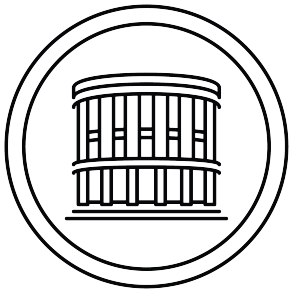
UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



KLASIFIKÁCIA SPRÁVANIA V ČASOVÝCH RADOCH KURY DOMÁCEJ

DIPLOMOVÁ PRÁCA

2022 Bc. Miroslav HLÁSNIK

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

KLASIFIKÁCIA SPRÁVANIA V ČASOVÝCH RADOCH KURY DOMÁCEJ

DIPLOMOVÁ PRÁCA

Študijný program:	Ekonomicko-finančná matematika a modelovanie
Študijný odbor:	Matematika
Školiace pracovisko:	Katedra aplikovanej matematiky a štatistiky
Vedúci práce:	Mgr. Katarína Boďová, PhD.

Bratislava 2022 **Bc. Miroslav Hlásnik**





Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Miroslav Hlásnik
Študijný program: ekonomicko-finančná matematika a modelovanie (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Klasifikácia správania v časových radoch kury domácej
Classification of poultry behavior based on temporal data

Anotácia: Tradičné štúdie správania zvierat sa opierajú o priame pozorovania alebo analýzu videozáznamov. Takýto prístup je však veľmi náročný na čas, je náchylný na chyby a je subjektívny. Vďaka rapídnemu rozvoju technológií je možné správanie zvierat autonómne zaznamenávať pomocou rôznych typov nositeľných senzorov bez narušenia prirodzeného správania zvierat. Dôležitým krokom pri následnej analýze takýchto veľkých dát je ich klasifikácia, na ktorú možno využiť metódy strojového učenia. Cieľom diplomovej práce je identifikovať základné typy správania v zaznamenaných časových dátach získaných z akcelerometrov upevnených na tele nosníc (ako sú príjem krmiva, čistenie, popolenie, a iné) a následne zaznamenané časové rady klasifikovať podľa týchto behaviorálnych stavov. V práci bude použitá metodika autorov Abdoli et al. 2018, vrátane ich implementácie v programe Matlab. Hlavným zámerom práce je vyskúšať funkčnosť prístupu autorov článku na vlastné dáta a v prípade úspešnosti tejto časti práce ďalej rozšíriť „slovník správania (behaviour dictionary) kury domácej o ďalšie prvky správania pomocou autormi navrhovaného algoritmu na budovanie slovníka a využiť metódy dátovej analýzy na štúdium konkrétnych biologických otázok, napríklad na štúdium ekonomicky významného poškodzujúceho správania nosníc. Práca bude prebiehať v spolupráci s laboratóriom Ľubora Košťála z Centra biovied SAV.

Vedúci: Mgr. Katarína Boďová, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Marek Fila, DrSc.

Dátum zadania: 13.01.2021

Dátum schválenia: 17.01.2021

prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
vedúci práce

Poďakovanie Touto cestou by som rád poďakoval svojej pani školiteľke Mgr. Kataríne Boďovej, PhD. za ochotu, povzbudenia a cenné rady, ktoré mi adresovala. Rovnako chcem poďakovať všetkým, ktorí ma počas môjho štúdia podporovali.

Abstrakt v štátnom jazyku

HLÁSNIK, Miroslav: Klasifikácia správania v časových radoch kury domácej [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: Mgr. Katarína Boďová, PhD., Bratislava, 2022, 63 s.

Táto práca sa zaoberá klasifikáciou behaviorálnych stavov kury domácej na reálnych akcelerometrických dátach z experimentov, ktoré vyhotovila Slovenská akadémia vied (SAV). Nadmerná početnosť alebo absencia určitého správania často popisuje aktuálny zdravotný stav kury domácej. Včasnou diagnostikou ochorenia pri jednotlivcovi je možné ochrániť chov pred šírením chorôb, ako aj eliminovať negatívne ekonomické následky pre chov. Najprv v práci uvedieme metodiku jednotlivých experimentov a štruktúru akcelerometrických dát. Ďalej vysvetlíme princíp identifikácie správaní a popíšeme metódy použité na ich klasifikovanie. Nakoniec porovnáme rôzne prístupy identifikácie správania na reálnych dátach SAV.

Kľúčové slová: klasifikácia behaviorálnych stavov, hľadanie podobnosti v časových radoch, Euklidovská vzdialenosť, z-normalizácia, strojové učenie.

Abstract

HLÁSNIK, Miroslav: Classification of poultry behavior based on temporal data [Master Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Katarína Boďová, PhD., Bratislava, 2022, 63 p.

The thesis focuses on the classification of poultry behaviour based on real accelerometric data from experiments, achieved by the Slovak Academy of Sciences (SAV). Excessive incidence or absence of certain behavior often describes the current state of poultry health. Early disease diagnosis of an individual can protect the breed from the spreading, as well as eliminate the negative economic consequences for poultry farming. The thesis includes the methodology of performed experiments and the structure of accelerometric data. Furthermore, we explain the principle of behavior identification and describe the methods used for their classification. Last but not least, we compare various approaches to categorise and identify poultry behaviors on legitimate data of SAV.

Keywords: behaviour classification, similarity search on time series, Euclidean distance, z-normalization, machine learning.

Obsah

Úvod **8**

1 Experimentálny protokol a štruktúra dát **10**

2 Typy správania kury domácej **17**

3 Použité metódy **21**

4 Aplikácia na SAV dáta **31**

4.1 Aplikácia na cvičné dáta (Experiment 1) 31

4.2 Aplikácia na koncové dáta (Experiment 2) 35

4.2.1 Naivný prístup 35

4.2.2 Prístup 1 - použitie hromadnej knižnice 39

4.2.3 Prístup 2 - použitie separátnej knižnice 48

4.2.4 Analýza prejdených vzdialeností 57

Záver **60**

Zoznam použitej literatúry **62**

Úvod

Správanie kolektívov sociálnych organizmov (napr. mravcov, rýb [3], atď.) môže byť ovplyvnené množstvom rôznych faktorov, napríklad napadnutím patogénmi alebo inými ochoreniami. V minulosti bolo monitorovanie a vyhodnocovanie správania zvierat časovo náročná úloha, vyžadujúca manuálnu klasifikáciu správaní odborníkom. Dnes už však existujú pokročilejšie technológie na kontinuálne zaznamenávanie správania kolektívov. Jedným z takýchto prístupov je vysokofrekvenčné zaznamenávanie pohybu v 3D priestore pomocou akcelerometra, ktorý je pripevnený na telo zvieraťa. Ďalším kľúčovým aspektom sú moderné počítačové algoritmy, ktoré dokážu dané vysokofrekvenčné dáta efektívne prechádzať a pracovať s nimi. Tento prístup využijeme aj v našej diplomovej práci, kde sa zaoberáme otázkou klasifikovania správania kury domácej.

Hospodárske zvieratá, ako kura domáca, sú chované vo väčších skupinách a včasná diagnostika jednotlivca môže prispieť k eliminácii šírenia choroby alebo množenia parazitov v kolektíve. Rýchla identifikácia výkyvov v správaní je teda aj hospodársky zaujímavá úloha. Pre chov je viac ako žiadúce monitorovať zdravie chovných sliepok na to, aby sa v chove nešírili choroby. Takto by sme zabezpečili čo najvyššiu kvalitu priemyselného produktu. K danej motivácií taktiež prispieva aj fakt každoročne sa zvyšujúcej globálnej produkcie kury domácej. Jedným zo spomínaných ektoparazitov, ktorý parazituje na tele kury domácej je *Ornithonyssus sylviarum*, ktorý môže zapríčiniť náhle zmeny v behaviorálnych stavoch (správaní) a veľakrát spomalí rast sliepky, ako aj zníži produkciu vajec [11], čo môže mať, ako bolo spomenuté, devastačné ekonomické následky pre chov.

V sociálnych kolektívoch je veľakrát možné identifikovať akúsi knižnicu typických behaviorálnych stavov, ktoré jednotlivci v kolektíve vykazujú. Na uvedenie príkladu, knižnica správaní pre kone [4] môže byť tvorená stavmi ako státie, chôdza, klus, atď., alebo pre ryby stavmi ako, akcelerácia so zatáčaním doprava, doľava a pasívny stav [3]. Pri kure domácej použijeme knižnicu tvorenú z troch správaní, konkrétne ďobanie do zeme, čistenie a oprašovanie, ktorých zvýšený, alebo znížený výskyt môže opisovať jej zdravotný stav. Tieto behaviorálne stavy bývajú ovplyvnené sociálnymi interakciami v kolektíve, povahou jednotlivca, ako aj stresovými faktormi (patogény, nedostatok jedla, atď.). Ďalším ovplyvňujúcim faktorom je náhodnosť, keďže pri dvoch rovnako nastavených nezávislých pokusoch s tým istým jednotlivcom by sme nedostali identický priebeh správania.

Dáta zozbierané z akcelerometra majú podobu časových radov, ktoré reprezentujú 3D polohu zvieraťa v čase. Na vytvorenie klasifikátora použijeme kombináciu matematických a informatických prístupov, ako hľadanie podobnosti medzi krivkami pomocou metriky Euklidovských vzdialeností [5], strojové učenie a algoritmy, ktoré dokážu efektívne prechádzať rozsiahle časové rady. Spomenuté algoritmy, ako aj ostatné doplnkové programy a vizualizácie, boli naprogramované v softvérovom systéme MATLAB.

Práca prebieha v spolupráci s výskumným pracoviskom Centra biovied SAV pod vedením RNDr. Ľubora Košťála, CSc.. Hlavnou motiváciou práce je pokúsiť sa zreprodukovať klasifikačné metódy autorov Abdoli et al. 2018 [1] na reálnych dátach z experimentov, ktoré vykonala SAV na vlastných nosniciach.

1 Experimentálny protokol a štruktúra dát

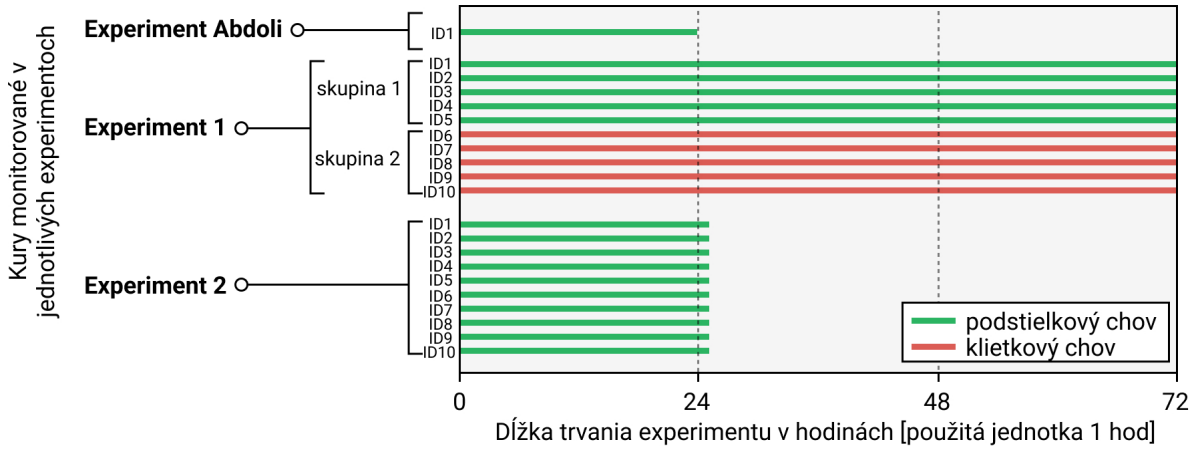
V našej práci pracujeme s dátami z troch experimentov. V tejto kapitole uvedieme v jednoduchosti metodiku daných experimentov spolu s ich popisnými názvami, ktoré budeme v danom tvare používať naprieč celou diplomovou prácou. Ďalej popíšeme, ako sú zaznamenávané dáta pomocou senzora, a čo tieto dáta reprezentujú. Nakoniec uvedieme jednoduché analýzy a kumulatívne štatistiky na dátach z prvého experimentu, ktorý vyhotovila Slovenská akadémia vied (ďalej v práci skr. SAV). Pomocou nich budeme vedieť lepšie porozumieť, aká je štruktúra dát, a čo sa v dátach nachádza.

V nasledovných riadkoch uvedieme popisné názvy jednotlivých experimentov, na ktoré sa odkazujeme v celej práci. Spolu s názvom uvedieme aj kým bol experiment vyhotovený, dátum a čas vykonania experimentu, ako aj konkretizáciu experimentálneho procesu.

- **Experiment Abdoli**
  - publikovaný v článku [1].
  - 24 hodinový záznam (od 00:00 do 24:00), dátum 11/22/2017.
  - monitorovanie 1 sliepky.
- **Experiment 1**
  - experiment vykonaný centrom biovied SAV.
  - 72 hodinový záznam (od 00:00 do 24:00), dátum od 10/12/2020 do 10/13/2020.
  - monitorovanie piatich sliepok na podstielke a piatich sliepok v klietke.
- **Experiment 2**
  - experiment vykonaný centrom biovied SAV.
  - 25 hodinový záznam (od 23:30 do 00:30), dátum od 9/20/2021 do 9/22/2021.
  - monitorovanie 10 sliepok na podstielke.

Ako doplnkový vizuál k popísaným experimentom prikladáme Obr. 1, v ktorom je zaznačený počet monitorovaných kúr v konkrétnom experimente spolu s celkovou dobou monitorovania v hodinách. Zelenou farbou sú vyznačené kury, ktoré boli monitorované

v podstielkovom chove. Naopak červenou farbou sú vyznačené kury z **Experimentu 1**, ktoré boli počas experimentu v klietkovom chove. Zdôrazňujeme, že každá kura použitá v experimentoch je jedinečná, resp. žiadna kura nebola použitá v dvoch experimentoch naraz.



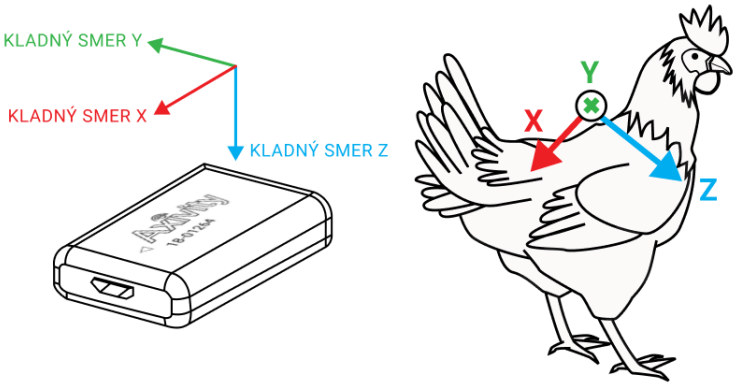
Obr. 1: Grafické znázornenie priebehu jednotlivých experimentov. (vlastné spracovanie)

Prvý zdroj dát z **Experimentu Abdoli** je z nedávno publikovaného článku [1]. Autori v ňom pracovali s pohybovými dátami a ukázali vysokú úspešnosť klasifikovania behaviorálnych stavov kury domácej. Dáta prislúchajúce danému článku sú voľne dostupné na internete [12]. Na dátach z daného experimentu a naprogramovaných algoritmoch sme pochopili princíp vytvorených klasifikačných algoritmov a získali prvý dotyk s časovými radmi reprezentujúcimi pohyb kury domácej. **Experiment 1** slúžil ako pilotný experiment vykonaný SAV, pomocou ktorého sa identifikovali limitácie a problémy pri zaznamenávaní pohybu kúr. Ďalej sme pracovali so surovými dátami zo senzora, ktoré bolo potrebné dostať do formátu, vhodného pre softvérové prostredie MATLAB. Porovnali sme výsledky z analyzovania skupiny monitorovanej na podstielke a v klietke a taktiež si predpripravili niektoré úkony potrebné na finálnu klasifikáciu behaviorálnych stavov. Posledný **Experiment 2** predstavuje hlavný experiment, ktorý čerpal z poznatkov získaných z **Experimentu 1**. V **Experimente 2** bola snaha minimalizovať problémy a limitácie objavené v pilotnom experimente, ktorým sa avšak často pri experimentoch so zvieratami predísť nedá.

Pre doplnenie informácií ohľadom prostredia, v ktorých prebiehali experimenty vykonané SAV, uvádzame v nasledovných riadkoch rozmer podstielky a klietky.

- rozmer podstielky: $100cm \times 150cm$ (šírka x hĺbka), výška presahujúca dva metre.
- rozmer klietky: $120cm \times 62,5cm \times 50cm$ (šírka x hĺbka x výška).

V každom z daných experimentov sa používali senzory **Axivity AX3** na monitorovanie pohybu, ktoré boli pripevnené na chrbát kury pomocou akéhosi batôžteka, kde bol senzor uložený. Senzor, ako aj poloha umiestnenia na chrbát kury sú naznačené na Obr. 2. Senzor zaznamenáva údaje o svojej polohe v troch osiach s výberovou frekvenciou 100 Hz, resp. zaznamenáva 100 údajov za sekundu. Spomeňme, že jednodňový záznam pri danej frekvencii predstavuje dáta, resp. maticu o rozmere $(3 \times 8\,640\,000)$, keďže disponujeme údajmi z troch osí. Pracujeme teda s rozsiahlymi časovými radmi. V Tabuľke 1 je ukážka výstupných dát zo senzora, ktoré majú podobu časových radov v každej z osí.



Obr. 2: (naľavo) Použitý senzor Axivity AX3 na snímanie pohybu, (napravo) uchytenie a umiestnenia senzora na kure domácej. (zdroj: [1], vlastné spracovanie)

os X	os Y	os Z
0.3594	-0.1875	-0.9062
0.3594	-0.2031	-0.9062
0.3594	-0.1875	-0.8906
⋮	⋮	⋮

Tabuľka 1: Náhľad výstupných dát zo senzora. Údaje v tabuľke predstavujú konkrétne dáta z **Experimentu Abdoli.**

Pri práci s dátami je dobré vedieť v akom prostredí sa experiment vykonával. Preto na Obr. 3 čitateľom ilustrujeme priebeh experimentov na podstielkach vykonaných SAV. Na Obr. 3b sú viditeľné čierne batohy s identifikačnými číslami, ktoré mali kury pripevnené na chrbte počas **Experimentu 2**. Ukážku batohov prikladáme na Obr. 4. Tieto batohy vyrobila SAV, za účelom pevnejšieho uchytenia, keďže batohy použité v pilotnom **Experimente 1** nedržali správne a kury si ich vedeli počas experimentu zhodiť. To je jeden zo spomínaných problémov, identifikovaných na pilotnom **Experimente 1**. Čísla pripevnené na batohoch sú potrebné kvôli rozpoznaniu kúr na videozázname. Viac k potrebe videozáznamu z experimentu povieme v kapitole 3.



(a) Experiment 1 - pilotná skupina kúr monitorovaná na podstielke.



(b) Experiment 2 - skupina 10 kúr monitorovaná na podstielke.

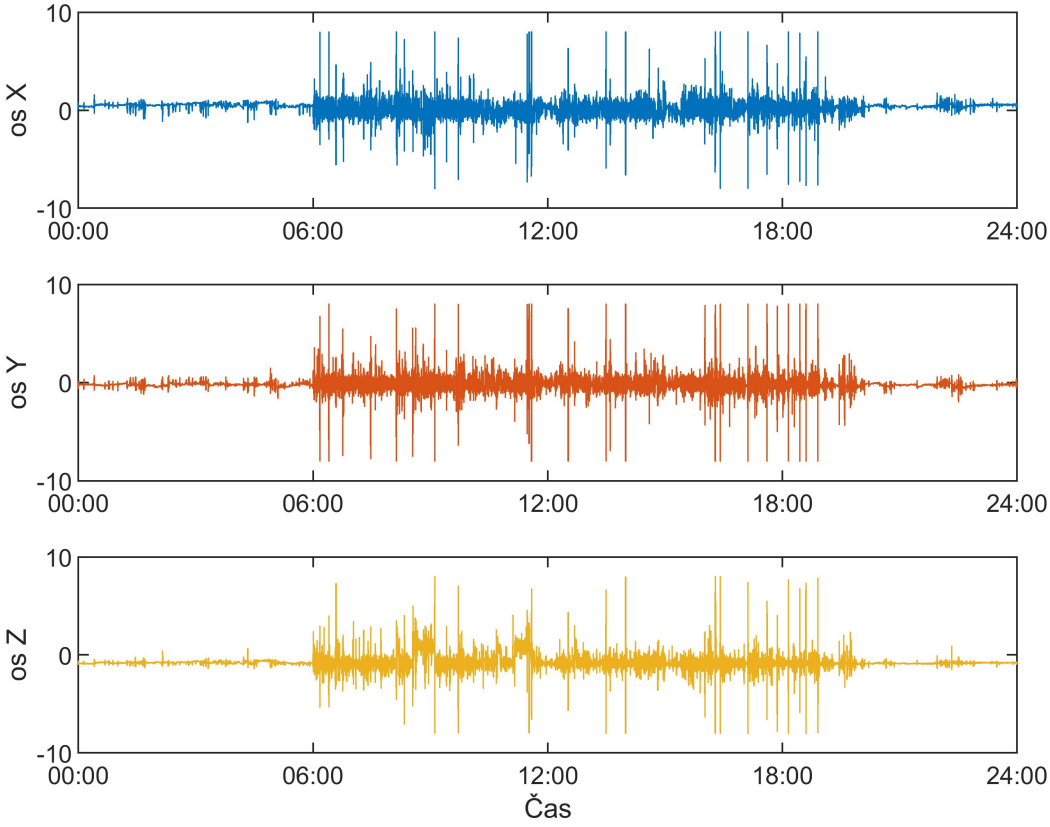
Obr. 3: Priebeh jednotlivých experimentov na podstielke, vykonaných centrom biovied SAV.



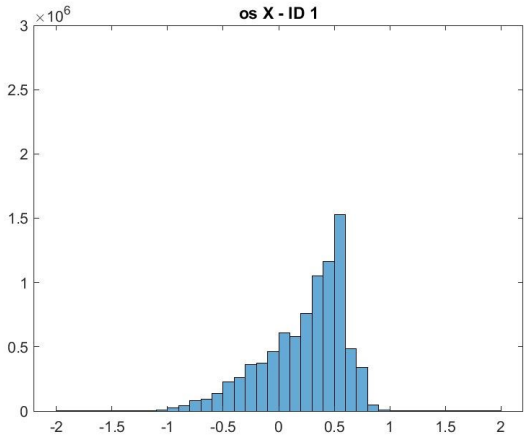
Obr. 4: Batohy s identifikačným číslom na uchytenie senzora použité v **Experimente 2**. Dané batohy vyrobila SAV z dôvodu pevnejšieho uchytenia na chrbát kury, z dôvodu nefunkčnosti pôvodných batohov použitých v **Experimente 1**.

Experimenty vykonané centrom biovied SAV, teda **Experiment 1** a **Experiment 2**, pracovali súčasne s 10 senzormi, keďže sa monitorovala vždy skupina piatich, resp. desiatich sliepok. To je hlavný rozdiel oproti prístupu z článku [1], v ktorom bola monitorovaná iba jedna kura. Pre jednoduchosť a prehľadnosť sme každému senzoru, resp. kure, priradili jedinečný identifikátor vo forme (ID číslo), korešpondujúci s číslami na batohoch vyobrazených na Obr. 4.

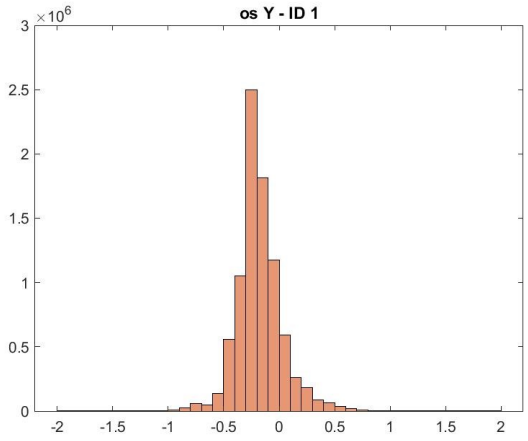
Záver tejto kapitoly venujeme popísaniu štruktúry dát, s ktorými pracujeme v diplomovej práci. Na Obr. 5 sme vizualizovali časové rady v jednotlivých osiach pre kuru ID 1 z **Experimentu 1**, pričom sme sledovali iba prvých 24 hodín experimentálneho záznamu. Môžeme vidieť variabilitu v časových radoch spôsobenú zníženou pohybovou aktivitou kury počas noci. Ďalej sme vizualizovali histogramy pohybu kury ID 1 z **Experimentu 1**. Danými histogramami sme objasnili, kde sa kura počas dňa najčastejšie nachádzala v rámci jednotlivých osí z pohľadu meracieho prístroja. Na Obr. 6a vidíme pre os X približne asymetrické rozdelenie, na Obr. 6b pre os Y dostávame rozdelenie jemne symetrické a nakoniec na Obr. 6c máme pre os Z rozdelenie bimodálne. Bimodálne rozdelenie v osi Z môžeme interpretovať ako možné prechádzanie medzi stavmi, keď kura stojí, alebo sedí, keďže os Z sleduje pohyb vo vertikálnom smere. Treba si však uvedomiť čisto informatívny charakter týchto štatistík, keďže senzor môže byť počas monitorovania jemne vychýlený.



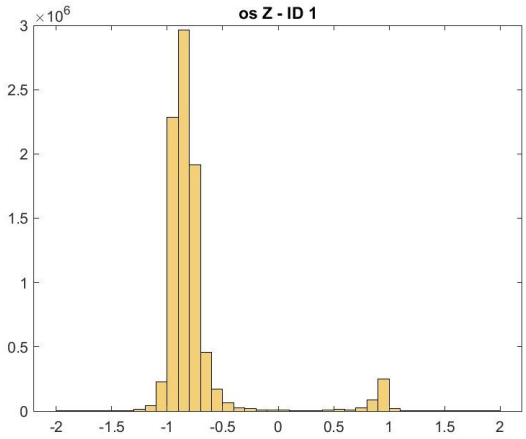
Obr. 5: Vizualizácia časových radov v jednotlivých osiach. Dáta použité na danú vizualizáciu sú z **Experimentu 1**, pričom bola analyzovaná kura ID 1 počas prvých 24 hodín experimentu. V jednotlivých osiach vidno menšiu pohybovú aktivitu kury počas noci.



(a) Histogram osi X



(b) Histogram osi Y



(c) Histogram osi Z

Obr. 6: Distribúcie pohybu v jednotlivých osiach na dátach z prvého dňa monitorovania **Experimentu 1**, pričom bola analyzovaná kura s identifikačným číslom ID 1. Vo všetkých histogramoch je horizontálna os rozdelená na dieliky veľkosti $0,1$. Sledujeme teda napr., ako často sa kura nachádzala v intervale $(i, i+0,1]$, pre $i = (-2, -1,9, \ldots, 1,9)$ v danej osi. Zdôraznujeme, že vertikálna os je preškálovaná na rovnaké hodnoty vo všetkých histogramoch. Na obrázkoch môžeme vidieť typ rozdelenia, ktorým sa riadi pohyb kury v daných osiach.

2 Typy správania kury domácej

Informácia o správaní kury domácej má pre hospodársky chov veľkú výpovednú hodnotu, pretože správanie kúr veľakrát vykazuje koreláciu s ich aktuálnym zdravotným stavom. Samozrejme, používanie senzorov na monitorovanie jednotlivcov je nákladné a preto je až nemožné monitorovať každú jednu kuru v chove. Oveľa výhodnejšie je vykonať experimenty na menšej skupine a zistiť, ako rôzne podmienky ovplyvňujú zdravie kury [1]. Následne môžu byť poznatky získané z týchto experimentov použité na celom chove.

Teraz uvedieme jednotlivé správania, ktoré podľa literatúry [1, 10] a odborníkov vykazujú koreláciu so zdravím kury domácej.

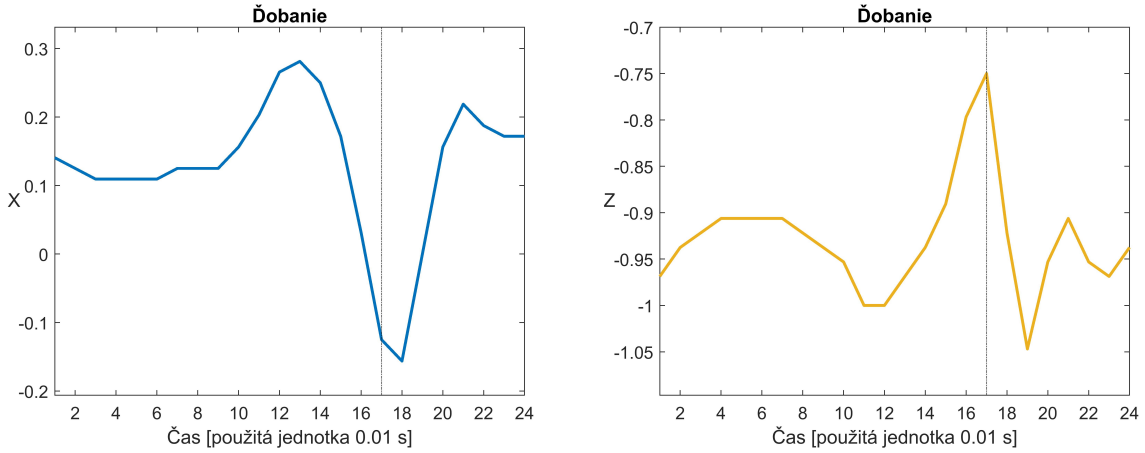
- **Feeding/Pecking**: ďobanie do zeme z dôvodu prijímania potravy, alebo interakcie s inou kurou, alebo ľubovoľným predmetom.
- **Preening**: čistenie peria zobákom.
- **Dustbathing**: oprašovanie, kura je v sede, alebo v ľahu, pričom trepe krídlami vo vertikálnom smere.

Keďže pracujeme s časovými radmi, je vhodné na začiatok uviesť príklad ako je napr. behaviorálny stav typu ďobanie, reprezentovaný pomocou časového radu. V článku [1] sledovali pri danom behaviorálnom stave pohyb v osi X, ktorá reprezentuje takmer horizontálny pohyb kury smerom dopredu a dozadu a pohyb v osi Z reprezentujúci takmer vertikálny pohyb kury smerom hore a dole. Výraz "takmer", spomenutý v predošlej vete, reflektuje mierne nakrivenie senzora na chrbte kury. Preto namiesto termínu horizontálny používame spojenie takmer horizontálny. Os Y je v tomto prípade možné zanedbať, keďže pri ďobaní do zeme sa kura pohybuje do strán iba v minimálnej miere oproti smerodajnému pohybu v osi X a osi Z. Pripomíname, že jednotlivé osi, ako aj ich približné smery, možno vidieť na Obr. 2 z kapitoly 1. V nasledovných riadkoch objasníme pohyb kury pri behaviorálnom stave typu ďobanie v smerodajných osiach X a Z. Na Obr. 7a, ktorý vizualizuje pohyb senzora v osi X pri správaní typu ďobanie, znamená rastúcosť pohyb senzora smerom dozadu a naopak klesajúcosť pohyb smerom dopredu. Na Obr. 7b sa nachádza pohyb v osi Z pri ďobaní, kde konkrétne rastúcosť znamená pohyb senzora smerom dole a naopak klesajúcosť pohyb smerom hore. Tieto dva obrázky dohromady teda reprezentujú

pohyb senzora v horizontálnom a vertikálnom smere, pričom sa jedná o jeden konkrétny príklad ďobania. Ako je možné vidieť na Obr. 7a, ďobanie v osi X začína jemným stúpaním (kladný smer X), čo predstavuje jemný záklon hlavy dozadu. Od bodu 13 na horizontálnej osi začína prudké klesanie (záporný smer X), čo v kontexte osi X znamená pohyb dopredu. Po prudkom klesaní pokračuje pohyb v osi X od bodu 18 prudkým stúpaním (kladný smer X), čo predstavuje vrátenie sa smerom dozadu do pôvodnej, neutrálnej polohy. Analogicky interpretujeme pohyb senzora vo vertikálnom smere, resp. osi Z na Obr. 7b. Zo začiatku na Obr. 7b vidíme akúsi predprípravu na samotné ďobnutie do zeme, t.j jemné stúpanie nasledované klesaním na osi Z. To v kontexte pohybu senzora predstavuje pohyb smerom dolu a následne smerom hore. Ďalej od bodu 12 na horizontálnej osi dostávame prudké stúpanie, teda pohyb senzora dolu k zemi. Nakoniec dostávame analogicky, ako pri osi X, vrátenie sa do neutrálnej polohy, t.j pohyb senzora smerom hore. Ako je možné postrehnúť, niekde v okolí bodu lokálneho minima časového radu na Obr. 7a a lokálneho maxima časového radu na Obr. 7b, t.j. v okolí bodu 17 na horizontálnej osi oboch vizualizácii sa nachádza samotný dotyk zobáka so zemou. Dodatočne, pre doplnenie interpretácie prikladáme Obr. 8, v ktorom sme farebne znázornili pohyb senzora vzájomne v osi X a osi Z počas správania typu ďobanie.

Rovnako prikladáme časový rad reprezentujúci behaviorálny stav typu čistenie na Obr. 9, ktorý je adekvátne reprezentovaný pohybom v osi Z [1]. Behaviorálny stav typu oprašovanie je vizualizovaný na Obr. 10 a je vhodne reprezentovaný osou X a osou Z [1]. Posledné dve spomenuté správania, čistenie a oprašovanie, sú však náročnejšie na interpretáciu, keďže, ako je vidno z Obr. 9 a Obr. 10, sa jedná o komplikované sekvencie.

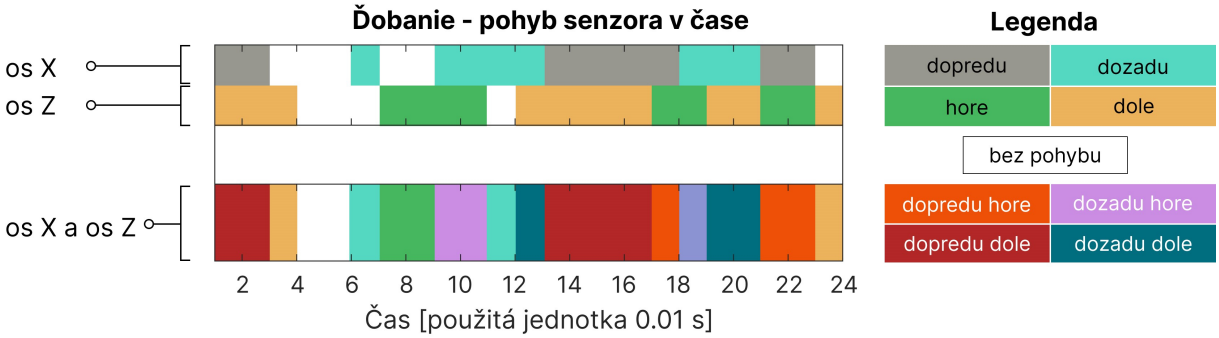
Podotýkame, že uvedené vizualizácie, reprezentujúce jednotlivé správania, sú akýmsi najkvalitnejším reprezentantom daného správania, ktorý identifikoval zložitý algoritmus popísaný v nasledovnej kapitole 3 na dátach z **Experimentu Abdoli** [1].



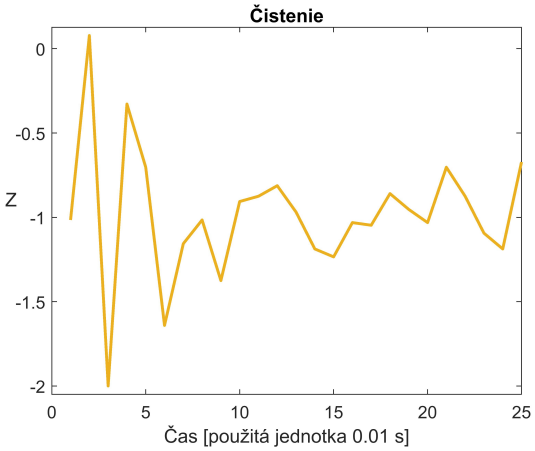
(a) Pohyb senzora v osi X.

(b) Pohyb senzora v osi Z.

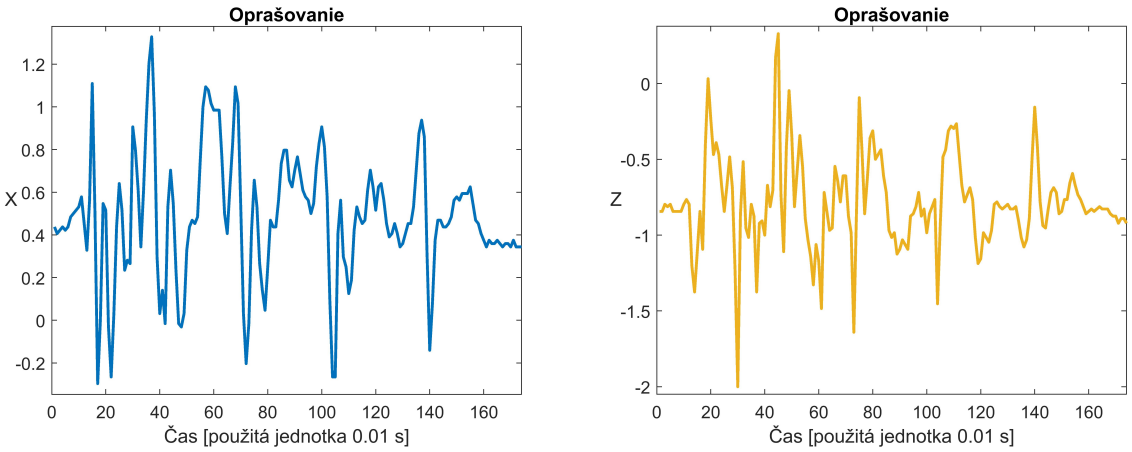
Obr. 7: Ďobanie do zeme reprezentované časovým radom v osi X a osi Z. Vykreslená vertikálna čiara predstavuje približný čas, v ktorom sa zobák dotkol zeme.



Obr. 8: Analýza pohybu separátne v osi X, osi Z a vzájomne v daných osiach pri behaviorálnom stave typu ďobanie, ktorý je vizualizovaný na predchádzajúcom Obr. 7.



Obr. 9: Čistenie reprezentované časovým radom v osi Z.



(a) Pohyb senzora v osi X.

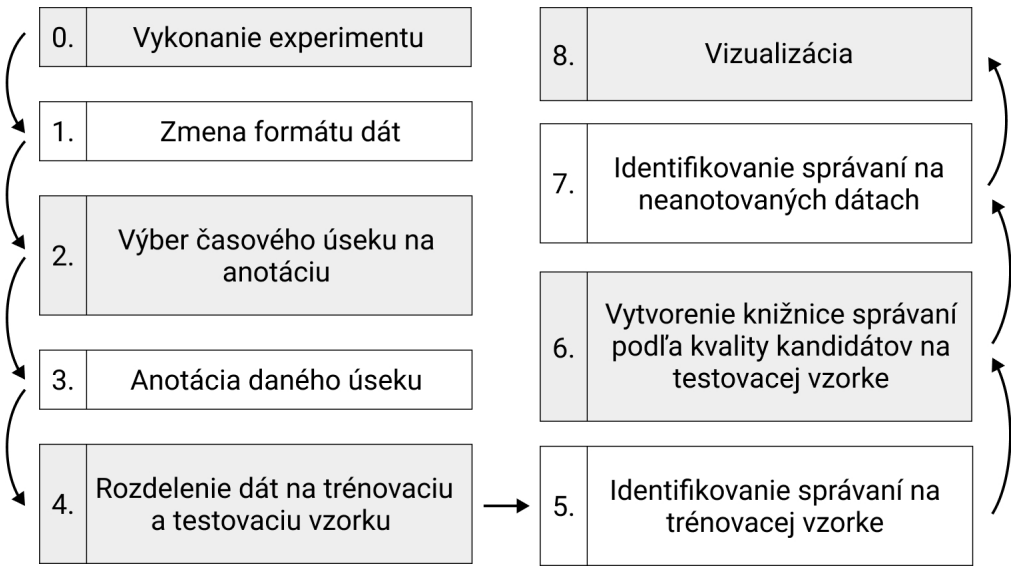
(b) Pohyb senzora v osi Z.

Obr. 10: Oprašovanie reprezentované časovým radom v osi X a osi Z.

3 Použité metódy

V tejto kapitole predstavíme komplexnú metódu na automatickú identifikáciu behaviorálnych stavov zo zaznamenaných rozsiahlych časových radov. K dosiahnutiu lepšej orientácie v klasifikačnom procese pre čitateľa uvedieme akúsi 'pipeline' algoritmu, resp. postupnosť krokov potrebných na identifikáciu a klasifikovanie behaviorálnych stavov, pričom jednotlivé kroky podrobnejšie vysvetlíme.

Na Obr. 11 vizualizujeme postupnosť základných krokov komplexného algoritmu, ktoré je nutné spraviť v procese klasifikácie behaviorálnych stavov kury domácej. Zdôrazňujeme, že jednotlivé kroky predstavujú iba základnú pointu, niektoré z daných krokov vysvetlíme podrobnejšie aj s potrebným teoretickým základom neskôr v tejto kapitole.

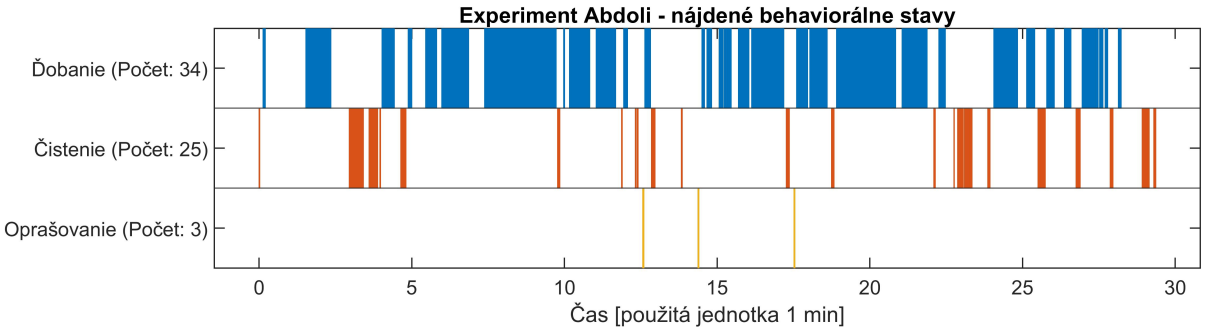


Obr. 11: Priebeh základných krokov, ktorými v danom poradí postupujeme pri identifikácii a klasifikovaní behaviorálnych stavov.

- Krok 0: **Vykonanie experimentu** - Obdržanie časových radov vo formáte *.cwa* z experimentu, ktorý je nastavený rovnako ako je popísané v kapitole 1.
- Krok 1: **Zmena formátu dát** - Zmena senzorového formátu dát z *.cwa* na *.mat* [13], keďže pracujeme v softvérovom prostredí MATLAB.
- Krok 2: **Výber časového úseku na anotáciu** - Prezretie videozáznamu z experimentu odborníkom a vybratie časového úseku dostatočnej dĺžky, napr. 30 minút

v ktorom sa nachádza dostatok z každého správania, ktoré potrebujeme klasifikovať.

- Krok 3: **Anotácia daného časového úseku** - Identifikácia jednotlivých správaní odborníkom na vybratom časovom úseku z predošlého kroku. Anotácia je v podobe tzv. 'weakly-labeled' dát. Odborník teda napríklad označí, že v týchto troch sekundách sa nachádza niekoľko príkladov správania typu ďobanie. Anotácia správaní je vytvorená odborníkom pomocou videozáznamu z vykonaného experimentu. Ako ukážku sme na Obr. 12 vizualizovali etogram anotovaných správaní na 30 minútovom zázname, ktoré anotovali odborníci na **Experimente Abdoli** [1].



Obr. 12: Anotované správania v **Experimente Abdoli**

- Krok 4: **Rozdelenie dát na trénovaciu a testovaciu vzorku** - Použitie tzv. 'cross-validation', čo predstavuje rozdelenie anotovaného datasetu v určenom pomere na trénovaciu a testovaciu vzorku. V článku [1] sa anotovaný dataset rozdelil na polovicu tak, aby anotovaný interval správania nepresahoval z trénovacej vzorky do testovacej vzorky a naopak.

- Krok 5: **Identifikovanie správaní na trénovacej vzorke** - Algoritmus je tzv. 'shape-based algorithm', ktorý sleduje tvar časového radu v danej osi a pomocou metriky Euklidovských vzdialeností meria akúsi vzdialenosť medzi dvomi časovými radmi rovnakej dĺžky. Vo všeobecnosti hľadáme takého kandidáta pre jednotlivé správanie, ktorý mal v trénovacej vzorke najväčší počet najbližších susedov, pričom všetci boli 'true positive', resp. boli obsiahnutý v anotovanom intervale pre dané správanie.

- Krok 6: **Vytvorenie knižnice správaní podľa kvality kandidátov na testovacej vzorke** - Knižnica správaní pozostáva z najlepších kandidátov jednotlivých správaní nájdených na trénovacej vzorke, ktorých kvalita je otestovaná na testovacej vzorke. Na vyhodnotenie kvality sa používali rôzne 'cross-validation' miery kvality, ako 'True-positives ratio', 'Precision' alebo 'Accuracy'. Daný krok vysvetlíme podrobnejšie neskôr v tejto kapitole.

- Krok 7: **Identifikovanie správaní na neanotovaných dátach** - Klasifikácia správaní na dátach, pričom ako predlohu najkvalitnejších správaní používame knižnicu správaní vytvorenú v predošlom kroku.

- Krok 8: **Vizualizácia** - Vizualizácia identifikovaných správaní na celých, neanotovaných dátach pomocou etogramu, čo je graf zobrazujúci jednotlivé správania v čase.

Ďalej je vhodné podrobnejšie vysvetliť techniky, s ktorými pracuje klasifikačný algoritmus. Ako bolo spomenuté v kroku 5, algoritmus je tzv. 'shape-based', čo znamená, že podobnosť medzi dvomi časovými radmi rovnakej dĺžky meriame pomocou ich vzájomnej Euklidovskej vzdialenosti. Treba podotknúť, že podobnosť meriame vždy medzi dvomi lokálnymi podsekvenciami, resp. výsekmi časových radov rovnakých dĺžok a nie na celých časových radoch. S lokálnymi podsekvenciami pracujeme preto, lebo sa snažíme zachytiť jednotlivé správania, ktoré trvajú relatívne krátku dobu. Spomenuté podsekvencie, resp. výseky časových radov rovnakej dĺžky je však vhodné najprv Z-normalizovať, pretože nás zaujímajú najmä relatívne zmeny v časových radoch. Na nasledovných riadkoch vo všeobecnosti uvedieme Z-normalizáciu časových radov, ako aj výpočet vzdialenosti medzi dvomi Z-normalizovanými časovými radmi.

Predpokladajme vo všeobecnosti 2 časové rady dĺžky n.

$$x = x_1, \ldots, x_n,$$

$$y = y_1, \ldots, y_n.$$

V ďalšom kroku aplikujeme na dané časové rady Z-normalizáciu.

$$\hat{x_i} = \frac{x_i - \mu_x}{\sigma_x}, \qquad \hat{y_i} = \frac{y_i - \mu_y}{\sigma_y},$$

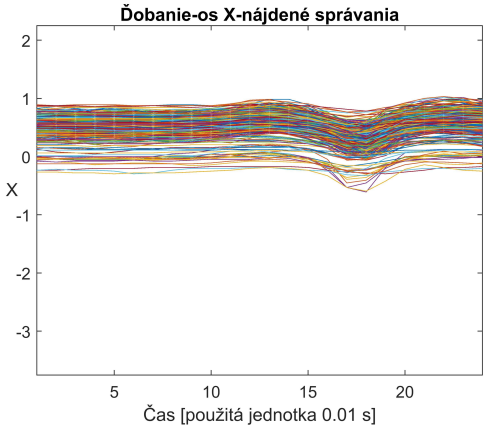
kde $i = 1, 2, \ldots, n$. Člen μ predstavuje strednú hodnotu daného časového radu a člen σ štandardnú odchýlku daného radu. Euklidovská vzdialenosť medzi danými Z-normalizovanými podsekvenciami je potom

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (\hat{x}_i - \hat{y}_i)^2}.$$

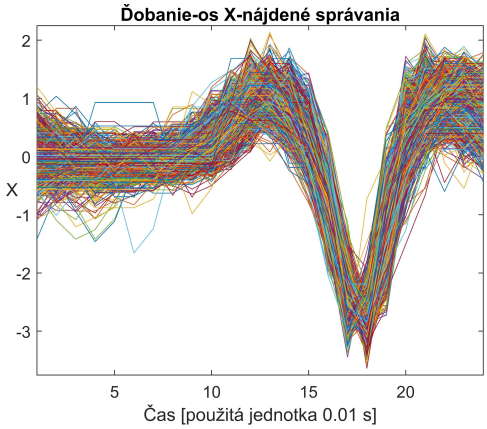
Pripomíname, že Z-normalizovaná podsekvencia má nulovú strednú hodnotu a štandardnú odchýlku rovnú jedna.

Euklidovská vzdialenosť medzi dvomi časovými podsekvenciami nám slúži na určenie ich vzájomnej podobnosti. Ak je teda vzájomná vzdialenosť dvoch Z-normalizovaných časových podsekvencii malá, tak je zrejmé, že sú podobné. V kapitole 1 sme spomenuli, že pracujeme s časovými radmi v troch osiach. Ďalej sme v závere kapitoly 2 objasnili, ktoré osi sú smerodajné pre jednotlivé správania. Zoberme si ako príklad behaviorálny stav typu ďobanie, pri ktorom je smerodajná os X a os Z, pričom os Y zanedbávame. Pre získanie Euklidovskej vzdialenosti medzi dvomi Z-normalizovanými podsekvenciami pre ďobanie musíme najprv spočítať Euklidovskú vzdialenosť časových podsekvencii separátne pre os X a následne pre os Z. Spočítaním spomenutých vzdialeností pre dané osi dostaneme celkovú Euklidovskú vzdialenosť daných dvoch podsekvencií.

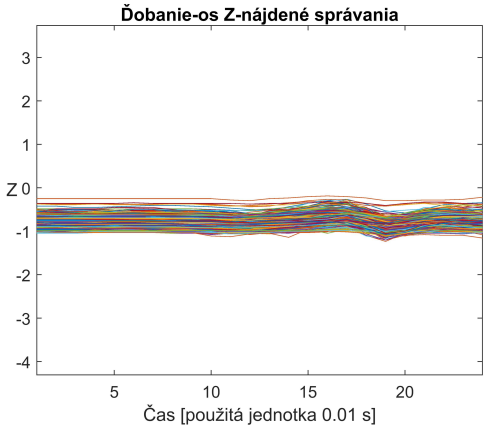
Pre porovnanie nenormalizovaných a Z-normalizovaných podsekvencii vizualizujeme na Obr. 13 všetky identifikované behaviorálne stavy typu ďobanie na neanotovanom datasete z **Experimentu Abdoli**, pričom knižnica správaní bola vytvorená na anotovanom datasete z rovnakého **Experimentu Abdoli**. Na Obr. 13a a Obr. 13b, resp. Obr. 13c a Obr. 13d, sú vykreslené dané správania v osi X, resp. osi Z. Ľavá strana Obr. 13 sa týka nenormalizovaných podsekvencií a na pravej strane sú identické podsekvencie, ale Z-normalizované. Na Obr. 13a a Obr. 13c vidíme, ako Z-normalizácia zdôrazní relatívne zmeny v pohybe, vďaka čomu vykazuje hľadanie podobnosti medzi jednotlivými podsekvenciami časového radu kvalitnejšie výsledky, ako bez použitia Z-normalizácie.



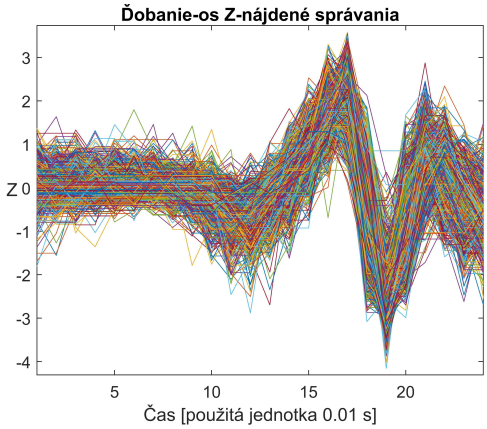
(a) Nenormalizované podsekvencie v osi X.



(b) Z-normalizované podsekvencie v osi X.



(c) Nenormalizované podsekvencie v osi Z.



(d) Z-normalizované podsekvencie v osi Z.

Obr. 13: Algoritmom nájdené behaviorálne stavy typu ďobanie v osi X a osi Z. Podsekvencie boli nájdené na **Experimente Abdoli** za použitia knižnice, vytvorenej rovnako na **Experimente Abdoli**. Naľavo na Obr. 13a a Obr. 13c sú vykreslené podsekvencie správania bez použitia normalizácie a napravo na Obr. 13b a Obr. 13d sú zobrazené identické správania, pričom každá podsekvencia je Z-normalizovaná.

Aby sme upresnili krok 5, dĺžka prechádzaných podsekvencií je nami zvolená hodnota, ktorá vstupuje do algoritmu. Empirické pozorovanie [1] udáva približnú dobu trvania jednotlivých správaní nasledovne.

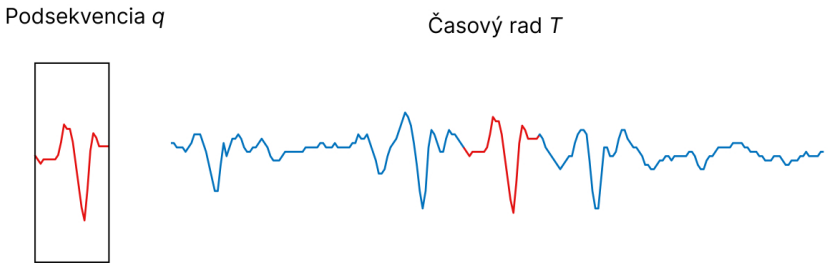
- ďobanie: doba trvania je približne $0,25$ sekundy,
- čistenie: doba trvania je približne $0,25$ sekundy,

- oprašovanie: doba trvania je približne $1,75$ sekundy.

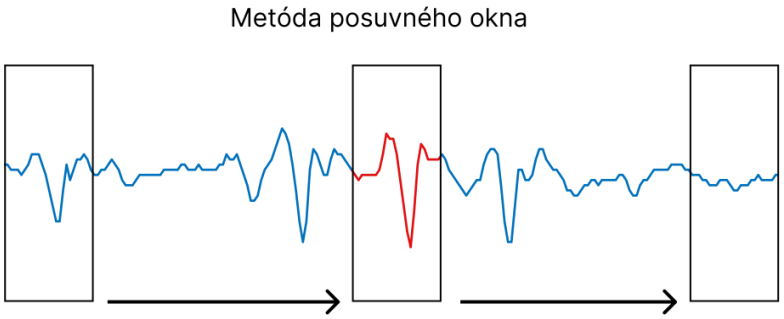
Ako bolo bližšie popísané v kapitole 1, senzor na chrbte kury zaznamenáva údaje o polohe s výberovou frekvenciou 100 Hz, resp. 100 údajov za sekundu. Pre správania typu ďobanie a čistenie hľadáme kandidáta, ktorého dĺžka je približne $0,25s$, takže postupne v algoritme prechádzame podsekvencie dĺžky $24, 25$ a 26. Naopak pre oprašovanie postupne prechádzame podsekvencie dĺžky $174, 175$ a 176, keďže hľadáme kandidáta, ktorého približná dĺžka je $1,75s$.

Ďalším dôležitým krokom je samotné počítanie vzdialeností. Na to, aby sme vypočítali vzdialenosť medzi jednou konkrétnou podsekvenciou časového radu a všetkými ostatnými podsekvenciami časového radu tej istej dĺžky, je potrebné použiť metódu posuvného okna. Pre uvedenie príkladu si zoberme časový rad T dĺžky n. Z daného časového radu T vyberieme ľubovoľnú podsekvenciu, označme ju q, ktorá má dĺžku m, pričom platí nerovnosť $n > m$. Postupne teda prechádzame každú podsekvenciu časového radu T dĺžky m a každej priradíme jej Euklidovskú vzdialenosť s podsekvenciou q. Ako príklad uvádzame na Obr. 14 grafické znázornenie metódy posuvného okna na reálnych dátach z **Experimentu Abdoli**, pričom, bez ujmy na všeobecnosti, sledujeme os X. Daná vizualizácia slúži ako príklad na pochopenie metódy pre čitateľa, a preto z dôvodu zanechania prehľadnosti zanedbávame pri počítaní vzájomnej vzdialenosti podsekvencií Z-normalizáciu každej podsekvencie.

Po vypočítaní Euklidovskej vzdialenosti medzi posekvenciou q a všetkými podsekvenciami rovnakej dĺžky časového radu T dostávame vektor vzdialeností, ktorý nazývame dištančný profil pre podsekvenciu q. Na Obr.15 je vizualizovaný rovnaký časový rad T so zvýraznenou podsekvenciou q, ako v Obr. 14, pričom je doplnený o vypočítaný dištančný profil pre podsekvenciu q. Je vhodné si všimnúť, že Euklidovská vzdialenosť medzi podsekvenciou q so sebou samou je rovná nule, čo je tzv. triviálna zhoda.



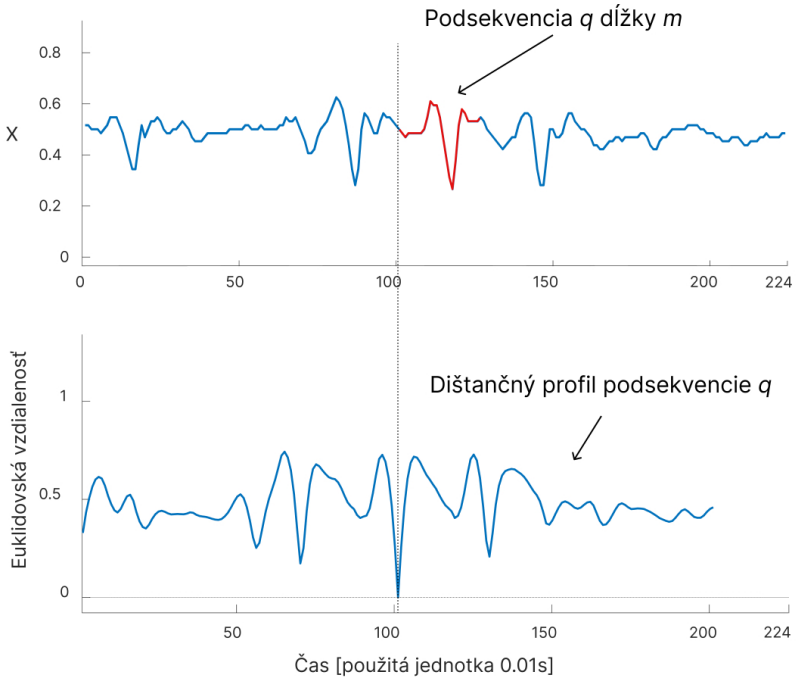
(a) Vizualizácia časového radu T dĺžky $n = 224$ v osi X z **Experimentu Abdoli**. Červenou farbou zvýraznená podsekvencia q dĺžky $m = 24$ znázorňuje behaviorálny stav typu ďobanie.



(b) Aplikovanie metódy posuvného okna na časový rad T pričom pri každom posune okna je počítaná Euklidovská vzdialenosť časovej podsekvencie vnútri posuvného okna so zvýraznenou podsekvenciou q.

Obr. 14: Znázornenie metódy posuvného okna na reálnych dátach z **Experimetu Abdoli** (vlastné spracovanie).

Pri práci s rozsiahlymi časovými radmi je vhodné brať do úvahy aj efektívnosť počítania dištančného profilu pomocou posuvného okna. Predpokladajme opäť časový rad T dĺžky n a z neho vybratú ľubovoľnú podsekvenciu q dĺžky m. Autor článku [5] Abdullah Mueen vytvoril so spoluautormi na danú manipuláciu s rozsiahlymi časovými radmi tzv. MASS algoritmus [6], v preklade 'Mueenov algoritmus hľadania podobnosti', ktorého výpočtová zložitosť je $\mathcal{O}(n \log n)$, čo je výrazné zefektívnenie oproti najintuitívnejšiemu 'brute force' algoritmu, ktorého výpočtová zložitosť je $\mathcal{O}(nm)$. Zmienený 'brute force' algoritmus musí prejsť celý časový rad pomocou posuvného okna dĺžky m, následne musí dané okno, resp. časovú podsekvenciu Z-normalizovať a nakoniec vypočítať Euklidovskú vzdialenosť medzi oknom a už Z-normalizovanou podsekvenciou q. Pracuje teda iteračne a v každom kroku robí výpočty, čo je pri rozsiahlych časových radoch časovo neefek-



Obr. 15: Vizualizácia dištančného profilu pre podsekvenciu q dĺžky $m = 24$, ktorá je vybratá z časového radu T dĺžky $n = 224$ z **Experimentu Abdoli**, pričom sledujeme os X. Dištančný profil pre podsekvenciu q je počítaný metódou posuvného okna, pričom zvýraznená podsekvencia q predstavuje ďobanie nájdené algoritmom na **Experimente Abdoli**. ([1], vlastné spracovanie)

tívne. Na rozdiel od toho sa v MASS algoritme vypočítajú pre Z-normalizáciu všetky potrebné hodnoty v lineárnom čase pomocou kumulatívnych súčtov podsekvencii, ktoré následne od seba odpočítavame [7]. Ďalej sa pomocou rýchlej Fourierovej transformácie pretransformuje časový rad a analyzovaná podsekvencia do frekvenčnej oblasti a vykoná sa medzi nimi operácia skalárneho súčinu. Inverznou Fourierovou transformáciou sa vrátime do časovej oblasti a vypočítame dištančný profil analyzovanej podsekvencie [8]. Týmto prístupom sa výrazne urýchli výpočet dištančného profilu.

Primárnym cieľom je krok 6, teda vytvorenie knižnice správaní. Daná knižnica správaní pozostáva z najlepších kandidátov z trénovacej vzorky, ktorí boli otestovaní na testovacej vzorke. Najlepší kandidát jednotlivého správania bol oproti ostatným kandidátom najviackrát nájdený v trénovacej vzorke, pričom všetky nájdené správania sú typu 'true positives', t.j nachádzali sa vrámci anotovaného intervalu pre dané správanie.

Pri vyhodnocovaní najdených správaní, resp. kandidátov, na trénovacej vzorke pomocou testovacej vzorky musíme zohľadniť, že anotácie sú tzv. 'weakly labeled', čo znamená,

že v jednom anotovanom intervale sa môže nachádzať niekoľko daného správania. Dané správania sa ale celkom s určitosťou nachádzajú aj mimo anotovaných intervalov, čo môže byť spôsobené chybnou anotáciou, napríklad kvôli prekryvu jednotlivých kúr na videozázname [1]. Z toho dôvodu využili autori článku [1] koncept 'Multiple instance learning', ktorý každý anotovaný interval správania považuje za akýsi zhluk, v ktorom sa nachádza niekoľko príkladov daného správania. Ak sa aspoň jedno klasifikované správanie vyskytuje vrámci daného zhluku identického správania, tak celý zhluk považujeme za 'true positive'. Naopak, ak nie je ani jedno klasifikované správanie vrámci daného zhluku, tak celý zhluk považujeme za 'false negative'. Ak je klasifikované správanie mimo anotovaného zhluku, alebo sa nachádza v zhluku prislúchajúcemu inému typu správania, tak celý zhluk označíme ako 'false positive'. Nakoniec, ak sa žiadne klasifikované správanie nenachádza v anotovanom zhluku iného správania, tak daný zhluk považujeme za 'true negative'.

Knižnica správaní konkrétne obsahuje zoznam správaní v daných osiach vo forme, kde každé správanie konkretizujeme, resp. vytvoríme stĺpec **správanie**, v ktorom bude zapísané o aké správanie sa jedná. V našom prípade teda vyberieme z behaviorálnych stavov typu ďobanie, čistenie a oprašovanie. Ďalší stĺpec **threshold** obsahuje pre jednotlivé správanie threshold hodnotu, čo je maximálna hraničná Z-normalizovaná Euklidovská vzdialenosť, pri ktorej sme v trénovacej vzorke nedostali ani jeden 'False positive' príklad, resp. všetky nájdené správania v anotovanej trénovacej vzorke boli 'True positives'. Inými slovami, nájdené správanie v trénovacej vzorke, ktoré presahovalo vzdialenosť thresholdu, už bolo 'False positive', resp. nebolo obsiahnute v anotovanom intervale pre dané správanie. Zdôraznujeme, že ak sme pri jednotlivom správaní sledovali viac ako jednu os, ako napríklad pri správaní typu ďobanie, tak threshold hodnota pre dané správanie je súčtom Z-normalizovaných Euklidovských vzdialeností v daných osiach. Tretí stĺpec už iba konkretizuje os, ktorú sledujeme pri jednotlivom správaní. Nakoniec sú pridané stĺpce obsahujúce dátové body jednotlivých správaní v daných osiach najlepšieho kandidáta, resp. časovej podsekvencie daného správania v danej osi. V tabuľke 2 je získaná knižnica správaní najlepších kandidátov z **Experimentu Abdoli**. Najkvalitnejší kandidát z anotovaného datasetu na **Experimente Abdoli** pre ďobanie dosahoval dĺžku časového radu 24, pre čistenie 25 a pre oprašovanie 175. Z toho dôvodu sú hodnoty presahujúce dĺžku najlepšieho kandidáta daného správania typu NA tak, ako je to znázornené v tabuľke 2.

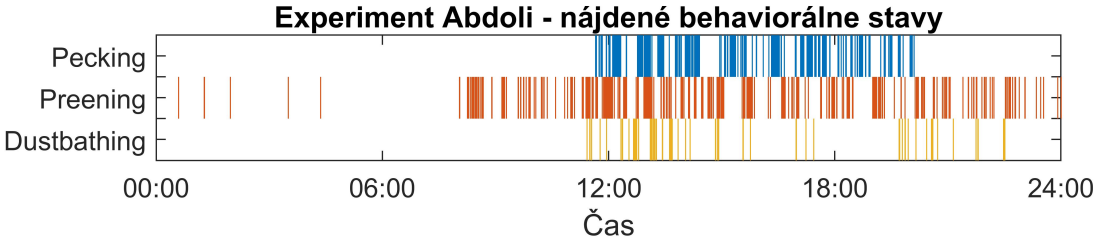
správanie	threshold	os	1	2	...	24	25	26	...	175
ďobanie	5,8599	X	0,141	0,125	...	0,172	NA	NA	...	NA
ďobanie	5,8599	Z	-0,969	-0,938	...	-0,938	NA	NA	...	NA
čistenie	4,0729	Z	-1,016	0,078	...	-1,188	-0,672	NA	...	NA
oprašovanie	29,1480	X	0,906	0,797	...	0,453	0,484	0,750	...	0,188
oprašovanie	29,1480	Z	-2,000	-0,859	...	-0,344	-0,547	-0,891	...	-0,844

Tabuľka 2: Knižnica správaní vytvorená na **Experimente Abdoli** [1]. Všetky dátové hodnoty, resp. stĺpce $1, 2, \ldots, 175$ sú pre účel čisto vizuálny zaokrúhlené na tri desatinné miesta.

Pri takto vytvorenej knižnici správaní predpokladáme, že existuje iba jeden spôsob akým je môžné vykonať správanie. Vo všeobecnosti však môže mať jednotlivé správanie zvieraťa viac rôznych príkladov. Keď hovoríme o kure domácej, mohli by sme všeobecné správanie typu ďobanie rozdeliť na ďobanie1, ďobanie2, atď. Taká knižnica správaní sa nazýva polymorfická [1].

Po úspešnom vytvorení knižnice pristupujeme ku kroku 7, čo je identifikovanie správaní na neanotovaných dátach za pomoci vytvorenej knižnice z predošlého kroku. Algoritmus opäť prechádza všetky podsekvencie posúvnym oknom, pričom počíta Euklidovskú vzdialenosť medzi Z-normalizovanou podsekvenciou posuvného okna a Z-normalizovanou podsekvenciou správania z vytvorenej knižnice správaní. Ak vzdialenosť nepresiahne hraničnú hodnotu, resp. threshold daného správania, označíme danú podsekvenciu ako korektne nájdené správanie.

Posledný krok 8 je vizualizácia nájdených správaní na celých dátach. Na Obr. 16 je vyhotovený etogram nájdených správaní kury z **Experimentu Abdoli** klasifikovaných pomocou knižnice správaní vytvorenej taktiež na **Experimente Abdoli**.



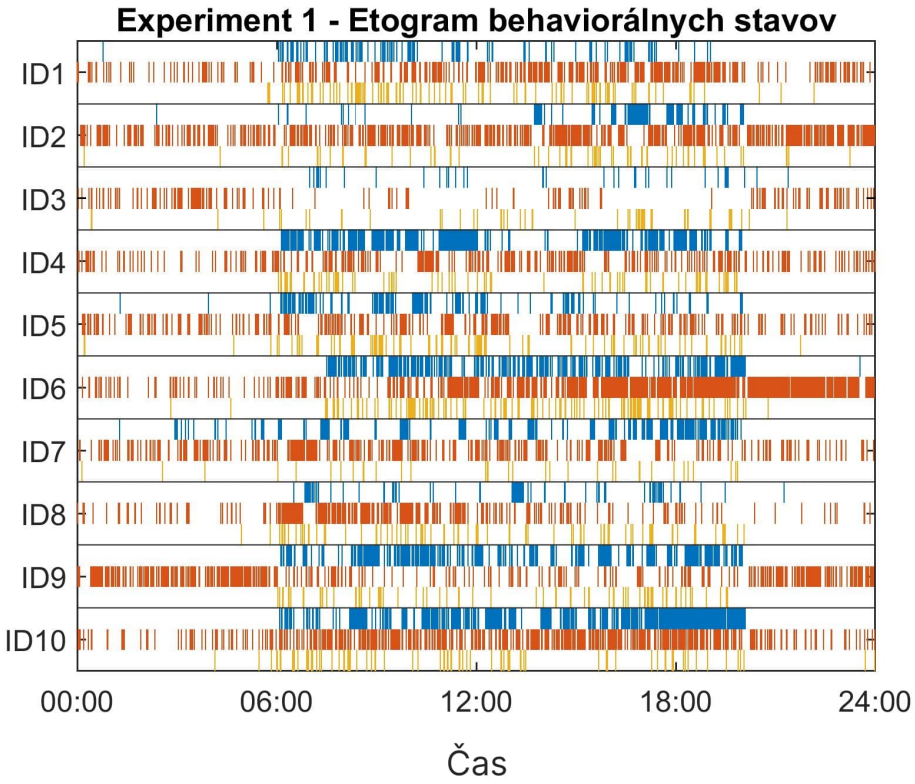
Obr. 16: Etogram jednotlivých správaní v **Experimente Abdoli**. Knižnica správaní bola vytvorená na trénovacích/testovacích dátach z **Experimentu Abdoli.**

4 Aplikácia na SAV dáta

Použité metódy vysvetlené v predchádzajúcej kapitole 3 aplikujeme na reálne dáta z **Experimentu 1** a **Experimentu 2**. Hlavným cieľom je vizualizácia identifikovaných správaní pomocou etogramov, ako aj rôzne ďalšie štatistiky a vizualizácie súvisiace s prejdenými vzdialenosťami kúr v jednotlivých experimentoch. Pripomíname, že konkretizáciu jednotlivých experimentov, resp. počet monitorovaných kúr a prostredie v ktorom boli monitorované, je znázornené v kapitole 1 na Obr. 1.

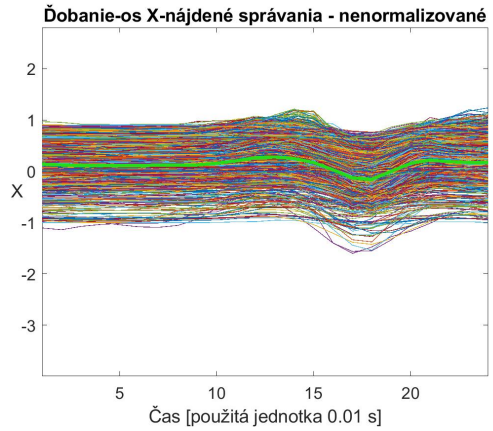
4.1 Aplikácia na cvičné dáta (Experiment 1)

V pilotnom **Experimente 1** sme sa primárne zameriavali na pochopenie dát a odstránenie limitácií spojených so samotným experimentom. Z toho dôvodu ani neboli spracované anotácie slúžiace na vytvorenie jedinečnej knižnice správaní pre daný **Experiment 1**. Využila sa teda knižnica správaní z publikovaného článku [1] (ďalej v práci "knižnica Abdoli"), ktorá nám aspoň naivne ukázala početnosť správaní v čase, pre kury z **Experimentu 1**. Je samozrejmé, že identifikované správania budú mať limitácie, keďže tento prístup vychádza z hypotézy, že sa kura z článku [1] pohybuje pri daných správaniach identicky ako kury z **Experimentu 1**. Interpretácia výsledkov pri takto nastavenom experimente, v ktorom používame knižnicu správaní nájdenú na iných dátach, môže byť iba vizuálneho charakteru, keďže nedisponujeme vlastnými anotáciami, vďaka ktorým by sme vedeli vypočítať presnosť a kvalitu klasifikátora. Etogram behaviorálnych stavov kúr z **Experimentu 1**, ktoré boli klasifikované pomocou knižnice správaní z **Experimentu Abdoli**, je vizualizovaný na Obr. 17. Ako je možné vidieť z daného Obr. 17, behaviorálny stav typu ďobanie (modrá farba) bol identifikovaný prevažne počas dňa naprieč všetkými kurami, čo súhlasí s časom, kedy kura reálne vykonáva dané správanie. Naopak behaviorálny stav typu čistenie (červená farba) je prítomný počas celého dňa, čo je bežný jav (pozn. diskutované so spolupracovníkmi zo SAV). Správanie typu oprašovanie sa korektne vyskytuje primárne počas dňa, rovnako ako ďobanie. Dané výsledky sú veľmi podobné výsledkom z článku [1], čo pripisujeme použitiu knižnice správaní zo spomenutého článku.

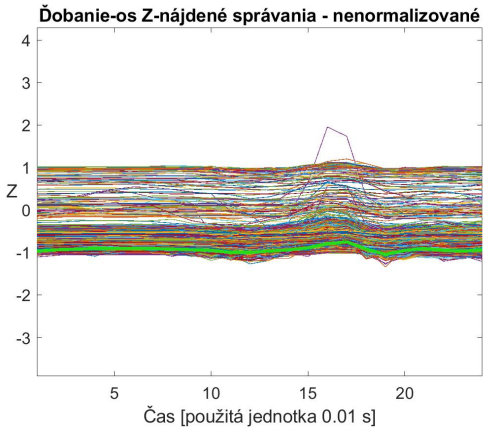


Obr. 17: Etogram jednotlivých správaní na kurách z **Experimentu 1**. Jednotlivé správania sú farebne rozlíšené, pričom modrá predstavuje ďobanie, červená čistenie a žltá oprašovanie. Knižnica správaní, použitá na identifikovanie správaní na neanotovaných dátach jednotlivých kúr z **Experimentu 1**, bola vytvorená na trénovacej a testovacej vzorke z **Experimentu Abdoli**.

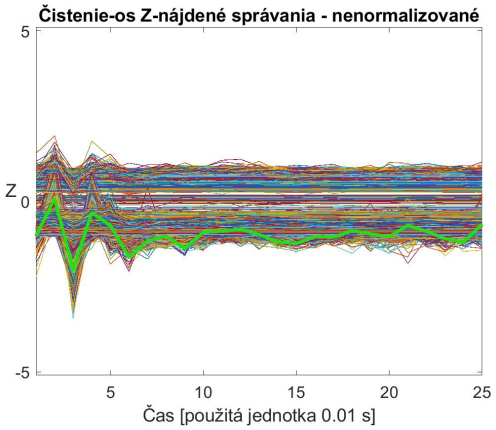
Ďalej na Obr. 18 a Obr. 19 vizualizujeme jednotlivé nájdené správania, ako pohyb v čase naprieč všetkými kurami, ktorých nájdené správania boli podobné so správaním obsiahnutým v knižnici z článku [1]. Na Obr. 18 vykreslujeme nenormalizované podsekvencie. Na Obr. 19 Z-normalizované podsekvencie nájdených správaní. Hrubšie a farebne zvýraznené podsekvencie reprezentujú najlepších kandidátov jednotlivých správaní v daných osiach z knižnice Abdoli.



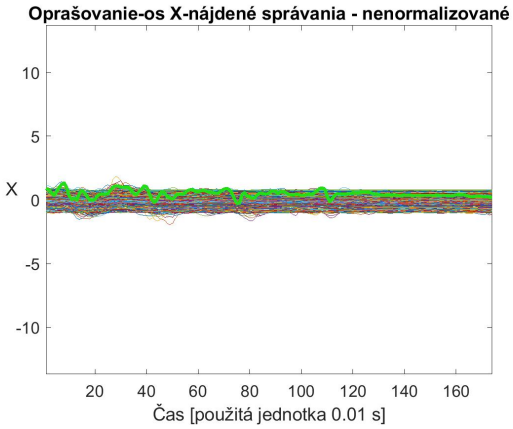
(a) Ďobanie v osi X.



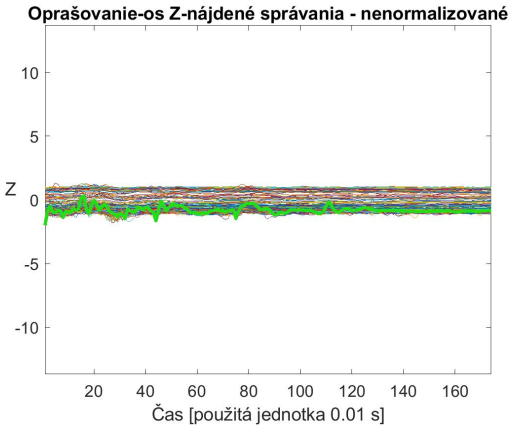
(b) Ďobanie v osi Z.



(c) Čistenie v osi Z.

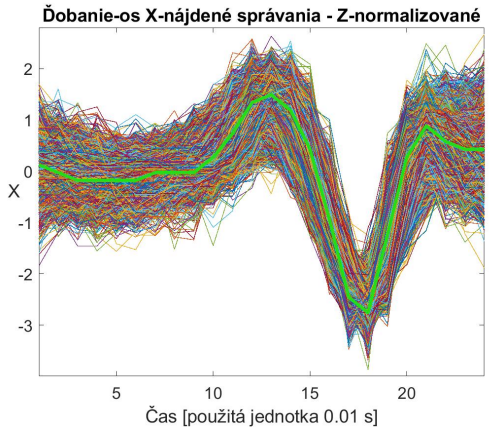


(d) Oprašovanie v osi X.

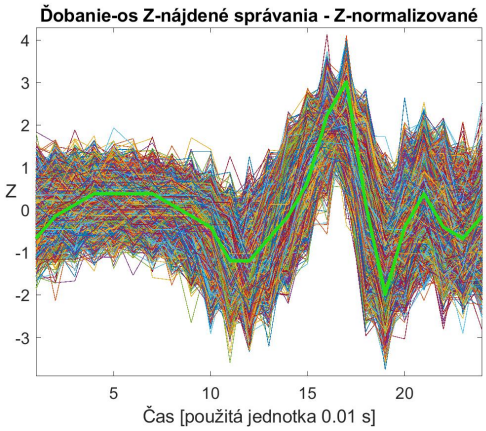


(e) Oprašovanie v osi Z.

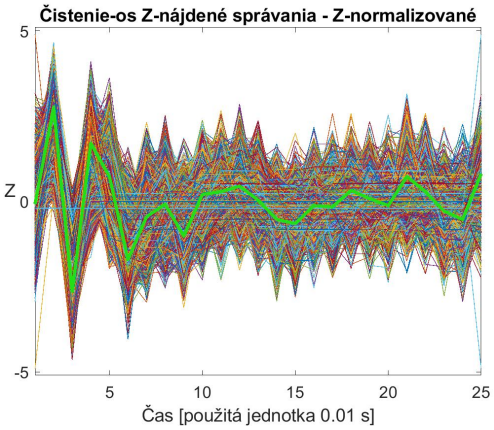
Obr. 18: Nájdené nenormalizované podsekvencie správaní všetkých 10 kúr z **Experimentu 1**. Knižnica použitá na nájdenie daných správaní je z **Experimentu Abdoli**. Zelenou farbou zvýraznené podsekvencie predstavujú najlepších kandidátov jednotlivých správaní v daných osiach z knižnice Abdoli.



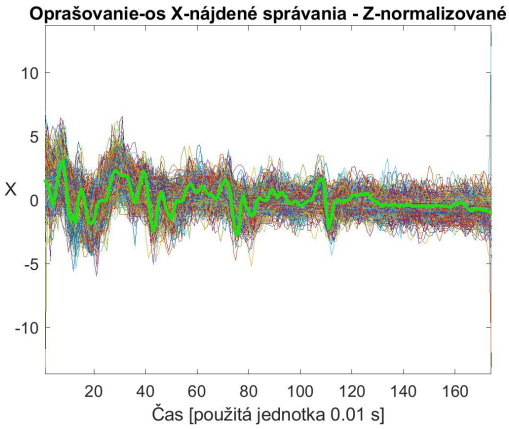
(a) Ďobanie v osi X.



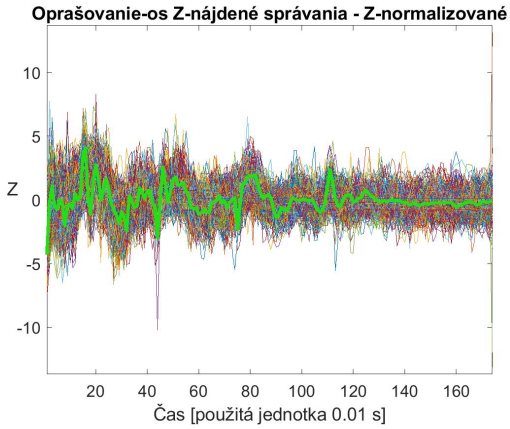
(b) Ďobanie v osi Z.



(c) Čistenie v osi Z.



(d) Oprašovanie v osi X.



(e) Oprašovanie v osi Z.

Obr. 19: Nájdené Z-normalizované podsekvencie správaní všetkých 10 kúr z **Experimentu 1**. Knižnica použitá na nájdenie daných správaní je z **Experimentu Abdoli**. Zelenou farbou zvýraznené podsekvencie predstavujú najlepších kandidátov jednotlivých správaní v daných osiach z knižnice Abdoli, na ktoré je rovnako aplikovaná Z-normalizácia.

4.2 Aplikácia na koncové dáta (Experiment 2)

Pri aplikácii na koncové dáta je hlavným cieľom vytvoriť vlastnú knižnicu správaní, pomocou ktorej môžeme klasifikovať jednotlivé správania na neanotovaných dátach. Vytvorenie jedinečnej knižnice správaní je možné vďaka anotáciam dodaných SAV. V prvom rade však zrealizujeme naivné identifikovanie správaní pomocou knižnice z článku [1], kde hlavnou motiváciou je porovnanie vizuálnych výsledkov s výsledkami identického prístupu na **Experimente 1**. Prvý prístup na vytvorenie vlastnej knižnice správaní, je spojenie anotácií jednotlivých kúr do jedného hromadného anotovaného datasetu, na ktorom nájdeme najkvalitnejšieho reprezentanta jednotlivého správania. Ďalší prístup je klasifikovanie na anotáciach jednotlivých kúr separátne, čo je identický prístup použitý v článku [1].

V **Experimente 2** bolo monitorovaných 10 kúr, no kure s identifikačným číslom ID3 sa počas priebehu experimentu vyvliekol postroj so senzorom. Z toho dôvodu sú akcelerometrické dáta kury ID3 skreslené a z analýz v tejto podkapitole ju vyraďujeme.

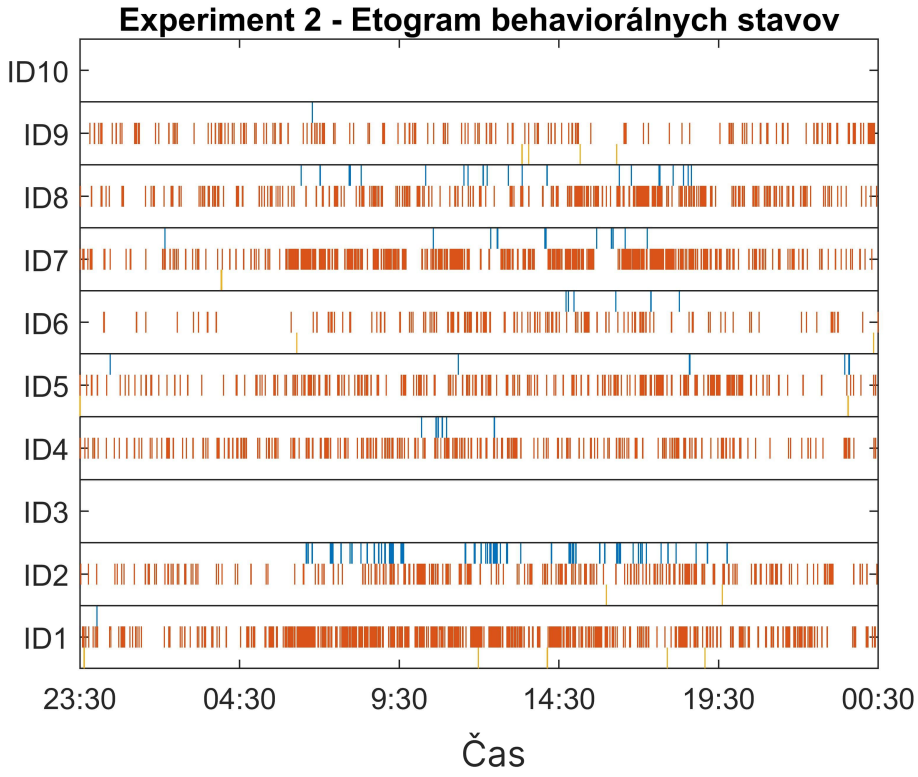
4.2.1 Naivný prístup

Naivný prístup je analogický k prístupu z podkapitoly 4.1. Použijeme knižnicu správaní vytvorenú na **Experimente Abdoli**, pomocou ktorej identifikujeme správania na neanotovaných dátach z **Experimentu 2**. Na Obr. 20 sme vykreslili etogram nájdených správaní kúr, použitím daného prístupu. Kuru ID10 vyraďujeme z dôvodu nájdenia dvoch odlišných správaní v rovnakom čase, čo je v rozpore s predpokladom exkluzivity správaní.

Hlavnou motiváciou vykonania daného prístupu je porovnanie výsledkov so spomenutým identickým prístupom z podkapitoly 4.1, ktorý bol aplikovaný na **Experiment 1**. Pri vizuálnom porovnaní etogramov na Obr. 20 a Obr. 17 vidíme markantne menej nájdeného správania typu ďobanie v **Experimente 2**. Pozitívnou správou je, že ak bolo správanie typu ďobanie nájdené, tak sa vyskytuje prevažne počas dňa. Nájdené správania typu čistenie v **Experimente 2** približne kopírujú početnosť nájdených správaní počas dňa aj noci v **Experimente 1**. Naopak správanie typu oprašovanie, sa na kurách z **Experimentu 2** takmer vôbec nenašlo, pričom v **Experimente 1** boli správania daného typu frekventované počas dňa.

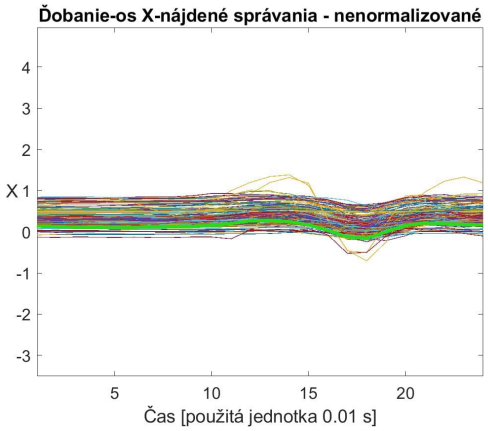
Nekonzistentné výsledky pri použití identickej knižnice Abdoli na dva odlišné expe-

rimenty, ktoré vykonala SAV, môžeme interpretovať mnohými spôsobmi. Početnosť jednotlivých správaní môže závisieť napríklad od prostredia alebo nálady sociálnej skupiny, na ktorej je experiment vykonaný.

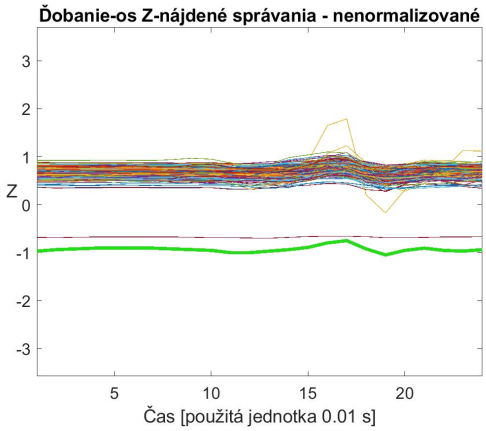


Obr. 20: Etogram jednotlivých správaní na kurách z **Experimentu 2**. Jednotlivé správania sú farebne rozlíšené, modrá predstavuje ďobanie, červená čistenie a žltá oprašovanie. Knižnica správaní, použitá na identifikovanie správaní na neanotovaných dátach jednotlivých kúr z **Experimentu 2**, bola vytvorená na trénovacej a testovacej vzorke z **Experimentu Abdoli**.

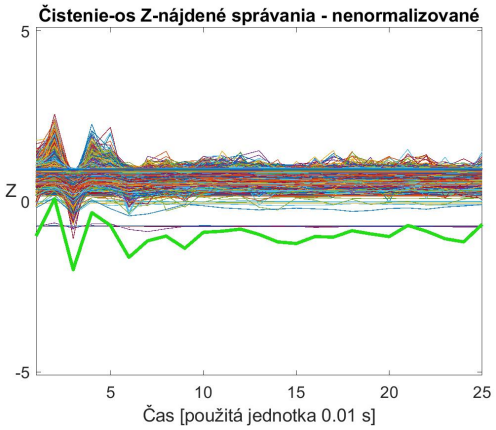
K vykreslenému etogramu prikladáme vizualizácie správaní na Obr. 21, ako pohyb v čase pre všetky kury z **Experimentu 2**. Hrubšie farebne zvýraznené podsekvencie predstavujú jednotlivé správania v daných osiach z knižnice Abdoli. Porovnať teraz môžeme vizualizácie z Obr. 21 s vizualizáciami z **Experimentu 1** na Obr. 18. Pri ich porovnaní vidíme, že sa správania v **Experimente 2** posunuli do kladnejších hodnôt na osi Z, čo v kontexte pohybu senzora značí posun smerom k zemi. To môže byť vysvetlené samotným nastavením **Experimentu 1**, v ktorom sú monitorované kury aj v klietkovom chove, pričom v **Experimente 2** iba v podstielkovom.



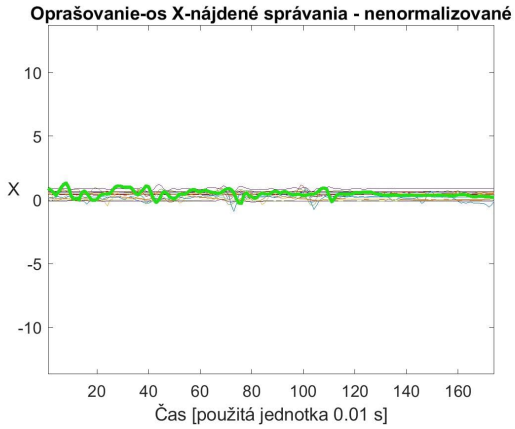
(a) Ďobanie v osi X.



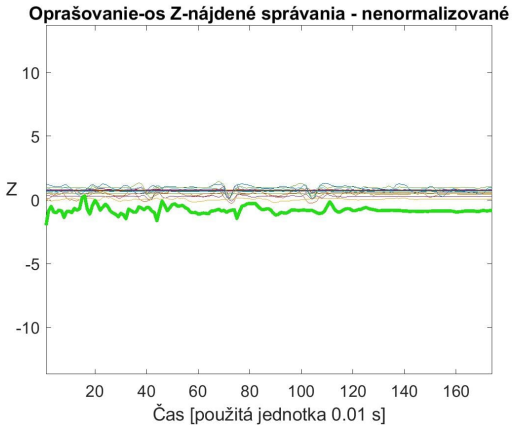
(b) Ďobanie v osi Z.



(c) Čistenie v osi Z.

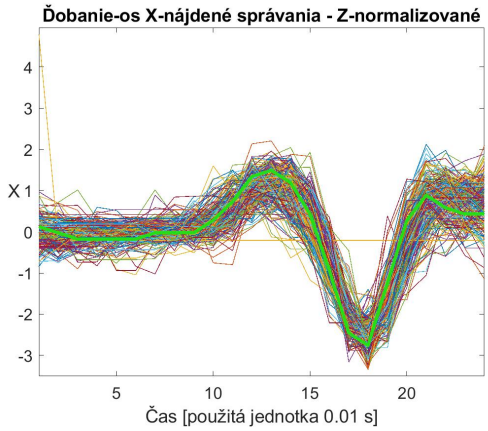


(d) Oprašovanie v osi X.

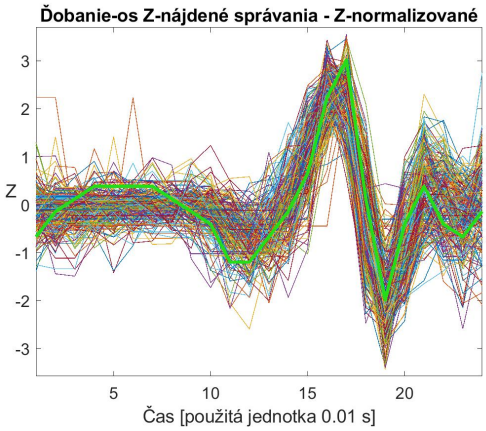


(e) Oprašovanie v osi Z.

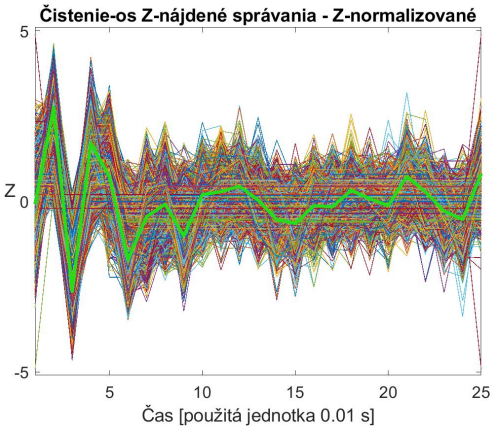
Obr. 21: Nájdené nenormalizované podsekvencie správaní všetkých kúr okrem ID3 a ID10 z **Experimentu 2**. Knižnica použitá na nájdenie daných správaní je z **Experimentu Abdoli**. Zelenou farbou zvýraznené podsekvencie predsavujú najlepších kandidátov jednotlivých správaní knižnice Abdoli.



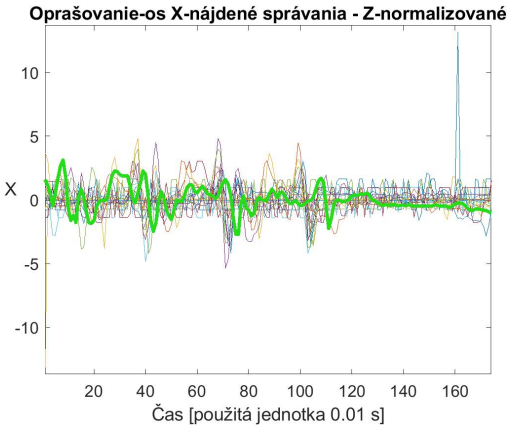
(a) Ďobanie v osi X.



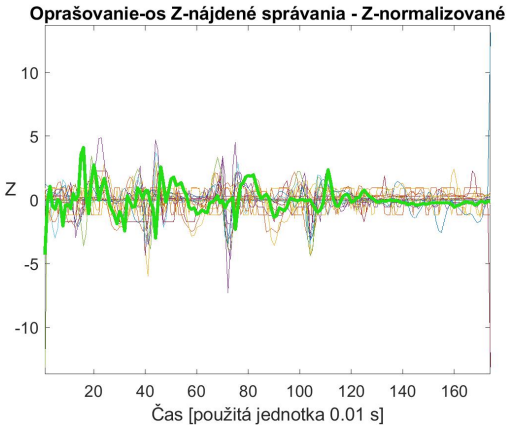
(b) Ďobanie v osi Z.



(c) Čistenie v osi Z.



(d) Oprašovanie v osi X.

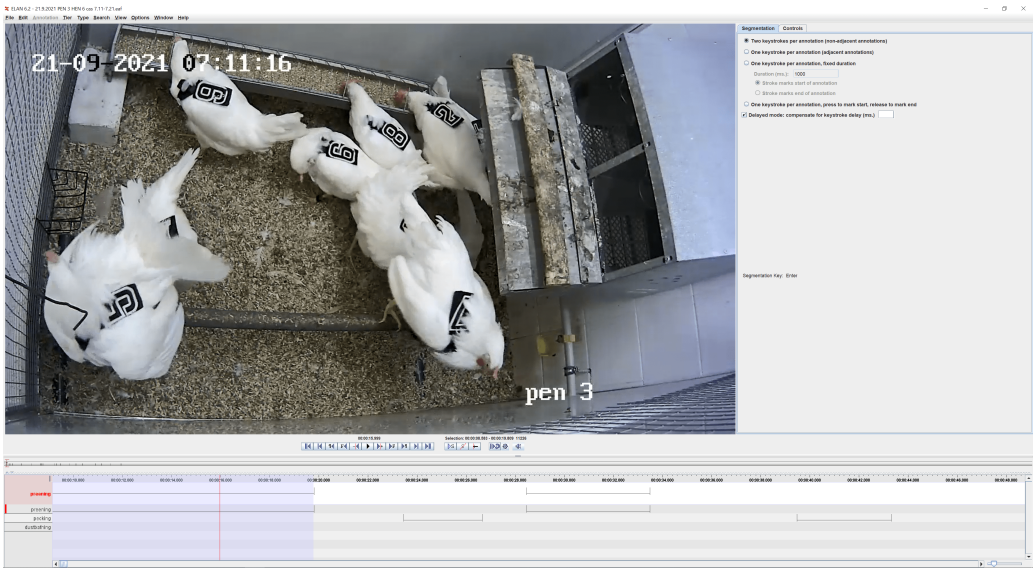


(e) Oprašovanie v osi Z.

Obr. 22: Nájdené Z-normalizované podsekvencie správaní všetkých kúr okrem ID3 a ID10 z **Experimentu 2**. Knižnica použitá na nájdenie daných správaní je z **Experimentu Abdoli**. Zelenou farbou zvýraznené podsekvencie predsavujú najlepších kandidátov jednotlivých správaní z knižnice Abdoli.

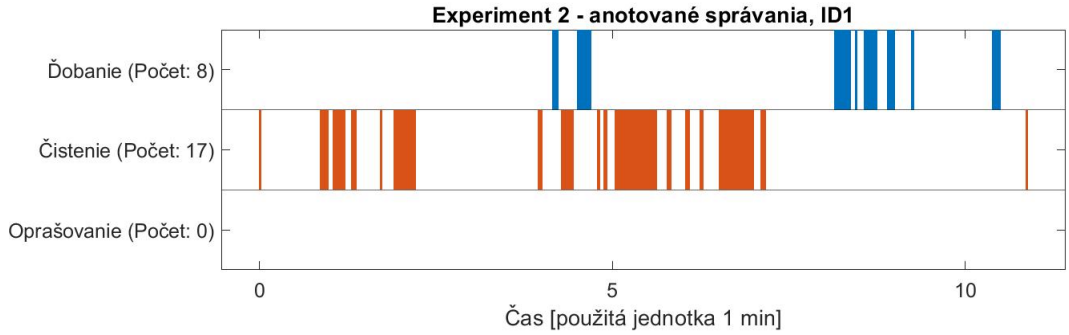
4.2.2 Prístup 1 - použitie hromadnej knižnice

V **Experimente 2**, rovnako ako v **Experimente 1** je naraz monitorovaná skupina 10 kúr, čo je rozdiel oproti prístupu v článku [1], kde je monitorovaná individuálne jedna kura. Návrh anotácie dát na natrénovanie klasifikátora je teda odlišný od anotácie dát na 30 minútach záznamu individuálnej kury z článku [1]. Namiesto toho sú anotácie navrhnuté tak, aby bol anotovaný približne 10 minutový úsek pre každú z desiatich kúr. Hlavný predpoklad je dostatok jednotlivých správaní v anotovanom úseku, s cieľom získať čo najkvalitnejších kandidátov identifikovaných na trénovacej vzorke. Anotácie vyhotovila SAV pomocou videozáznamu z **Experimentu 2** využitím softvéru ELAN [4], ktorý slúži na daný typ anotácii. Na Obr. 23 je znázornený proces vyhotovenia anotácii jednotlivých správaní sliepok SAV pomocou softvéru ELAN na **Experimente 2**.

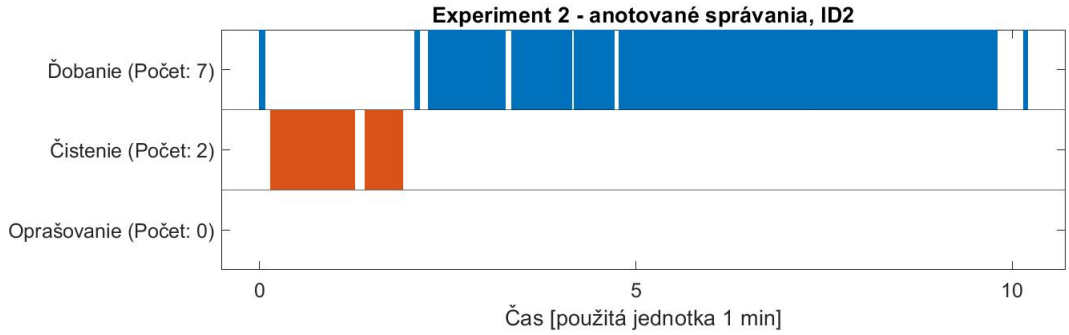


Obr. 23: Anotovanie správaní kúr z videozáznamu **Experimentu 2**, ktoré vykonala Slovenská akadémia vied pomocou softvéru ELAN.

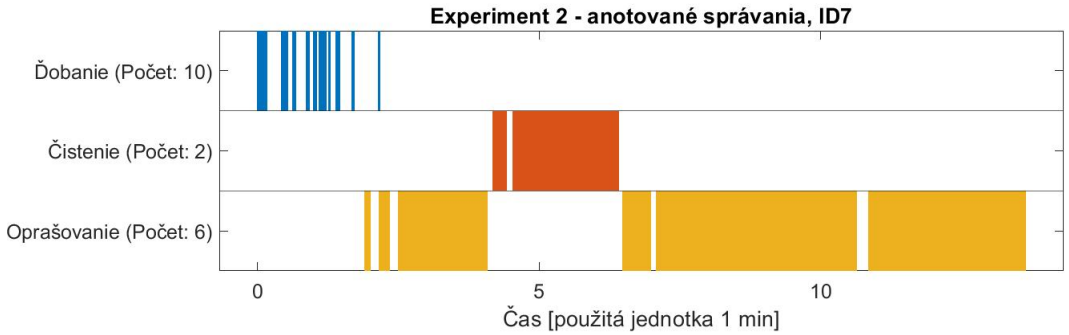
Na ukážku vytvorených anotácii prikladáme na Obr. 24 etogramy anotovaných úsekov pre niekoľko kúr z **Experimentu 2**. Nie každý anotovaný úsek pre jednotlivú kuru sa dal zvoliť tak, aby obsahoval príklady z každého správania. Z toho dôvodu je napríklad prítomná absencia behaviorálneho stavu oprašovanie na Obr. 24a, resp. Obr. 24b.



(a) Anotácie pre kuru ID1.



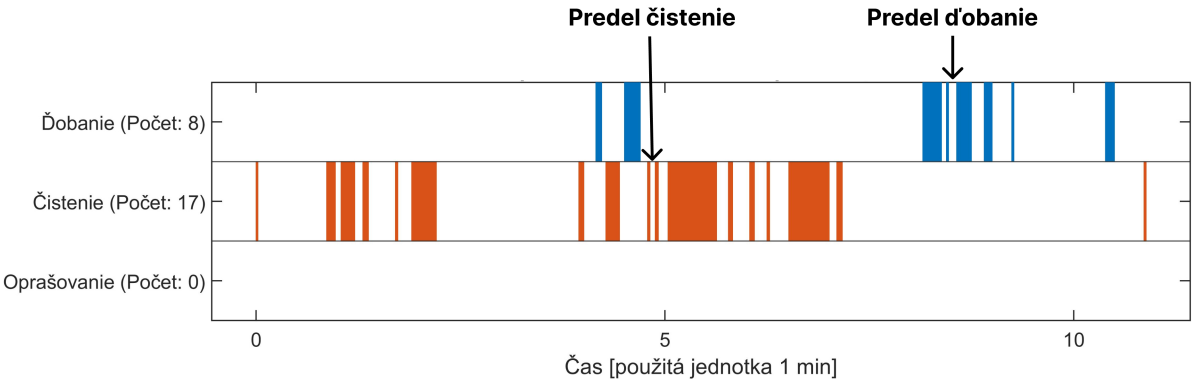
(b) Anotácie pre kuru ID2.



(c) Anotácie pre kuru ID7.

Obr. 24: Anotované správania pre kury ID1, ID2 a ID7 z **Experimentu 2**.

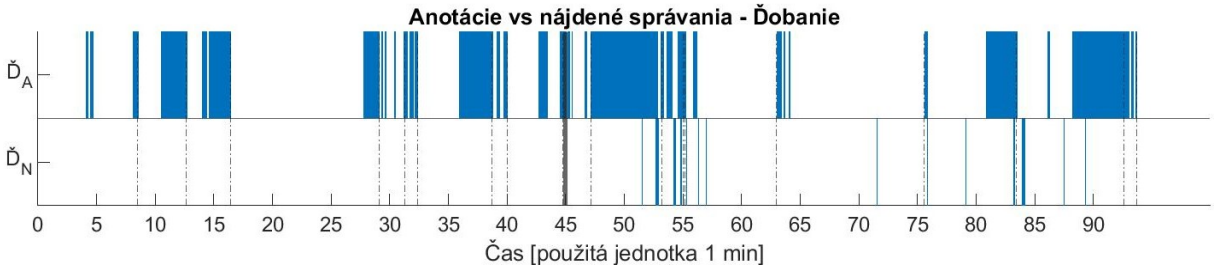
Pripomíname, že anotácie sú vytvorené pre 9 z 10 kúr, kvôli zistenej komplikácii zberu dát kury s identifikačným číslom ID3. Pri pohľade na anotované úseky je viditeľné, že intervaly anotovaných správaní nie sú rovnomerne rozmiestnené v okolí stredu vrámci daného časového úseku. Z toho dôvodu je vhodné rozdeliť datasety na trénovaciu a testovaciu vzorku pre každú kuru a každé správanie tak, aby bol v trénovacej a testovacej vzorke rovnaký počet správaní. To je odlišnosť oproti prístupu v **Experimente Abdoli**, kde je predel na trénovaciu a testovaciu vzorku v polovici anotovaného datasetu. Na Obr. 25 je ukážka predelov v anotáciach jednotlivých správaní pre kuru ID1 z **Experimentu 2**.



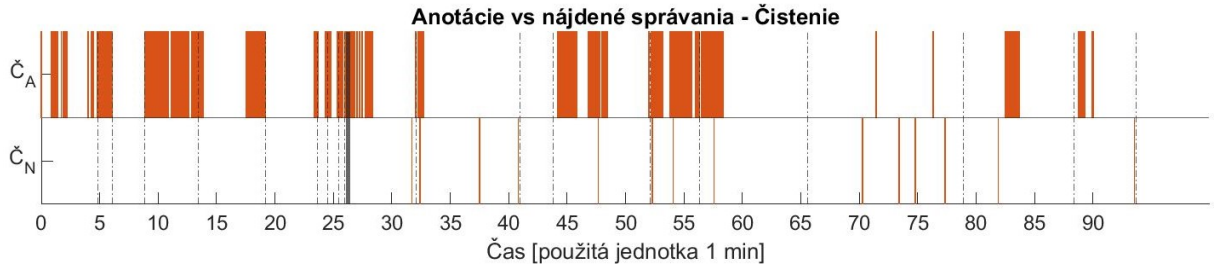
Obr. 25: Rozdelenie datasetu na trénovaciu a testovaciu vzorku spôsobom, aby počet správaní v trénovacej vzorke zodpovedal počtu správaní v testovacej vzorke. Pri nepárnom počte správaní vrámci anotovaného časového radu je vždy v trénovacej vzorke o jedno správanie viac ako v testovacej. Trénovacia vzorka je vždy naľavo od nájdeného predelu, resp. od začiatku po daný predel vrátane. Testovacia vzorka je napravo od nájdeného predelu, resp. od predelu po koniec časového úseku. Ak z jednotlivého správania nebol anotovaný ani jeden príklad, rozdelíme dataset presne na polovicu.

Takýmto spôsobom rozdelíme anotovaný dataset na trénovaciu a testovaciu vzorku pre každú kuru a pre jej každé jedno správanie zvlášť. Následne vytvoríme akúsi hromadnú trénovaciu a testovaciu vzorku pre každé správanie. Postup ukážeme na behaviorálnom stave typu ďobanie.

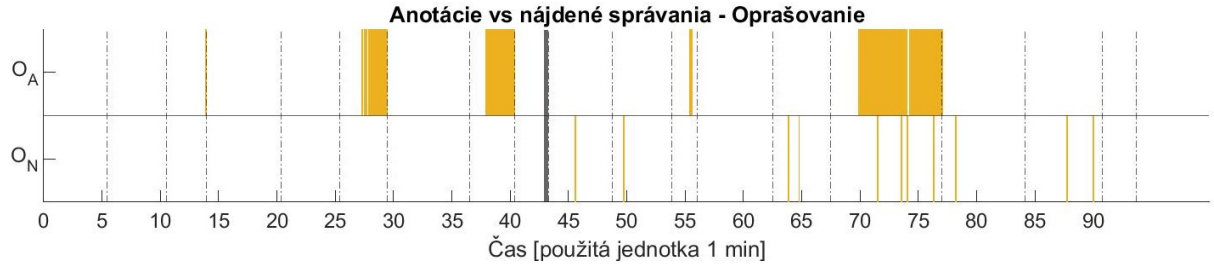
V prvom kroku nájdeme pre každú kuru tzv. predel ďobanie, ako je vizualizované na Obr. 25. Ďalej spojíme, resp. poukladáme za sebou všetky trénovacie vzorky jednotlivých kúr pre behaviorálny stav typu ďobanie. Vytvoríme teda akúsi hromadnú trénovaciu vzorku. Rovnako postupujeme pri testovacích vzorkách daného správania jednotlivých kúr, z ktorých vytvoríme hromadnú testovaciu vzorku. Nakoniec tieto dva hromadné datasety spojíme a vytvoríme hromadnú trénovaciu a testovaciu vzorku pre správanie typu ďobanie. Na danom anotovanom datasete nájdeme najlepšieho kandidáta správania typu ďobanie, čím chceme získať akýsi všeobecný príklad ďobania, ktorý vykonáva každá kura identicky. Analogicky sme postupovali pri behaviorálnom stave čistenie a oprašovanie. Na Obr. 26 vizualizujeme porovnanie anotovaných a nájdených správaní pre jednotlivé typy správania pomocou hromadnej knižnice správaní.



(a) Porovnanie anotovaných a nájdených správaní pre behaviorálny stav typu ďobanie.



(b) Porovnanie anotovaných a nájdených správaní pre behaviorálny stav typu čistenie.



(c) Porovnanie anotovaných a nájdených správaní pre behaviorálny stav typu oprašovanie.

Obr. 26: Vizuálne porovnanie anotovaných a nájdených správaní podobných najlepšiemu kandidátovi jednotlivého typu správania, ktorý bol nájdený na hromadnej trénovacej vzorke z **Experimentu 2**. Na Obr. 26a sú v riadku $\check{\mathrm{D}}_{\mathrm{A}}$ vykreslené anotované intervaly pre ďobanie a v riadku $\check{\mathrm{D}}_{\mathrm{N}}$ nájdené správania typu ďobanie. Analogicky sú riadky označené v Obr. 26b, kde $\check{\mathrm{C}}_{\mathrm{A}}$ sú anotované intervaly pre čistenie a $\check{\mathrm{C}}_{\mathrm{N}}$ nájdené správania typu čistenie. Na poslednom Obr. 26c sú v riadku O_{A} anotované intervaly pre oprašovanie a v riadku O_{N} nájdené správania typu oprašovanie. Prerušované vertikálne čiary naprieč danými obrázkami, predstavujú predely medzi jednotlivými kurami, pričom hrubšie znázornená vertikálna čiara predstavuje rozdelenie na trénovaciu (vľavo) a testovaciu vzorku (vpravo). V trénovacej vzorke sa teda nachádzajú postupne trénovacie vzorky jednotlivých kúr, t.j. $\mathrm{ID1}_{\mathrm{Train}}$, $\mathrm{ID2}_{\mathrm{Train}}$, $\mathrm{ID4}_{\mathrm{Train}}$, ..., $\mathrm{ID10}_{\mathrm{Train}}$ a v testovacej vzorke $\mathrm{ID1}_{\mathrm{Test}}$, $\mathrm{ID2}_{\mathrm{Test}}$, $\mathrm{ID4}_{\mathrm{Test}}$, ..., $\mathrm{ID10}_{\mathrm{Test}}$, pre jednotlivé správania.

Na ohodnotenie výsledkov použijeme jemne odlišný prístup, ako bol spomenutý v kapitole 3, resp. v článku [1]. Každý anotovaný interval správania však stále považujeme

za akýsi zhluk, v ktorom sa nachádza niekoľko príkladov daného správania. Keďže sú správania exkluzívne, vykonáme validáciu pre každé správanie osobitne tak, aby zhluky konkrétneho správania neboli doplnené o zhluky odlišných typov správaní. Napríklad, pri validácii správania typu ďobanie pracujeme s zhlukmi typu ďobanie a zhlukmi, kde kura nevykonávala žiadne správanie, resp. neďobala. Tak je to možné vidieť na Obr. 26a v riadku anotovaných intervalov $\check{\mathrm{D}}_{\mathrm{A}}$. Narozdiel od toho, pri validácii správania typu ďobanie v článku [1], pracovali s zhlukmi typu ďobania, čistenia, oprašovania doplnenými o zhluky, kde kura nevykazovala žiadne správanie.

Ak sa aspoň jedno klasifikované správanie vyskytuje vrámci daného zhluku identického správania, tak celý zhluk považujeme za 'true positive' (TP). Naopak, ak nie je ani jedno klasifikované správanie vrámci daného zhluku, tak celý zhluk považujeme za 'false negative' (FN). Ak je klasifikované správanie v zhluku, ktorý nezodpovedá danému správaniu, tak celý zhluk označíme ako 'false positive' (FP). Nakoniec, ak nie je nájdené ani jedno správanie vnútri neanotovaného zhluku, tak celý zhluk označíme za 'true negative' (TN).

V tabuľkách 3, 4, 5 sme zobrazili výsledky testovania najlepších kandidátov pre jednotlivé správanie. Nakoniec v tabuľke 6 uvádzame miery kvality klasifikátora, na základe hodnôt z tabuliek 3, 4, 5. Miery kvality klasifikátora pre jednotlivé správanie počítame nasledovne:

$$\text{Precision}_{\text{správanie}} = \frac{TP}{TP + FP},$$

$$\text{Recall}_{\text{správanie}} = \frac{TP}{TP + FN},$$

$$\text{Accuracy}_{\text{správanie}} = \frac{TP + TN}{TP + FP + TN + FN}.$$

		Anotované	
		ďobanie	neďobanie
Nájdené	ďobanie	$TP = 4$	$FP = 9$
	neďobanie	$FN = 26$	$TN = 20$

Tabuľka 3: Confusion matrix pre behaviorálny stav typu ďobanie, získaný z testovacej vzorky hromadného datasetu z **Experimentu 2**.

		Anotované	
		čistenie	nečistenie
Nájdené	čistenie	$TP = 5$	$FP = 5$
	nečistenie	$FN = 21$	$TN = 20$

Tabuľka 4: Confusion matrix pre behaviorálny stav typu čistenie, získaný z testovacej vzorky hromadného datasetu z **Experimentu 2**.

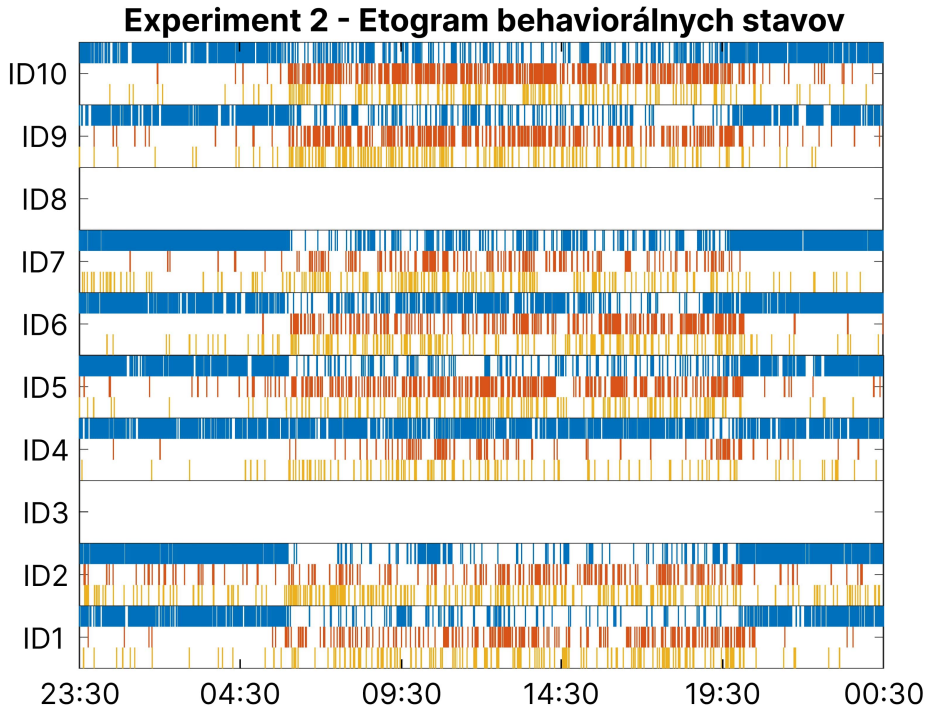
		Anotované	
		oprašovanie	neoprašovanie
Nájdené	oprašovanie	$TP = 2$	$FP = 2$
	neoprašovanie	$FN = 1$	$TN = 1$

Tabuľka 5: Confusion matrix pre behaviorálny stav typu oprašovanie, získaný z testovacej vzorky hromadného datasetu z **Experimentu 2**.

	Ďobanie	Čistenie	Oprašovanie
Precision	0,31	0,5	0,5
Recall	0,13	0,19	0,5
Accuracy	0,41	0,49	0,43

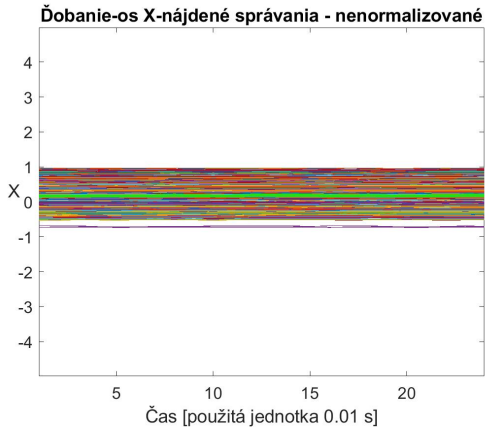
Tabuľka 6: Hodnoty jednotlivých ukazovateľov kvality klasifikačného modelu na testovacej vzorke hromadného datasetu z **Experimentu 2**.

Jednotlivé hodnoty z tabuľky 6 nám ukazujú limitácie kvality klasifikátora, resp. najlepšieho kandidáta daného správania, natrénovaného na hromadnej trénovacej vzorke a otestovaného na hromadnej testovacej vzorke. Klasifikátor nemá ani pre jedno správanie precision hodnotu väčšiu ako $0,5$. Najlepšie výsledky vykazuje klasifikátor pre správanie typu čistenie a oprašovanie. Správanie typu oprašovanie ale musíme brať s rezervou, keďže sme disponovali malým počtom zhlukov typu oprašovania v trénovacej a testovacej vzorke. Najhoršiu kvalitu vykazuje klasifikátor pri správaní typu ďobanie, čo bude vidno aj na priloženom etograme, nájdených správaní, Obr. 27, ktoré bolo frekventované počas noci.

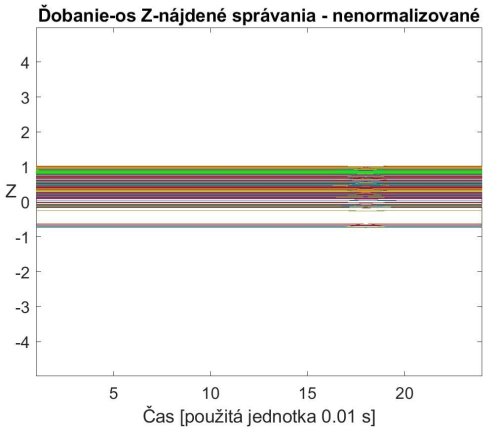


Obr. 27: Etogram nájdených správaní na kurách z **Experimentu 2**. Jednotlivé správania sú farebne rozlíšené, pričom modrá predstavuje ďobanie, červená čistenie a žltá oprašovanie. Knižnica správaní bola vytvorená na hromadnej trénovacej a testovacej vzorke z **Experimentu 2**.

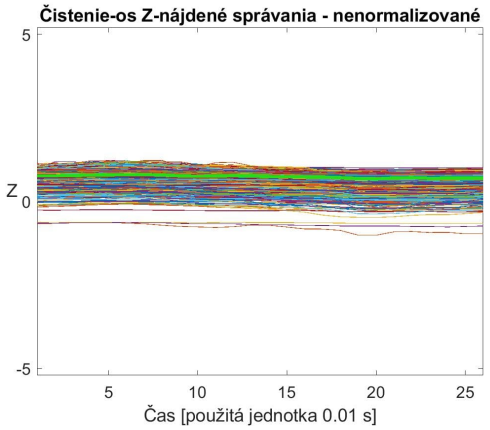
Vizuálna kontrola etogramu pri správaní typu ďobanie vykazuje limitácie klasifikátora nami vytvorenej hromadnej knižnice správaní. Pri pohľade na behaviorálny stav typu ďobanie dostávame nájdené správania na neanotovanom datasete aj počas noci, pričom sú dokonca početnejšie ako cez deň. Správania typu čistenie a oprašovanie sú prítomné hlavne počas dňa, pričom sa zriedkavo vyskytujú aj v noci, čo je bežný jav. Rovnako ako v predchádzajúcich prístupoch vizualizujeme na Obr. 28, resp. Obr. 29, nájdené nenormalizované podsekvencie, resp., nájdené Z-normalizované podsekvencie daných správaní.



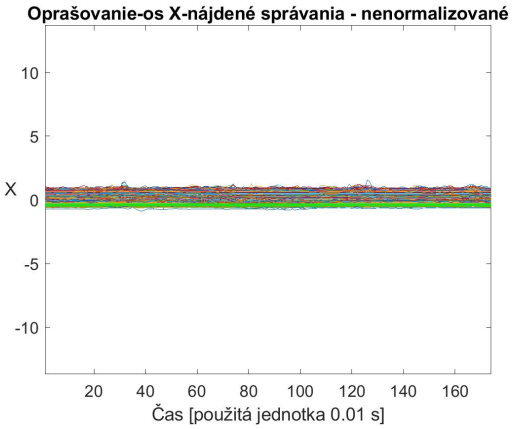
(a) Ďobanie v osi X



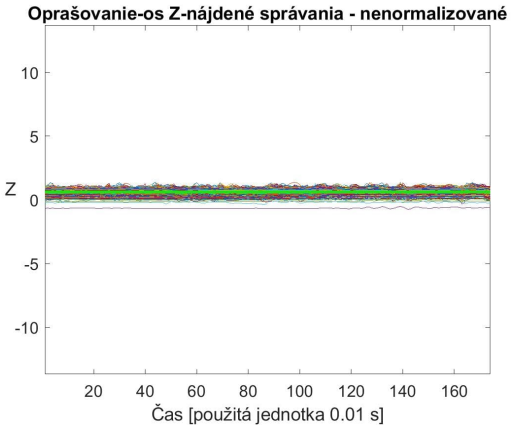
(b) Ďobanie v osi Z



(c) Čistenie v osi Z

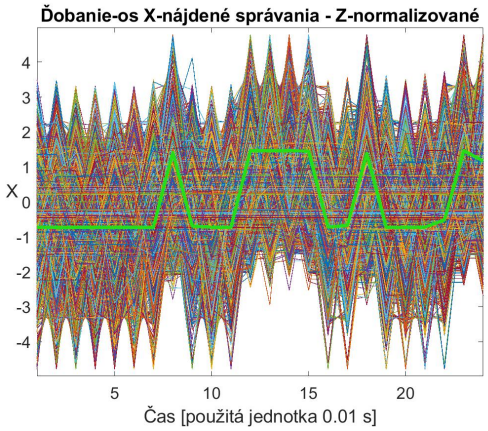


(d) Oprašovanie v osi X

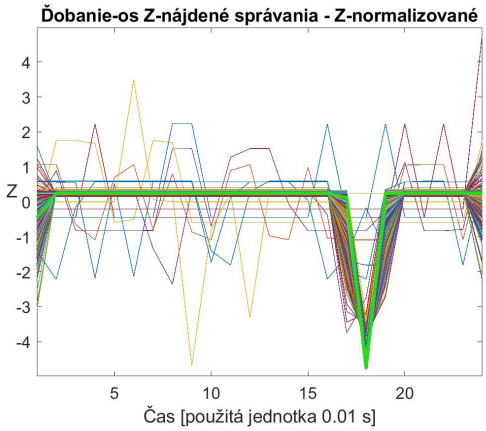


(e) Oprašovanie v osi Z

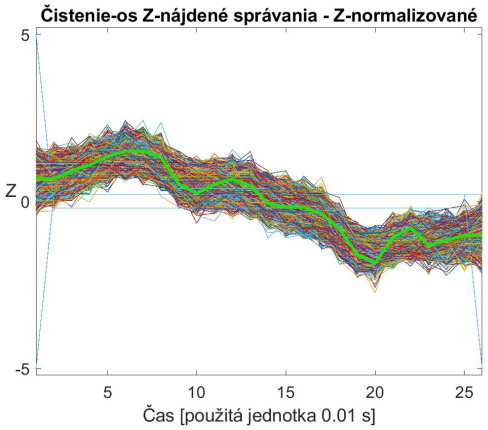
Obr. 28: Nájdené nenormalizované podsekvencie správaní všetkých kúr okrem ID3 a ID8 z **Experimentu 2**. Knižnica použitá na nájdenie daných správaní je vytvorená na hromadnej trénovacej vzorke z **Experimentu 2**. Zelenou zvýraznené podsekvencie predsavujú najlepších kandidátov jednotlivých správaní z natrénovanej hromadnej knižnice.



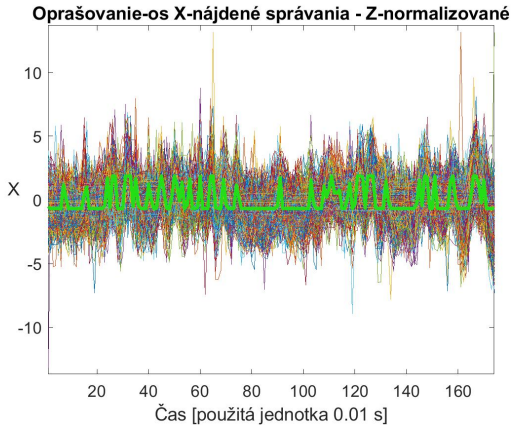
(a) Ďobanie v osi X



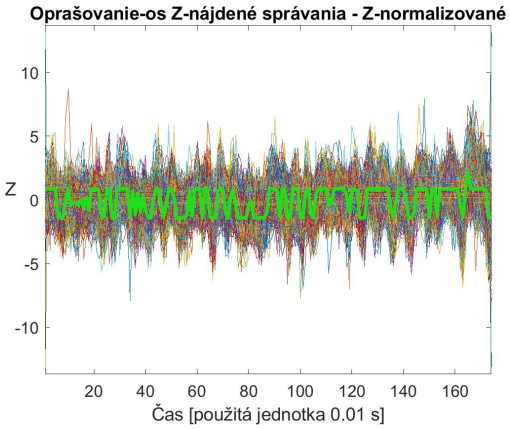
(b) Ďobanie v osi Z



(c) Čistenie v osi Z



(d) Oprašovanie v osi X



(e) Oprašovanie v osi Z

Obr. 29: Nájdené Z-normalizované podsekvencie správaní všetkých kúr okrem ID3 a ID8 z **Experimentu 2**. Knižnica použitá na nájdenie daných správaní je vytvorená na hromadnej trénovacej vzorke z **Experimentu 2**. Zelenou zvýraznené podsekvencie predsavujú najlepších kandidátov jednotlivých správaní z natrénovanej hromadnej knižnice.

Nájdené Z-normalizované správania na Obr. 29 je teraz možné porovnať s nájdenými správaniami pri použití Abdoliho knižnice z naivného prístupu, ktoré sú vizualizované na Obr. 22. Pripomíname, že oba prístupy sú aplikované na dáta z **Experimentu 2**. Vidíme, že zvýraznené podsekvencie, ktoré reprezentujú najlepších kandidátov jednotlivých správaní sa od seba výrazne líšia. To najmä pri správaní typu ďobanie, kde nájdený (zvýraznený) kandidát z hromadnej knižnice obsahuje úplne odlišný pohyb aj v osi X aj v osi Z. Špeciálne ďobanie v osi Z na Obr. 29b vykazuje zvláštny pohyb. Správania typu čistenie sa na seba tiež nepodobajú, avšak kandidát z hromadnej knižnice obsahuje dobrý konzistentný vzor naprieč všetkými nájdenými správaniami. Kandidáti správania typu oprašovanie sa podobajú v jednotlivých prístupoch najviac, pričom kandidát z hromadnej knižnice obsahuje kolísavejší pohyb ako kandidát z knižnice Abdoli.

Daný prístup, v ktorom sme vytvorili hromadnú knižnicu na základe anotovaných dát celej sociálnej skupiny 9 kúr, nevykazuje až tak kvalitné výsledky. Spôsobené je to zrejme veľmi veľkou rozmanitosťou správaní medzi jednotlivými kurami. Je možné, že identický typ správania sa značne odlišuje konkrétnym pohybom kury, alebo časovou dĺžkou vykonania daného správania. Tým pádom je priam nemožné nájsť takého reprezentanta jednotlivého správania, ktorý by generalizoval dané správanie naprieč celou sociálnou skupinou.

4.2.3 Prístup 2 - použitie separátnej knižnice

Ďalší a posledný prístup je separátne trénovanie a testovanie na anotovanom datasete každej kury zvlášť. Jedná sa teda o identický prístup k článku [1]. Jediný rozdiel je kratší úsek anotovaných datasetov pre jednotlivé kury, čo môže ovplyvniť kvalitu nájdeného najlepšieho kandidáta jednotlivého správania. Pracujeme teda s rovnakými anotáciami, ktoré boli vizualizované v predošlom prístupe na Obr. 24. Pripomíname, že anotované úseky pre jednotlivé kury väčšinou neobsahovali všetky typy správania. Preto klasifikujeme oprašovanie iba na kurách s identifikačným číslom ID4 a ID7. Niektoré anotácie pre kury obsahovali iba dva anotované intervaly konkrétneho správania, čo je nedostatočné na natrénovanie a otestovanie. Z toho dôvodu je potreba niektoré výsledky brať s rezervou.

Pre každú kuru a jednotlivé správanie sme vytvorili špecifický predel anotovaného datasetu na trénovaciu a testovaciu vzorku spôsobom, aby bol v trénovacej a testovacej

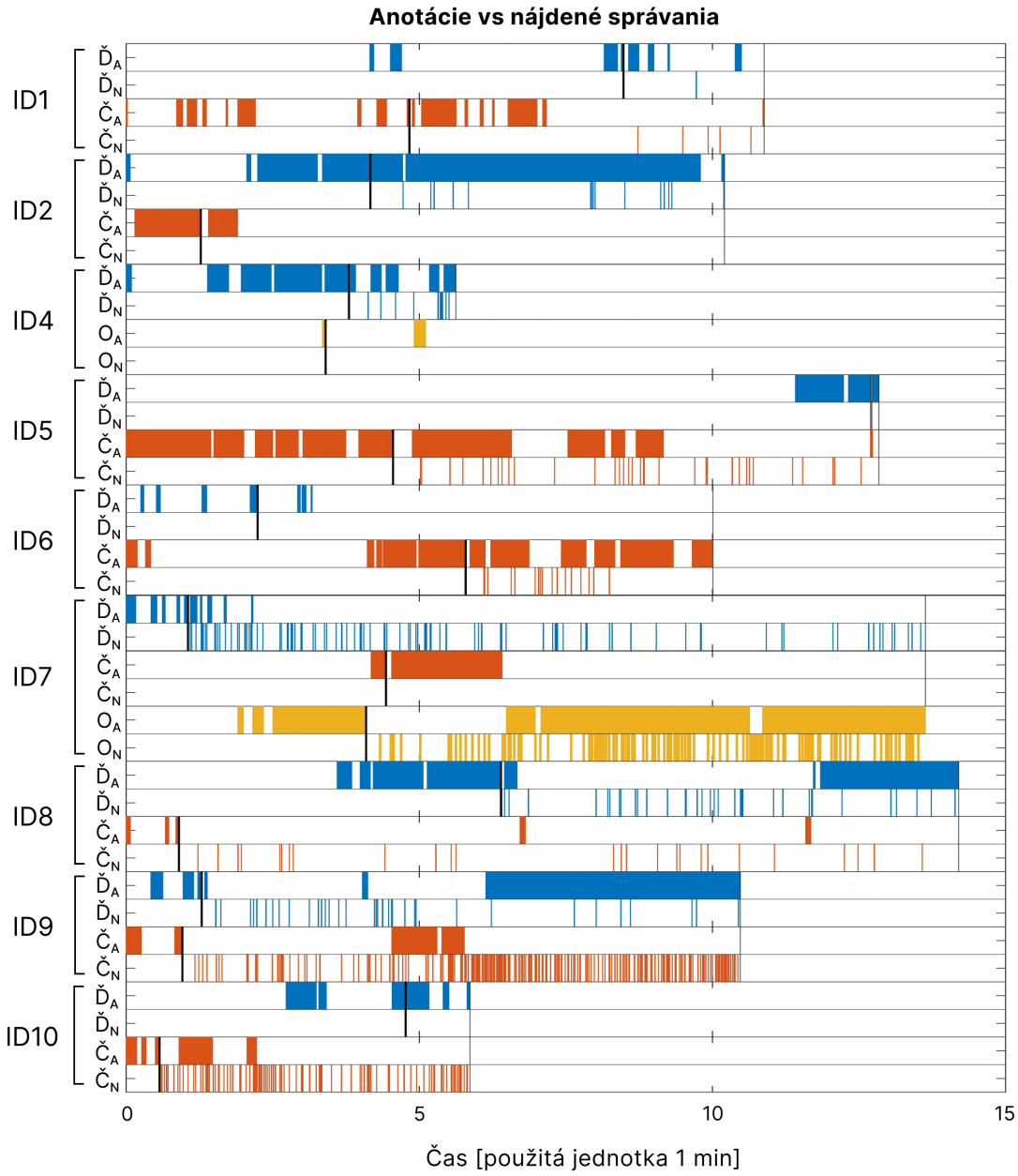
vzorke rovnaký počet anotovaných intervalov daného správania. Na takto rozdelených anotovaných datasetoch sme pre každú kuru a jednotlivé správania vytvorili jedinečnú knižnicu správaní, ktorú sme následne použili na identifikáciu správaní v neanotovanom datasete pre danú kuru a dané správanie. Na Obr. 30 vizualizujeme porovnanie anotovaných a nájdených správaní, pričom pri každej kure a každom správaní bola použitá jedinečná knižnica správaní pre danú kombináciu kury a správania.

Mieru kvality klasifikátora pre jednotlivo natrénované knižnice správaní otestujeme identickým spôsobom, ako v predošlej podkapitole 4.2.2. Na Obr. 31a, resp. Obr. 31b vizualizujeme confusion matice pre jednotlivé kury a správania typu ďobanie a čistenie, resp. oprašovanie. Zložky daných matíc obsahujú počty 'true positive', 'false positive', 'false negative' a 'true negative tak, ako je to znázornené v tabuľke 7. Kvôli uľahčeniu porovnania daných matíc je každá matica preškálovaná svojou najvyššou hodnotou. Tým pádom sú zložky v maticiach z intervalu $[0, 1]$. Pre správanie typu oprašovanie máme iba 2 také matice, pretože iba kury ID4 a ID7 disponovali anotáciami pre dané správanie.

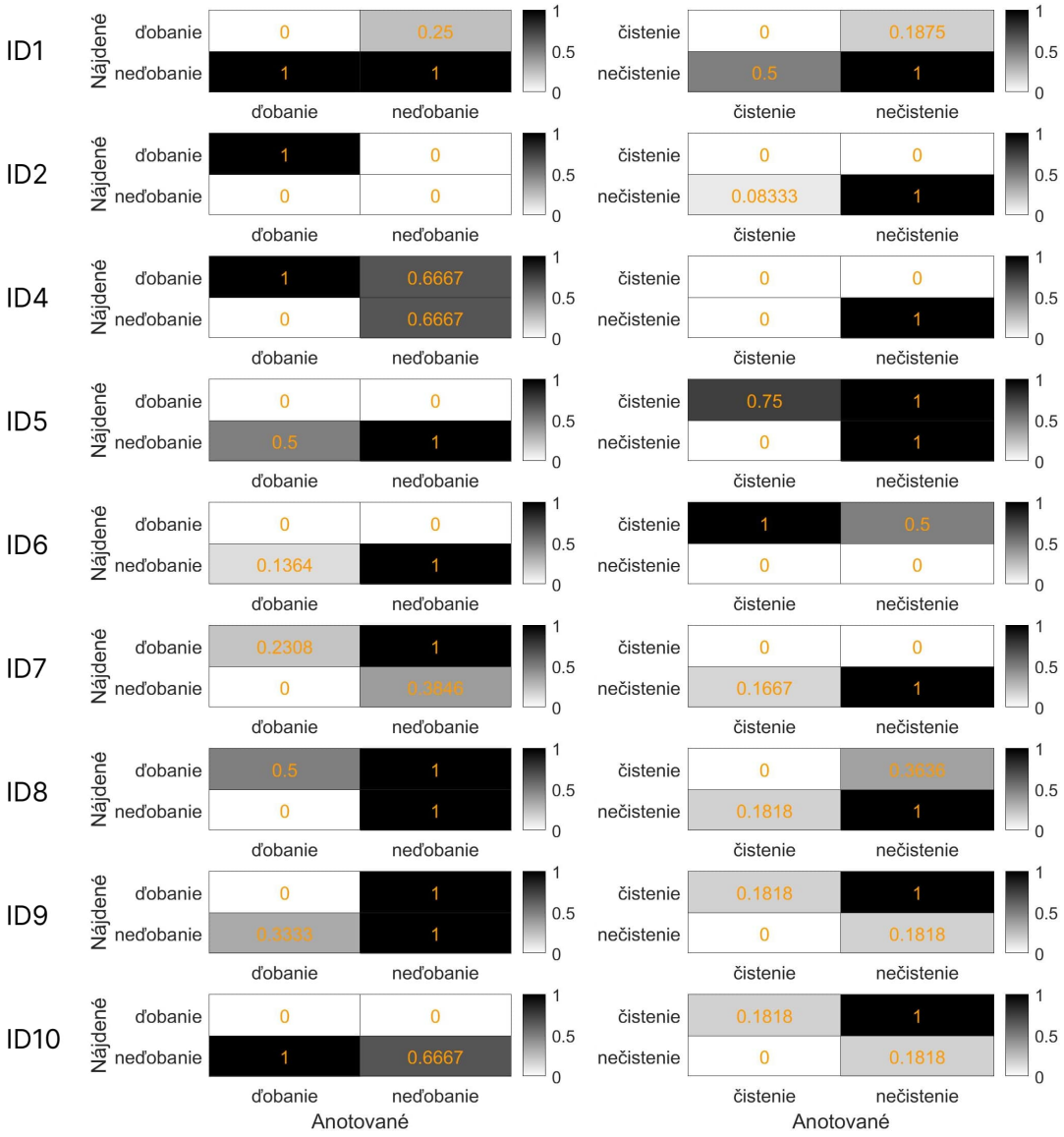
Doplnkovo uvádzame aj hromadné confusion matice správania typu ďobanie a oprašovanie, kde sledujeme hodnoty confusion matíc pre dané správanie naprieč všetkými kurami. Následne z dôvodu ľahšieho porovnania preškálujeme hromadné confusion matice ich najvyššou hodnotou. Hromadná confusion matica pre správanie typu ďobanie je vyobrazená na Obr. 32 a pre čistenie na Obr. 33. Najviac nás však zaujímajú hodnoty precision pre jednotlivé správania a kury, ostatné miery kvality teraz vynechávame. Precision hodnoty vyčíslené pre confusion matice uvedené na Obr. 31 sú pre jednotlivé správania a kury vyzobrazené na Obr. 34. Precision hodnoty typu NA sú dôsledkom nula klasifikovaných 'true positive' a 'false positive' ánotovaných intervalov. Jediné, ako tak zmysluplné výsledky pre správanie typu ďobanie vykazovali kury ID2 a ID4, pre správanie typu čistenie kury ID5 a ID6 a pre správanie typu oprašovanie kura ID7.

Nájdené	správanie	počet 'true positives'	počet 'false positives'
	ne-správanie	počet 'false negatives'	počet 'true negatives'
		správanie	ne-správanie
		Anotované	

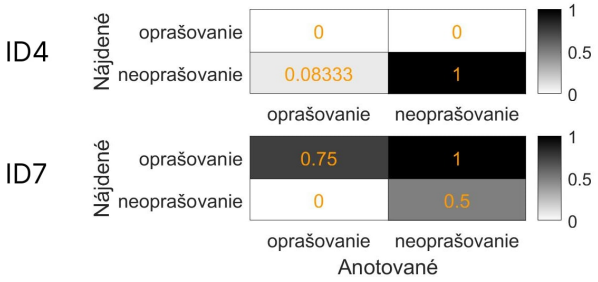
Tabuľka 7: Vzor Confusion matice pre všeobecný typ správania.



Obr. 30: Porovnanie anotovaných a nájdených správaní na anotovaných datasetoch jednotlivých kúr z **Experimentu 2**, kde trénovanie a testovanie prebiehalo separátne pre každú jednu kuru zvlášť. V riadkoch $\check{D}_A$ sú vykreslené anotované intervaly pre ďobanie a v riadku $\check{D}_N$ nájdené správania typu ďobanie. Identicky sú použité označenia pre správanie typu čistenie, resp. oprašovanie, t.j. $\check{C}_A$ a $\check{C}_N$, resp. t.j. O_A a O_N. Jednotlivé správania sú zároveň aj farebne odlíšiteľné, modrá predstavuje ďobanie, červená čistenie a žltá oprašovanie. Čiernou farbou zvýraznené vertikálne čiary predstavujú predel na trénovaciu a testovaciu vzorku pre jednotlivé správanie a kuru.

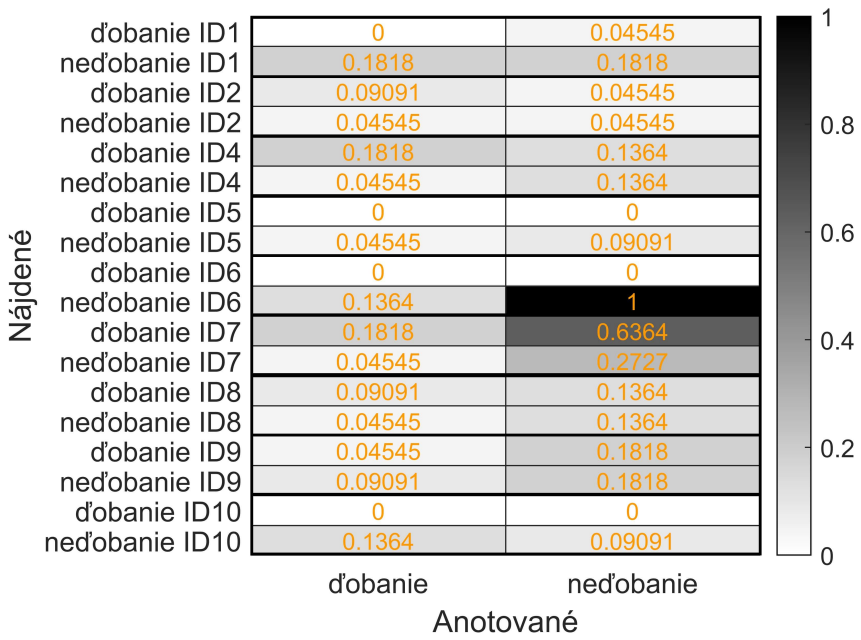


(a) Confusion matice pre správania typu ďobanie a čistenie

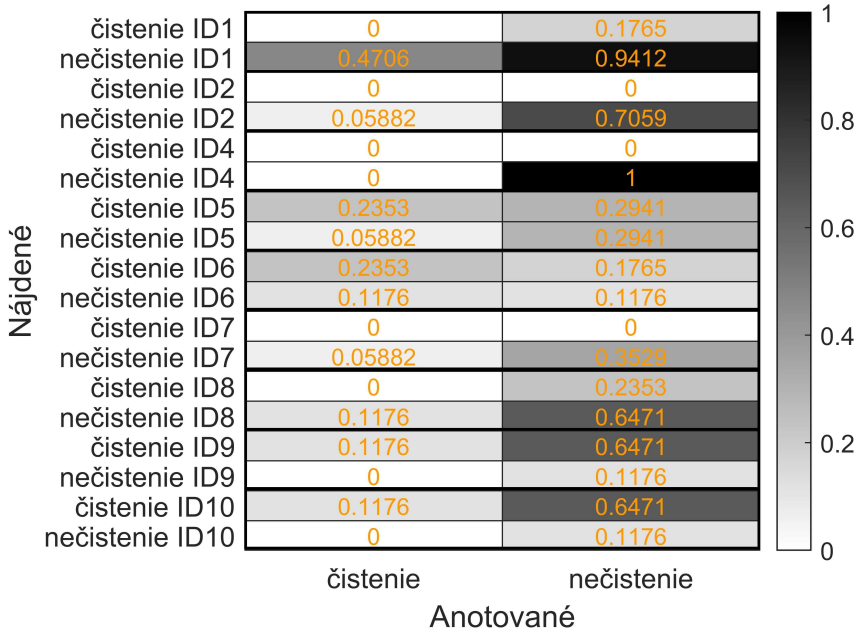


(b) Confusion matice pre správanie typu oprašovanie

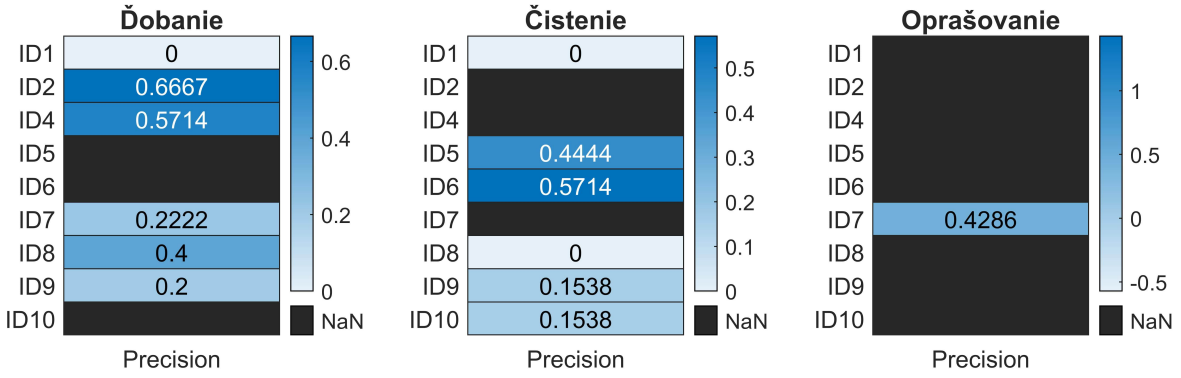
Obr. 31: Confusion matice pre jednotlivé správania kúr z **Experimentu 2**, pričom pre každú kuru a každé správanie bola vytvorená jedinečná knižnica správaní. Každá matica je preškálovaná najvyššou hodnotou jednotlivej matice.



Obr. 32: Hromadná confusion matica pre správanie typu ďobanie pre kury z **Experimentu 2**, pričom pre každú kuru a každé správanie bola vytvorená jedinečná knižnica správaní. Hromadná matica je preškálovaná jej najvyššou hodnotou.



Obr. 33: Hromadná confusion matica pre správanie typu čistenie pre kury z **Experimentu 2**, pričom pre každú kuru a každé správanie bola vytvorená jedinečná knižnica správaní. Hromadná matica je preškálovaná jej najvyššou hodnotou.

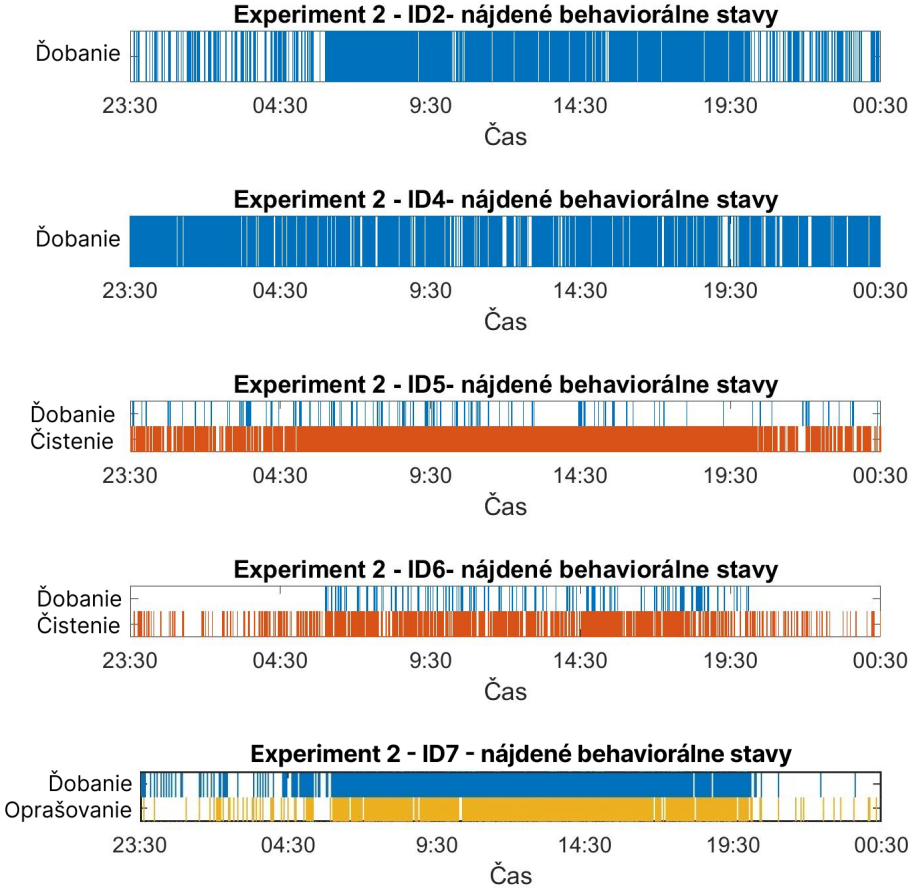


(a) Precision pre správanie typu ďobanie

(b) Precision pre správanie typu čistenie

(c) Precision pre správanie typu oprašovanie

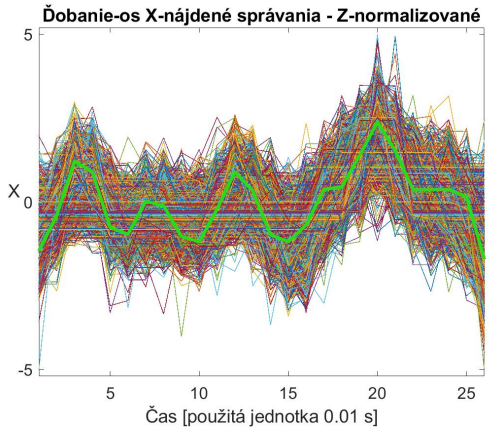
Obr. 34: Precision hodnoty pre jednotlivé správania a jednotlivé kury.



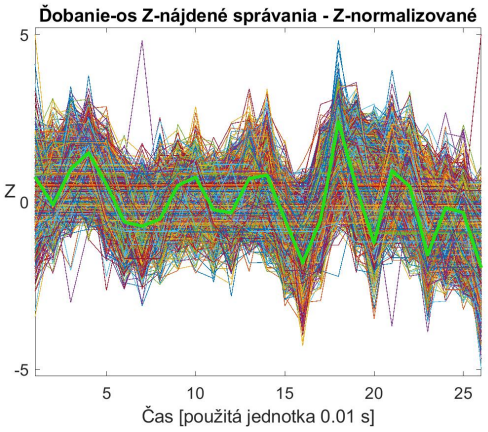
Obr. 35: Etogramy nájdených správaní pre jednotlivé správania a kury, ktoré vykazovali najlepšiu precision hodnotu.

Na Obr. 35 sme vykreslili etogramy nájdených správaní pre jednotlivé správania a kury,

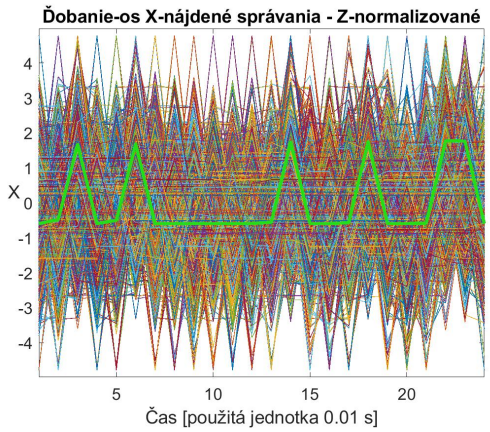
ktorých precision z tabuľky 34 vykazovala spomedzi kúr a správaní najlepšie hodnoty. Na Obr. 36 a Obr. 37 ďalej vizualizujeme nájdené správania na neanotovaných dátach z celého dňa taktiež pre kury, ktorých precision hodnota vykazovala najlepšie hodnoty. Uvádzame iba Z-normalizované podsekvencie, pričom nenormalizované z dúvodu zvýšenia prehľadnosti nateraz vynechávame. Pri ďobaní teda sledujeme kury ID2 a ID4, pri čistení kury ID5 a ID6 a pri oprašovaní kuru ID7.



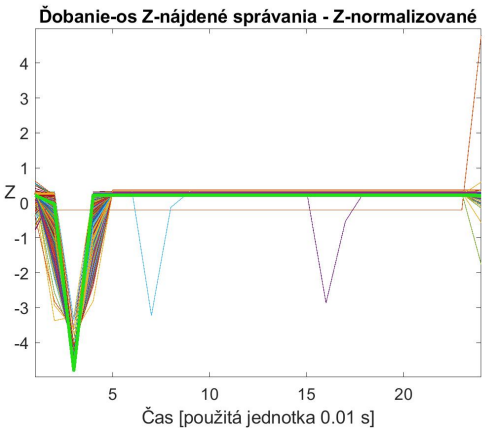
(a) Ďobanie v osi X pre kuru ID2



(b) Ďobanie v osi Z pre kuru ID2

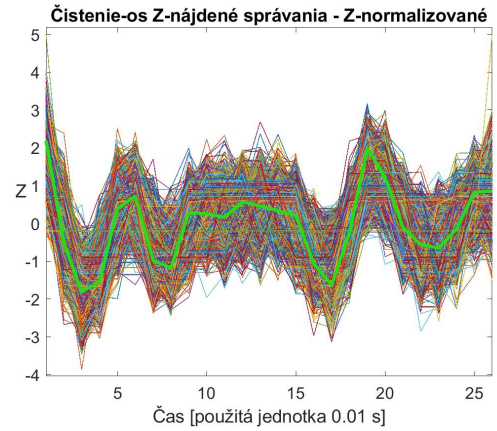


(c) Ďobanie v osi X pre kuru ID4

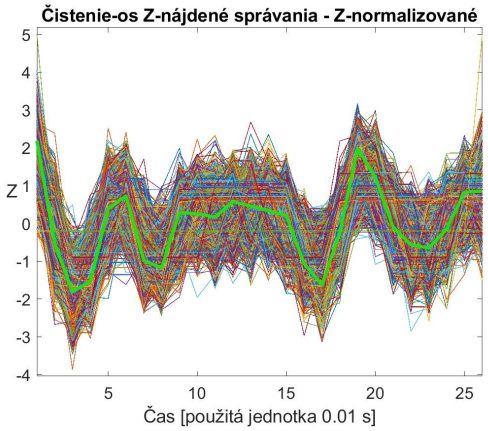


(d) Ďobanie v osi Z pre kuru ID4

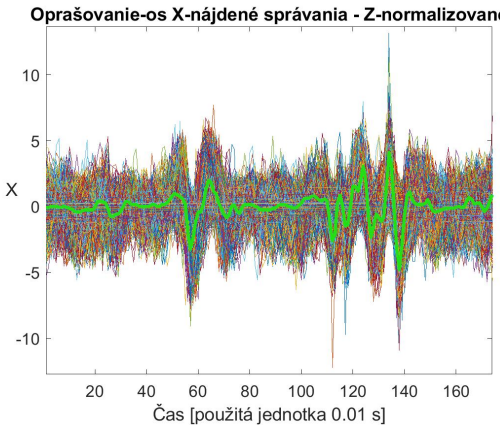
Obr. 36: Nájdené Z-normalizované podsekvencie správania typu ďobanie pre kury ID2 a ID4 z **Experimentu 2**. Knižnica použitá na nájdenie daných správaní je pre každú kuru a každé správanie natrénovaná zvlášť. Zelenou zvýraznené podsekvencie predsavujú najlepších kandidátov jednotlivých správaní zo spomenutej knižnice.



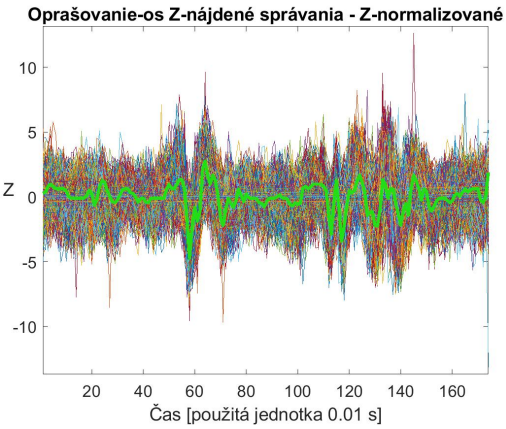
(a) Čistenie v osi Z pre kuru ID5



(b) Čistenie v osi Z pre kuru ID6



(c) Oprašovanie v osi X pre kuru ID7



(d) Oprašovanie v osi Z pre kuru ID7

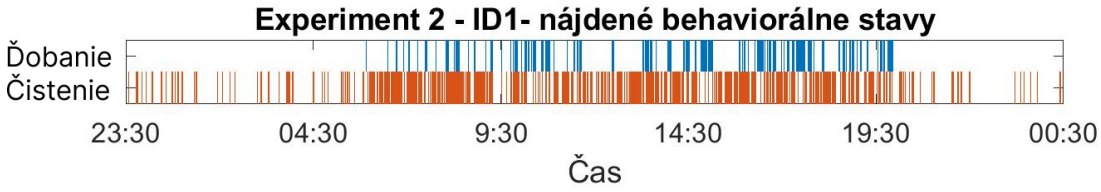
Obr. 37: Nájdené Z-normalizované podsekvencie správaní typu čistenie (kury ID5 a ID6) a oprašovanie (kura ID7) z **Experimentu 2**. Knižnica použitá na nájdenie daných správaní je pre každú kuru a každé správanie natrénovaná zvlášť. Zelenou zvýraznené podsekvencie predsavujú najlepších kandidátov jednotlivých správaní zo spomenutej knižnice.

Ako je môžné vidieť z etogramov na Obr. 35, reprezentanti typu ďobanie s najlepšiou hodnotou precision, konkrétne ID2 a ID4, klasifikovali na neanotovaných dátach ďobanie výrazne počas noci, čo nie je vhodné. Pri pohľade na čistenie (ID5 a ID6) a oprašovanie (ID7) sú nájdené správania klasifikované lepšie, keďže čistenie a oprašovanie sú správania, ktorých výskyt je bežný cez deň aj noc. Takisto pri pohľade na vizualizáciu nájdených správaní typu ďobanie, pomocou Z-normalizovaných časových podsekvencií na Obr. 36 vidíme, že pohyb senzora veľmi nenapovedá, že by sa malo jednať o ďobanie. Konkrétne na podsekvenciách ďobania v osi Z pre kuru ID4 na Obr. 36d vidíme podobnú zvlášnosť

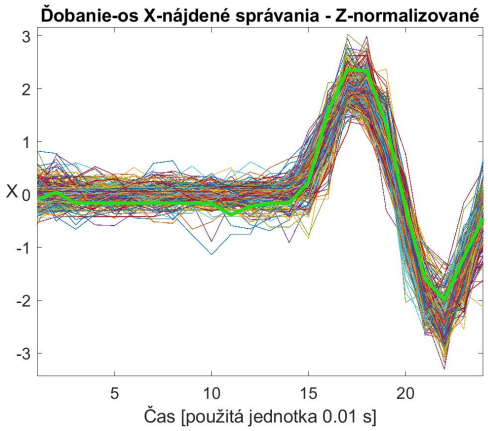
tvaru podsekvencií, akú sme spozorovali aj pri hromadne trénovanej knižnici z podkapitoly 4.2.2, konkrétne na Obr. 29b.

Klasifikácia jednotlivých správaní pomocou separátneho trénovania a testovania na každej kure zvlášť vykazuje kladné výsledky iba pri správaní typu čistenie a oprašovanie. Limitácia pri danej klasifikácii môže byť spôsobená netriviálnosťou anotovania, ako aj relatívne krátkym anotovaným časovým úsekom (okolo 5 min), na ktorom klasifikátor trénujeme, čo je rozdiel oproti 15 min trénovacieho úseku použitého v článku [1]. Takisto mohol nastať problém pri synchronizácii videa s časovým radom pri vytváraní anotácií.

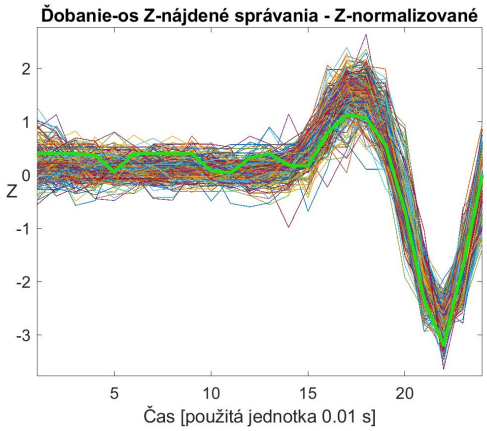
Limitácia klasifikátora je zrejmá, čo môžeme ukázať na správaní typu ďobanie a čistenie kury ID1, ktorých precision na Obr. 34 vykazuje pre obe správania hodnotu 0. Na druhú stranu, pre kuru ID1 vizualizujeme etogram na Obr. 38, ktorý je vizuálne v poriadku, resp. ďobanie sa vyskytuje počas dňa a čistenie primárne počas dňa a málo početne počas noci. Na Obr. 39 sme vykreslili aj Z-normalizované podsekvencie nájdených správaní, pričom zvýraznená zelená podsekvencia reprezentuje najlepšieho kandidáta z trénovacej a testovacej vzorky. Pri pohľade na správanie typu ďobanie vidíme, že sa nájdený najlepší kandidát (zvýraznený) relatívne podobá s kandidátom nájdeným na dátach z článku [1], ktorý je vizualizovaný na Obr. 7 v kapitole 2.



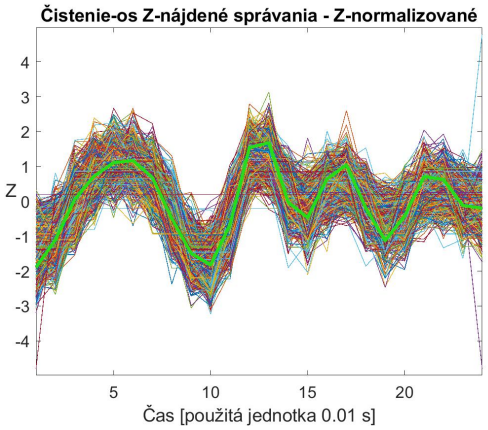
Obr. 38: Etogram nájdených správaní pre kuru ID1 využitím knižnice vyhotovenej na anotovaných dátach kury ID1



(a) Ďobanie v osi X pre kuru ID1



(b) Ďobanie v osi X pre kuru ID1



(c) Čistenie v osi Z pre kuru ID1

Obr. 39: Nájdené Z-normalizované podsekvencie správaní typu ďobanie a čistenie pre kuru ID1 z **Experimentu 2**. Knižnica použitá na nájdenie daných správaní je natrénovaná na anotáciach kury ID1 z rovnakého experimentu. Zelenou farbou zvýraznené podsekvencie predstavujú najlepšieho kandidáta jednotlivého správania z danej knižnice.

4.2.4 Analýza prejdených vzdialeností

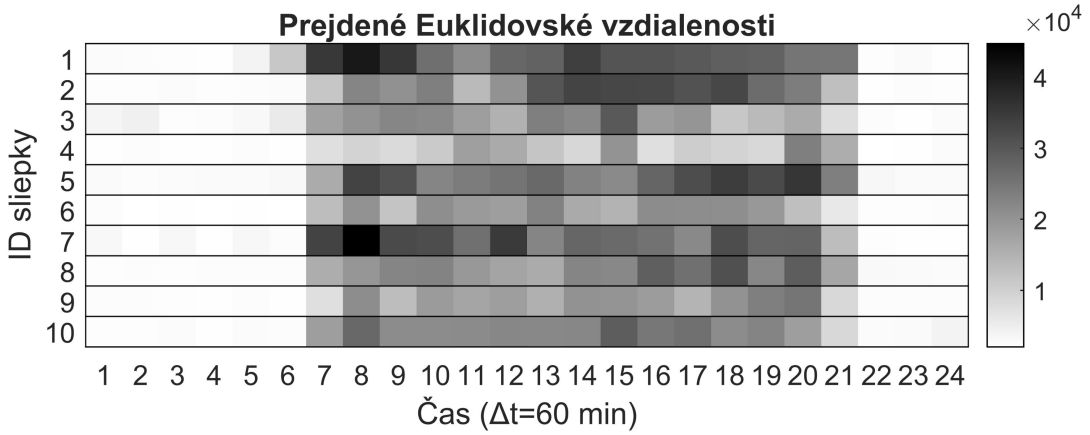
Vysoká frekvencia záznamu dát, v našej práci konkrétne 100 záznamov za sekundu, nemusí úplne presne odzrkadľovať prejdenú vzdialenosť kury domácej počas experimentu. Pri sledovaní každej stotiny sekundy záznamu, resp. ak výberová frekvencia je rovná jednej, zahŕňame do prejdenej vzdialenosti aj mikro pohyby, ako otočenie sa, naklonenie, atď., ktoré nereflektujú samotný pohyb kury. Ak však zvýšime výberovú frekvenciu napr. na hodnotu 100, t.j. uvažujeme každú sekundu záznamu, tak väčšinu spomenutých mikro pohybov odstránime a dostaneme približnú prejdenú vzdialenosť jednotlivej kury. Prej-

denú vzdialenosť meriame pomocou metriky Euklidovskej vzdialenosti, ktorá je meraná ako vzdialenosť medzi dvomi bodmi v 3D priestore pomocou vzorca

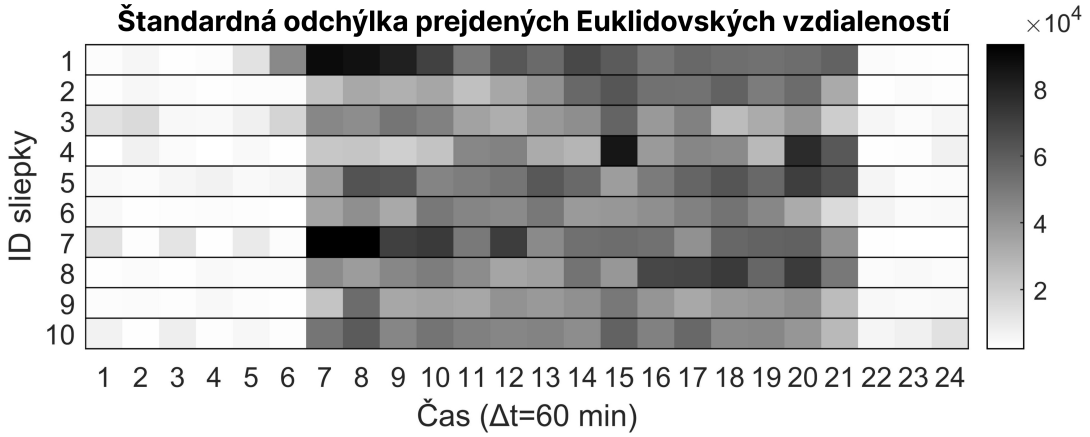
$$AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2},$$

kde bod A je tvorený súradnicami $(x1, y1, z1)$ a bod B súradnicami $(x2, y2, z2)$.

Na Obr. 40, resp. Obr. 41 je vykreslená heatmapa prejdených Euklidovských vzdialeností spolu s heatmapou štandardnej odchýlky prejdenej Euklidovskej vzdialenosti jednotlivých kúr z **Experimentu 2** pri výberovej frekvencii rovnej jedna, resp. 100. Pri porovnaní Obr. 40b a Obr. 41 vidíme omnoho menšiu kolísavosť hodnôt pri výberovej frekvencii 100. Aj to naznačuje kvalitnejšiu interpretáciu pohybu kury domácej, ak berieme do úvahy každý 100. dátový bod z akcelerometrických dát.

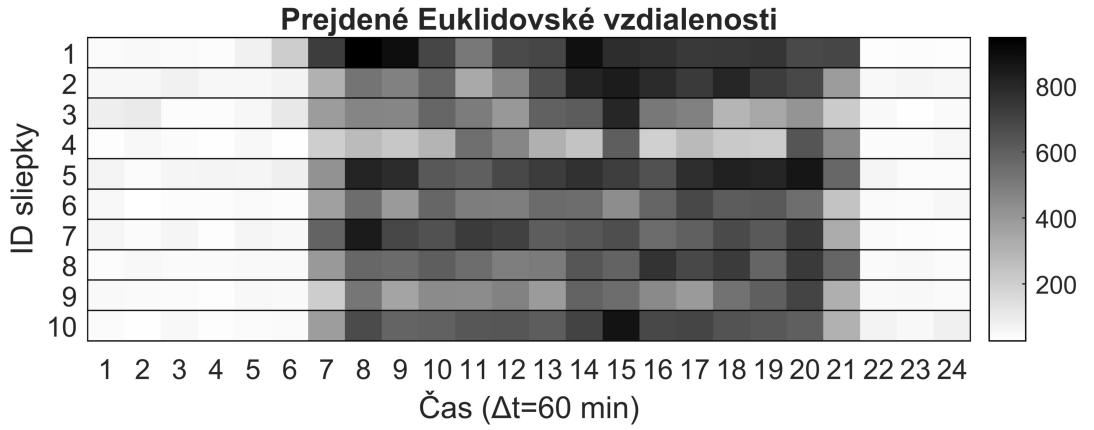


(a) Heatmapa prejdených Euklidovských vzdialeností

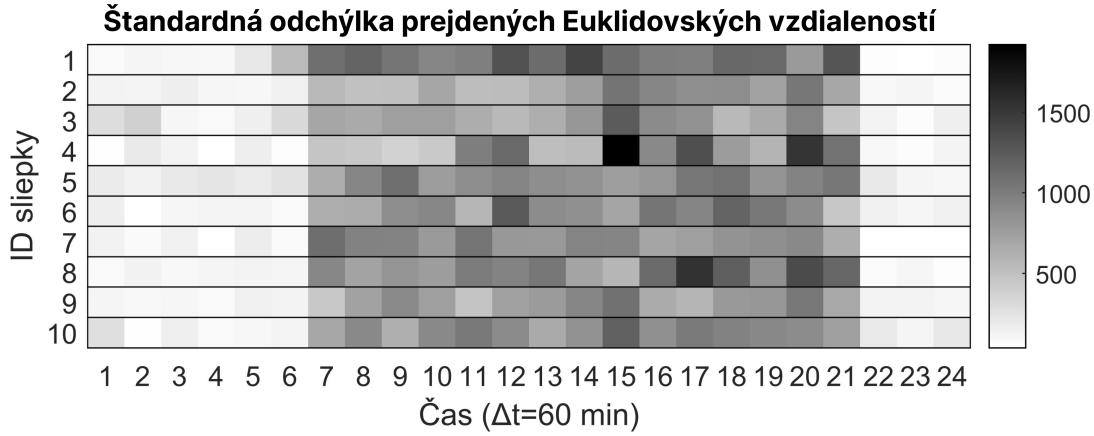


(b) Heatmapa štandartnej odchýlky prejdených Euklidovských vzdialeností

Obr. 40: Heatmapa prejdených Euklidovských vzdialeností počas jednotlivých hodín v **Experimente 2**. Výberová frekvencia je jedna, resp. uvažujeme každú stotinu sekundy záznamu a z daných bodov vypočítame celkovú prejdenú Euklidovskú vzdialenosť.



(a) Heatmapa prejdených Euklidovských vzdialeností



(b) Heatmapa štandartnej odchýlky prejdených Euklidovských vzdialeností

Obr. 41: Heatmapa prejdených Euklidovských vzdialeností počas jednotlivých hodín v **Experimente 2**. Výberová frekvencia je 100, resp. uvažujeme každú sekundu záznamu a z daných bodov vypočítame celkovú prejdenú Euklidovskú vzdialenosť.

Záver

Cieľom práce bolo primárne vyskúšať funkčnosť prístupu autorov článku [1] na dáta, ktoré zozbierala Slovenská akadémia vied (SAV) na vlastných experimentoch. Súčasťou práce bolo využitie viacerých prístupov, ktorých výsledky sme vzájomne porovnali.

V prvej kapitole sme najprv uviedli metodiku a priebeh jednotlivých experimentov, ktoré vyhotovila SAV. Rovnako sme objasnili aj technickú stránku ohľadom zbierania a zaznamenávania dát. Následne sme popísali štruktúru akcelerometrických dát a uviedli kumulatívne štatistiky pomocou histogramov.

V druhej kapitole sme vysvetlili, ako môžu byť jednotlivé správania reprezentovateľné pomocou časových radov. Jednotlivé správania v podobe časových radov sme vizualizovali, pričom niektoré aj dôkladne interpretovali. V danej kapitole sme čerpali prevažne zo zdrojov [1, 10].

Tretia kapitola slúžila na vysvetlenie a objasnenie použitých metód, pričom sme postup klasifikovania behaviorálnych stavov rozdelili do sledu krokov. Predstavili sme potrebnú anotáciu dát, ktorá hrá hlavnú úlohu pri trénovaní a testovaní klasifikátora. Príklad anotovaného datasetu správaní sme vykreslili na Obr. 12. Ďalej sme objasnili použitú metriku Euklidovskej vzdialenosti pri hľadaní podobnosti medzi časovými podsekvenciami spolu s odôvodnením potreby ich Z-normalizácie. Identické správania typu ďobanie, reprezentované pomocou časových radov, sme vizualizovali nenormalizované a Z-normalizované na Obr.13. Okrem toho sme znázornili zhotovenie knižnice správaní na anotovaných dátach kury domácej, pomocou ktorej prebieha finálna klasifikácia na neanotovaných dátach. Čerpali sme hlavne zo zdrojov [1, 5, 6, 7, 8].

V záverečnej kapitole sme aplikovali spomenuté metódy na reálne dáta z experimentov, ktoré vyhotovila SAV. Najprv sme na klasifikáciu správaní v dátach z **Experimentu 1** a **Experimentu 2** použili knižnicu správaní z článku [1], pričom sme porovnali dosiahnuté výsledky. Ďalej sme využili dva odlišné prístupy vytvárania knižnice správaní, pričom sme ich porovnali s knižnicou vytvorenou autormi v článku [1]. Pre jednotlivé prístupy sme klasifikovali správania na dátach SAV, kde sme klasifikátor natrénovali na anotáciach vyhotovených SAV. Pri každom prístupe bola vyhodnocovaná kvalita klasifikátora pomocou mier, akými sú precision alebo accuracy. Následne sme nájdené správania vizualizovali v čase pomocou etogramov, ktoré slúžili ako vizuálna kontrola. Rovnako sme vykreslili aj

priebeh klasifikovaných správaní pomocou časových radov. Pre prístupy, v ktorých sme vytvárali vlastnú knižnicu správaní pomocou anotácií vytvorených SAV, sme identifikovali limitácie klasifikátora, ktorého nepriaznivé výsledky môžu byť zapríčinené netriviálnosťou anotovania dát, nedostatočne dlhým anotovaným úsekom, alebo chybou pri synchronizácii videa s časovým radom počas vytvárania anotácií.

Za možný prínos diplomovej práce považujeme identifikáciu limitácii pri klasifikačnom procese, ako aj aplikáciu moderných metód na reálne dáta, vďaka ktorým bude môcť centrum biovied SAV identifikovať a analyzovať správania kury domácej.

Motivácia do budúcna je pomocou identifikovaných správaní jednotlivcov overiť prítomnosť vzájomnej behaviorálnej interakcie kúr v rámci monitorovanej skupiny. Inými slovami, bolo by zaujímavé zistiť, či konkrétne správanie jednej kury ovplyvňuje správanie inej kury. Taktiež by sme chceli zlepšiť klasifikačný proces a využiť tzv. 'feature-based' algoritmy na klasifiáciu behaviorálnych stavov z článku [2].

Zoznam použitej literatúry

[1] Abdoli, A., et al.: *Time Series Classification to Improve Poultry Welfare*, 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2018

[2] Abdoli, A., et al.: *Fitbit for Chickens? Time Series Data Mining Can Increase the Productivity of Poultry Farms*, University of California, Riverside, 2020

[3] Boďová K., Mitchell G.J., Harpaz R., Schneidman E., Tkačik G.: *Probabilistic models of individual and collective animal behavior*, 2018, PLOS ONE 13(3): e0193049. Dostupné na internete (12.4.2022) https://doi.org/10.1371/journal.pone.0193049

[4] Eerdekens, A.: *Automatic detection of abnormal behaviour of equines*, Faculty of engineering and architecture, Ghent university, Ghent, 2018, dostupné na internete (5.5.2022) https://www.ugent.be/ea/bme/en/program/bachelor/masterthesis/eerdekens-lq.pdf

[5] Mueen A., et al.: *The fastest similarity search algorithm for time series subsequences under euclidean distance*, 2015, dostupné na internete (2.4.2022)https://www.cs.unm.edu/~mueen/FastestSimilaritySearch.html

[6] Mueen A., et al.: *Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets.*, 1317-1322, 2016

[7] Mueen A., Hamooni H., Estrada T.: *Time Series Join on Subsequence Correlation.*, 2014 IEEE International Conference on Data Mining, 2014, pp. 450-459, doi: 10.1109/ICDM.2014.52.

[8] Mueen A., Keogh E., Young N.: *Logical-shapelets: an expressive primitive for time series classification.*, In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '11). Association for Computing Machinery, New York, USA, 1154–1162. dostupné na internete (5.5.2022):https://doi.org/10.1145/2020408.2020587

[9] Murillo A.C., Abdoli, A., Blatchford, R.A. et al.: *Parasitic mites alter chicken behaviour and negatively impact animal welfare*, Sci Rep 10, 8236 (2020)

[10] Vezzoli, G., Mullens, B.A., Mench, J.A..:*Relationships between beak condition, preening behavior and ectoparasite infestation levels in laying hens*, 2015, Poultry Science94, 1997–2007, dostupné na internete (2.5.2022) https://doi.org/10.3382/ps/pev171

[11] Murillo A.C., Abdoli, A., Blatchford, R.A. et al.: *Parasitic mites alter chicken behaviour and negatively impact animal welfare*, Sci Rep 10, 8236 (2020)

[12] Dáta k článku [1], dostupné na internete (2.5.2022): https://sites.google.com/site/chickenbehavioricmla18/

[13] Zmena formátu akcelerometrických dát zo senzora Axivity AX3, dostupné na internete (2.5.2022): https://github.com/digitalinteraction/openmovement/tree/master/Software/Analysis/Matlab