AZƏRBAYCAN RESPUBLİKASININ DÖVLƏT STANDARTI

AZS ISO/IEC 24029-2 2024

Birinci nəşr
2024

Süni intellekt (Sİ) – Neyron şəbəkələrinin dayanıqlığının qiymətləndirilməsi – Hissə 2: Formal metodların istifadəsi metodologiyası

Artificial intelligence (AI) – Assessment of the robustness of neural networks – Part 2: Methodology for the use of formal methods

AZƏRBAYCAN STANDARTLAŞDIRMA İNSTİTUTU

İstinad nömrəsi
AZS ISO/IEC 24029-2:2024

Elçin İsaqzadə küç., 7-ci köndələn
Qaynar xətt: +994125149308
Email: office@azstand.gov.az

MÜQƏDDİMƏ

1. Bu standart Azərbaycan Elm Fondunun qrant layihəsi (Qrant №AEF-MQM-QA-1-2021-4(41)-8/03/1) çərçivəsində işlənib hazırlanıb və “İnformasiya-kommunikasiya texnologiyaları” standartlaşdırma üzrə Texniki Komitə (AZSTAND/TK 05) tərəfindən təqdim edilib.
2. Azərbaycan Standartlaşdırma İnstitutunun ..... 2024-cü il tarixli ..... saylı qərarı ilə təsdiq edilib.
3. Bu standart Beynəlxalq Standart ISO/IEC 24029-2 nəşr 1.0 (2023-08) ilə eynidir. This standart is identical (IDT) to the International Standard ISO/IEC 24029-2, (2023-08).
4. İlk dəfə tətbiq edilir.
5. Dövlət standartında müəyyən edilən tələblərin beynəlxalq standartlara, norma, qayda və tövsiyələrə və digər dövlətlərin müvafiq mütərəqqi milli standartlarına, elm, texnika və texnologiyanın müasir nailiyyətlərinə əsaslanmasını müəyyən etmək üçün standartın dövri yoxlama müddəti ildə 1 dəfədir.

MÜNDƏRİCAT

ÖN SÖZ

ISO (Beynəlxalq Standartlaşdırma Təşkilatı) və IEC (Beynəlxalq Elektrotexniki Komissiya) dünya üzrə standartlaşdırma sahəsində ixtisaslaşmış sistemi formalaşdırırlar. ISO və ya IEC üzvü olan milli orqanlar texniki fəaliyyətin konkret sahələri ilə məşğul olmaq üçün müvafiq təşkilat tərəfindən yaradılmış texniki komitələr vasitəsilə beynəlxalq standartların hazırlanmasında iştirak edirlər. ISO və IEC texniki komitələri qarşılıqlı maraq doğuran sahələrdə əməkdaşlıq edirlər. ISO və IEC ilə əməkdaşlıq edən digər beynəlxalq təşkilatlar, dövlət və qeyri-hökumət təşkilatları da bu işdə iştirak edirlər.

Bu standartın hazırlanması üçün istifadə olunan və həmçinin sonrakı texniki xidmət üçün nəzərdə tutulan prosedurlar ISO/IEC Direktivlərinin 1-ci hissəsində təsvir edilmişdir. Müxtəlif növ sənədlər üçün tələb olunan fərqli təsdiq meyarlarına xüsusilə diqqət yetirilməlidir. Bu sənəd ISO/IEC Direktivlərinin 2-ci hissəsində (www.iso.org/directives və ya www.iec.ch/members_experts/refdocs) verilmiş qaydalara uyğun olaraq hazırlanmışdır.

ISO və IEC qeyd edir ki, bu standartın tətbiqi patent(lər)in istifadəsini əhatə edə bilər. ISO və IEC bu standartla bağlı iddia edilən hər hansı patent hüquqlarının sübutu, həqiqiliyinin yoxlanılması və ya tətbiq oluna bilməsi ilə əlaqədar heç bir mövqe tutmur. Standartın dərc edildiyi tarixə ISO və IEC bu sənədin tətbiqi üçün tələb oluna bilən patent(lər) haqqında heç bir bildiriş almamışdır. Bununla yanaşı, xəbərdarlıq edilir ki, bu, “www.iso.org/patents” və “https://patents.iec.ch” ünvanlarında yerləşən patent verilənlər bazasında mövcud olan ən son məlumatları əks etdirməyə bilər. ISO və IEC bu patent hüquqlarının hər hansı birinin və ya hamısının müəyyən edilməsinə görə məsuliyyət daşımır.

Bu standartdakı ticarət adları (“trade name”) haqqında məlumatlar istifadəçilərin rahat istifadəsi üçün təqdim olunur və bu təqdimat tövsiyə xarakteri daşımır.

Standartların könüllü xarakter daşıması, uyğunluğun qiymətləndirilməsi üzrə ISO-nun xüsusi termin və ifadələrinin mənası ilə bağlı izahlar, eləcə də Ticarətdə Texniki Maneələrin (Technical Barriers to Trade, TBT) aradan qaldırılması ilə əlaqədar ISO-nun Ümumdünya Ticarət Təşkilatının (ÜTT) prinsiplərinə sadiqliyi haqqında məlumat “www.iso.org/iso/foreword.html” internet informasiya ehtiyatından əldə edilə bilər. IEC ilə bağlı “www.iec.ch/understanding-standards” internet informasiya ehtiyatına müraciət etmək olar.

Bu sənəd ISO və CEN arasında texniki əməkdaşlıq haqqında Sazişə (Vyana Sazişi) uyğun olaraq, Avropa Standartlaşdırma Komitəsinin (CEN) “CEN/CLC/JTC 21, Süni intellekt” Texniki Komitəsi ilə əməkdaşlıq çırçivəsində ISO/IEC JTC 1 “İnformasiya texnologiyaları” Birgə Texniki Komitəsinin “SC 42, Süni intellekt” Altkomitəsi tərəfindən hazırlanmışdır.

ISO/IEC 24029 standartının hissələrinin siyahısı ISO və IEC təşkilatlarının internet informasiya ehtiyatlarında yerləşir.

Bu sənədlə bağlı istənilən rəy və suallar milli standartlaşdırma qurumuna yönəldilməlidir. Bu qurumların tam siyahısı ilə “www.iso.org/members.html” və “www.iec.ch/national-committees” internet informasiya ehtiyatlarında tanış olmaq olar.

GİRİŞ

Neyron şəbəkələri təsvirlərin və ya təbii dilin emalı, proqnozlaşdırılan (profilaktik) texniki xidmət kimi mürəkkəb məsələlərin həllini əhatə edən müxtəlif sahələrdə geniş istifadə olunur. Süni intellekt (Sİ) sistemlərinin keyfiyyət modelləri dayanıqlıq da daxil olmaqla müəyyən xüsusiyyətlərə malikdir. Məsələn, SQuaRE [2] beynəlxalq standartlarını Sİ sistemlərinə genişləndirərək şamil edən ISO/IEC 25059:2023 [1] standartı keyfiyyət modelində dayanıqlığı etibarlılığın alt xüsusiyyəti kimi nəzərdən keçirir. Müxtəlif şəraitlərdə sistemin öz performans səviyyəsini saxlamaq qabiliyyəti statistik təhlildən istifadə olunmaqla nümayiş etdirilə bilər, lakin bu qabiliyyətin mövcudluğunun sübutu formal təhlilin aparılmasını tələb edir. Bu kontekstdə formal metodlar neyron şəbəkələrinin dayanıqlığına inamın artırılması üçün digər metodlara əlavə olaraq birgə istifadə edilə bilər.

Formal metodlar proqram və aparat sistemlərinin düzgünlüyünün (korrektliyinin) sübutu məqsədilə onların ciddi şəkildə spesifikasiyası və verifikasiyası üçün istifadə olunan riyazi üsullar, texnikalardır. Neyron şəbəkələri haqqında formal mülahizələr (əsaslandırmalar) aparmaq və onların tələb olunan dayanıqlıq xüsusiyyətlərinə malik olduğunu sübut etmək üçün formal metodlardan istifadə edilə bilər. Məsələn, giriş verilənləri təsvir, çıxış verilənləri isə verilmiş siniflər çoxluğundakı nişan (məsələn, avtomobil və ya təyyarə) olan neyron şəbəkəsi əsasında klassifikasiyanı nəzərdən keçirək. Belə bir klassifikator giriş verilənləri qismində təsvirin piksel intensivliyini qəbul edən, verilmiş çoxluqdan hər bir mümkün sinif üçün ehtimalları hesablayan və çıxış veriləni kimi ən böyük ehtimala uyğun nişanı seçən riyazi funksiya kimi göstərilə bilər. Bu formal model (klassifikator) daha sonra giriş təsvirində dəyişiklik edildikdə neyron şəbəkəsinin işinin riyazi əsaslandırılması üçün istifadə edilə bilər. Məsələn, tutaq ki, konkret təsvir üçün (onunla bağlı neyron şəbəkəsi “avtomobil” nişanını generasiya edir) belə bir sual verilə bilər: “Şəkildəki ixtiyari pikselin qiyməti dəyişdirilərsə, şəbəkə başqa nişan generasiya edəcək?" Bu sual verilmiş neyron şəbəkəsi və təsvir üçün doğru və ya yalan qiymət alan formal riyazi ifadə şəklində göstərilə bilər.

Formal metodlardan istifadə ilə bağlı klassik yanaşma bu standartda təsvir olunan üç əsas addımdan ibarətdir. İlk növbədə, təhlil edilən sistem onun bütün mümkün davranışlarını dəqiq şəkildə əks etdirən modeldə formal olaraq müəyyən edilir. İkinci addımda riyazi şəkildə tələb müəyyən edilir. Nəhayət üçüncü addımda, sistemin verilmiş tələbə cavab verib-vermədiyini qiymətləndirmək üçün “çözücü” (“solver”), mücərrəd interpretasiya və ya modelin yoxlanılması kimi formal metodlardan istifadə edilir ki, bu da sübutla, kontr-nümunə ilə və ya qeyri-müəyyənliklə nəticələnir.

Bu standartda əlçatan bir neçə formal metod texnikası nəzərdən keçirilir. Standart Sİ sistemlərinin həyat dövrünün hər bir mərhələsində neyron şəbəkələrinin dayanıqlığını qiymətləndirmək və formal metodlardan istifadə etməklə neyron şəbəkələrinin yoxlanılması imkanlarını müəyyən etmək üçün tətbiq olunan meyarlar təqdim edir. Formal metodların istifadəsi zamanı miqyaslama baxımından çətinliklər yarana bilər, lakin buna baxmayaraq, onlar hələ də fəqli verilənlər növləri üzrə müxtəlif tapşırıqları yerinə yetirən istənilən neyron şəbəkələrinə tətbiq olunur. Ənənəvi proqram sistemlərində formal metodlar çoxdan istifadə olunsa da, neyron şəbəkələrində formal metodların istifadəsi nisbətən yenidir və aktiv tədqiqat sahəsidir.

Bu standart neyron şəbəkələrindən istifadə edən və Sİ sistemlərinin həyat dövrünün müvafiq mərhələlərində onların dayanıqlığını qiymətləndirməli olan Sİ tərtibatçılarına kömək məqsədi daşıyır. ISO/IEC TR 24029-1 standartında bu sənəddə təsvir edilən formal metodlardan əlavə neyron şəbəkələrin dayanıqlığını qiymətləndirmək üçün mövcud metodların daha ətraflı icmalı təqdim edilir.

AZƏRBAYCAN RESPUBLİKASININ DÖVLƏT STANDARTI

Süni intellekt (Sİ) – Neyron şəbəkələrinin dayanıqlığının qiymətləndirilməsi – Hissə 2: Formal metodların istifadəsi metodologiyası

AZE ISO/IEC 24029-2:2024

Artificial intelligence – Assessment of the robustness of neural networks – Part 2: Methodology for the use of formal methods

Tətbiq edilmə tarixi ... . 2024-cü il

1 TƏTBİQ SAHƏSİ

Bu standart neyron şəbəkələrinin dayanıqlıq xassəsini qiymətləndirmək üçün formal metodlardan istifadə metodologiyasını müəyyən edir. Əsas diqqət dayanıqlıq xassələrinin sübut edilməsi məqsədilə formal metodların seçilməsi, tətbiqi və idarə edilməsinə yönəldilir.

2 NORMATİV İSTİNADLAR

Aşağıdakı sənədlərə mətn boyu istinad onun məzmununun tam və ya qismən bu sənədin tələblərinə uyğun gəldiyi təqdirdə edilmişdir. Tarix qeyd edilmiş istinadlarda yalnız istinad olunan nəşrlər istifadə edilir. Tarix qeyd edilməmiş istinadlarda istinad olunan sənədin (düzəlişlər daxil olmaqla) sonuncu nəşri istifadə edilir.

AZE ISO/IEC 22989:2023, *İnformasiya texnologiyaları — Süni intellekt — Süni intellekt üzrə anlayışlar və terminologiya*

AZS ISO/IEC 23053:2024, *Maşın öyrənməsindən (ML) istifadə edən süni intellekt (Sİ) sistemləri üçün çərçivə sənədi*

3 TERMİN VƏ ANLAYIŞLAR

Bu sənədin məqsədləri üçün AZE ISO/IEC 22989:2023 və AZS ISO/IEC 23053:2024 standartlarındakı terminlər və anlayışlar istifadə edilir.

ISO и IEC-in standartlaşdırma sahəsində istifadə olunan terminoloji məlumat bazaları aşağıdakı ünvanlarda saxlanılır:

— ISO Onlayn baxış platforması: https://www.iso.org/obp;

— IEC Electropedia: http://www.electropedia.org/.

3.1
domen sahəsi (domain)
neyron şəbəkəsinin mühitin atributları ilə xarakterizə olunan mümkün giriş verilənləri çoxluğu

NÜMUNƏ 1. Təbii dilin emalı üzrə tapşırığı yerinə yetirən neyron şəbəkəsi sözlərdən ibarət mətnlər üzərində işləyir. Mümkün müxtəlif mətnlərin sayı qeyri-məhdud olsa da, hər bir cümlənin maksimum uzunluğu həmişə məhduddur. Ona görə də, bu domen sahəsini xarakterizə edən atribut hər cümlənin icazə verilən maksimum uzunluğu ola bilər.

NÜMUNƏ 2. Üzün tanınması üzrə domen sahəsi, məsələn, üzün şəklinin ölçülərinin ən azı 40×40 piksel olmasını tələb edən atributlara əsaslana bilər. Əksər üz cizgilərinin görünməsi şərti ilə yarımprofildən çəkilmiş üzlər daha aşağı dəqiqliklə aşkarlana bilər. Analoji olaraq, qismən okklüziyalar da (qismən üst-üstə düşən şəkillər də) müəyyən dərəcədə emal edilir. Tanınma prosesi, ümumiyyətlə, üzün 70%-dən çoxunun görünməsini tələb edir. Kameranın üzlə eyni hündürlükdə olduğu görüntülər ən yaxşı keyfiyyətə malik olur və görüntünün mövqeyi 30 dərəcədən yuxarı və ya 20 dərəcədən aşağı dəyişdikdə isə keyfiyyət aşağı düşür.

Qeyd 1: Atribut, domen sahəsi qeyr-məhdud olsa belə, məhdud saylı atributa malik obyekti təsvir etmək üçün istifadə olunur.

3.2
atribut (attribute)
Obyektin insanlar tərəfindən və ya avtomatik vasitələrlə kəmiyyət və ya keyfiyyətcə tanına (fərqləndirilə) bilən xassəsi və ya xüsusiyyəti.

[MƏNBƏ: ISO/IEC TR 15939:2017, 3.2, düzəliş – “element” sözü “obyekt” sözü ilə əvəz olunub.]

3.3
məhdud domen sahəsi (bounded domain)
məhdud sayda obyekt ehtiva edən çoxluq

NÜMUNƏ 1. n-pikselli bütün mümkün 8 bitlik RGB şəkillərinin domen sahəsinin ölçüsü ən çox $256^{3\times n}$ ilə məhdudlaşır.

NÜMUNƏ 2. İngilis dilində bütün mümkün cümlələrin sayı sonsuzdur, ona görə də bu domen sahəsi qeyri-məhduddur.

Qeyd 1: Qeyri-məhdud domen sahəsində obyektlərin sayı sonsuzdur.

3.4
məhdud obyekt (bounded object)
Sonlu sayda atributlarla təmsil olunan obyekt.

Qeyd 1: Məhdud obyektdən fərqli olaraq, qeyri-məhdud obyektin sonsuz sayda atributları var.

3.5 stabillik (stability)
Neyron şəbəkəsinin giriş verilənləri dəyişdirilərkən çıxış verilənlərinin dəyişməz qalmasını təmin edən xüsusiyyəti.

Qeyd 1: Daha stabil neyron şəbəkəsində giriş verilənlərinin dəyişiklikləri küy olduqda, çıxış verilənlərinin dəyişikliyə məruz qalması ehtimalı azdır.

3.6 həssaslıq (sensitivity)
Neyron şəbəkəsinin giriş verilənləri dəyişdirilərkən çıxış verilənlərinin də dəyişməsini təmin edən xüsusiyyəti.

Qeyd 1: Daha həssas neyron şəbəkəsində giriş verilənlərinin dəyişiklikləri informativ olduqda, çıxış verilənlərinin dəyişikliyə məruz qalması ehtimalı azdır.

3.7
arxitektura (architecture)
sistemin (öz mühitində) elementlərində, əlaqələrində, həmçinin layihələndirmə və inkişaf prinsiplərində əks olunan əsas anlayışları və ya xüsusiyyətləri

3.8
relevantlıq (relevance)
Bütün digər girişlərlə müqayisədə neyron şəbəkəsinin müəyyən giriş verilənlərinin çıxış verilənlərinə təsirinin nizamlanmış nisbi əhəmiyyəti

3.9
meyar (criterion)
Mülahizənin və ya qərarın əsaslana biləcəyi və ya məhsul, xidmət, nəticə və ya prosesin qiymətləndirilə biləcəyi qayda.

[MƏNBƏ: ISO/IEC/IEEE 15289:2019, 3.1.6]

3.10
zaman sıraları (time series)
ardıcıl zaman nöqtələrində götürülmüş qiymətlər ardıcıllığı

[MƏNBƏ: ISO/IEC 19794-1:2011, 3.54]

3.11
əlçatarlıq (reachability)
Sİ agentinin müəyyən mühitdəki vəziyyətlər çoxluğunda olmağının mümkünlüyünü təsvir edən xüsusiyyət

3.12
hissə-hissə xətti neyron şəbəkəsi (piecewise neural network)
hissə-hissə xətti aktivləşdirmə funksiyaları istifadə edən neyron şəbəkəsi

Qeyd 1: Xətti aktivləşdirmə funksiyalarına nümunə olaraq “Rectify linear unit” və ya “MaxOut” göstərmək olar.

3.13
binarlaşdırılmış neyron şəbəkəsi (binarized neural network)
əsasən ikilik parametrlərə malik olan neyron şəbəkəsi

3.14
rekkurent neyron şəbəkəsi (recurrent neural network)
giriş verilənlərinin altardıcıllığını emal etdikdən sonra öyrəndiklərini kodlayaraq daxili vəziyyəti qoruyub saxlayan neyron şəbəkəsi

3.15
transformer neyron şəbəkəsi (transformer neural network)
transformer
emal zamanı giriş verilənlərinin müxtəlif hissələrinin təsirini müəyyən etmək üçün "özünədiqqət" mexanizmini istifadə edən neyron şəbəkəsi

3.16
model yoxlaması (model checking)
nəzəriyyənin formal təsdiqinin isbatı

3.17
struktur əsaslı testetmə (structural-based testing)
"şüşə qutu" testi (glass-box testing)
"ağ qutu" testi (white-box testing)
struktur testetmə (structural testing)
testlərin test edilən obyektin strukturunun tədqiqi nətisəsində əldə edildiyi dinamik testetmə

Qeyd 1: Struktur əsaslı testetmə komponentlər səviyyəsində istifadə ilə məhdudlaşmır və bütün səviyyələrdə istifadə oluna bilər, məsələn, sistem testinin bir hissəsi kimi menyu elementinin əhatə olunması zamanı.

Qeyd 2: Texnikalara budaqların test olunması, qərarların və operatorların test edilməsi daxildir.

[MƏNBƏ: ISO/IEC/IEEE 29119-1:2022, 3.80]

3.18
"qapalı qutu" testi (closed-box testing)
spesifikasiyaya əsaslı testetmə (specification-based testing)
"qara qutu" testi (black-box testing)
əsas test bazası test edilən obyektin (test elementinin) xarici giriş və çıxışlarından ibarət olan, adətən (spesifikasiyanın mənbə kodunda və ya icra edilən proqram təminatında yerinə yetirilməsinə deyil) spesifikasiyaya əsaslanan testetmə.

[MƏNBƏ: ISO/IEC/IEEE 29119-1:2022, 3.75]

4 QISALTMALAR

AI	artificial intelligence	süni intellekt
BNN	binarized neural network	binarlaşdırılmış neyron şəbəkəsi

GNN	graph neural network	qraf neyron şəbəkəsi
MILP	mixed-integer linear programming	qismən tamqiymətli xətti proqlaşdırma
MRI	magnetic resonance imaging	maqnit rezonans tomoqrafiyası
PLNN	piecewise linear neural networks	hissə-hissə xətti neyron şəbəkələri
ReLU	rectified linear unit	düzləndirilmiş xətti vahid (blok)
RNN	recurrent neural network	rekurrent neyron şəbəkəsi
SAR	synthetic aperture radar	sintez olunan aperturlu radar
SMC	satisfiability modulo convex	qabarıqlıq modulu üzrə qaneetmə
SMT	satisfiability modulo theories	qaneetmə modulu nəzəriyyələri

5 DAYANIQLIĞIN QİYMƏTLƏNDİRİLMƏSİ

5.1. Ümumi müddəalar

Neyron şəbəkələri kontekstində dayanıqlıq spesifikasiyasları adətən neyron şəbəkəsinin tətbiq olunduğu domen sahəsindən (bax 5.2) asılı olaraq, təbii və ya qeyri-təbii (rəqabətli) olaraq dəyişə bilən müxtəlif şərtlərdən ibarətdir.

NÜMUNƏ 1. Tibbi təsvirləri emal edən neyron şəbəkəsini nəzərdən keçirək, burada neyron şəbəkəsinin giriş verilənləri xəstələrin skan edildiyi tibbi qurğular vasitəsilə toplanır. Bir xəstə üçün bir neçə təsvir çəkilərkən, təbii ki, identik təsvirlər əldə edilmir. Bunun səbəbi xəstənin skan zamanı pozisiyasının cüzi də olsa dəyişməsi, otağın fərqli işıqlandırılması, obyektin əks olunması və ya təsvirin sonrakı emalı prosesində təsadüfi küylərin əlavə edilməsi ola bilər.

NÜMUNƏ 2. Pilotsuz idarə olunan nəqliyyat vasitəsinin sensorlarının və bort kameralarının çıxış verilənlərini emal edən neyron şəbəkəsini nəzərdən keçirək. Ətraf mühitin hava şəraiti, çirklənmə və işıqlandırılma şəraiti kimi dinamik xüsusiyyətlərə malik olması neyron şəbəkəsinin giriş verilənlərini müxtəlif atributlarını çox dəyişə bilər.

Vacib məqam odur ki, ətraf mühitin bu variasiyaları adətən neyron şəbəkəsinin dayanıqlığını dəyişmir. Neyron şəbəkəsinin dayanıqlığı sonradan NN-in tətbiq sahəsində müvafiq proksi spesifikasiyalardan istifadə etməklə ətraf mühitin bu cür dəyişikliklərinə nəzərən yoxlanıla bilər.

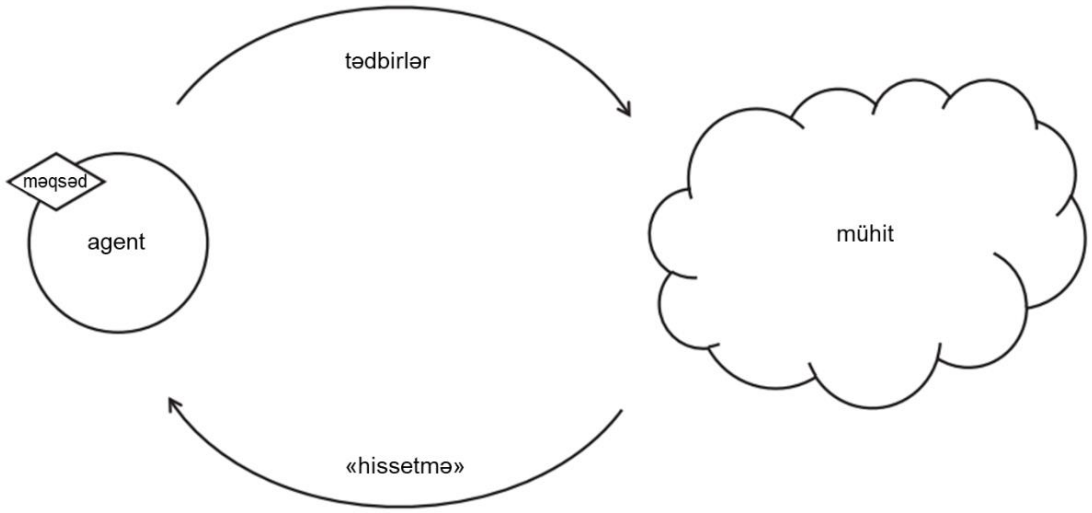
Dayanıqlıq xassələri lokal və ya qlobal ola bilər [10]. Lokal dayanıqlıq xassələri daha asan təyin olunduğu üçün onlar qlobal dayanıqlıq xassələri ilə müqayisədə daha tez-tez yoxlanılır. Lokal dayanıqlıq xassələri test verilənləri çoxluğundan olan giriş verilənləri nümunəsinə əsasən müəyyən edilir. Məsələn, əgər ilkin təsvir avtomobil kimi düzgün təsnif edilibsə, lokal dayanıqlıq xassəsinə əsasən ilkin təsvirin 5 dərəcə ətrafında fırlanması ilə

alınan bütün şəkillərin də avtomobil kimi təsnif edildiyi göstərilə bilər. Lokal dayanıqlıq xassələrinin yoxlanılmasının çatışmazlığı ondan ibarətdir ki, zəmanətlər təqdim edilmiş test nümunəsi üçün lokal xarakter daşıyır və verilənlər çoxluğundakı digər nümunələrə şamil edilmir. Bunun əksinə olaraq, qlobal dayanıqlıq xassələri bütün mümkün giriş verilənləri üçün deterministik olaraq saxlanılan zəmanətləri müəyyən edir [11]. Giriş xüsusiyyətlərinin semantik məna daşıdığı, məsələn, hava nəqliyyatı toqquşmalarının qarşısının alınması sistemləri üçün qlobal xassələr real dünyada gözlənilən giriş xüsusiyyətləri üçün mümkün giriş qiymətlərini müəyyən etməklə təyin oluna bilər. Fərdi xüsusiyyətlərin semantik mənası olmadığı hallarda önəmli giriş qiymətlərini müəyyən etmək daha çətindir. Bu bölmədə təsvir edilən dayanıqlıq xassələri çoxluğu tam deyil və gələcəkdə yeni dayanıqlıq xassələrinin meydana çıxması mümkündür.

5.2. Domen sahəsi anlayışı

Neyron şəbəkələri də daxil olmaqla, əksər Sİ sistemləri onların performans (məhsuldarlıq) xüsusiyyətlərinin müəyyən edilə və qiymətləndirilə biləcəyi konkret bir mühitdə fəaliyyət göstərmək üçün nəzərdə tutulur (tipik qiymətləndirmə metrikaları ISO/IEC TR 24029-1:2021 standartındakı “Cədvəl 1”-də verilir). Neyron şəbəkəsinin əsas performans xüsusiyyətlərindən olan dayanıqlıq neyron şəbəkəsinin fəaliyyət göstərdiyi domen sahəsinin də ayrılmaz xarakteristikalarından biridir. Neyron şəbəkələrinin bir çox tətbiqlərində məhdud domen sahəsinin mövcudluğu nəzərdə tutulur (məsələn, təsvirin klassifikasiyası şəkillərin müəyyən keyfiyyətə və formata malik olmasını tələb edir).

Şəkil 1-də (bax: AZE ISO/IEC 22989:2023, Şəkil 1) təqdim olunan agent paradiqması agentin müəyyən məqsədlərə çatmaq üçün öz mühitini “hiss” etməsini və bu mühitdə tədbirlər gördüyünü təsvir edir. Təqdim olunan paradiqmada müxtəlif mühit və Sİ agenti konsepsiyalarına xüsusi diqqət yetirilir. Domen sahəsi anlayışı müasir texnologiyaların məhdudiyyətlərini əks etdirir, belə ki, burada neyron şəbəkəsi müəyyən növ Sİ agenti olaraq, texniki cəhətdən öz məqsədinə yalnız müvafiq giriş verilənləri ilə işlədiyi halda nail ola bilir.



Şəkil 1 – Agent paradiqması

Domen sahəsi konsepsiyası aşağıdakı prinsiplərə əsaslanır:

- domen sahəsi dəqiq təyin olunmuş atributlar çoxluğu ilə müəyyən edilməlidir (yəni domen sahəsi məhdud obyektləri ehtiva edir);
- domen sahəsinin spesifikasiyası Sİ sistemi tərəfindən bir və ya bir neçə tapşırığın təyinatına uyğun yerinə yetirməsi üçün kifayət olmalıdır;
- təlim üçün istifadə edilən verilənlər nəticə çıxarmaq üçün istifadə edilməsi nəzərdə tutulan verilənlərə nəzərən reprezentativ olmalıdır.

Domen sahəsini müəyyən etmək neyron şəbəkəsinin öz məqsədinə çatması üçün lazım olan verilənlərin bütün atributlarının təyin edilməsini nəzərdə tutur.

Neyron şəbəkələri tətbiqlərinin geniş yayılmış bəzi sahələrinə görmə, nitq emalı və robototexnika üzrə tətbiqlər misal göstərilə bilər. Bu domen sahələrini, lakin ən əsası, onların dəyişkənliyini təsvir etmək üçün adətən ədədi atributlardan istifadə olunur. Bu atributlara nümunə olaraq, təsvirdəki obyektin formasını, bəzi piksellərin intensivliyini və ya audio siqnalın amplitudasını göstərmək olar.

Bununla belə, digər domen sahələri təbii dilin emalı, qraflar və “Böyük Kod” da (mövcud kod əsasında avtomatik öyrənmə istifadə edərək) daxil olmaqla, ədədi olmayan atributlar vasitəsilə ifadə edilə bilər. Bu hallarda, atributlar ədədi olmaya bilər (məsələn, cümlədəki sözlər və ya qrafın tilləri).

Atributlar Sİ tərtibatçısına domen sahəsində mövcud bir nümunə əsasında başqa bir nümunə yaratmağa imkan verir. Atributlar dayanıqlıq spesifikasiyasında məhdudlaşdırılmalıdır.

5.3. Stabillik

5.3.1. Stabillik xassəsi

Stabillik xassəsi müəyyən domen sahəsində neyron şəbəkəsinin giriş verilənləri dəyişdirildikdə onun çıxış verilənlərinin dəyişməz qalma dərəcəsini ifadə edir. Davranışın sabit saxlanılması nəzərdə tutulan domen sahəsində stabilliyin yoxlanılması performansın da saxlanıla biləcəyini yoxlamağa imkan verir. Stabillik xassəsi ya qapalı formada (məsələn, “dəyişkənlik hər hansı sərhəd qiyməti çərçivəsindədirmi?”) və ya açıq formada (məsələn, “ən böyük stabil domen sahəsi hansıdır?”) ifadə edilə bilər.

Küylü giriş verilənləri ilə neyron şəbəkəsinin öz performansını saxladığını sübut etmək üçün stabillik xassəsi ifadə edilməlidir. Stabillik xassəsi gözlənilən davranış baxımından müəyyən müntəzəmlik xassələrinə malik domen sahələrində istifadə edilməlidir. Əksinə, stabillik xassəsi xaotik sistemlərdə məna kəsb etmədiyi üçün istifadə edilməməlidir. Hətta domen sahəsinin müntəzəmlik xassəsinin asanlıqla təsdiqlənməsi mümkün olmadıqda belə (məsələn, xaotik sistemdə), bu xassədən neyron şəbəkələrini müqayisə etmək üçün istifadə etmək olar.

5.3.2. Stabillik meyarı

Stabillik meyarı stabilliyin (xassə kimi) yalnız hər hansı nümunələr çoxluğu və ya domen sahəsinin altçoxluğunda (məsələn, təlim və ya validasiya verilənləri çoxluğu kimi) deyil, konkret domen sahəsi çərçivəsində saxlanıb-saxlanmadığını müəyyən edir. Stabillik meyarı 6.2-ci bəndə təsvir olunan formal metodlardan istifadə etməklə yoxlanıla bilər.

Stabillik meyarı ən azı ölçüldüyü domen sahəsinin qiymətlər fəzasını və çıxış qiymətləri fəzasını, həmçinin gözlənilən stabillik xassəsini müəyyən etməlidir.

Stabillik meyarı modellərin müqayisəsi meyarı kimi istifadə edilə bilər.

Modellərin müqayisəsində dəqiqliyin təmin olunması üçün aşağıdakı tələblər yerinə yetirilməlidir:

- neyron şəbəkələri eyni tapşırığı yerinə yetirir;
- stabillik meyarı eyni domen sahəsində istifadə olunur;
- stabillik meyarı eyni məqsədi təsdiq edir.

Məsələn, klassifikasiyanı həyata keçirən neyron şəbəkəsi üçün stabillik meyarı domen sahəsinin hər bir girişi üçün konkret həllin yerinə yetirildiyini qiymətləndirir. Reqressiyanı həyata keçirən neyron şəbəkəsi üçün stabillik meyarı reqressiyanın domen sahəsində sabit olub-olmamasını qiymətləndirir.

Stabillik meyarı tətbiq olunması üçün neyron şəbəkəsinin gözlənilən çıxış verilənləri haqqında əvvəlcədən mövcud məlumatlara əsaslanır. Bu məlumatlar Sİ tərtibatçısına məlum ola və ya başqa bir üsulla (simulasiya və ya “çözücü” sistemlərdən istifadə etməklə) müəyyən edilə bilər. Stabillik meyarı gözlənilən cavabın analoji olduğu domen sahəsində dayanıqlığın qiymətləndirilməsi üçün daha uyğundur. Bu səbəbdən, neyron şəbəkəsi tərəfindən emal edilən istənilən qərar qəbuletmə prosesində (məsələn, klassifikasiya və identifikasiya) stabillik meyarından istifadə etmək tövsiyə olunur.

5.4. Həssaslıq

5.4.1. Həssaslıq xassəsi

Neyron şəbəkəsinin həssaslıq xassəsi giriş verilənləri dəyişdirildikdə çıxış verilənlərinin dəyişmə dərəcəsini ifadə edir. Domen sahəsindəki dayanıqlığı qiymətləndirmək üçün bəzən sistemin dəyişkənliyini yoxlamaq lazımdır. Sistemin dəyişkənlik dərəcəsini və bu dəyişikliyə təsir edə biləcək giriş verilənlərini müəyyən etmək üçün həssaslıq təhlili aparıla bilər. Bu təhlil daha sonra sistemin gözlənilən performansına dair əvvəlcədən mövcud olan mühakimələrlə müqayisə edilir.

Həssaslıq analizi neyron şəbəkəsinin məhdud qaldığını müəyyən etmək üçün tətbiq olunarsa, domen sahəsində istifadə edilməlidir. Stabillik xassəsi ilə bağlı qeyd olunduğu kimi, həssaslıq analizi də bəzi müntəzəmlik xüsusiyyətlərinə malik olan domen sahələri üçün daha uyğundur.

5.4.2. Həssaslıq meyarı

Həssaslıq meyarı (hər hansı nümunələr çoxluğunda deyil) domen sahəsi üzrə xassəni ifadə etdiyi üçün o, 6.2-ci bənddə təsvir olunan formal metodlardan istifadə etməklə yoxlanıla bilər.

Həssaslıq meyarı ən azı onun ölçüldüyü domen sahəsini və yoxlanılmalı olan həssaslıq sərhədlərini müəyyən etməlidir.

Həssaslıq meyarı müxtəlif neyron şəbəkələri arxitekturalarını və ya təlim keçmiş modelləri müqayisə etmək üçün istifadə edilə bilər. Dəqiq müqayisə üçün aşağıdakı tələblər yerinə yetirilməlidir:

— neyron şəbəkələri eyni tapşırığı yerinə yetirir;

— həssaslıq meyarı eyni domen sahəsində istifadə edilir;

— həssaslıq meyarı eyni məqsədi təsdiqləyir.

Həssaslıq meyarı xüsusilə interpolyasiya və ya reqressiya tapşırıqlarını yerinə yetirən neyron şəbəkələri üçün daha uyğundur. Bu tip tapşırıqlar üçün həssaslıq meyarı bütün domen sahəsi üzrə “etalon/mütləq həqiqət”in (“ground truth”) əksinə olaraq, birbaşa isbat əldə etməyə imkan verir.

Həssaslıq meyarı adətən giriş verilənlərinin dəyişkənliyinin müəyyən domen sahəsində variasiya sərhədi kimi qapalı formada ifadə edilir.

5.5. Relevantlıq

5.5.1. Relevantlıq xassəsi

Neyron şəbəkəsində relevantlıq xassəsi giriş verilənlərinin çıxış verilənlərinə təsirinin sıralanmasını ifadə edir. Hər bir çıxış üçün relevantlıq hesablana bilər. Relevantlıq hər bir girişin çıxış əsasında əldə edilən nəticəyə fərdi təsirini ifadə edir. Hər bir girişin hər bir çıxışa fərdi təsiri nizamlı şəkildə çeşidlənə bilər. Relevantlıq xassəsi nəticədə alınmış sıralamanın Sİ tərtibatçı tərəfindən təqdim olunan sıralama tələblərinə cavab verdiyini yoxlayır. Hər bir girişin təsirini qiymətləndirmək üçün relevantlıq xassəsi müxtəlif üsullardan istifadə etməklə yoxlanıla bilər. Stabillik və həssaslıq xassələrindən fərqli olaraq, relevantlıq xassəsi onun qiymətləndirilməsinə cavabdeh olan ekspertlər arasında mübahisəyə səbəb ola bilər. Həqiqətən də, iki neyron şəbəkəsi relevantlıq xassəsi üzrə tamamilə fərqli qiymətlərə malik ola bilər, lakin hər iki nəticə məqbul hesab ediləcək. Ziddiyyətli nəticələrin həlli metodu müqayisə protokoluna daxil edilməlidir. Məsələn, protokol müəyyən situasiyada ziddiyyəti aradan qaldırmaq üçün səsvermə sistemindən istifadə edə bilər.

Neyron şəbəkəsinin insanın yerinə yetirə biləcəyi hər hansı bir işi icra etdiyi hallarda relevantlıq xassəsindən istifadə edilməlidir. Bu hallarda neyron şəbəkəsinin çıxışındakı nəticələrin əsaslandırmasının başa düşülməsi və yoxlanılması zəruridir. Relevantlıq xassəsi sistemin işinə düzgün mühakimələrlə zəmanət verilib-verilməyəcəyini müəyyənləşdirir. Bu zaman sistemin dayanıqlığı yalnız bəyan edilməyəcək, həm də əsaslandırılacaqdır. Bu yoxlama insan-operator tərəfindən əl ilə və ya əvvəllər yoxlanılmış etalonlardan istifadə etməklə avtomatik olaraq həyata keçirilə bilər.

5.5.2. Relevantlıq meyarı

Relevantlıq meyarı domen sahəsində hər bir giriş və çıxış arasındakı əlaqənin “nümayişini” tələb edən relevantlıq xassəsini ifadə edir. Bunun üçün hər bir girişin təsirini ayıra bilən bir üsul tələb olunur. Bu məqsədə çatmaq üçün simvolik hesablamalara, məntiqi hesablamalara və ya hesablama metodlarına əsaslanan formal metodlardan istifadə edilə bilər. Relevantlıq meyarını yoxlamaq üçün 6.2-ci bənddə təsvir olunan formal metodların nümunələrindən istifadə etmək olar.

Relevantlıq meyarı ölçüldüyü domen sahəsini və gözlənilən nəticələri əks etdirməlidir. Əgər gözlənilən nəticələri əvvəlcədən (apriori) müəyyən etmək mümkün deyilsə, relevantlıq meyarı ən azı nəticələrin qiymətləndirilməsi metodologiyasını təmin etməlidir.

Relevantlıq meyarı müxtəlif neyron şəbəkə arxitekturalarını və ya təlim nəticələrini müqayisə etmək üçün istifadə edilə bilər. Dəqiq müqayisə üçün aşağıdakı tələblər yerinə yetirilməlidir:

— neyron şəbəkələri eyni tapşırığı yerinə yetirir;

— relevantlıq meyarı eyni domen sahəsində istifadə edilir;

— relevantlıq meyarı eyni məqsədi təsdiqləyir.

Nümunə: Klassifikasiya tapşırığını yerinə yetirən neyron şəbəkəsində ən relevant piksellərin identifikasiya olunan obyektin konkret hissəsində (məsələn, nəqliyyat vasitəsini identifikasiya etmək üçün onun təkərləri) yerləşib-yerləşmədiyini yoxlamaq üçün relevantlıq meyarı istifadə edilə bilər. Zaman seriyasının proqnozlaşdırıcı təhlilini həyata keçirən neyron şəbəkəsində proqnozlaşdırılan hadisənin Sİ tərtibatçısı üçün məqbul olan məntiqi nəticələrə uyğun olub-olmadığını yoxlamaq üçün relevantlıq meyarı istifadə edilə bilər (məsələn, tezliklə sıradan cıxacaq mühərriyin həddən artıq qızması barədə xəbərdarlıq siqnalı).

Nəticənin Sİ tərtibatçısı tərəfindən təhlil olunması şərtilə, relevantlıq meyarı müxtəlif tapşırıqlar üçün ifadə edilə bilər. Relevantlıq meyarı, məsələn, klassifikasiya, aşkarlama, interpolyasiya və ya reqressiya tapşırıqlarında istifadə edilə bilər. Relevantlıq meyarının yoxlanılması avtomatlaşdırıla və ya əldə edilən nəticənin məqbul olduğunun müəyyənləşdirilməsi üçün insan tərəfindən qiymətləndirməyə əsaslana bilər. Yoxlama insan tərəfindən qiymətləndirməyə əsaslandıqda, verilmiş qərar testlərin mümkün dərəcədə avtomatlaşdırılmasını nəzərdə tutan yeni tələb kimi təqdim oluna bilər.

5.6 Əlçatarlıq

5.6.1. Əlçatarlıq xassəsi

Neyron şəbəkəsinin əlçatarlıq xassəsi şəbəkənin çoxmərhələli performansını (məhsuldarlıq) əməliyyat (istismar) mühiti ilə birlikdə ifadə edir. Bu tip xassə Şəkil 1-də təqdim olunan agent paradiqmasında işləyən sistemlərə tətbiq olunur. Əlçatarlıq xassəsi verilmiş mühütdə neyron şəbəkəsindən istifadə edərkən Sİ agentinin özünü idarə etmək üçün bir sıra vəziyyətlərdə olmağın mümkünlüyünü yoxlayır. Əlçatarlıq xassəsi Sİ agentinin qarşılaşmalı olmadığı uğursuz vəziyyətlər çoxluğunu və ya Sİ agentinin çatmalı olduğu hədəf vəziyyətlər çoxluğunu müəyyən edə bilər.

Bu tip xassənin ifadə edilməsi Sİ agentinin növbəti vəziyyətə təsirini təsvir edən mühit modelinin müəyyənləşdirilməsini tələb edir. Ətraf mühit ya deterministik, ya da stoxastik

şəkildə inkişaf edə bilər. Deterministik mühit üçün əlçatarlıq xassəsi Sİ agentinin müəyyən vəziyyətlər çoxluğuna çatmasının mümkün olduğunu ifadə edir.

5.6.2. Əlçatarlıq meyarı

Əlçatarlıq meyarı verilmiş ilkin vəziyyətlər çoxluğu üzərində əlçatarlıq xassəsini ifadə edir. Deterministik mühitlər üçün bu meyar 6.2.4-cü yarımbənddə təsvir olunan metodlardan istifadə etməklə yoxlanıla bilər. Stoxastik mühit üçün isə əlçatarlıq meyarı vəziyyətlər çoxluğunda olmağın mümkünlüyünün ehtimalını ifadə edir. Bu ehtimal 6.2.5-ci yarımbənddə təsvir olunan metodlardan istifadə etməklə müəyyən edilə bilər.

Əlçatarlıq meyarı verilmiş ilkin vəziyyətlər çoxluğu üçün ödəlinməlidir. İlkin vəziyyətlər çoxluğu isə meyarın bir hissəsi kimi verilə bilər. Neyron şəbəkəsinin meyarın ödənildiyi ilkin vəziyyətlər çoxluğunu müəyyən etmək üçün, alternativ olaraq, formal metodlardan istifadə etmək olar. Neyron şəbəkəsinin qiymətləndirilməsi üçün əlçatarlıq meyarından istifadənin üstünlüyü ondan ibarətdir ki, bu meyar qapalı mühitdə şəbəkənin performansının metrikasını təqdim edir. Ona görə də giriş-çıxış xassələrinin çərçivəsindən kənara çıxan yüksək səviyyəli təhlükəsizlik xassələrini ifadə etmək üçün bu meyardan istifadə edilə bilər.

Məsələn, təyyarələrin toqquşmasının qarşısının alınması sistemlərində tətbiq olunan neyron şəbəkələrində əlçatarlıq meyarı ətraf mühitin konkret modelini nəzərə alaraq, toqquşma vəziyyətləri çoxluğu ilə qarşılaşmamaq tələbini ifadə edə bilər.

6 NEYRON ŞƏBƏKƏLƏRİNDƏ FORMAL METODLARIN TƏTBİQİ

6.1. Nəzərdən keçirilən neyron şəbəkələrinin növləri

6.1.1. Neyron şəbəkələrinin arxitekturaları

6.1.1.1. Ümumi müddəalar

Neyron şəbəkələri müxtəlif növ arxitekturalardan istifadə etməklə layihələndirilə və işlənə bilər. Neyron şəbəkələrinin formal verifikasiya metodları onların arxitekturasından asılıdır. 6.1.1-ci yarımbənd aşağıdakı arxitekturalar üçün işlənmiş formal metodları təsvir edir: hissə-hissə xətti neyron şəbəkələri, binarlaşdırılmış neyron şəbəkələri, rekkurent neyron şəbəkələri və transformer şəbəkələr. Bu siyahı tam olmasa da, yeni arxitekturalar və müvafiq formal verifikasiya metodları yarana bilsə də, bu siyahı çoxsaylı mövcud neyron şəbəkə arxitekturalarını və tətbiq olunan metodları əhatə edir. Qeyd olunan metodlar haqqında daha ətraflı məlumat 6.2-ci bənddə təqdim olunur.

6.1.1.2. Hissə-hissə xətti neyron şəbəkələri

Hissə-hissə xətti neyron şəbəkələri (PLNN) [12] siqmoid və ya tangens kimi qeyri-xətti funksiyalardan istifadə etmir. PLNN tam əlaqəli və ya konvolyusiya layları kimi xətti transformasiyalardan, MaxPooling kimi birləşdirmə laylarından və hissə-hissə xəttiliyi saxlayan paket normallaşdırılması və ya “ələnmə” (dropout) kimi əməliyyatlardan istifadə edə bilər. Müasir neyron şəbəkələrin əksəriyyəti PLNN-dir.

[13]-də əvvəlcə PLNN-i riyazi ekvivalent xətti klassifikatorlar çoxluğuna çevirən və sonra hər bir xətti klassifikatoru onun proqnozunda üstünlük təşkil edən xüsusiyyətlərə görə interpretasiya edən formal verifikasiya metodları təklif olunur. Digər verifikasiya metodları PLNN-ə qlobal optimallaşdırma məsələsi kimi baxır və bu metoddan “modul üzrə qaneolma nəzəriyyəsinin” (satisfiability modulo theories, SMT) çözücüsü kimi istifadə edir. [14]-də dayanıqlığın formal verifikasiyası hətta qarışıq tam ədədli xətti proqram şəklində təqdim olunur. Digər metodlar ISO/IEC TR 24029-1 standartında təqdim olunur. Əlavə verifikasiya metodlarına Fast-Lin–Fast-Lip [15], CROWN [16] və formal təhlükəsizlik təhlili [17] daxildir.

6.1.1.3. Binarlaşdırılmış neyron şəbəkələri

Binarlaşdırlmış neyron şəbəkələrində (BNN) bütün əməliyyatlar binardır və bu da ikilik matrislər üzərində sürətli vurma əməliyyatını aparmaq üçün xüsusi alqoritmlərdən istifadə etməyə imkan verməklə, bu şəbəkələri yaddaş və hesablama gücü baxımından səmərəli edir [18]. Belə bir arxitekturadan istifadə etməklə təsvirin klassifikasiyasından tutmuş obyektlərin aşkarlanmasına qədər müxtəlif quraşdırılmış tətbiqlər yaradılmışdır.

Belə BNN-lərin formal verifikasiyası BNN-in Bul düsturu şəklində dəqiq təqdimatını yaratmaqla əldə edilmişdir, belə ki, verilmiş şəbəkənin bütün mümkün giriş və çıxış cütləri Bul düsturunun həlləridir [19]. Daha sonra verifikasiya Bul qənaətbəxşliyi (“Boolean satisfiability”) və tamqiymətli xətti proqramlaşdırma kimi metodlardan istifadə etməklə həyata keçirilir [18].

6.1.1.4. Rekkurent neyron şəbəkələri

Rekkurent neyron şəbəkələri (RNN) nitq, maliyyə və mətn daxil olmaqla bir çox domen sahələrində ardıcıl verilənlərin dəqiq və səmərəli işlənməsinə imkan verir. Hər bir zaman pilləsində RNN bu addımdakı giriş verilənləri və əvvəlki addımlardakı daxili vəziyyəti əsasında öz daxili vəziyyətini yeniləyir. Son nəticə bütün giriş verilənləri ardıcıllıqla emal etdikdən sonra əldə edilir.

Son klassifikator kimi istifadə edilən rekkurent neyron şəbəkəsinə sonsuz sayda vəziyyətə malik maşın kimi baxmaq olar [20]. Sonsuz sayda vəziyyətə malik belə bir sistem üçün sonlu avtomat mövcud sistemi aproksimasiya edən kölgə modeli kimi avtomatlaşdırılmış təlim metodlarından istifadə etməklə öyrədilə bilər. Daha sonra kölgə modeli, məsələn, modelin yoxlanılması üsullarından istifadə etməklə, RNN-in öz spesifikasiyasına uyğunluğunu yoxlamaq üçün istifadə edilə bilər. Modelin yoxlanılmasına əlavə olaraq, təsvir, audio və hərəkət sensoru məlumatlarının təsnifatlaşdırılmasında istifadə olunan RNN-lərin lokal dayanıqlığını sübut etmək üçün abstrakt interpretasiyadan istifadə edilə bilər [21].

6.1.1.5. Transformer şəbəkələri

Transformer şəbəkələri koder-dekoder arxitekturasına malik dərin öyrənmə şəbəkələri ola bilər [22]. Transformer ilkin olaraq koder vasitəsilə giriş verilənlərinin hər bir fərqli hissəsi üçün təqdimatlar (“representations”) və ya ardıcıllıqlar generasiya edir. Bu zaman transformer giriş üçün yeni daxili təqdimatın generasiyası məqsədilə girişin bütün digər hissələrindən məlumatları aqreqasiya etmək üçün “özünədiqqət” ("self-attention”) mexanizmindən istifadə edir. Sonra bu addım ardıcıl olaraq yeni daxili təqdimatlar generasiya etməklə paralel olaraq giriş verilənlərinin bütün hissələri üçün dəfələrlə

təkrarlanır. Dekoder analoji şəkildə işləyir, bir dəfəyə çıxış verilənlərinin bir hissəsini generasiya edir. Bu zaman dekoder əvvəl generasiya edilmiş çıxış verilənlərinin digər hissələrinə diqqət yetirir və həmçinin koderin generasiya etdiyi daxili təqdimatları da nəzərə alır. Beləliklə, transformerlər çarpaz qeyri-xəttilik və çarpaz mövqe asılılığı da daxil olmaqla, verifikasiya prosesi üçün çətinliklər yaradan mürəkkəb "özünədiqqət" laylarına malikdir. "Özünədiqqət" layları transformerlərin dayanıqlığının verifikasiyası baxımından ən mürəkkəb hissələrdir.

[23]-də transformerlərin dayanıqlılığını formal olaraq verifikasiya etmək üçün metod təklif olunur. Transformer layı altlaylara ayrılır, hər altlayda həmin altlaydakı neyronlar üzərində müəyyən əməliyyatlar icra edilir. İcra olunan əməliyyatlar üç kateqoriyaya bölünür:

— xətti çevirmələr;

— unar qeyri-xətti funksiyalar;

— "özünədiqqət" mexanizminə dair əməliyyatlar.

Hər altlay *n* mövqedən ibarət ardıcıllıq kimi nəzərdə tutulur, belə ki, hər mövqe neyronlar qrupunu ehtiva edir. Bu mövqelərin hər biri üçün sərhədlər birinci altlaydan sonuncu altlaya qədər hesablanır.

6.1.2. Neyron şəbəkələrinin giriş verilənlərinin tipləri

6.1.2.1. Ümumi müddəalar

Neyron şəbəkələri bir neçə mümkün çıxış verilənlərinin tiplərini əldə etmək məqsədilə müxtəlif giriş verilənlərinin tiplərinin emalı ilə bağlı tapşırıqları yerinə yetirə bilər (bax: 6.1.1). Neyron şəbəkələri tətbiqləri təsvir, zaman sıraları, təbii dil, qraflar və ya cədvəl kimi verilənlər tipləri ilə işləyir. Təqdim olunan siyahı tam olmasa da, həmçinin yeni tətbiqlər yarana bilsə də, bu siyahı neyron şəbəkələri tərəfindən emal edilə bilən verilənlər tiplərinin əksər hissəsini əhatə edir.

Bununla belə, formal metodlarda verilənlərin effektiv və səmərəli şəkildə təhlil oluna bilən tipinin müəyyənləşdirilməsi ilə bağlı məhdudiyyətlər ola bilər. Ümumiyyətlə, məhdudiyyətlərə aşağıdakılar aiddir:

— giriş verilənlərinin (və nəticə etibarilə, şəbəkənin) ölçüsündən asılı olaraq, onlara dair hesablamaların miqyaslılığı;

— perturbasiyaların modelləşdirilməsi üçün giriş verilənlərinin mahiyyəti.

Formal metodların istifadəsi zamanı miqyaslılıqla bağlı məhdudiyyətlərin tətbiqi geniş yayılmışdır. Neyron şəbəkəsi eyni zamanda bir neçə giriş vektorunun emalı üçün layihələndirilmişdir. Lakin bütün domen sahəsi üzrə (yəni hər bir giriş vektoru üzrə) müəyyən bir xassənin riyazi isbatı bu sahə daxilində bəzi nöqtədəki nəticələri hesablamaqdan mahiyyətcə daha çətindir.

İkinci məhdudiyyət giriş verilənləri əsasında təqdim olunan domen sahəsinin mahiyyətindən və bu sahə üzrə formal isbatları modelləşdirmək imkanından irəli gəlir. Bu məhdudiyyət domen sahəsini təsvir etmək üçün istifadə olunan atributlar anlayışından asılıdır (bax: 5.2). Bəzi hallarda atributlar ədədlərlə ifadə olunduğu üçün atributların bəzi önəmli variasiyalarını asan modelləşdirilə bilər.

6.1.2.2. Təsvir verilənləri

Neyron şəbəkələrinin əldə etdiyi son uğurlarının səbəblərindən biri də onların təsvirləri emal etmək imkanının mövcud olmasıdır. Müxtəlif ayırdetmə dəqiqliklərinə malik bəzi növ təsvirləri (məsələn, kamera, MRT, radar, sonar və SAR) və hətta video axınlarını emal imkanı onların geniş miqyasda tətbiqinə şərait yaratmışdır.

Formal metod nöqteyi-nəzərindən, əhatə olunan giriş fəzası massivin ölçüsünün hər pikselin ölçülərinin sayına vurulması ilə müəyyən edilir. Böyük təsvirlərin emalı bu baxımdan (hər girişin hər bir ölçüsü ilə konkret simvol əlaqələndirən) bir çox formal metodlar üçün çətin ola bilər.

Giriş fəzasında variasiyaları ifadə etmək üçün təsvirlər üçün bir neçə atribut müəyyən edilə bilər. Məsələn, təsvirin işıqlandırılması piksellərin intensivliyinin dəyişməsi kimi ifadə edilə bilər. Ətraf mühitdəki dəyişikliklər daha mürəkkəb olsa da, analitik təyinetmənin mümkün olduğu təqdirdə, dəyişikliklər birbaşa piksellərin qiymətləri ilə ifadə edilə bilər ki, bu da formal metodların birbaşa tətbiqinə imkan verir. Analitik təyinetmə mümkün olmadıqda, dəyişiklik təsvirə tətbiq olunan modelin approksimasiyası vasitəsilə (məsələn, təsvirdə maskadan istifadə etməklə) ifadə edilə bilər.

6.1.2.3. Zaman sıraları verilənləri

Proqnozlaşdırma texnologiyaları üzrə son nailiyyətlər proqnozların verilməsi və ya klassifikasiyanın aparılması məqsədilə zaman sıraları verilənlərinə neyron şəbəkəsinin tətbiq edilməsini nümayiş etdirir. Hər bir zaman sırası (adətən) eyni tipli verilənlərə aid məlumatları qeydə alan (“record”) bir neçə elementdən (nümunədən) ibarət olur. Formal metodlar hər bir elementdə saxlanılan verilənlərin tipini təhlil etməyin mümkünlüyü şərti daxilində zaman sıralarına tətbiq oluna bilər. Bu halda, giriş verilənlərinin ölçüsü hər bir zaman sırasının uzunluğunun verilənlərin hər bir elementinin ölçüsünə vurulmasının hasilinə bərabərdir.

Zaman sıraları verilənlərinə tətbiq edilməsi üçün formal metodlar hər bir elementdəki məlumatların manipulyasiya oluna bilməsini tələb edir. Giriş verilənlərinin sayını idarə etmək çətindir, çünki hər bir element müstəqil olaraq nəzərdən keçirilərsə, giriş verilənlərinin sayı ixtiyari qədər böyük ola bilər.

6.1.2.4. Təbii dil verilənləri

Mətn və nitqə əsaslanan təbii dil verilənlərinin tipləri neyron şəbəkələri tərəfindən emal edilə bilər. Məsələn, ağıllı audio qurğularının geniş miqyasda tətbiqi və dil modellərinin asanlıqla başa düşülən mətn yaratmaq imkanı bunu nümayiş etdirir. Təbii dil verilənləri çox vaxt neyron şəbəkəsinə ötürülməzdən əvvəl ilkin emal prosesindən keçir.

Giriş verilənlərinin bəzi variasiyalarını formal olaraq (məsələn, yazılara (“record”) küy əlavə etməklə) asanlıqla yaratmaq olar. Digər variasiyaları ifadə etmək daha çətin ola bilər (məsələn, semantikanı dəyişdirmədən cümlədən bir sözü çıxarmaq və ya söz əlavə etmək yaxud eyni dilin müxtəlif dialektləri üçün cümlədə fərqli semantikanı nəzərə almaq). Bu

zaman tətbiq olunan formal metodlar həm ilkin emal konveyeri, həm də neyron şəbəkəsi ilə əlaqədardır [21].

6.1.2.5. Qraf verilənləri

Qraf neyron şəbəkələri (GNN) molekulyar biologiyada, dələduzluğun aşkarlanmasında və sosial elmlərdə təpələrin klassifikasiyası, əlaqələrin proqnozlaşdırılması və qrafların klassifikasiyası kimi müxtəlif tapşırıqların yerinə yetirilməsi məqsədilə qraf verilənlərinin emal edilməsi üçün geniş tətbiq edilir. GNN-lərin bəzi dayanıqlıq xassələri təpələrin xüsusiyyətlərinin perturbasiyası, eləcə də tillərin əlavə edilməsi və ya silinməsi kimi strukturlaşdırılmış informasiyanın perturbasiyası əsasında müəyyən edilir. Xüsusiyyətlərə əsaslanan perturbasiyalar kəsilməz olsa da və təsvirlərdəki piksel intensivliyinin dəyişmələrinə analoji olaraq formal emal oluna bilsə də, strukturlaşdırılmış perturbasiyalar diskretdir və buna görə də, xüsusi formal üsulların işlənməsini tələb edir.

6.1.2.6. Cədvəl verilənləri

Maliyyə, səhiyyə və logistika kimi bir çox tətbiq sahələri bu tətbiqlərə müxtəlif tip verilənləri (məsələn, ədədi, simvol, mətn, kateqorial) kombinə etməyə və elementlər arasında əlaqələri ifadə etməyə imkan verən cədvəl verilənlərinə daha çox əsaslanır. Cədvəl verilənlərində sətirlərin sayının həddən çox olması mümükündür, həm də bəzən hər sətir çətin təxmin edilə bilən dispersiyaya malik ola bilər.

6.1.2.2-6.1.2.5-ci yarımbəndlərdə təsvir edilmiş heterogen verilənlər tipli cədvəllərə formal metodların tətbiqi yuxarıda qeyd olunan məhdudiyyətlərə səbəb ola bilər.

6.2. Tətbiq olunan formal metodların növləri

6.2.1. Ümumi müddəalar

6.2.1.1. Tətbiq edilən formal metodların növlərinin icmalı

6.2-ci bənddə neyron şəbəkələrinin dayanıqlığının qiymətləndirilməsi üçün tətbiq oluna bilən mövcud formal metodlar təsvir edilir. Bu metodlar aşağıdakı meyarlar əsasında klassifikasiya edilə bilər:

— onlar tam və ya natamam ola bilər;

— onlar deterministik və ya qeyri-deterministik ola bilər;

— onlar “şüşə qutu” və ya “qapalı qutu” test metodlarından istifadə edə bilər;

— onlar həqiqi ədədlər hesabı və ya (həqiqi ədədlərin) kompüter hesabına əsaslana bilər.

6.2.1.2. Tam və natamam verifikatorlar

Tam verifikatorlar dəqiq cavablar təmin edə bilər. Onlar ya dayanıqlıq xassəsini sübut edir, ya da xassənin pozulmasını nümayiş etdirən konkret kontr-nümunə təqdim edir. Tam verifikatorların məhdudiyyəti ondan ibarətdir ki, onlar mürəkkəb verilənlər çoxluqları üçün yüksək dəqiqliyi təmin edən neyron şəbəkələrinin dayanıqlığını verifikasiya etməkdə effektiv deyil. Bunun əksinə olaraq, natamam verifikatorlar yüksək dəqiqlikli neyron şəbəkələrinə

qədər miqyaslanan abstraksiya texnikalarından istifadə edir. Lakin natamam verifikatorlar dayanıqlıq xassəsinin həqiqətən də saxlandığını sübut etməyə bilər.

6.2.1.3. Deterministik və qeyri-deterministik verifikatorlar

Deterministik verifikator dayanıqlıq xassəsini sübut etdikdə konkret giriş sahəsi çərçivəsində hər giriş üçün bu xassə saxlanılır. Bununla belə, ehtiyatların proqnozlaşdırılması, nitqin tanınması və təsvirlərin generasiyası kimi müxtəlif domen sahələrində tətbiq edilən qarışıq sıxlıq şəbəkələri və ya variasiya avtokodlaşdırıcıları kimi müəyyən modellər deterministik çıxış deyil, paylanma generasiya edir. Bu cür şəbəkələr üçün formal metodlardan ya bütün giriş verilənləri üçün ödənilən çıxış verilənləri paylanmasının parametrlərini deterministik qaydada hesablamaq (məsələn, orta qiymət və ya standart sapma) və ya onların dayanıqlığına yüksək ehtimalla rəsmi zəmanət vermək məqsədilə istifadə edilə bilər.

6.2.1.4. "Şüşə qutu" testindən istifadə edən verifikatorlar (model nəzərə alınmaqla) və ya "qapalı qutu" testindən istifadə edən verifikatorlar (model nəzərə alınmamaqla)

"Şüşə qutu" testindən istifadə edən verifikatorlar üçün arxitektura və öyrənilmiş parametrlər də daxil olmaqla, modelin (yəni şəbəkənin daxili təqdimatına) əlçatan olması tələb edilir. Lakin bu verifikatorlar təlim verilənlərinin və ya neyron şəbəkəsini təlimləndirmək üçün istifadə olunan alqoritm in əlçatan olmasını tələb etmirlər. Quraşdırılmış modelin əlçatan olmadığı (məsələn, şifrlənmiş olduğu) domen sahələrində "şüşə qutu" testindən istifadə edən verifikatorlar tətbiq edilmir. Belə hallarda "qapalı qutu" testindən istifadə edən verifikatorlardan istifadə edilə bilər. "Qapalı qutu" testindən istifadə edən verifikatorlar modelin yalnız seçilmiş giriş verilənləri üzrə işləmə qabiliyyətinin olmasını tələb edir. Bu isə "qapalı qutu" testindən istifadə edən verifikatorların dəqiqliyinin "şüşə qutu" testindən istifadə edən verifikatorlara nisbətən daha az olmasına səbəb ola bilər.

6.2.1.5. Həqiqi ədədlər hesabı və ya kompüter hesabı verifikatorları

Əksər verifikatorlar neyron şəbəkəsində hesablamaların ideal həqiqi ədədlər hesabı əsasında (yəni yuvarlaqlaşdırma xətası olmadan) aparıldığını fərz edir. Buna görə də, verifikatorların dayanıqlığa dair zəmanətləri sürüşən vergüllü hesab qaydalarına və ya digər qeyri-standart kompüter hesabı əsasında həyata keçirilən faktiki hesablamalara şamil edilmir. Bunun əksinə olaraq, səs verifikatorları (tətbiq olunan hesab qaydalarına nəzərən) kompüter hesabının semantikasını nəzərə alır və onların çıxış verilənlərinin neyron şəbəkəsinin (bu semantika çərçivəsində mümkün olan) çıxış verilənlərini əks etdirdiyinə zəmanət verir. Bəzi hallarda verifikatorlar hesablamalar ardıcıllığındakı dəyişiklikləri də nəzərə ala bilər (məsələn, IEEE 754:2019 [24] standartına uyğun olaraq, yalnız düzgün yuvarlaqlaşdırılmış operatorlar istifadə edildikdə). Düzgün yuvarlaqlaşdırılmış operatorlardan (IEEE 754:2019) istifadə edilmədikdə, verifikator hər bir operator üçün həyata keçirilən yuvarlaqlaşdırmanı approksimasiya edə bilər.

6.2.2. Çözücü

Qismən tamqiymətli xətti proqlaşdırma (MILP) çözücüləri [25] və qaneetmə modulu nəzəriyyələri (SMT) çözücüləri [11] [26] deterministik, "şüşə qutu" və adətən tam verifikasiya metodlarına aid edilir. Onlar verilən neyron şəbəkəsinin bütün hesablamalarını

məhdudiyyətlər çoxluğu kimi kodlaşdırır və sonra bu məhdudiyyətlərdən dayanıqlıq xassəsini sübut etmək üçün istifadə edir.

Tam verifikasiya metodları əlçatan olmadıqda, natamam verifikasiya metodlarından istifadə oluna bilər. Bəzi qeyri-xətti aktivləşdirmə funksiyaları (məsələn, siqmoid və tangens də daxil olmaqla hiperbolik funksiyalar) dəqiq kodlaşdırılma üçün çox mürəkkəbdir. Buna görə də, çözücülər onları səs abstraksiyaları ilə approksimasiya edirlər. Digər qeyri-xətti aktivləşdirmə funksiyaları (məsələn, ReLU) dəqiq kodlaşdırıla bilər.

Verilmiş dayanıqlıq xassəsini sübut etmək üçün neyron şəbəkəsi və girişdəki məhdudiyyətlər qismən tamqiymətli xətti proqlaşdırma məsələsi kimi kodlaşdırılır ki, bu da daha sonra dayanıqlığın məhdudiyyətini optimallaşdırmaq üçün istifadə oluna bilər. Əgər dayanıqlığın məhdudiyyət sərhədləri məhdudiyyətləri ödəyirsə, dayanıqlıq xassəsi sübut olunmuş hesab edilir. SMT çözücüləri verifikasiya problemini məhdudiyyətin ödənilməsi məsələsi (ya təmin olunur, ya da olunmur) kimi qoyur.

Bəzi üsullar neyron şəbəkəsinin giriş verilənləri arasında zəif asılılıqları izləməklə çıxış verilənləri üçün daha sərt sərhədləri hesablayan simvolik xətti relaksasiyadan, sonra isə təhlükəsizlik xassəsinin verifikasiyası üçün istiqamətlənmiş məhdudiyyət dəqiqləşdirməsindən (ilkin və ya aralıq neyronlar çoxluğunun bölünməsi yolu ilə çıxış relaksasiyasının dəqiqləşdirilməsi) istifadə edir [27]. Digər üsullar neyron şəbəkəsinin idarə etdiyi avtonom sistemlərin sonlu abstraksiyalarını hesablamaq üçün SMT-əsaslı ilkin emal ilə birgə qabarıqlıq modulu üzrə qaneetmə (SMT) əsasında alqoritm təklif edir [28].

6.2.3. Absrakt interpretasiya

Absrakt interpretasiya böyük və mürəkkəb deterministik [29], ehtimal [30] sistemləri miqyaslana bilən şəkildə təhlil etmək üçün ümumi strukturdur. Neyron şəbəkələri kontekstində ondan böyük neyron şəbəkələrinin dayanıqlığını verifikasiya edə bilən natamam, deterministik və “şüşə qutu” test metodunu təmin etmək üçün istifadə olunur. Verifikasiya prosesi aşağıdakı kimidir:

— ilk növbədə təqdim edilən test giriş verilənləri və dayanıqlıq spesifikasiyası birlikdə dayanıqlıq spesifikasiyası əsasında giriş verilənlərini dəyişdirməklə əldə edilə bilən bütün mümkün perturbasiyalı girişləri ehtiva edən sahəni (“region”) müəyyən edir. Bu sahə düzbucaqlı paralelepipedlər (“box”), zonotoplar və çoxüzlülər kimi müəyyən həndəsi fiqurlardan istifadə etməklə, ya da neyron şəbəkələrinin fərdi abstrakt domen sahələri kimi dəqiq və ya təqribi şəkildə təsvir oluna bilər [31];

— bu sahə daha sonra neyron şəbəkəsi üzrə elə paylanır ki, hər bir lay ardıcıl olaraq giriş sahəsinə tətbiq edilir. Giriş sahəsi bu sahədən əlçatan olan bütün çıxış verilənlərini ehtiva edən çıxış sahəsinə çevrilir. Laydan asılı olaraq bu, approksimasiyalara (giriş sahəsindən əlçatan olmayan çıxış verilənləri) səbəb ola bilər;

— nəticədə dayanıqlıq spesifikasiyasına uyğun olaraq, çıxış sahəsi perturbasiyalı giriş verilənləri üçün şəbəkənin bütün mümkün çıxış verilənlərini əhatə edir.

Abstrakt interpretasiyada dəqiqlik və miqyaslılıq arasında özünəməxsus bir kompromis mövcuddur. Məsələn, düzbucaqlı paralelepipedlər kimi sadə abstrakt domen sahələri bir neçə saniyə ərzində milyonlarla neyrona malik şəbəkələri verifikasiya edə bilsə də, bunlar

arzuolunan dayanıqlıq xassələrinin verifikasiyası üçün çox vaxt dəqiq nəticə vermir. Digər tərəfdən, yarı-müəyyən relaksasiyalar daha dəqiq olsalar da, böyük şəbəkələrə qədər miqyaslana bilmir. Ona görə də effektiv verifikasiyaya nail olmağın açarı bu kompromisin balanslaşdırılmasından ibarətdir.

6.2.4. Deterministik mühitlərdə əlçatarlığın təhlili

Neyron şəbəkəsinin əlçatarlığa əsaslanan verifikasiya üsulları verilmiş mühitdə işləyən neyron şəbəkələrinin qapalı dövrlü perfomansını təmin etmək üçün 6.2.2-ci yarımbənddə təsvir edilən çözücülərin çıxış verilənlərini əlçatarlığın təhlili metodları ilə birləşdirir. Bu təhlildə ilk addım giriş fəzasını xana adlanan çoxsaylı daha kiçik sahələrə bölməkdir. Hər bir xana üçün şəbəkənin təyin etdiyi sahədə mümkün idarəetmə çıxışlarını müəyyənləşdirmək məqsədilə 6.2.2-ci yarımbənddə təsvir olunan çözücülərdən istifadə oluna bilər. Bu məlumatı ətraf mühit modeli ilə birlikdə istifadə edərək, istənilən verilmiş xana üçün mümkün növbəti vəziyyətlər diapazonunun izafi approksimasiyasını ("overapproximation") müəyyən etməyə imkan verir. İlkin vəziyyət sahəsindəki bütün xanalar üçün bəzi zaman addımları ərzində bunu təkrarlamaqla, əlçatar vəziyyətlər çoxluğunun izafi approksimasiyasını müəyyən etmək olar [32]. Bu problemə digər yanaşma isə ətraf mühitin dinamikasının izafi approksimasiyasını qismən tamqiymətli xətti proqlaşdırma məsələsində məhdudiyyətlər kimi kodlaşdırmaqdan və əlçatar vəziyyətlər çoxluğunun izafi approksimasiyası məsələsini həll etmək üçün 6.2.2-ci yarımbənddəki qismən tamqiymətli verifikasiya üsulundan istifadə etməkdən ibarətdir [33].

6.2.5. Qeyri-deterministik mühitlərdə əlçatarlığın təhlili

Ətraf mühit stoxastik olduqda, 6.2.2-ci yarımbənddəki çözücülər vəziyyətlər çoxluğunun əlçatarlığının ehtimalını müəyyən etmək üçün ehtimal modelinin yoxlanılması üsulları ilə birləşdirilə bilər. 6.2.4-ci yarımbənddə təsvir edilən üsula anoloji olaraq, giriş verilənləri fəzası xanalar çoxluğuna bölünür və neyron şəbəkəsinin mümkün çıxış verilənlərini müəyyən etmək üçün hər bir xana çözücüdən "keçirilir". Ehtimal modelinin yoxlanılması dinamik proqramlaşdırmadan istifadə edərək verilmiş ilkin vəziyyətdən müəyyən vəziyyətlər çoxluğunun əlçatar olmasının ehtimalını müəyyən edir [34]. Bu yanaşmanı ayrı-ayrı giriş vəziyyətləri ilə deyil, xanalarla işləmək üçün uyğunlaşdırmaqla, neyron şəbəkəsindən istifadə edərkən vəziyyətlər çoxluğunun əlçatarlığının izafi approksimasiya ehtimalını əldə etmək mümkündür [35].

6.2.6 Model yoxlaması

Model yoxlaması – müəyyən bir interpretasiya çərçivəsində nəzəriyyənin formal ifadəsinin həqiqiliyinin ("valid") isbatı metodudur. Daha ətraflı məlumat ISO/IEC/IEEE 24765:2017 standartı və [36]-da təqdim edilir. Nəzəriyyə ideyanın nəzərdə tutulan semantikası haqqında müddəaları ərz edən cümlələr qurmaq üçün sabitlər, funksiyalar və predikatlardan ibarət simvollar lüğəti vasitəsilə ifadə edilir. Nəzəriyyə ya predikatlar məntiqinin cümlələri ilə, ya da verilənlərin obrazları (şablonları) ilə ifadə oluna bilər. Neyron şəbəkələri verilənlər obrazlarının modellərinin aşkarlanması və istifadəsi üçün nəzərdə tutulmuş alqoritmlər kimi başa düşülür. Verilənlər obrazlarının modeli onun giriş verilənlərinə uyğunluğu baxımından yoxlanılır.

Model yoxlamasının həqiqiliyinin əsaslandırılması üçün bütün modellər yoxlanılmalıdır. Neyron şəbəkələrində model yoxlaması müəyyən münasibətlərə tabe olan müxtəlif növ çoxluqlar arasında qarşılıqlı əlaqələri sübut etmək üçün istifadə edilə bilər.

Nümunə 1: “Ailə nəzəriyyəsi” [37] ailəyə mənsub şəxslərin ailə üzvü olmasını həyata keçirən interpretasiya əsasında qurulur. Beləliklə, iki ixtiyari şəxsin ailə üzvü olması və ya olmaması isbat edilir. Sonra “bir şəxs digər şəxsin valideynidir” cümləsi bütün mümkün cütlüklər üçün yoxlanılır.

Nümunə 2: Neyron şəbəkəsində rəqabətli giriş verilənlərinin mövcudluğunu isbat etmək üçün [38]-də model yoxlamasından istifadə edilir. Nəzəriyyə - hərflərdən tərtib edilən dilin və neyron şəbəkəsinin çəki əmsallarının və qərəzliliyinin təsviridir. İnterpretasiya hərfin təsvirinə əlavə edilmiş nişan əsasında təyin edilir. Əlifba üzrə hər bir mümkün hərf cütü arasındakı məsafəni hesablamaq mümkündür. Daha sonra model hər bir məsafənin Sİ tərtibatçısı tərəfindən müəyyən edilmiş həddən böyük olmasını ifadə edən predikata uyğunluğu baxımından yoxlanıla bilər. Sİ tərtibatçısı tərəfindən əvvəlcədən təyin edilmiş predikatların neyron şəbəkə vasitəsilə nəzəriyyəyə uyğunluğu yoxlanılır.

6.3. Xülasə

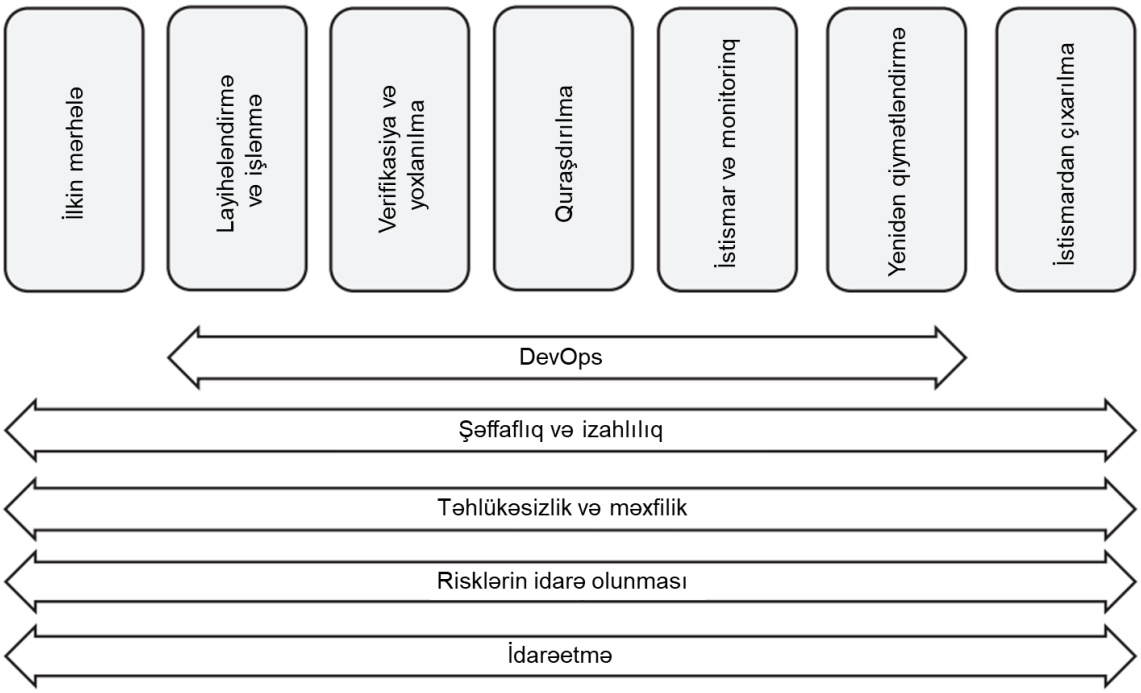
Neyron şəbəkələrinin dayanıqlığını qiymətləndirmək üçün formal metodların tətbiqinin mümkünlüyü bəzi amillərin nəzərə alınmasını tələb edir. Bir tərəfdən, neyron şəbəkəsinin arxitekturasının təsiri mövcuddur, çünki neyron şəbəkəsində istifadə olunan hər bir riyazi funksiyanı emal etmək üçün formal metodların hər biri güclü və zəif cəhətlərə malikdir. Digər tərəfdən, neyron şəbəkəsində giriş kimi istifadə edilən verilənlər növləri də müəyyən təsirə malik ola bilər, çünki giriş verilənlərinin dəyişkənliyi, ədədi və ya kateqorial mahiyyəti və ölçüsü hesablama xərclərinə və formal təhlilin asanlığına birbaşa təsir göstərir. Bu amillərin nəzərə alınması üçün bir neçə formal metod mövcuddur: “çözücü”dən istifadə edən yanaşmalar, abstrakt interpretasiya, vəziyyətin əlçatarlığının təhlili və ya model yoxlaması.

Hazırda ən çox yayılmış arxitekturalar və neyron şəbəkələri tərəfindən emal edilən verilənlər növləri ən azı bir formal metodla təhlil edilə bilər. Hər bir metod üstünlüklərə və məhdudiyyətlərə (məsələn, miqyaslanma qabiliyyəti) malikdir və 5-ci bölmədə təsvir olunan bir və ya bir neçə meyara cavab verə bilər.

7 HƏYAT DÖVRÜ ƏRZİNDƏ DAYANIQLIQ

7.1. Ümumi müddəalar

Sİ sisteminin AZE ISO/IEC 22989:2023 standartında təsvir edilmiş 7 mərhələdən ibarət həyat dövrü Şəkil 2-də təqdim olunur.



Şəkil 2 – Sİ sisteminin həyat dövrü modelinin mərhələləri və yüksək səviyyəli proseslərə dair nümunə

AZE ISO/IEC 22989:2023 standartının 5.19-cu bəndində Sİ provayderi, Sİ istehsalçısı, Sİ tərtibatçısı, Sİ müştərisi və s. daxil olmaqla, Sİ üzrə maraqlı tərəflərin rolları və altrolları çoxluğunu müəyyən edilir. Bu bənd həmin rollara istinad edir, layihələndirmə və işlənmə, verifikasiya və validasiya, quraşdırılma və istismar monitorinqi zamanı neyron şəbəkəsinin dayanıqlığının qiymətləndirilməsi ilə bağlı ətraflı məlumat verir.

7.2. Layihələndirmə və işlənmə mərhələsində dayanıqlığın qiymətləndirməsi

7.2.1. Ümumi müddəalar

Neyron şəbəkəsinin dayanıqlığının yoxlanılması hətta işlənmənin erkən mərhələsində belə, şəbəkənin layihələndirilməsinə kömək edə bilər. Dayanıqlıqla bağlı potensial çatışmazlıqları erkən mərhələdə müəyyən etməklə, Sİ tərtibatçısı daha sonra onları aradan qaldırmaq üçün lazımi addımlar ataraq, sonrakı yetərsiz dayanıqlıqdan yayına bilər (bu sənəddə dayanıqlıqla bağlı çatışmazlıqların aradan qaldırılması prosesi nəzərdən keçirilmir). Bu addımda təlim verilənlərinin və neyron şəbəkəsi arxitekturasının hələ də modifikasiyalara açıq olduğu fərz edilir. Formal metodlar nəinki dayanıqlığı ölçmə qabiliyyətinə malikdir, həmçinin Sİ tərtibatçısını bəzi mühüm geridönüş məlumatları ilə təmin etmək üçün dayanıqlığın itirilməsinə səbəb olan mənbələri də əks etdirir. Məsələn, formal metodlar kompüter görməsi və ya zaman sıralarının emalı üçün neyron şəbəkəsi tərəfindən öyrənilən xüsusiyyətləri ön plana çəkə bilər. Formal metodlar, həmçinin neyron şəbəkəsinin təsnifatlaşdırıla bilmədiyi sinifləri də xüsusi qeyd edə bilər.

7.2.2. Tanınmış xüsusiyyətlərin identifikasiyası

Neyron şəbəkəsi tərəfindən tanınmış xüsusiyyətlərin identifikasiyası Sİ tərtibatçısına neyron şəbəkəsinin davranışını daha yaxşı başa düşməyə, izah yaxud interpretasiya etməyə imkan verir. Beləliklə, neyron şəbəkəsinin hansı dayanıqlığa malik olacağını daha yaxşı başa düşmək olar. Daha asan identifikasiya edilən xüsusiyyətlər haqqında məlumat Sİ tərtibatçısına (produksiya verilənlərinin təqdim edilməsi zamanı) neyron şəbəkəsinin öz tapşırıqlarını nə dərəcədə yerinə yetirə biləcəyini anlamağa imkan verir.

Neyron şəbəkəsi müəllimlə, müəllimsiz və ya möhkəmləndirici öyrənmə yolu ilə təlimlənməsindən asılı olmayaraq, ona təqdim edilən verilənlərdən əldə edə biləcəyi bəzi xüsusiyyətlərə əsaslanır. Bu xüsusiyyətlər, ümumiyyətlə, Sİ tərtibatçısı üçün birbaşa əlçatan deyil və onların strukturunda oxuna bilən şəkildə təsvir olunmur. Əvəzində, onlar təlim nəticəsində yaradılan riyazi modelə daxil edilir. Bu o deməkdir ki, xüsusiyyətlər birbaşa insan tərəfindən oxuna bilən şəkildə ifadə edilə bilməz. Xüsusiyyətlər çoxölçülü fəzada ifadə olunan riyazi artefaktlardır.

Formal metodlar model vasitəsilə neyron şəbəkəsinin giriş verilənlərindən çıxış verilənlərinə doğru domen sahəsində əlaqə yaratmaq üçün simvolik və ya relyasiyalı yanaşmalardan istifadə edə bilər (daha ətraflı məlumat üçün 6.2-ci bəndə baxın). Bu əlaqə Sİ tərtibatçısına hər bir giriş verilənlərinin çıxış verilənlərinə nə dərəcədə təsir göstərdiyini müəyyənləşdirməyə imkan verir. Model daxilində öyrənilmiş xüsusiyyətlər giriş və çıxış verilənləri arasındakı hər bir əlaqənin gücünə və ya zəifliyinə görə məsuldur. Əlaqələri müşahidə etməklə, öyrənilmiş xüsusiyyətlərin nəticələrini müşahidə etmək və bunun əsasında da, onların neyron şəbəkəsinin dayanıqlığına təsirini daha yaxşı başa düşmək mümkündür.

Bəzi öyrənilmiş xüsusiyyətləri identifikasiya etmək üçün Sİ tərtibatçısı relevantlıq meyarından istifadə edir. Relevantlıq meyarına görə nəticənin təsdiqi ya manual (birbaşa təsdiqləmə vasitəsilə), ya da avtomatik qaydada (hədəf relevantlıq göstəricilərinə uyğunluğun qiymətləndirilməsi yolu ilə) həyata keçirilə bilər.

— Manual təsdiqləmə zamanı ekspert meyara əsasən nəticələri birbaşa qiymətləndirir. Ekspertin bu qiymətləndirməsini daha yaxşı başa düşmək üçün qiymətləndirmə hesabatına ekspertin əsaslandırması da əlavə edilə bilər.

— Avtomatik təsdiqləmə zamanı qiymətləndirmə verilənlərin relevantlığı üzrə dəqiq hədəf göstəricilərə əsaslanmalıdır. İxtiyari verilənlər üzrə ölçülən relevantlıq və hədəf relevantlıq göstəricisi arasındakı uyğunluq (“correspondence”) səviyyəsini ölçmək üçün konkret metod müəyyən edilməlidir. Uyğunluq səviyyəsinin kifayət qədər yüksək olduğunu yoxlamaq üçün sərhəd qiyməti təyin olunmalıdır. Həmçinin, hədəf relevantlıq göstəricisi təmin edilməlidir.

Qeyd: Təsdiqləmə prosesində məsul şəxsin şəxsiyyəti və kompetensiya səviyyəsi izlənmə və ya diaqnostika məqsədləri üçün identifikasiya edilə bilər. Avtomatik təsdiqləmə halında hədəf relevantlıq mənbəyi də izlənmə və ya diaqnostika məqsədləri üçün istifadə oluna bilər.

7.2.3 Siniflərə bölmə qabiliyyətinin yoxlanılması

Siniflərə bölmə qabiliyyətinin yoxlanılması klassifikasiyanı həyata keçirən neyron şəbəkələrində istifadə edilə bilən bir üsuldur. Bu neyron şəbəkələri üçün modelin rolu giriş verilənlərinə əsaslanaraq siniflərin proqnozlaşdırılmasından ibarətdir. Bunun üçün,

klassifikasiya modeli təlim keçdiyi (və təlimdən kənarda qalan) verilənlər nöqtələri arasındakı məlumatı ümumiləşdirilir. Modelin siniflərə bölmə qabiliyyəti nə qədər yaxşı olarsa, klassifikatorun nəticəsi bir o qədər effektiv olar. Beləliklə, modelin dayanıqlığı onun effektiv şəkildə siniflərə bölmək qabiliyyətindən asılıdır.

Klassifikatoru layihələndirərkən, hansı siniflərin digərlərindən daha çox və ya az bölündüyünü identifikasiya etmək üçün həssaslıq meyarından istifadə edilə bilər. Bunun üçün, həssaslıq üzrə təhlildə test verilənlərindəki verilənlər nöqtələri ətrafında qurulmuş domen sahələrindən istifadə edilməlidir. Hansı siniflərin üst-üstə düşməyə başladığını (kəsişmə) müəyyənləşdirmək üçün atributların qiymətlərinin səpələnməsi tədricən artırılmalıdır. Kəsişmə bir sinif üzrə neyron şəbəkəsinin çıxış verilənlərinin digər sinfin çıxış verilənlərini üstələməyə başlaması kimi başa düşülür. Bütün siniflərin çıxış verilənləri digər siniflərin çıxış verilənləri ilə kəsişdikdə proses dayanır.

Siniflərə bölmə qabiliyyətinin təhlilinin nəticələri siniflərin kəsişməyə başlama sırasına və bu kəsişmənin baş vediyi domen sahəsinin ölçüsünə əsaslanır. Həssaslıq üzrə təhlilin nəticələrinə əsasən, ya təlim verilənləri, ya da neyron şəbəkəsinin arxitekturası ilə bağlı tədbirlər görmək mümkündür. Məqsəd isə həssaslıq üzrə təhlilləri müqayisəli şəkildə qiymətləndirməklə, hər bir sinfin bölünmə qabiliyyətini artırmaqdır.

7.3. Verifikasiya və validasiya mərhələsində dayanıqlığın qiymətləndirməsi

7.3.1. Ümumi müddəalar

Verifikasiya və validasiya mərhələsində neyron şəbəkəsi onun tələb və məqsədlərə cavab verməsinin yoxlanılması məqsədilə test edilir. Bu mərhələdə formal metodlardan istifadə digər verifikasiya və validasiya vasitələrini (məsələn, statistik testetmə və ya sahəvi sınaqlar) əvəz etmir. Bununla belə, formal metodlar konkret domen sahəsi çərçivəsində neyron şəbəkəsinin dayanıqlığı barədə yeni məlumat təqdim edə bilər. Bu mərhələdə formal metodların əsas üstünlüyü onların domen sahəsi üzrə tətbiq olunduğu üçün dayanıqlığın daha ümumi formada sübutuna imkan verməsindən ibarətdir.

7.3.2. Giriş domen sahəsinin hissələrinin əhatə olunması

Neyron şəbəkəsinin fəaliyyət göstərməsi üçün nəzərdə tutulan giriş verilənlərinin domen sahəsi müxtəlif mürəkkəblik dərəcələri ilə müəyyən oluna bilər. Onların bəzilərini müəyyənləşdirmək (məsələn, müəyyən sərhədlər daxilində mövcud olan verilənlər çoxluğu üzrə reqressiya tapşırıqlarını yerinə yetirən neyron şəbəkəsi üçün) çox asandır. Digər hallarda, bu daha mürəkkəb ola bilər. Məsələn, təsvirlərin emalı tapşırıqlarında giriş verilənlərinin domen sahəsi bəzi atributlarla xarakterizə edilə bilər (5.2-ci bəndə baxın), lakin domen sahəsi asanlıqla müəyyən edilən riyazi obyekt olmaya bilər. Formal metodlardan çıxış verilənləri üzrə sərhəd hesablamalarının bəzi formalarında istifadə olunur, buna görə də giriş verilənlərinin domen sahəsinin müəyyən edilməsi texnikası metoda əhəmiyyətli dərəcədə təsir göstərir.

Giriş verilənlərinin domen sahələri aşkar formada məhdudlaşdırılmalı olan atributların variasiyaları ilə validasiya ediləcək fəzanı müəyyən edən atributlar əsasında təyin edilir. Daha sonra, 5-ci bölmədə təsvir olunan dayanıqlıq meyarlarından istifadə etməklə domen sahələri və ya bu sahələrin hissələrində formal metodlar tətbiq olunur. Validasiyanın məna

kəsb etdiyi və qiymətləndirməni aparanı faydalı informasiya ilə təmin edən domen sahəsinin hissəsinin müəyyən edilməsinə ehtiyac yarana bilər. Giriş verilənlərinin domen sahəsinin bu cür bölünməsi əsaslandırılmalıdır. Xüsusilə, əsaslandırmada seçilmiş meyarın giriş verilənlərinin domen sahəsinin konkret bölgüdə dayanıqlığı qiymətləndirə bilmə imkanı vurğulanmalıdır.

Domen sahəsinin müəyyən hissəsində formal metodlardan istifadənin məqsədi validasiyanın aparılmadığı və ya natamam aparıldığı hissələrdə dayanıqlığın qiymətləndirilməsini genişləndirməkdən ibarətdir. İdeal halda qiymətləndirmə bütün domen sahəsi üzrə aparıla bilər. Lakin praktikada bu, domen sahəsinin bütövlükdə asanlıqla müəyyən edilə bilməməsi səbəbindən və ya domen sahəsinin ölçüsü ilə əlaqədar çox vaxt mümkün olmur.

Buna görə də, ilk addım validasiyanın məna kəsb etdiyi və qiymətləndirməni aparanı faydalı informasiya ilə təmin edən domen sahəsi hissəsini müəyyən etməkdən ibarətdir.

Bu konsepti ən yaxşı şəkildə iki fərqli nümunə ilə (asan və çətin müəyyən edilən domen sahəsi üzrə) izah etmək olar:

Birinci nümunəyə uyğun olaraq, iki girişi və bir çıxışı olan riyazi funksiyanın davranışını interpolyasiya etmək üçün təlim keçmiş neyron şəbəkəsini nəzərdən keçirək. Giriş verilənlərinin domen sahəsini təyin edən sərhədlər məlumdur və neyron şəbəkəsinin imitaisya edəcəyi funksiya həmişə bu sərhədlər daxilində təyin edilir. Burada yalnız giriş verilənlərinin sərhədlərini məhdudlaşdırmaqla giriş verilənlərinin bölgüsünü müəyyən etmək asandır. Sonra çıxış verilənlərinin sərhədlərini yoxlamaq üçün formal metodlardan istifadə etmək olar. Funksiya burada dəqiq təyin olunduğu üçün neyron şəbəkəsinin kifayət qədər dayanıqlığa malik olduğunu (həssaslıq meyarından istifadə etməklə) yoxlamaq asandır. Daha sonra domen sahəsi üzrə dayanıqlığın qiymətləndirilməsini genişləndirmək üçün bütün fəza ayrıca yoxlanıla bilən bir neçə hissəyə bölünə bilər.

İkinci nümunə üçün insanın konkret bir orqanının sağlam olub-olmadığını təyin etmək məqsədilə tibbi təsvirlərin klassifikasiyası üzrə təlim keçmiş neyron şəbəkəsini nəzərdən keçirək. Təsvirlərin ölçüsü 100x100-dir, təsvirlər eyni məsafədən çəkilir, orqanın əyilmə bucağı həmişə təsvirin mərkəzindədir və təsvirlərin hamısı eyni cihazda çəkilir. İnsanların orqanların ölçüsü və forması müxtəlif olduğu üçün giriş verilənlərinin domen sahəsini müəyyən etmək asan deyil. Həmçinin orqanın ətrafındakı təsvirin hissələri də fərqli ola bilər. Bu ssenaridə giriş verilənlərinin domen sahəsinin hər hansı hissəsində gözlənilən davranışı əvvəlcədən bilmək də asan olmur. Validasiyanı aparan tərəfindən başa düşülə bilən parametrin dəyişməsi (məsələn, orqanın həcmi və ya təsvirdəki fonun parlaqlığı) ilə bağlı domen sahəsinin bəzi hissələrini nəzərdən keçirmək üçün formal metodlardan istifadə edilə bilər.

7.3.3. Perturbasiyanın təsirinin ölçülməsi

Neyron şəbəkəsinin istifadəsi üçün nəzərdə tutulan domen sahəsinin təsvirinə əsaslanaraq, neyron şəbəkəsinin giriş verilənlərinin məruz qala biləcəyi perturbasiya növlərini identifikasiya etmək mümkündür. Hər bir perturbasiya neyron şəbəkəsinin performans səviyyəsinə fərqli təsir göstərə bilər. Onların kombinasiyası da neyron şəbəkəsinin dayanıqlığına müxtəlif təsir göstərə bilər. Verifikasiya və validasiya mərhələsində sistemin

dayanıqlığını bu kimi perturbasiya (kombinə edilmiş və ya edilməmiş) nümunələri üzrə qiymətləndirmək mümkündür. Formal metodlardan istifadə etməklə, bu kimi perturbasiyalara qarşı sistemin dayanıqlığını daha ümumi şəkildə qiymətləndirmək mümkündür.

Perturbasiyalar faydalı (“beneficial”) və ya zərərli (“detrimental”) ola bilər. Onlar, həmçinin qəsdən (məsələn, rəqabətli hücum zamanı) və ya qərəzsiz (məsələn, sensor defektləri və ya ətraf mühitin dəyişiklikləri ilə bağlı) ola bilər. Perturbasiyalar ya riyazi olaraq təsvir edilə, ya da yalnız illüstrasiya oluna bilər.

Nümunə: Təsvirdəki bulanıqlıq perturbasiyası riyazi olaraq hər bir pikselə tətbiq edilməklə bulanıqlıq yaradan konkret nüvənin konvolyusiyası kimi müəyyən edilə bilər. Bununla belə, linzada təsvirin qüsurlu olmasına səbəb olan damcılar sadəcə olaraq təsvirə süni şəkildə damcı əlavə edən bir və ya bir neçə maskanın istifadəsini tövsiyə etməklə illüstrasiya oluna bilər. Birinci halda, ekvivalentlik aşkardır, çünki perturbasiya riyazi funksiyadır. İkinci halda, funksiyanın tətbiqi daha çox kontekstdən asılıdır və iki verilənin (məsələn, təsvir və maskanın) birləşməsinə uyğun ola bilər.

Perturbasiyalı giriş verilənlərinin generasiyası prosesinin sazlanması bir istifadəçi digərinə dəyişdikdə fərqlənisə, bu proses izah edilməlidir. Neyron şəbəkəsinin giriş verilənlərinə perturbasiyanın tətbiqi onları dəyişdirmək üçün giriş verilənlərinə funksiyanın tətbiqi kimi nəzərdə tutulur.

Formal metodlardan istifadə etməklə neyron şəbəkəsinin konkret perturbasiyaya qarşı dayanıqlığını qiymətləndirmək üçün ilkin şərtlərdən biri perturbasiyanın tətbiqi prosesini təsvir edən funksiyanın mövcud olmasıdır. Bu funksiya ən azı bir məhdudlaşdırılmış parametrə əsaslanmalıdır. Məhdudlaşdırılmış parametrlər minimum və maksimum qiymətlərlə müəyyən edilir. Müvafiq parametrlər üçün minimum və maksimum variasiyalar müəyyən edilməlidir. Domen sahəsi üzrə bir və ya bir neçə meyar (5-ci bölməyə baxın) təyin olunmalıdır. Daha sonra, təqdim olunan domen sahəsindəki meyarları (müxtəlif mümkün perturbasiyalarla) qiymətləndirmək üçün formal metoddan istifadə edilməlidir.

Bir neçə konkret perturbasiyaya qarşı neyron şəbəkəsinin dayanıqlığını qiymətləndirmək üçün hər bir perturbasiyanı təmsil edən funksiyalar bir-biri ilə birləşdirilə (“compose”) bilər. Məcburi olmasa da, funksiyaların kommutativ kompozisiyasına üstünlük verilir. Bu yarımbənddə təsvir olunan eyni proses funksiyaların kompozisiyasına da şamil edilir.

7.4. Quraşdırma mərhələsində dayanıqlığın qiymətləndirməsi

Neyron şəbəkələri qeyri-xətti sistemlər olduğundan giriş verilənlərindəki kiçik dəyişikliklərə həssasdırlar. Bu dəyişikliklər neyron şəbəkəsinin iş prosesində baş verən hesablama (ədədi) dəqiqlikləri ilə bağlı ola bilər. Hesablama dəqiqlikləri ilə bağlı problemlərə aşağıdakılar səbəb ola bilər:

— əməliyyatların yerini dəyişdirən (“rearranging”) və ya onları digər əməliyyatlarla əvəz edən kompilyatorlar (məsələn, birləşdirilmiş vurma-toplama əməliyyatından istifadə zamanı [39]);

— baza avadanlıqlarının yerdəyişməsi əməliyyatları (məsələn, konveyer əməliyyatları nəticəsində üstünlük (mənfəət) əldə etmək üçün);

— hesablama dəqiqliyini azaltmaq üçün optimallaşdırma (məsələn, kvantlama, sürüşən vergüllü əməliyyatlardan və ya sabit vergüllü hesab qaydalarından daha az istifadə etməklə);

— yuvarlaqlaşdırma prosesindəki dəyişikliklər;

— aşağı səviyyəli ədədi operatorların tətbiqi zamanı dəyişikliklər (məsələn, IEEE 754:2019 standartından fərqli standartlara uyğun operatorlardan, yanlış yuvarlaqlaşdırma və ya fərqli interpolyasiyaya malik operatorlardan istifadə [24]).

Neyron şəbəkəsini hesablamalarla bağlı bu səbəblərdən birinin və ya bir neçəsinin baş verə biləcəyi sistemə inteqrasiya edərkən sadalanan bu problemlər nəzərə alınmalıdır. Xüsusilə, onların təsirlərini yoxlamaq üçün formal metodlardan istifadə edilməlidir. Bu məqsədlə yerdəyişmə əməliyyatları və ya baza hesab qaydalarının dəyişdirilməsi nəticəsində yaranan maksimum yuvarlaqlaşdırma xətasının sərhədlərini ölçmək üçün formal metodlardan istifadə edilə bilər. Təcrübədə seçilmiş formal metodlar onların hələ də tətbiqinin mümkünlüyü yoxlamaq üçün əvvəllər istifadə edilmiş hər bir meyar əsasında qiymətləndirilir.

Bu problemləri həll etmək üçün Sİ tərtibatçısı aşağıdakı addımları ata bilər:

— Birincisi, Sİ tərtibatçısı seçilmiş hesab qaydalarının təsirini yoxlamalıdır. Bunun üçün əvvəlcə hər bir əsas əməliyyat üçün onun yuvarlaqlaşdırma xətalarının tətbiq sahəsində statik olaraq məhdudlaşdırıla biləcəyini müəyyən etmək lazımdır.

    — Operatorun yuvarlaqlaşdırma xətasını statik olaraq məhdudlaşdırmaq mümkün olduqda, formal verifikatorlar onları neyron şəbəkəsinin verifikasiyası üçün istifadə olunan semantikada nəzərə almalıdırlar. Məsələn, toplama və ya bölmə kimi standart sürüşən vergüllü hesab əməliyyatları (nəticədə son mövqedə duran ölçü vahidinin qiyməti ilə məhdudlaşdırıla bilən) yuvarlaqlaşdırma xətasına malikdir.

    — Yuvarlaqlaşdırma xətasını statik olaraq məhdudlaşdırmaq mümkün olmadıqda, formal verifikatorlar Sİ tərtibatçısı tərəfindən təmin edilən zəruri əməliyyatların izafi approksimasiyasına etibar etməlidir. Sİ tərtibatçısı bu cür izafi approksimasiya üçün istifadə olunan hipotezi aydın şəkildə sənədləşdirməlidir.

— İkincisi, Sİ tərtibatçısı inteqrasiya prosesinin əməliyyatları (istər statik, istərsə də dinamik) dəyişdirə və ya yerini dəyişə biləcəyi mümkün yolları identifikasiya etməlidir. İdentifikasiyadan sonra verifikator bu mümkün modifikasiyaları nəzərə almalıdır. Bu modifikasiyalar müəyyən və ya qeyri-müəyyən ola bilər.

    — Modifikasiyalar müəyyəndirsə, verifikator neyron şəbəkəsini onun quraşdırıldıqdan sonrakı davranışının uyğunluğu (reprezentativliyi) nöqteyi-nəzərindən təhlil etməlidir.

    — Modifikasiyalar qeyri-müəyyəndirsə, verifikator onların baş verə biləcəyini və ümumilikdə təsirini nəzərə almalıdır. Bəzən ən pis halda ixtiyari əməliyyatların dəqiqliyini müəyyən etmək mümkün olur (məsələn, yalnız düzgün yuvarlaqlaşdırılmış standart sürüşən vergüllü əməliyyatlardan istifadə edildikdə). Bu mümkün olmadıqda isə verifikator Sİ tərtibatçısının təqdim etdiyi bu dəyişikliklərin səbəb olduğu dəqiqliyin izafi approksimasiyasına etibar etməlidir.

Sİ tərtibatçısı bu cür izafi approksimasiya üçün istifadə olunan hipotezi aydın şəkildə sənədləşdirməlidir.

7.5. İstismar və monitorinq mərhələsində dayanıqlığın qiymətləndirməsi

7.5.1. Ümumi müddəalar

Neyron şəbəkəsi quraşdırıldıqdan və istismara verildikdən sonra onun dayanıqlığı monitorinq edilə bilər. Bu zaman ya neyron şəbəkəsi dəyişdirilməməlidir (hər bir neyron üzrə çəkilər və sapma qiymətləri sabitdir), ya da neyron şəbəkəsi fasiləsiz öyrənmədən istifadə etdiyi üçün dəyişə bilər. Neyron şəbəkəsi üzrə dəyişikliklərdən asılı olaraq, dayanıqlığı qiymətləndirmək üçün formal metodlar müxtəlif yollarla tətbiq oluna bilər. Bir tərəfdən, əgər neyron şəbəkəsi sabitdirsə, onun emal etdiyi yeni giriş verilənlərinə nəzərən dayanıqlığı qiymətləndirilə bilər. Digər tərəfdən, neyron şəbəkəsi dəyişirsə, bu dəyişikliyin təsirini ölçmək üçün onun dayanıqlığı qiymətləndirilə bilər.

Neyron şəbəkəsinin necə fəaliyyət göstərməsindən asılı olmayaraq, qeyd etmək vacibdir ki, formal metodlar həmişə neyron şəbəkəsinin davranışının monitorinqi üçün bu şəbəkənin işi zamanı istifadə edilən resurslarla müqayisədə daha böyük resurslar (yəni hrsablam gücü, yaddaş və enerji) tələb edilir. Monitorinqin nə zaman həyata keçirilməsindən asılı olaraq, resurs artımı formal metodların yalnız kiçik neyron şəbəkələrində tətbiqini tələb edə bilər. Həmçinin formal metodların neyron şəbəkəsinin bir nəticəçıxarma prosesinə nisbətən daha baha başa gəldiyini nəzərə alsaq, onları eyni tezliklə tətbiq etmək çətindir. Məsələn, proqramlaşdırılmaqla saniyədə 10 dəfə ventil (qapı) matrisini çıxış kimi generasiya edən neyron şəbəkəsi eyni tezliklə onu təhlil edilə bilməz, bunun əvəzinə təhlil yalnız onun nəticələrinin altçoxluğunda aparıla bilər. 6.2-ci bənddə təsvir edilmiş formal metodlar üçün resursların azaldılması üzrə yanaşmalar [40], [41] və [42]-də təqdim olunur.

7.5.2. İstismar mərhələsi üzrə dayanıqlıq

Sistem istismara verildikdə, onun ilkin olaraq təyin olunduğu və validasiya edilməsi nəzərdə tutulan istifadə sahəsində istismar ediləcəyinə zəmanət vermək çətin ola bilər. İstismar şərtləri əvvəlcədən gözlənilməyən qaydada dəyişə bilər. Bu, xüsusilə, “açıq dünya” (“open-world”) mühitində istismar olunan sistemlərə aiddir. Bu mərhələdə dayanıqlığın qiymətləndirilməsi sistemin planlandırılan istismar sahəsindən nə dərəcədə kənara çıxdığını identifikasiya etməyə kömək edə bilər. Tələb olunan dayanıqlıq səviyyəsinə nail olunmadıqda, korreksiyaedici tədbirlər (məsələn, operatoru xəbərdar etmək və ya imtinalara dayanıqlı olan rejimə keçmək) görülə bilər.

Neyron şəbəkələri analoji giriş verilənləri üzrə analoji performansa malik olduğundan, onların dayanıqlığının monitorinqi giriş verilənlərinin hələ də neyron şəbəkələrinin təlim keçdiyi domen sahəsinin tərkib hissəsi olduğunu identifikasiya etməyə kömək edə bilər. Lokal dayanıqlıq xassəsi analoji giriş verilənləri üzrə analoji nəticələr verə bilər. Məsələn, stabillik xassəsi (5.3.1-ci yarımbənd) yanlış klassifikasiya edilmiş giriş verilənləri üzrə maksimum qiymətlərlə müqayisədə, düzgün klassifikasiya edilmiş giriş verilənləri üzrə maksimum qiymətlərə malik ola bilər. Bu yanaşmalar neyron şəbəkələrinin cari domen sahəsi üzrə dayanıqlığının qiymətləndirilməsi üçün istifadə edilə bilər.

Yeni domen sahəsi üzrə istismar edilən neyron şəbəkələrinin dayanıqlığını qiymətləndirərkən, yeni giriş verilənlərinin meyarlar üzrə nəticələrini həyat dövrünün əvvəlki mərhələsində eyni meyarlar üzrə nəticələri ilə mütəmadi olaraq müqayisə etmək lazımdır.

7.5.3. Dayanıqlığa dair dəyişikliklər

Fəsiləsiz öyrənmədən istifadə neyron şəbəkələrində çəkilər və sapma qiymətləri dəyişə bildiyi üçün şəbəkələrin davranışı da dəyişir. Neyron şəbəkələrinin daxili strukturunun dəyişməsi onun dayanıqlığına da təsir (müsbət və ya mənfi) göstərə bilər.

İstismar mərhələsində neyron şəbəkələrinin dayanıqlığına dair dəyişikliklər qiymətləndirilərkən əvvəlcədən müəyyən edilmiş meyarlar çoxluğundan istifadə edilməlidir. Sonra müəyyən edilmiş meyarlar üzrə neyron şəbəkəsinin yeni versiyası ilə əvvəlki versiyası arasında nəticələr fərqlərinin müqayisəsi əsasında dayanıqlığın qiymətləndirilməsi aparılır. Dayanıqlığın aşağı düşməsinin məqbul olub-olmadığını yoxlamaq üçün müşahidə edilən fərqlər üzrə əlavə meyarlar müəyyən edilməlidir.

ƏDƏBİYYAT

[1] ISO/IEC 25059:2023, *Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems*

[2] ISO/IEC 25000, *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Guide to SQuaRE*

[3] ISO/IEC/TR 24029-1, *Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview*

[4] ISO/IEC/IEEE 42010, *Software, systems and enterprise — Architecture description*

[5] ISO/IEC/IEEE 15939:2017, *Systems and software engineering — Measurement process*

[6] ISO/IEC/IEEE 15289:2019, *Systems and software engineering — Content of life-cycle information items (documentation)*

[7] ISO/IEC 19794-1:2011, *Information technology — Biometric data interchange formats — Part 1: Framework*

[8] ISO/IEC/IEEE 29119-1:2022, *Software and systems engineering — Software testing — Part 1: General concepts*

[9] ISO/IEC/IEEE 24765:2017, *Systems and software engineering — Vocabulary*

[10] Leino K., Wang Z., Fredrikson M. Globally-Robust Neural Networks. *Proceedings of the 38th International Conference on Machine Learning, ICML.* 2021, **139**, 6212-6222

[11] Katz G. Barrett C., Dill D., Julian K., Kochenderfer M. Reluplex: An Efficient SMT Solver for Verifying Deep Neural Networks. 2017, *International Conference on Computer Aided Verification*, **10426**, 97–117. doi:10.1007/978-3-319-63387-9_5

[12] Szandala T., Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks. *CoRR*. 2020. arXiv: abs/ 2010 .09458

[13] Chu L., Hu X., Hu J., Wang L., Pei J. Exact and Consistent Interpretation for Piecewise Linear Neural Networks: A Closed Form Solution. *International Conference on Knowledge Discovery & Data Mining*, 2018, **24**, 1244–1253. doi:10.1145/3219819.3220063

[14] Bunel R., Turkaslan I., Torr P.H.S., Kohli P., Kumar M.P. A Unified View of Piecewise Linear Neural Network Verification. *International Conference on Neural Information Processing Systems*, 2018, **32**, pp. 4795–4804

[15] Weng T.-W., Zhang H., Chen H., Song Z., Hsieh C.-J., Boning D. et al. Towards fast computation of certified robustness for ReLU networks. *Proceedings of the 35th International Conference on Machine Learning*. 2018, **80**, 5276–5285

[16] Zhang H., Weng T.-W., Chen P.-Y., Hsieh C.-J., Daniel L. Efficient Neural Network Robustness Certification with General Activation Functions. *Neural Information Processing Systems Conference*. 2018, **31**, 4944–4953

[17] Wang S., Pei K., Whitehouse J., Yang J., Jana S. Efficient formal safety analysis of neural networks. *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. 2018, **80**, 6369–6379

[18] Narodytska N. Formal Analysis of Deep Binarized Neural Networks. *International Joint Conference on Artificial Intelligence*. 2018, **27**, 5692–5696. doi:10.24963/ijcai.2018/811

[19] Jia K., Rinard M., Efficient Exact Verification of Binarized Neural Networks. Neural Information Processing Systems. 2020, **33**, 1782–1795

[20] Khmelnitsky I., Neider D., Roy R., Barbot B., Bollig B., Finkel A. et al. Property-directed verification of recurrent neural networks. *International Symposium on Automated Technology for Verification and Analysis*. 2021. arXiv: 2009 .10610

[21] Ryou W., Chen J., Balunovic M., Singh G., Dan A., Vechev M., Scalable Polyhedral Verification of Recurrent Neural Networks. Computer-Aided Verification. 2021, **12759**, 225–248. doi:10.1007/978-3-030-81685-8_10

[22] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N. et al. Polosukhin I. Attention is All you Need. *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems, NIPS*. 2017, **30**, 5998-6008

[23] Shi Z., Zhang H., Chang K.-W., Huang M., Hsieh C.-J. Robustness Verification for Transformers. *International Conference on Learning Representations*. 2020. arXiv: 2002 .06622

[24] IEEE 754:2019, *IEEE Standard for Floating-Point Arithmetic*

[25] Tjeng V., Xiao K., Tedrake R. Evaluating Robustness of Neural Networks with Mixed Integer Programming. *The International Conference on Learning Representations*. 2019

[26] Katz G., Huang D. A., Ibeling D., Julian K., Lazarus C., Lim R., Shah P., Thakoor S., Wu H., Zeljic A., Dill D. L., Kochenderfer M., Barrett C., The Marabou Framework for Verification and Analysis of Deep Neural Network. Computer Aided Verification. 2019, **11561**, 443–452. doi:10.1007/978-3-030-25540-4_26

[27] Wang S., Pei K., Whitehouse J., Yang J., Jana S. Efficient Formal Safety Analysis of Neural Networks. *International Conference on Neural Information Processing Systems*. 2018, **32**, 6369–6379

[28] Sun X., Khedr H., Shoukry Y. Formal Verification of Neural Network Controlled Autonomous Systems. *International Conference on Hybrid Systems: Computation and Control*. 2019, **22**, 147–156. doi:10.1145/3302504.3311802

[29] Cousot P., Cousot R. Abstract interpretation: A unified lattice model for static analysis of programs by construction or approximation. *Principles of programming languages*. 1977, 238–252. doi:10.1145/512950.512973

[30] Probabilistic Abstract Interpretation Cousot P., Programming Languages and Systems. 2012, **7211**, 169–193. doi:10.1007/978-3-642-28869-2_9

[31] Singh G., Gehr T., Püschel M., Vechev M. An Abstract Domain for Certifying Neural Networks. *Programming Languages*. 2019, **3**, 41:1–41:30. doi:10.1145/3290354

[32] Julian K., Kochenderfer M. Guaranteeing safety for neural network-based aircraft collision avoidance systems. *IEEE/AIAA 38th Digital Avionics Systems Conference*. 2019, 603–612. doi:10.1109/DASC43569.2019.9081748

[33] Sidrane C., Kochenderfer M. OVERT: Verification of nonlinear dynamical systems with neural network controllers via overapproximation. *Workshop on Safe Machine Learning*. 2019

[34] Bouton M., Tumova J., Kochenderfer M. Point-based methods for model checking in partially observable Markov decision processes. *AAAI Conference on Artificial Intelligence*. 2020, **24**, 10061–10068

[35] Katz S., Strong C., Julian K., Kochenderfer M., Generating probabilistic safety guarantees for neural network controllers. *CoRR.* 2021. arXiv: abs/2103.01203

[36] Ehrig H., Mahr B., Fundamentals of Algebraic Specification 1: Equations and Initial Semantics. Berlin: Springer, 1985

[37] Manna Z., Waldinger R., The deductive foundations of computer programming - a one-volume version of "The logical basis for computer programming". Boston, Massachusetts: Addison-Wesley, 1993

[38] Sena L.H., Bessa I.V., Gadelha M.R., Cordeiro L.C., Mota E. Incremental Bouns.ded Model Checking of Artificial Neural Networks in CUDA. *Brazilian Symposium on Computing System Engineering*. 2019, 1–8. doi:10.1109/SBESC49506.2019.9046094

[39] Higham N.J. The Mathematics of Floating-Point Arithmetic. *London Mathematical Society* [online], London, UK, March 2021 [viewed 02 September 2021]. Available from: https://www.lms.ac.uk/sites/lms.ac.uk/files/files/NLMS_493_for%20web2.pdf

[40] Hymans C., Design and Implementation of an Abstract Interpreter for VHDL. Correct Hardware Design and Verification Methods. 2003, **2860**, 263–269. doi:10.1007/978-3-540-39724-3_23

[41] Banterle F., Giacobazzi R. A Fast Implementation of the Octagon Abstract Domain on Graphics Hardware. *International Static Analysis Symposium*. 2007, **4634**, 315–332. doi: 10.1007/978-3-540-74061-2_20

[42] Mirman M., Gehr T., Vechev M. Differentiable Abstract Interpretation for Provably Robust Neural Networks. *Proceedings of the 35th International Conference on Machine Learning*. 2018, **80**, 3575–3583