

МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІ ЖӘНЕ БОЛЖАУ ЖАСАУДАҒЫ МАҢЫЗЫ

Зарлыкова Меруерт Азаматовна

z.meruert.a@mail.ru

Л.Н. Гумилев атындағы ЕҰУ-нің Механика-математика факультеті, Математикалық және компьютерлік модельдеу, 1-ші курс магистранты, Нұр-Сұлтан, Қазақстан
Ғылыми жетекші – Г.К. Абдрашева

Адамзатты әрқашан оның болашағы қызықтырды. Дәл осы себептен болжау жасау қоғам өмірінің түрлі аспектілеріне байланысты көптеген үрдістердің даму перспективаларын зерттеу саласында әрдайым маңызды орынға ие. Ғылыми дәлелдер негізінде болжамдар – белгілі бір қоғамдық оқиғалардың дамуына кепілдік болды. Машиналық оқыту (Machine Learning) – бұл жүйелерге, бағдарламаларға адамның көмегінсіз белгілі бір міндеттерді шешуге мүмкіндік беретін технологиялардан тұратын конструкция. Қазіргі уақытта машиналық оқытудың өзекті тақырыбы болжау болып табылады.

Машиналық оқыту компьютерлердің көмегімен болжамдарды жасауға бағытталған есептеуіш статистикамен тығыз байланысты. Математикалық оңтайландыруды зерттеу әдістері, теориялары мен қолданбалы салаларын машиналық оқыту саласына жетілдіреді. Деректерді интеллектуалды талдау машиналық оқыту аясында зерттеу саласы болып табылады және бақыланбайтын оқыту (unsupervised learning) арқылы деректерді зерттеу талдауына бағытталады. Өзінің бизнес-міндеттеріне қолдануында машиналық оқытуды болжамдық талдау деп атайды [1].

Машиналық оқыту үшін көптеген модельдер бар, бірақ оларды әдетте үш типке бөледі: біріншісі бақыланатын оқыту (supervised learning), екіншісі бақыланбайтын оқыту немесе өзін-өзі оқыту және де үшіншісі бекітумен не күшпен оқыту (reinforcement learning). Орындалатын міндеттерге байланысты бір модель басқаларға қарағанда анағұрлым қолайлы және неғұрлым тиімді болуы мүмкін.

Бақыланатын оқыту моделі оқыту кезінде дұрыс нәтиже деректер жиынтығындағы әрбір сәйкестендірілген элемент үшін анық белгіленеді. Бұл дегеніміз, деректерді оқу кезінде алгоритмде дұрыс жауабы бар. Сондықтан жауапты іздеудің орнына ол одан әрі белгіленбеген деректерді енгізу кезінде дұрыс жіктеу немесе болжам алу үшін байланыс табуға ұмтылады. Жіктеу контекстінде оқыту алгоритмі, мысалы, кредиттік карталар бойынша транзакция тарихымен жабдықталуы мүмкін, олардың әрқайсысы қауіпсіз немесе күдікті ретінде белгіленеді. Ол осы екі жіктеменің арасындағы қатынастарды зерделеп, содан кейін жіктеу параметрлеріне байланысты жаңа операцияларды тиісті түрде таңбалай алады. Деректер бір-бірімен үздіксіз байланысқан жағдайда, мысалы, уақыт бойынша акциялар курсы өзгерту сияқты, оқытудың регрессиялық алгоритмі деректер жиынтығындағы келесі мәнді болжау үшін пайдаланылуы мүмкін.

Бақыланбайтын оқыту барысында алгоритмде алдын ала белгіленген жауаптар жоқ. Оның мақсаты- жеке деректер арасындағы мағыналық байланыстарды табу, үлгілер мен заңдылықтарды анықтау. Мысалы, кластерлеу – бұл ұсынымдық жүйелерде бақыланбайтын оқытуды пайдалану.

Бекітумен оқытудың түрі – бұл алғашқы екі модельдің қоспасы. Әдетте ол күрделі есептерді шешу үшін қолданылады және қоршаған ортамен өзара әрекеттесуді талап етеді. Деректер ортамен беріледі және әрекет етуге және оқуға мүмкіндік береді. Мұндай әдісті қолдану аймағы кең: роботталған қолды бақылаудан және қозғалыстардың неғұрлым тиімді комбинациясын іздеуден бастап, роботтардың навигация жүйесін әзірлегенге дейін, онда мінез-

құлық алгоритмі “соқтығысудан аулақ болу” тәжірибелік жолмен оқытылады, кедергілермен соқтығысу кезінде кері байланыс алады. Логикалық ойындар логикалық шешімдер тізбегін қамтиды. Бұл оқыту әдісі логистикада, кестелерді құрастыруда және міндеттерді тактикалық жоспарлауда жиі қолданылады [2].

Бизнес-техникалық оқытуды қолданудың үш саласын қарастыруға болады: сипаттама, болжамдау және нормативтік.

Сипаттамалық қолдану бизнес-талдаушылардың мүмкіндіктерін кеңейту үшін статистикалық деректерді жазуға және талдауға жатады. Басшылар өткен іс-қимылдар мен шешімдердің нәтижелері мен салдарларының сипаттамасын және барынша ақпараттық талдауын алады. Бұл процесс қазіргі уақытта бүкіл әлем бойынша көптеген ірі компаниялар үшін, олардың нәтижелері мен рентабельділігін анықтау үшін, сату мен жарнамалық жобаларды талдау үшін қолданады.

Машиналық оқытуды екінші қолдану саласы – болжау. Деректерді жинау және оларды нақты нәтижені болжау үшін пайдалану реакция жылдамдығын арттыруға және дұрыс шешімдерді жылдам қабылдауға мүмкіндік береді. Мысалы, клиенттердің кетуін болжамдап, оны болдырмауға көмектесе алады. Бүгінде бұл процесс көптеген ірі компанияларда қолданылады.

Машиналық оқытуды үшінші және неғұрлым озық қолдану бұрыннан бар компаниялармен енгізілуде және жетілдірілуде. Нәтижелерді болжау бизнесті тиімді жүргізу үшін жеткіліксіз. Себептерді, уәждерді және қоршаған жағдайды түсіну, оңтайлы шешім қабылдау үшін қажетті шарт. Бұл әдіс адам мен машина күш біріктірсе ғана тиімді. Машиналық оқыту маңызды тәуелділіктерді іздеу және нәтижелерді болжау үшін пайдаланылады, ал мамандар мәліметтер бойынша мұндай байланыс неге бар екенін түсіну үшін нәтижені түсіндіреді. Нәтижесінде нақты және дұрыс шешімдерді қабылдау мүмкін болады [3].

Машиналық оқытудың кез келген міндеті талданатын объект туралы қандай да бір ақпарат бар тізімдер немесе кестелер түрінде ұсынылатын деректер жиынтығы – іріктеменің болуын білдіреді. Есеп қолда бар деректерді талдау негізінде талданатын объект қалай құрылған модельді құрудан басталады. Модельді құру кезінде тағы бір маңызды міндет оның болжау сапасын бағалау болып табылады, ол құрылған модельдің көмегімен алынған болжамға қаншалықты барабар сүйенетінін сипаттайды.

Нейрондық желілер мидың құрылымын имитациялайды: әрбір жасанды нейрон бірнеше басқа нейрондармен қосылады. Нейрожелілер жалпы бірқатар салаларда қолданылады. Денсаулық сақтауда оларды диагностикалық процедураларды жеделдету және дәрі-дәрмектерді іздеу мақсатында медициналық суреттерді талдау кезінде пайдаланады. Телекоммуникация саласында және медиаиндустрияда нейрожелілерді машиналық аудару, алаяқтықты тану және виртуалды ассистенттердің қызметтерін ұсыну үшін қолдануға болады. Қаржы саласында оларды алаяқтықты тану, портфельдерді басқару және тәуекелді талдау үшін пайдаланады. Бөлшек саудада-кассандағы кезектерден арылу үшін және сатып алушыларға қызмет көрсетуді дербестендіру үшін қолданады [4].

Келесі шешім ағашы, оның алгоритмі тораптық нүктелерде орналасқан атрибуттары туралы сұрақтарға жауап бере отырып, объектілерді жіктейді. Жауаптарға байланысты тармақтардың бірі таңдалады. Шешім ағашын қолданылуы клиенттік қызмет көрсетуге арналған. Сол сияқты білімді басқару платформасы, бағаның болжамды мақсаты және өнім шығаруды жоспарлауға қолданылады. Сақтандыру компаниясында шешім ағашы сақтандыру өнімдері мен сыйлықтардың қандай түрлерін ықтимал тәуекелді ескере отырып іске қосу керек екенін анықтауға көмектеседі. Ауа райы жағдайларын ескере отырып, орналасқан жері туралы деректерді және сақтандыру жағдайлары туралы мәліметтерді пайдалана отырып, жүйе берілген талаптар мен жұмсалған сомалар негізінде тәуекел санаттарын анықтай алады [5].

Тағы бір түрі ол кездейсоқ орман-эмбебап, тез оқитын механизм деректер жиынтығы ішіндегі байланыстарды анықтау үшін қолданылады. Мысалы, пайдаланушыларға ғана емес, спам салдарынан серверлерге жоғары жүктемемен жұмыс істеуге тура келетін Интернет провайдерлеріне де проблемалар тудыратын жағымсыз бұқаралық таратылымдарды келтіруге болады. Проблемамен күресу үшін спамды сүзудің автоматтандырылған әдістері әзірленді, олар шешетін ағаштар ансамблінің көмегімен жағымсыз хаттарды тез және тиімді анықтайды. Басқа қолданулардың арасында-пациенттің медициналық картасын талдау жолымен ауруларды диагностикалау, банк алаяқтығын тану, колл-орталықтардағы қоңырау санын болжау және белгілі бір акцияларды сатып алу кезінде пайда мен шығынның ықтималдығын болжауға өту ыңғайлы жұмыс жасайды.

Ал кластерлеу – статистикалық алгоритмдердің көмегімен ұқсас сипаттамалары бар деректер элементтерін топтастыру. Бұл сыныптама есептерін шешу үшін қолдануға болатын бақыланбайтын оқыту әдісі. Мысалға, маркетингтік науқандардың адресациясын анықтау үшін сипаттамаларына байланысты сатып алу аудиториясын саралау, нақты оқырмандарға жаңалықтар ұсыну, құқық қорғау органдарына жұмысқа көмек көрсету үшін қолданылады. Кластерлеу, сондай-ақ, күрделі деректер жиынтықтарында арнайы құралдарсыз байқауға қиын топтарды табу қажет болған кезде тиімді [6].

Қорытындылай келгенде, алгоритмді таңдау кезінде талданатын мәліметтердің ерекшелігін ескеру қажет. Логистикалық регрессия (Logistic Regression) нәтижелерді интерпретациялаудың ауқымдылығы мен ашықтығының арқасында болжаушы міндеттерді шешу үшін қолайлы. Тірек векторлар әдісі (Support Vector Methods) болжауда жоғары дәлдікті көрсете алады, бірақ ол есептеу жоспарында әлдеқайда қымбат және болжау нәтижелерін түсіндіруде қиын. Ең жақын көршілер әдісі (Nearest-Neighbors Methods) іске асыру оңай, бірақ барлық оқыту іріктмесін сақтау қажеттілігі және жіктелетін объектіні іріктеменің барлық объектілерімен желілік салыстыру қажеттілігі салдарынан есептеу қуатын тиімсіз жұмсайды.

Қолданылған әдебиеттер тізімі

1. Бринк Х., Ричардс Д., Феверолф М. Машинное обучение.-СПб.: Питер, 2017. -336 с.
2. Крюкова Я., Кручинин И. Обзор способов применения методов машинного обучения для прогнозирования. М.: Международный студенческий научный вестник, 2019. – № 2.
3. Вьюгин В. Математические основы машинного обучения и прогнозирования. – М.:МЦНМО, 2018. – 384 с.
4. Bob Violino. Machine learning: When to use each method and technique. InfoWorld, 2018.
5. Записи лекций Сенько О. Задачи прогнозирования, обобщающая способность, байесовский классификатор, скользящий контроль.
6. Записи лекций Власов В., Духовный А., Малахова И. Прогнозирование с помощью методов машинного обучения.

УДК 517.951, 517.957

КАДОМЦЕВ-ПЕТВИАШВИЛИ ТЕНДЕУІНІҢ ОҢАШАЛАНҒАН НАҚТЫ ШЕШІМДЕРІ

Картджанова Гульнур Режепбаевна

gulnurk02@gmail.com

Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Нұр-Сұлтан, Қазақстан

Ғылыми жетекші – Г.Н. Нугманова