты уақыт режимінде бақылауға мүмкіндік береді. Сонымен қатар, кешенге атмосфералық қысымды, CO₂ деңгейін, ұшпа органикалық қосылыстардың концентрациясын (TVOC), электр параметрлерін (қуат, ток, кернеу), жарық деңгейін және ультракүлгін сәулеленуді

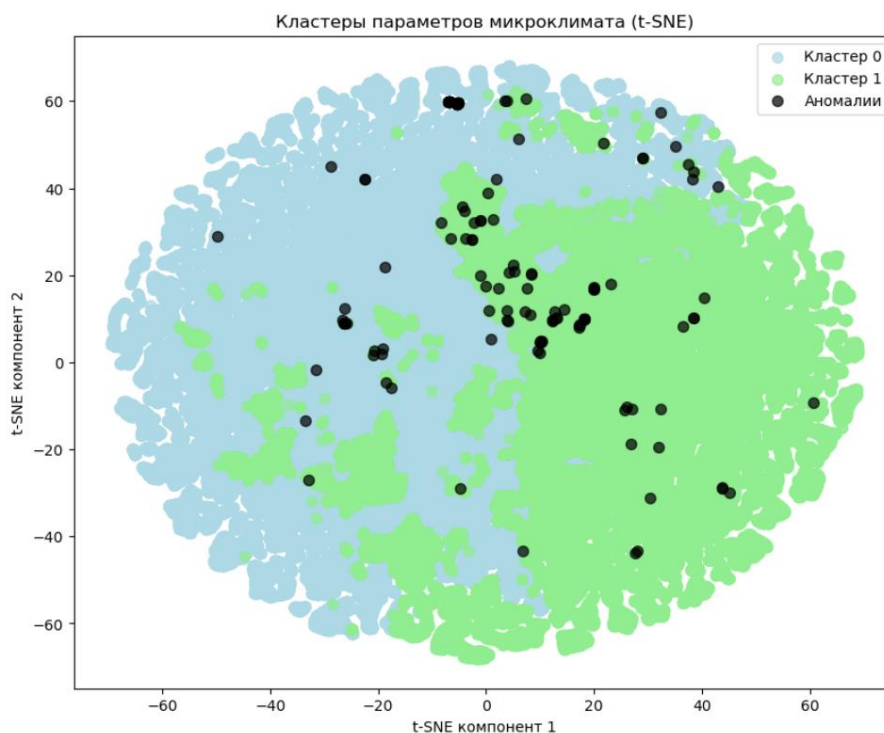
және соққыдан кейінгі белсенділік деректерін өлшеуге арналған датчиктер кіреді. Жиналған деректер климаттық факторлардың өмір сүру ыңғайлылығына және жылыту және кондиционерлеу жүйелерінің тиімділігіне әсерін одан әрі талдау үшін деректер жиынтығын құрайды.



1-сурет. Саяжай мен балабақшаға арналған K-means әдісі көмегімен кластерлеу



2-сурет. Балабақша мен саяжайға арналған DBSCAN әдісі көмегімен кластерлеу



3-сурет. K-Means алгоритмімен бірге VAE вариациялық автоэнкодерді қолдану

Жүйедегі ауытқуларды анықтау үшін машиналық оқыту және терең оқыту әдістері қолданылды: кластеризация (DBSCAN және K-means) (1, 2, 3 суреттерде Ручарт ортасында тұрғызылған графиктері көрсетілген). DBSCAN (шулы деректермен кеңістіктік кластеризациялауға негізделген алгоритм) – күрделі құрылымдары бар деректер жиынтықтарындағы ауытқуларды немесе аномалияларды анықтауға ыңғайлы, кеңінен қолданылатын кластеризация алгоритмі. K-means сияқты дәстүрлі әдістерден айырмашылығы, DBSCAN алдын ала кластерлер санын белгілеуді талап етпейді және кез келген пішін мен өлшемдегі кластерлерді автоматты түрде анықтай алады.

DBSCAN-нің басты артықшылықтарының бірі – кластерлерге жатпайтын ауытқулар мен шу нүктелерін анықтау мүмкіндігі. Алгоритм кластерлерді жоғары тығыздықтағы аймақтар ретінде анықтап, оларды төмен тығыздықтағы аймақтардан бөліп қарастырады. Бұл DBSCAN-ге тығыздықтағы өзгерістерге төзімді болып, әртүрлі пішіндер мен тығыздықтары бар деректер жиынтықтарын тиімді өңдеуге мүмкіндік береді.

Зерттеу барысында нақты уақыт режимінде микроклимат параметрлерін жинау және талдау үшін саяжай мен балабақшаға орнатылған аппараттық кешен пайдаланылды. CRISP-DM әдіснамасын қолдану деректерді өңдеу мен аномалияларды анықтауда жүйелі тәсілді қамтамасыз етті.

Микроклимат жүйесінің жұмыс режимдерін сипаттайтын негізгі кластерлерді K-means әдісі тиімді жіктегенімен, DBSCAN әдісі кластерлердің алдын ала белгіленген санынсыз, кез келген пішіндегі кластерлер мен шу нүктелерін анықтауда тиімділігін көрсетті.

Осылайша, DBSCAN әртүрлі жағдайларда микроклиматты сенімді және нақты басқару үшін оңтайлы таңдау екенін дәлелдеді, бұл оны ғимараттардағы микроклиматты басқару жүйелерінде қолданудың тиімділігін растайды. Зерттеу нәтижелері микроклиматтық бақылаудың сапасын арттыру және үй-жайларда жайлылық пен қауіпсіздікті жақсарту үшін заманауи машиналық оқыту әдістерін қолданудың маңыздылығын көрсетеді.

Қорытынды

Зерттеу барысында жүргізілген талдау көрсеткендей, DBSCAN алгоритмі микроклимат параметрлерін кластерлеу және аномалияларды анықтау барысында K-means әдісіне қарағанда бірқатар артықшылықтарға ие. DBSCAN алдын ала белгіленген кластерлер санын талап етпей, әртүрлі пішіндер мен тығыздықтағы кластерлерді дәл анықтайды. Сонымен қатар, шу нүктелерін тиімді түрде анықтау мүмкіндігі оны микроклиматты басқару жүйелерінде қолданудың тиімділігін арттырады.

K-means алгоритмі жүйенің жұмыс режимдерін жіктеуде жақсы нәтижелер көрсетті, бірақ DBSCAN әдісі аномалияларды анықтауда және деректердегі ауытқуларды зерттеуде айқын артықшылықтар ұсынады. Нәтижесінде, DBSCAN микроклиматты сенімді және дәл басқару үшін ең жақсы таңдау болып табылады.

Алдағы зерттеулер VAE моделінің гиперпараметрлерін оңтайландыруға, нейрондық желілердің әртүрлі архитектураларын қарастыруға, сондай-ақ аномалияларды анықтаудың альтернативті әдістері, мысалы, Isolation Forest немесе One-Class SVM, нәтижелерді одан әрі жақсартуға мүмкіндік береді. Аномалиялардың себептерін анықтау және микроклиматты бақылау жүйесінің сенімділігін арттыру үшін анықталған ауытқуларға терең талдау жүргізу маңызды.

Авторлардың қосқан үлесі.

Дауренбаева Н.А. – DBSCAN әдісінің ерекшеліктері мен артықшылықтары, сондай-ақ оның микроклиматтық деректердегі кластерлер мен аномалияларды анықтады.

Нұрланұлы А. – DBSCAN-нің басты артықшылықтарының бірі – кластерлерге жатпайтын ауытқулар мен шу нүктелерін анықтады.

Атымтаева Л.Б. – CRISP-DM әдіснамасын қолдану деректерді өңдеу мен аномалияларды анықтауда жүйелі тәсілді қамтамасыз етті.

Быков А.А. – DBSCAN әдісі аномалияларды анықтауда және деректердегі ауытқуларды зерттеуде айқын артықшылықтар ұсынды.

Ергалиев Д.С. – VAE моделінің гиперпараметрлерін оңтайландыруға, нейрондық желілердің әртүрлі архитектураларын қарастыруға, сондай-ақ аномалияларды анықтады.

Әбдірашев Ө.К. – DBSCAN кластеризация әдістерін қолдану арқылы жүйедегі аномалияларды анықтау және алынған нәтижелер негізінде микроклиматты тиімді басқару жолдарын ұсынды.

Әдебиеттер тізімі

1. Asperti, A., Evangelista, D., & Loli Piccolomini, E. (2021). A Survey on Variational Autoencoders from a Green AI Perspective. *SN Computer Science*, 2, 301. <https://doi.org/10.1007/s42979-021-00702-9>
2. Scikit-learn. (n.d.). RobustScaler. Retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html>
3. Sharma, N., & Sharma, S. (2023). Optimization of t-SNE by Tuning Perplexity for Dimensionality Reduction in NLP. In S. Kumar, S. Hiranwal, S. Purohit, & M. Prasad (Eds.), *Proceedings of the International Conference on Communication and Computational Technologies (ICCCT 2023)* (pp. 41). Springer, Singapore. https://doi.org/10.1007/978-981-99-3485-0_41
4. Daurenbayeva, N., Nurlanuly, A., & Atymtayeva, L. (2023). Choosing the Intelligent Thermostats for Effective Decision Making in BEMS. In *17th International Conference on Electronics Computer and Computation (ICECCO)* (pp. 1-4).
5. Daurenbayeva, N., Nurlanuly, A., Atymtayeva, L., & Mendes, M. (2023). Survey of Applications of Machine Learning for Fault Detection, Diagnosis and Prediction in Microclimate Control Systems. *Energies*, 16, 3508.
6. Mateus, B.; Mendes, M.; Farinha, J.T.; Martins, A.B.; Cardoso, A.M. Data Analysis for Predictive Maintenance Using Time Series and Deep Learning Models—A Case Study in a Pulp Paper Industry. In *Proceedings of IncoME-VI and TEPEN 2021*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 11–25.
7. Osipov, A., Pleshakova, E., Bykov, A., Kuzichkin, O., Surzhik, D., Suvorov, S., Gataullin, S. Machine Learning Methods Based on Geophysical Monitoring Data in Low Time Delay Mode for Drilling Optimization. *IEEE Access* 2023, 11, 60349–60364

Дауренбаева Н.А.¹, Нұрланұлы А.², Атымтаева Л.Б.³, Быков А.А.¹, Ергалиев Д.С.²,
Абдирашев О.К.*⁴

¹Международный университет информационных технологий

²Академия гражданской авиации

³Университет имени Сулеймана Демиреля

⁴Евразийский национальный университет имени Л.Н. Гумилева

(E-mail: *omeke_92@mail.ru)

Кластеризация параметров микроклимата: методы и математические характеристики

Аннотация. В настоящее время контроль и анализ параметров микроклимата играют важную роль в различных областях, включая производство, экологические исследования и управление зданиями. Параметры микроклимата-температура, влажность, давление воздуха и другие физические показатели-имеют большое значение при определении эффективности производственных процессов и качества продукции. При работе с большими объемами данных, особенно при анализе микроклиматических данных с большим количеством измерений и параметров, особое значение приобретают методы кластеризации. Кластеризация-это процесс разделения данных на группы или кластеры на основе сходств и различий между ними. Такие методы, как K-Means и DBSCAN (Density-Based Spatial Clustering of Applications with Noise), помогают обнаруживать аномалии в данных и выявлять сбои или сбои в работе системы.

В этой статье рассматриваются методы, используемые для кластеризации параметров микроклимата, и их математические модели. Анализируются особенности и преимущества

метода DBSCAN, а также его эффективность в обнаружении кластеров и аномалий в данных микроклимата. В статье приводятся конкретные примеры применения метода DBSCAN, с помощью которого предлагаются математические формулы и расчеты для эффективного анализа и управления параметрами микроклимата.

Углубляя представления о математических основах методов кластеризации и их роли в анализе параметров микроклимата, мы стремимся предложить новые подходы и решения в управлении и оптимизации микроклиматических систем.

Ключевые слова: Микроклимат, кластеризация, DBSCAN, VAE, K-means, обнаружение аномалий, машинное обучение, параметры микроклимата, кластеризация на основе плотности, анализ данных, обнаружение аномалии.

Daurenbayeva N.A.¹, Nurlanuly A.², Atymtaeva L.B.³, Bykov A.A.¹, Yergaliyev D.S.³,
Abdirashev O.K.*⁴

¹International University of Information Technologies

²Academy of Civil Aviation

³Suleyman Demirel University

⁴L.N. Gumilyov Eurasian National University

(E-mail: *omeke_92@mail.ru)

Clustering of microclimate parameters: methods and mathematical characteristics

Abstract. Currently, the control and analysis of the parameters of the microclimate plays a particularly important role in various fields, including production, laboratory work and Environmental Research. Microclimate parameters — temperature, humidity, air pressure and other physical indicators—are of great importance in determining the efficiency of production processes and product quality.

When working with large amounts of data, especially when analyzing microclimatic data, which is characterized by a large number of dimensions and parameters, clustering methods acquire particular importance. Clustering is the process of dividing data into groups or clusters based on similarities and differences within them. Such methods, including such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise), can help detect anomalies in the data and detect malfunctions or violations in the operation of the system.

This article discusses the methods used for clustering microclimate parameters and their mathematical models. The features and advantages of the DBSCAN method are analyzed, as well as its effectiveness in identifying clusters and anomalies in microclimatic data. The article will give specific examples of using the DBSCAN method, with the help of which mathematical formulas and calculations are proposed for the effective analysis and management of microclimate parameters.

By deepening the understanding of the mathematical foundations of clustering methods and their role in the analysis of microclimate parameters, we strive to offer new approaches and solutions in the management and optimization of microclimate systems.

Keywords: microclimate, clustering, DBSCAN, VAE, K-means, anomaly detection, machine learning, microclimate parameters, density-based clustering, data analysis, fault detection.

References

1. Asperti, A., Evangelista, D., & Loli Piccolomini, E. (2021). A Survey on Variational Autoencoders from a Green AI Perspective. *SN Computer Science*, 2, 301. <https://doi.org/10.1007/s42979-021-00702-9>
2. Scikit-learn. (n.d.). RobustScaler. Retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html>
3. Sharma, N., & Sharma, S. (2023). Optimization of t-SNE by Tuning Perplexity for Dimensionality Reduction in NLP. In S. Kumar, S. Hiranwal, S. Purohit, & M. Prasad (Eds.), *Proceedings of the International Conference on Communication and Computational Technologies (ICCCT 2023)* (pp. 41). Springer, Singapore. https://doi.org/10.1007/978-981-99-3485-0_41
4. Daurenbayeva, N., Nurlanuly, A., & Atymtayeva, L. (2023). Choosing the Intelligent Thermostats for Effective Decision Making in BEMS. In *17th International Conference on Electronics Computer and Computation (ICECCO)* (pp. 1-4).
5. Daurenbayeva, N., Nurlanuly, A., Atymtayeva, L., & Mendes, M. (2023). Survey of Applications of Machine Learning for Fault Detection, Diagnosis and Prediction in Microclimate Control Systems. *Energies*, 16, 3508.
6. Mateus, B.; Mendes, M.; Farinha, J.T.; Martins, A.B.; Cardoso, A.M. Data Analysis for Predictive Maintenance Using Time Series and Deep Learning Models – A Case Study in a Pulp Paper Industry. In *Proceedings of IncoME-VI and TEPEN 2021*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 11–25.
7. Osipov, A., Pleshakova, E., Bykov, A., Kuzichkin, O., Surzhik, D., Suvorov, S., Gataullin, S. Machine Learning Methods Based on Geophysical Monitoring Data in Low Time Delay Mode for Drilling Optimization. *IEEE Access* 2023, 11, 60349–60364

Авторлар туралы мәлімет:

Н.А. Дауренбаева – докторант, Халықаралық ақпараттық технологиялар университеті, (ХАТУ), «компьютерлік инженерия» кафедрасының сениор лекторы, Манас көшесі, 34/1, 050040, Алматы, Қазақстан, n.daurenbayeva@iitu.edu.kz

А. Нұрланұлы – докторант, техника ғылымдарының магистрі, Азаматтық авиация академиясы, авиациялық техника және технологиялар кафедрасының сениор лекторы, Ахметов көшесі 44, 050039, Алматы, Қазақстан, a.nurlanuly@agakaz.kz

Л.Б. Атымтаева – физика-математика ғылымдарының докторы, доцент, SDU University, ақпараттық жүйелер кафедрасының қауымдастырылған профессоры, Абылайхан 1/1, 040000, Қаскелең қ., Қазақстан., lyazzat.atymtayeva@sdu.edu.kz

А.А. Быков – техника ғылымдарының кандидаты, доцент, Халықаралық ақпараттық технологиялар университеті, (ХАТУ), «компьютерлік инженерия» кафедрасының қауымдастырылған профессоры, Манас көшесі, 34/1, 050040, Алматы, Қазақстан, Bykov_a_a@list.ru

Д.С. Ергалиев – «авиациялық техника және технологиялар» кафедрасының профессоры, Азаматтық Авиация Академиясы, Ахметов көшесі 44, Алматы, Қазақстан, 8-701-749-58-54, des-67@yandex.kz

Ө.К. Әбдірашев – «ғарыштық техника және технологиялар» кафедрасының доцент м.а., PhD, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Сәтбаев көшесі 2, Астана, Қазақстан, 8-702-682-16-92, omeke_92@mail.ru

Н.А. Дауренбаева – докторант, магистр, Международный университет информационных технологий (МУИТ), сениор лектор кафедры «Компьютерная инженерия», Манаса 34/1, 050040, Алматы, Казахстан, n.daurenbayeva@iitu.edu.kz

А.Нұрланұлы – докторант, магистр технических наук, Академия гражданской авиаций (АГА), сениор лектор кафедры «Авиационная техника и технология», Ахметова 44, 050039, Алматы, Қазақстан. a.nurlanuly@agakaz.kz

Л.Б.Атымтаева – Доктор физико -математических наук, доцент, ассоциированный профессор кафедры Информационных систем SDU University, Аблайхана 1/1, 040000, г Каскелен, Казахстан, lyazzat.atymtayeva@sdu.edu.kz

А.А. Быков – Кандидат технических наук, доцент, Международный университет информационных технологий (МУИТ), ассоциированный профессор кафедры «Компьютерная инженерия», Манаса 34/1, 050040, Алматы, Казахстан, Bykov_a_a@list.ru

Д.С. Ергалиев – профессор кафедры «Авиационная техника и технологии», Академия Гражданской Авиации, ул. Ахметова 44, Алматы, Казахстан, 8-701-749-58-54, des-67@yandex.kz

О.К. Абдірашев – PhD, и.о., доцент кафедры «Космическая техника и технологии», Евразийский национальный университет имени Л.Н.Гумилева, ул. Сатпаева 2, Астана, Казахстан, 8-702-682-16-92, omeke_92@mail.ru

N.A. Daurenbayeva – doctoral student, Master's degree, International University of Information Technology (IIT), senior Lecturer of Department of Computer Engineering, Manasa 34/1, 050040, Almaty, Kazakhstan, n.daurenbayeva@iitu.edu.kz

A. Nurlanuly – doctoral student, Master of Technical sciences, Civil Aviation Academy, senior lecturer of the Department of Aviation Engineering and Technology, Akhmetova 44, 050039, Almaty, Kazakhstan, a.nurlanuly@agakaz.kz

L.B.Atymtayeva – Doctor of Physical and Mathematical Sciences, Associate Professor, Department of Information Systems, SDU University, Ablaihan 1/1, 040000, Kaskelen, Kazakhstan, lyazzat.atymtayeva@sdu.edu.kz

A.A. Bykov – Candidate of Technical Sciences, International University of Information Technology (IIT), Associate Professor, Department of Computer Engineering Manasa 34/1, 050040, Almaty, Kazakhstan, Bykov_a_a@list.ru

D.S. Yergaliev – Professor of the Department of Aviation Engineering and Technology, Academy of Civil Aviation, st. Akhmetova 44, Almaty, Kazakhstan, 8-701-749-58-54, des-67@yandex.kz

O.K. Abdirashev – PhD, acting, associate professor of the Department of Space Engineering and Technologies, Eurasian National University named after L.N. Gumilyov, st. Satpayeva 2, Astana, Kazakhstan, 8-702-682-16-92, omeke_92@mail.ru



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).