

FELHASZNÁLÓI TEVÉKENYSÉGEK MONITOROZÁSA

MILEFF PÉTER

Az ERP projekt által megfogalmazott egyik fontos cél, hogy a rendelkezésre álló, valamilyen információs rendszerből kinyert aktivitásnaplók alapján bizonyos események előre megjósolhatóak legyenek valamilyen hatékonyan konfigurálható mesterséges neurális hálózatot alkalmazó modell segítségével. A mesterséges intelligencián alapuló informatikai rendszerek egyik kulcskérdése a megfelelő adathalmaz, hiszen ez teszi lehetővé a ráépülő folyamatok, a tanuló algoritmusok tervezését, a megfelelő betanulást. A magas minőségű adathalmaz tehát kulcsfontosságú, ezért kiemelt figyelemmel kell megközelíteni a problémát.

Jelen projekt során az adatokat az üzemeltetett rendszerben elvégzett felhasználói tevékenységek rögzítése, monitorozása fogja nyújtani. A rögzített adatok, a munkafolyamat-lépések egy magasabb szintű szekvenciába, tranzakcióba szerveződnek. Mivel a rendszert egyszerre több felhasználó is használja, így egy olyan naplófájl fog létrejönni, amelyekben véletlenül keverednek a különböző tranzakciókhoz tartozó tevékenységek.

Jelen dokumentum áttekinti a felhasználói aktivitás naplózásának legfontosabb kulcskérdéseit, problémáit és egyaránt gyakorlati oldalról is bemutatja a naplófájlok egy lehetséges feloldozásának kérdéseit.

1. A kutatás célja és lépései

Napjaink modern világában a vállalatok ma már komplex informatikai rendszereket használnak az ügyviteli folyamataik támogatására. Nem ritka az olyan vállalkozás, ahol akár több mint 100 ember végzi a napi tevékenységeket. Általános jellemzője ezeknek a rendszereknek, hogy több tucat különböző típusú folyamatot végeznek az alkalmazottak. Idővel a vállalat fejlődése a folyamatok komplexitásának növekedését eredményezi. Csökken az átláthatóság, az érthetőség és idővel sokszor észrevehetetlen szűk keresztmetszetek alakulhatnak ki. Ilyen komplex rendszerek irányítása egyre több problémát vet fel mind biztonságtechnikailag, mint pedig a folyamatok követhetősége szempontjából. Az utóbbi idők eredményeképpen létrejött felhasználói aktivitásfigyelés (User Activity Monitoring – UAM) és az RPA-megoldások ezekre a problémákra próbálnak megoldást találni. A felhasználóiaktivitás-figyelés (UAM) megoldások olyan szoftvereszközök alkalmazását jelentik, amelyek figyelemmel kísérik és nyomon követik a végfelhasználói viselkedést az eszközökön, hálózatokon és más, a vállalat tulajdonában lévő informatikai erőforrásokon.

A kutatási munka célja, hogy részletesen áttekintse a felhasználói aktivitás szerepét, megmutassa azokat a normákat, amik rendelkezésre állnak és egy olyan irányt javasoljon, amely megfelelő eredményt jelenthet a gyakorlati alkalmazás során is.

A tervezett kutatási célok és lépések a következők:

- UAM-rendszerek főbb csoportjainak áttekintése.
- A felhasználóitevékenység-felügyelet működésének általános bemutatása: mit jelent az, hogy naplózás. Milyen főbb megközelítéseket szokás alkalmazni.
- A felhasználóitevékenység-felügyelet legfontosabb eszközeinek feltárása és rövid ismertetése.
- Célszerű megvizsgálni az aktivitásfigyelés gyakorlati megközelítését. Magát a naplókat, az eseményelőzményeket biztosító rendszereket, valamint az olyan rendszereket, amelyekhez nekünk kell összeállítani az előzményeket.
- További fontos terület a gyakorlati megvalósításhoz kapcsolódó problémák és akadályok feltérképezése és áttekintése.
- Aktivitási naplók hiánya és a naplók hiányosságai.
- Az esetazonosítóval kapcsolatban felmerülő problémák.
- Az aktivitással kapcsolatos problémák.
- Az időbélyeggel kapcsolatos problémák.
- Naplók ismertebb hibái: hiányzó aktivitások, hiányzó időbélyeg, befejezetlen folyamatok stb.
- A probléma neurális hálózattal való megoldásának rövid ismertetése, figyelembe véve a tanítóminta teljességét és az erőforrásszükséglet

2. Kutatási eredmények összesítése

2.1. Elvégzett kísérletek bemutatása

Az UAM-rendszereket célszerű két csoportba sorolni az alkalmazási céloknak megfelelően:

Felhasználói monitorozás információmegőrzés, biztonság, rosszindulatú visszaélések megakadályozása érdekében:

Számos szervezet a felhasználóiaktivitás-figyelő eszközöket alkalmazza a bennfentes fenyegetések észlelésének és megakadályozásának, akár akaratlanul, akár rosszindulatú szándékkal. A monitorozás és az alkalmazott módszerek köre a vállalat célkitűzéseitől függ. Az aktivitáskövetés megvalósításával a szervezetek hatékonyabban követhetik nyomon a munkavállaló online tevékenységeit, biztosíthatják az erőforrások megfelelő felhasználását, és megállíthatnak minden

olyan gyanús tevékenységet, amely a termelékenység csökkenéséhez vezet, és veszélyezteti a szervezetet. Könnyebben azonosíthatják a gyanús magatartást és mérsékelhetik a kockázatokat, mielőtt azok adatvédelmi jogsértést eredményeznének, vagy legalább időben a károk minimalizálása érdekében.

Felhasználói aktivitások monitorozása RPA-folyamatok támogatására:

Ebben az esetben az információ gyűjtésének célja egy olyan információs adathalmaz létrehozása, amely alapja lehet egy későbbi RPA- vagy Process Mining elemzésnek, gyakorlati megvalósításának. Legfőbb jellemzője az információk pontos és jól strukturált tárolása. A kinyert adathalmaz lehetőséget nyújt a szűk keresztmetszetek gráfelméleti módszerekkel való feltárására, a folyamatok automatikus reprodukálására, mesterséges intelligenciával való támogatására, továbbá esetleges bekövetkező események előrejelzésére. A továbbiakban a projekt keretein belül ezzel a témacsoporttal foglalkozunk mélyebben.

2.1.1. Felhasználóitevékenység-felügyelet működése

A felhasználói aktivitások figyelésére általában valamilyen aktivitásfigyelő szoftvert alkalmazunk. Az aktivitásfigyelő szoftver rögzíti az alkalmazások és programok használatát a felügyelt munkaállomáson. A képernyőn megjelenő felhasználói tevékenységek egy előre kidolgozott és jól strukturált naplóba kerülnek. A naplók tehát információs adatbázisok, amelyek minden olyan tevékenységet tárolnak, amelyek aznap történtek. A mai modern technológiának köszönhetően számos lehetőség, megoldás áll rendelkezésre a monitorozásra, a tevékenységek figyelemmel kísérésére és kezelésére. Néhány fontosabb technika:

- **Naplózás és elemzés:** a felhasználóiaktivitás-figyelés leginkább elterjedt formája. Ebben az esetben egy klasszikus szöveges fájlban, esetleg adatbázisban tároljuk a történt eseményeket. Minden esemény és folyamat jól definiált. A készülő naplófájl jól strukturált formában tárolja az információkat a későbbi feldolgozhatóság és analizálhatóság érdekében. Talán azt mondhatjuk, hogy több más alkalmazott technika alapjául szolgál.
- **A munkamenetekről készült videófelvetelek:** a felhasználói aktivitásról videó stream formájában felvétel készül. Fontos cél a későbbi elemezhetőség, az esetleges biztonsági visszaélések megelőzése, valamint az elévült interakciók esetleges reprodukciója.

- **Képernyőkép rögzítése:** hasonló megfontolások alkotják a technika célját, mint a videóalapú megközelítésnél. Ennél a módszernél képeket rögzítünk, amely alapot ad a további elemezhetőségre. Több gyakorlati megvalósításban a rögzített képernyőképeken az UAM-eszközök az adott aktivitás egy dobozzal való bekeretezésére. Például a felhasználó által kiválasztott elem, nyomott gombok stb.
- **Billentőzethasználat naplózása:** a lenyomott gombok naplózása olyan esetekben lehet fontos, amikor a felhasználók által használt vállalati vagy egyéb rendszer kevesebb grafikus interakciót vár el, több folyamat esetleg kézi adatbevitelt / gépelést vár el.
- **Hálózati csomagellenőrzés:** talán inkább az UAM-rendszerek biztonsági csoportjába tartozó megoldás. A felhasználók, a rendszer használói által küldött és fogadott csomagok naplózása. A forgalom ellenőrzése számos adatvédelmi lehetőséget nyújt a vállalat számára, megvédhet akár az esetleges visszaélésektől is.
- **Kernel monitoring:** az operációs rendszer kerneljének alacsony szintű monitorozása. Alkalmazása főként szintén biztonsági kérdésekben indokolt.

Az összes összegyűjtött információt a vállalati irányelvek és a felhasználói szerepkör határain belül kell megvizsgálni, hogy meghatározhatók legyenek az összefüggő folyamatok, a későbbiekben elemezni kívánt tevékenységek, valamint az esetleges nem kívánt felhasználói tevékenységek. Az, hogy mit jelent valójában a „nem megfelelő felhasználói tevékenység”, az mindig az adott vállalattól függ. A vállalatnak kell kiválasztani azokat az UAM ellenőrzési megoldásokat, amelyek bármire kiterjedhetnek a személyes webhelyek felkeresésétől vagy a munkaidőben történő vásárlástól a kényes vállalati adatok, például szellemi tulajdon vagy pénzügyi információk lopásáig.

2.1.1.1. Felhasználóitevékenység-felügyelet eszközei

Számos eszköz használható a felhasználói aktivitás figyelemmel kísérésére vagy támogatására. Ezek az eszközök az általános biztonsági szoftveralkalmazásoktól a munkamenetek és tevékenységek nyomon követésére szolgáló célzott eszközökig terjednek, és minden felhasználó számára teljes naplózási nyomvonalat hoznak létre. Vannak privilegizált fiókbiztonsági megoldásokként is ismert eszközök, amelyek célja a privilegizált fióktevékenységek figyelemmel kísérése és biztosítása, valamint a házirendek kezelésének központosítása.

A legjobb felhasználóiaktivitás-figyelő eszközök közé tartoznak a *valós idejű* figyelőrendszerek. Ezek az eszközök valós időben (gyakorlatilag azonnal) naplózzák a felhasználói tevékenységeket a háttérben, figyelik a felhasználói képernyőt a háttérben, és amikor célzott tevékenységre kerül sor, figyelmezteti az adminisztrátort vagy a menedzsert, vagy elvégzik a megfelelő tevékenységeket. Ezek lehetnek egyszerű, például az operátort támogató javaslatok, esetleg valamilyen szinten automatizált háttérfolyamatok elvégzése. Más esetekben pedig akár gyanús tevékenységek riportálása az informatikai és biztonsági csoportoknak. A valós idejűség nélkül a kockázatok bizonyos esetekben észrevétlenek maradhatnak, miközben a felelős informatikai csoport más ismert kérdésekkel foglalkozik. A mai technológiának köszönhetően nem szükséges, hogy egész informatikai csapat foglalkozzon az aktivitások monitorozásával és elemzésével. Egy jó UAM-megoldás kidolgozása, amely támogatja a felhasználói aktivitás figyelemmel kísérését, elemzését, riportálását és esetleges beavatkozását, nagymértékben egyszerűsíti a problémát.

2.2. Aktivitásfigyelés gyakorlati megközelítése

Bármely olyan tevékenységnek, amely a későbbiekben a felhasználói aktivitás eredményére szeretne építeni, legfontosabb kiindulópontja minden esetben az úgynevezett eseménynapló lesz. A leginkább gyakorlatban is alkalmazott technika az úgynevezett *Process Mining*, amelynek célja az adatok folyamatszempontú elemzése. Olyan kérdésekre keresi a választ, mint például a „A folyamat jelenleg milyen állapotban van?“, „Vannak-e felesleges lépések, amelyeket ki lehetne küszöbölni?“, „Hol vannak a szűk keresztmetszetek?“ És „Vannak-e eltérések a lefektetett és előírt folyamatszabályoktól?“

Annak érdekében, hogy az eseménynaplóban lévő adatokat bármilyen formában is elemezni lehessen, azokat valamilyen folyamatspecifikus megközelítéssel kell a logba írni. Ez lehetővé fogja tenni, hogy a későbbiekben az eseménynaplóban szereplő lépésről el lehessen dönteni, melyik folyamathoz tartozik és azon belül pedig az adott folyamat melyik lépése volt.

Talán a témakör egyik legfontosabb kérdése lehet az, hogy „*Honnan érkeznek az adatok az eseménynaplóba?*“

A kérdés válasza rögtön két csoportra bontja az információforrást és az aktivitásfigyelés módszerét.

2.2.1. Naplókat, eseményelőzményeket biztosító rendszerek

Olyan rendszerek, amelyek már önmagukban tartalmazzák az eseménynaplózást, melyek sok esetben akár nagyon jól konfigurálhatók is. Például az ügyfélkapcsolat (CRM), az IT-szolgáltatáskezelés (ITSM), a rendeléskezelés és a munkafolyamat-rendszerek általában ebbe a kategóriába tartoznak. Az aktivitásfigyelés és folyamatbányászat szempontjából ezek a rendszerek a megvalósítási spektrum könnyebb oldalához tartoznak, hiszen az adatok nagy valószínűséggel megfelelő minőségben már rendelkezésre állnak naplók, adattáblák (előzménytáblák) és egyéb megoldások formájában, azok sokszor közvetlenül felhasználhatók. Például egy ITSM-rendszerben minden esetben van jegyszám, továbbá minden állapotváltozást rögzítenek.

Sok esetben ezek az adatok könnyen exportálhatók CSV-fájlként és közvetlenül importálhatók a folyamatbányász eszközbe előzetes feldolgozás nélkül.

2.2.2. Rendszerek, amelyekhez nekünk kell összeállítani az előzményeket

Itt, a spektrum nehezebb végén olyan adatbázis-központú rendszereket találunk, mint az Enterprise Resource Planning (ERP) és bizonyos egyedi vagy legacy típusú rendszerek. Ezekben az esettörténetek nem állnak rendelkezésre feldolgozáskész formában, mert nem így lettek tervezve és kialakítva már a kezdetektől fogva. Ezért minden esetben valamilyen egyedi megoldást kell készíteni arra vonatkozóan, hogy a releváns adatokat az üzleti adatbázis-táblákban megtaláljuk, és létrehozunk egy saját eseménynaplót az elemzéshez.

Ezen rendszerek további kihívása, ha a folyamatot több informatikai rendszer is támogatja, használja. Tipikus példa lehet, amikor a vásárlási igényeket az ERP-rendszerben kezeljük, a számlákat pedig külön, pénzügyi rendszeren keresztül támogatjuk. Ha két (vagy több) rendszer adatait kívánjuk egyesíteni, akkor a legfontosabb dolog, amire figyelni kell, hogy miként tudjuk követni az esetet ezeken a különböző rendszereken. Például, ha az ERP-rendszer minden esetben vásárlási rendelési számot használ, a pénzügyi rendszer pedig minden esetben számlaszámot használ, akkor a számlában találni kell egy olyan hivatkozást, amelyre a megrendelés vonatkozik.

A megoldást ezen rendszerek esetében két útra lehet bontani:

- **A rendszer forráskódjának kiegészítése:** egyik legjobb megoldás lehet, ha rendelkezünk a rendszer forráskódjával és a folyamatelemzés szempontjából igényelt és kritikus pontokra eseménynaplózást helyezünk el a kódban. Mivel a teljes forráskód módosítható, így az események naplózása maradéktalanul

megoldható. A módszer meglehetősen nagy hátrányát jelenti, hogy egy jól működő rendszer forráskódjába ismételten belenyúlni sokszor veszélyes és a rendszer teljes újratestelése időigényes lehet. Ez függ attól, hogy a korábbi fejlesztőkörnyezet hogyan lett kialakítva, esetleg kidolgozásra került-e valamilyen automatikus tesztelési eljárás, stb. Tovább nehezítheti a fejlesztést az is, hogy sok esetben bár rendelkezésre áll a forráskód, de a rendszer fejlesztése óta számos év telt el, így a forráskód és az egyes támogató függvénykönyvtárak frissítése elengedhetetlen. Végül a módosított rendszert (ismét) üzembe kell helyezni.

- **Felhasználóiaktivitás-figyelő szoftver alkalmazása:** egy másik alternatívát képvisel az, ha létező aktivitásfigyelő szoftvert alkalmazunk. A megoldás előnye, hogy nem kell a működő rendszer kódját vagy bármely részét módosítani, hanem a felügyelni kívánt rendszerekre, állomásokra csak az UAM-szoftvert kell telepíteni és konfigurálni. A mai modern UAM-szoftverek a felhasználói interakció már számos fajtáját képesek kezelni/naplózni többé-kevésbé. Sajnos mégsem mondhatjuk ki, hogy a probléma maradéktalanul megoldásra került, a következő okok miatt: bár a felhasználók interakciói naplózásra kerülnek, az események folyamatlépésekhez való kötése a legtöbb esetben nem tud maradéktalanul megvalósulni. Ennek oka, hogy a megtörtént esemény – legyen az egy megnyomott gomb – olyan egyedi azonosítóval kell legyen ellátva, amely köthető a folyamathoz, mint annak egy lépése. Mivel a felhasználói interfész elemeinek azonosítói a fejlesztéskor kerültek meghatározásra, nem biztos, hogy úgy lettek kialakítva, hogy azok alkalmasak legyenek a későbbi naplózáshoz.

A gyakorlatban a fent vázoltak miatt leginkább valamilyen UAM-szoftvert próbálnak alkalmazni a megoldás megvalósításához azért, mert a forráskód nem áll rendelkezésre, esetleg más cég készítette a szoftvert. Fontos szót ejteni arról, hogy annak megállapítása, hogy milyen UAM-szoftvert alkalmazunk, függ a naplózni kívánt szoftver megvalósításától, amelyek lehetnek vastag, illetve vékony kliensalappúak. Míg a korábbi években az informatikai szoftverek jellemzően vastag kliens módszerekkel lettek fejlesztve, ma már leginkább vékony kliens, web-alapú megoldásokkal találkozunk. A két megközelítés eltérő UAM-megoldást kíván meg. Vastag kliensek esetében alig található a piacon bármilyen UAM szoftver. Ebben az esetben a szoftver erősen függ az operációs rendszertől, amely nagymértékben megnehezíti azt, hogy egy külső szoftver naplózást készítsen a kívánt másik szoftverről. Sok esetben a képernyőképmintések jelentik a megoldást, de ez is számos további problémát vehet (felbontás, UI-azonosítók hiány stb.) fel.

Szerencsére a világ vékony kliens irányba való elmozdulása új lehetőségeket nyitott meg a naplózás területén is. Számos szoftvercsomag, beépülő modul áll rendelkezésre a kívánt eredmény elérésére.

2.3. A megvalósításhoz kapcsolódó problémák és akadályok

Egy jól megvalósított RPA rendszer általában komplex megvalósítási folyamatot igényel. Számos részfeladat, amelyek önmagukban is bonyolultak lehetnek, természetüknél fogva bizonytalanságokat, illetve problémákat rejtenek magukban. A következőkben néhány ilyen jellegű fontosabb problémát emelünk ki.

2.3.1. Aktivitási naplók hiánya

Mint ahogyan az korábban említésre került, a megfelelő adathalmaz a mesterséges intelligencián alapuló informatikai rendszerek egyik kulcskérdése, hiszen ez teszi lehetővé a a tanuló algoritmusok tervezését, a megfelelő betanulást.

Az egyik legfontosabb kérdés talán az, hogy lesz-e megfelelő minőségű adathalmaz a projekt során? A kérdés jogos aggályt fogalmaz meg, hiszen egy már évekkel ezelőtt üzembe állított rendszer vélhetően nem úgy lett tervezve, hogy a felhasználói aktivitások egyszerűen kinyerhetők belőle. Ráadásul a több komponensből álló rendszer különböző komponensei nem ugyanúgy viselkednek, eltérő felépítéssel rendelkeznek, melyek a naplózási aktivitásban is minden bizonytalansággal eltérnek. Természetes követelmény, hogy a projekt során a partner rendszereiben megtörténjen a megfelelő részletezettségű aktivitásnaplók készítésének beépítése, azonban mivel ez önmagában véve is időigényes feladat, nem várható, hogy már a kezdetben is rendelkezésre álljanak használható aktivitásnaplók.

A mintarendszert ezért olyan módon kell megtervezni, hogy képes legyen a valósághoz közelítő, vagy azzal megegyező adatokat generálni. Ezáltal a kezdeti adathalmaz hiánya feloldható, a kutatási feladatok, az algoritmusok tervezése nem blokkolódik.

2.3.2. Aktivitási naplók hiányosságai

Egy aktivitásnapló alapvetően három kötelező elemből kell álljon: esetazonosító (case id), aktivitás (activity) és egy időbélyeg (timestamp). Minden **eset** egy folyamat végrehajtási példányának egy lépését jelenti. Például egy rendelési folyamatban, egy megrendelés kezelése egy esetet jelent. Nagyon fontos kritérium, hogy minden esetenél tudni kell azt, hogy melyik az a folyamat, amihez globálisan tartozik. Az **aktivitás** a folyamat egyik lépése. Például egy dokumentumkészítési folyamat a következő lépésekből állhat: „Létrehozás”, „Frissítés”, „Küldés”,

„Jóváhagyás”, „Átdolgozás kérése”, „Felülvizsgálat”, „Közzététel”, „Elvetés” (különböző személyek, mint pl. szerzők és szerkesztők). Ezen lépések némelyike egy esetben többször is megtörténhet, de nem kell mindegyiknek minden alkalommal megtörténnie. A folyamat során végrehajtott különböző folyamatlépéseket vagy állapotváltozásokat megfelelő elnevezésekkel kell ellátni. Ha minden esetben csak egy bejegyzés (egy sor) van, akkor az adatai nem elég részletesek. Az adatoknak tranzakciós szinten kell lenniük (az egyes esetek előzményeihez hozzá kell férnie), és azokat nem szabad egy esetben aggregálni. Továbbá legyünk tisztában azzal, hogy a választott tevékenység befolyásolja azt a részletességi szintet, amellyel folyamatunkra tekintünk. Itt is előfordulhat, a tevékenység nevét akár több oszlop kombinációjaként érdemes meghatározni, és több alternatív nézet is lehet arra vonatkozóan, hogy mi minősül tevékenységnek. Végül a folyamatbányászathoz harmadik fontos előfeltétele, hogy legyen legalább egy **időbélyeg** oszlop, amely jelzi, hogy az egyes tevékenységek mikor történtek. Ez nemcsak a folyamat időzítési viselkedésének elemzéséhez fontos, hanem a tevékenységek sorrendjének megállapításához is az eseménynaplóban. Néha a folyamat minden tevékenységéhez van egy *start* és egy *end* időbélyeg. Ez jó, mert lehetővé teszi egy tevékenység feldolgozási idejének elemzését (az az idő, amelyet valaki aktívan az adott feladat elvégzésére fordított), amelyet végrehajtási időnek vagy tevékenységkezelési időnek is neveznek.

	Case ID	Timestamp	Activity			
1	CaseID	Timestamp	Medium	Status	Service Line	Urgency
2	case9700	20.8.09 11:46	Phone	Registered	1st line	0
3	case9700	20.8.09 11:50	Phone	Completed	1st line	0
4	case9701	23.9.09 12:23	Phone	Registered	1st line	0
5	case9701	23.9.09 12:27	Phone	Completed	1st line	0
6	case9705	20.10.09 14:21	Phone	Registered	Specialist	2
7	case9705	20.10.09 16:48	Phone	At specialist	Specialist	2
8	case9705	19.11.09 10:31	Phone	In progress	Specialist	2
9	case9705	19.11.09 10:32	Phone	Completed	Specialist	2
10	case3939	15.10.09 11:48	Mail	Registered	Specialist	2
11	case3939	15.10.09 11:48	Mail	Offered	Specialist	2
12	case3939	20.10.09 17:18	Mail	In progress	Specialist	2
13	case3939	20.10.09 17:19	Mail	At specialist	Specialist	2
14	case3939	21.10.09 14:49	Mail	In progress	Specialist	2
15	case3939	21.10.09 14:49	Mail	In progress	Specialist	2
16	case3939	28.10.09 10:17	Mail	In progress	Specialist	2
17	case3939	28.10.09 10:18	Mail	Completed	Specialist	2
18	case9704	20.10.09 14:19	Mail	Registered	1st line	0
19	case9704	20.10.09 14:24	Mail	Completed	1st line	0
20	case9703	20.10.09 14:40	Phone	Registered	1st line	0
21	case9703	20.10.09 14:58	Phone	Completed	1st line	0
22	case9702	24.8.09 12:24	Mail	Registered	2nd line	2
23	case9702	24.8.09 12:30	Mail	Offered	2nd line	2
24	case9702	24.8.09 12:31	Mail	Scheduled	2nd line	2
25	case9702	26.8.09 9:05	Mail	In progress	2nd line	2
26	case9702	26.8.09 9:19	Mail	Completed	2nd line	2
27	case9709	20.10.09 14:26	Mail	Registered	Specialist	2
28	case9709	20.10.09 14:26	Mail	Offered	Specialist	2

1. ábra. Mintanapló fájlra, amely tartalmazza a három kötelező elemet

2.3.2.1. Az esetazonosítóval kapcsolatban felmerülő problémák:

- Az esetazonosítót több mezőből kell kombinálni: Ez akkor történik, ha a fő esetazonosító önmagában nem azonosítja egyedileg a folyamatpéldányt. Például, ha elemezzük az adóbevallási folyamatot az adóhivatalnál, akkor minden állampolgárt a társadalombiztosítási száma alapján azonosítanak. Az adóbevallásokat azonban minden évben elkészítik. Ha több év adatai vannak, akkor a társadalombiztosítási szám mellett össze kell kapcsolnia azt az évet is, amelyre a nyilatkozatot benyújtották. Az aktivitásnapló készítésekor tehát nem szabad elfeledkezni az egyedi eset azonosításához szükséges összes mezőről.
- Különböző esetazonosítók használata a folyamat különböző részein: Ha a folyamat során több azonosítót használnak, akkor képesnek kell lenni egy eset láncolatának követésére az elejétől a végéig. Például, ha egy vásárlási folyamat során a beérkező kérelmeket PO-számmal azonosítják, és később a fizetési tevékenységek egy számlaszámhoz kapcsolódnak, akkor szükség van arra, hogy a számlaazonosító össze legyen kapcsolva a megrendelés számával (PO), legyen valami referencia, amely segít a folyamatok összerendelésében. Az aktivitás kinyerése során a naplóban egy folyamatos esetazonosítóra van szükség, amely referenciaként szolgál a folyamat minden lépésére.
- Több esetazonosító több-több kapcsolattal rendelkezik: Ha a folyamat során különböző esetazonosítókat használnak, előfordulhat, hogy nem mindig rendelkeznek 1-1 leképezéssel (például két megrendelés kombinálva egy kézbesítésben, vagy egy megrendelés két darabban). Ilyenkor meg kell határozni, hogy melyik szempontból kívánjuk vizsgálni a folyamatot. Előfordulhat, hogy több naplót/adatexportot kell létrehozni, ha több nézőpontot szeretnénk elemezni.

2.3.2.2. Az aktivitással kapcsolatos problémák

A második minimumkövetelmény a naplókkal szemben az említett tevékenység. A tevékenységek a folyamat különböző lépései vagy állapotváltozások. Az informatikai rendszerek nemcsak a számunkra fontos tevékenységeket rögzíthetik, hanem kevésbé érdekes hibakeresési információkat is. Ezért fontos arról meggyőződni, hogy a naplózás a releváns aktivitásokat is rögzíti. A kevésbé releváns tevékenységek naplóban való megléte önmagában nem probléma, később ezek kiszűrhetők.

Nagyon fontos az aktivitások megfelelő elnevezése már a kezdetektől fogva. Ha több jelölt is van a tevékenység elnevezéséhez, akkor célszerű megtartani

mindet, mert az elemzés során így lehetővé válik különböző szempontok vizsgálata. Például egy banki hiteligenylési folyamat során a rendszer rögzítheti mind a belső állapot (a belső folyamat szakaszának bemutatása), mind a külső állapot változását (megmutatja az ügyfél vagy az értékesítési csatorna állapotát). Mindkét szempont hasznos lehet. Ha mindkét mezőt megtartjuk az adatokban, akkor később teljes rugalmasságot élvezhetünk az elemzésben.

Fontos szempont, hogy a naplóban rögzített aktivitások nevei az ember által jól olvashatók legyenek. A legtöbb folyamat összetett, és az elemzés során nem jutunk túl messzire, ha a folyamatátré képi lépései pusztán technikai állapotszámokat vagy műveleti kódokat mutatnak. Az adatokban természetesen rögzíthetünk technikai műveleti kódokat is, amelyek később hasznosak lehetnek az adatokkal kapcsolatos technikai kérdések visszakeresésére. Az ember által olvasható név azonban nem elhanyagolható, mert az adja meg pontosan, hogy mi a státusz jelentése. Ezeket az ember által olvasható állapotneveket külön oszlopként kell felvenni az adatokba.

A tevékenységek több helyen rejtve lehetnek. Az adatbázis-központú rendszerekben gyakran a különféle üzleti adattáblákból nyerhetők ki bizonyos tevékenységek. Ebben az esetben automatikusan nem láthatók a folyamat összes eseményei. Ilyenkor meg kell határozni a legfontosabb mérföldkő-tevékenységeket, majd ezek alapján kigyűjteni az aktivitásokat.

2.3.2.3. Az időbélyeggel kapcsolatos problémák

Minden tevékenységhez legalább egy időbélyegre van szükség, hogy az egyes eseményeket megfelelő sorrendbe lehessen rendezni. Ha pedig elemezni szeretnénk a tevékenység időtartamát, akkor minden tevékenységhez meg kell adni egy kezdeti és egy befejezési időbélyeget.

Törekedni kell az időbélyegek lehető legpontosabb rögzítésére. Például, ha rendelkezésre állnak az órák és percek adatok is, akkor nem elégséges csupán a dátumot rögzíteni, hanem az időt is célszerű tárolni. Ha pedig másodperc vagy milliszekundum is rendelkezésre áll, az még sikeresebb, még akkor is, ha úgy tűnik, hogy nincs szükség annyi részletre az elemzéshez. A részletes időbélyegek készítése hasznos lehet a tevékenységek sorrendjének megfelelő meghatározásához.

Tevékenységenként több mint két időbélyeg: bizonyos esetekben nemcsak egy vagy két időbélyeg áll rendelkezésre tevékenységenként, hanem több is. Például lehet egy időbélyeg, amely jelzi, hogy egy feladat készen áll-e a felvételre, egy második, amely azt jelzi, hogy az illető mikor kezdett dolgozni rajta, és egy harmadik, amely a feladat befejezését jelzi. Célszerű ezeket mind rögzíteni. A

későbbiekben hasznos lehet különböző megválaszolendő kérdésekben. Például, hogy mennyi az átlagos időtartam a kész és a befejezés között.

A dátum és az idő külön oszlopokban szerepelnek: Mivel egyszerre több időbélyeget is rögzíthetünk, egyetlen időbélyeg dátumának és időpontjának célszerű egy mezőként szerepelnie. Ez segít a későbbi elemzésben.

Különböző időbélyegminták: A különböző adatforrásokból származó időbélyegek formátuma eltérő lehet. Például egy tevékenység dátum-időbélyeg formátuma lehet a 01SEP2013, míg egy másik tevékenysége a 2013-09-10 07.36.51.711899 formátum. A naplóban egységes időbélyeg-formátum alkalmazás az elvárt. Amennyiben többféle formátum is rendelkezésre áll, úgy vagy egységesre hozzuk őket, vagy külön adatként kezeljük.

Időbélyeg-felülírás: Ismétlődő aktivitások során bizonyos időbélyegek felülíródhatnak. Amikor a tevékenység időbélyegeit adatbázisban tároljuk, gyakran feltételezik, hogy a folyamatot az ideális sorrendben hajtják végre, és nem történik újrafeldolgozás. Természetesen ez a legtöbb helyzetben helytelen: Az átdolgozás történhet például azért, mert ezt a feladatot nem először hajtották végre, és a megfelelő folyamatlépést később meg kell ismételni. Ha a rendszer nem rögzíti a tevékenység ismétlésének összes időbélyegét (gyakran csak a legfrissebb időbélyeget őrzi meg), akkor értékes információk vesznek el.

2.3.3. Naplók ismertebb hibái

Bármilyen az inputadatokban megjelenő hiányosság, az úgynevezett „lyukak” alapjaiban befolyásolják a végeredményt. A leggyakrabban felmerülő problémák a következők:

- **Formázási hibák:** Az egyik leggyakrabban előforduló hiba, hogy bizonyos esetekben a napló adott sorának formátuma eltér az előírttól. Ez nagymértékben befolyásolhatja az elemzés eredményét, jobb esetben csak feldolgozási hibát eredményez. Speciális karakterek, egy felesleges idézőjel számos fejtörést okozhat.
- **Hiányzó események:** attól még, hogy az adatok hibátlanul importálódtak, továbbra is problémák lehetnek az adatokkal. Egyik tipikus probléma az adatok, események hiánya. Két típusa a következő:

Lyukak az idővonalban: néha szokatlan hiányosságok fedezhetők fel a napló bizonyos időkeretein belül bekövetkező események számában.

Váratlan adatmennyiség: szükség van arra, hogy nagyjából ismerjük azt, hogy az egyes folyamatok milyen nagyságú tevékenységszámból áll. A váratlan nagy adatmennyiség miatt annak egy része a feldolgozó bizonyos részein elveszhet bármilyen nemű jelzés nélkül egyszerű tervezési hiba miatt.

- Hiányzó aktivitások: Előfordulhat, hogy a folyamat egyes tevékenységeit nem rögzítik az adatok. Lehetnek olyan manuális tevékenységek (például telefonhívások), amelyeket az emberek az asztalunknál végeznek. Ezek a tevékenységek a folyamat során történnek, de nem láthatók az adatokban.
- Hiányzó időbélyeg: Bizonyos helyzetekben tudomásunk van arról, hogy történt-e tevékenység vagy sem, de nincs időbélyeg. Esetleg az aktivitás egy más rendszerből érkezett. Az adatfeldolgozó modul ilyen esetekben nem tudja meghatározni az esemény helyét az idővonalban.
- Befejezetlen folyamatok: a napló olyan folyamatokat is tartalmaz, amelyek rendelkeznek „start” jelzéssel, de „vég” jelzéssel nem. A probléma oka a következőkben keresendő:
 - a. Az adatkinyerési módszer csak bizonyos időintervallumon belüli eseményeket vizsgál. Például egyes folyamatok az előző évben kezdődtek (január előtt) és a következő évben folytatódtak. Ebben a helyzetben ezeknek a folyamatoknak csak azt a részét fogja látni, amelyek abban az évben történtek, amelyet elemez.
 - b. Néhány folyamat még nem fejeződött be. Még akkor is, ha minden adatot kinyertünk, előfordulhat, hogy az esetek egy része még nem fejeződött be. Még mindig „valahol a közepén vannak”.
 - c. Egyes folyamatok soha nem fejeződnek be. Nem minden folyamat fog rendelkezni „vég” jelzéssel. Bizonyos esetekben, akár technikai hiba miatt a folyamat megszakad. Például egy webshopban elvégzett fizetés után az ügyfél nem irányítódik vissza a bolt felületére. Ezek az esetek nem érnek véget a várható végpontok egyikén sem, soha nem fejeződnek be.

2.4. Problémamegoldás neurális hálózattal

A mesterséges neurális hálózatok napjaink egyik leginkább kutatott területe. Segítségével olyan problémák válnak megoldhatóvá, amelyek korábban nem, vagy csak részlegesen. Természetesen a hardverek fejlődése ehhez nagy támpontot nyújtott. Ma már számos modellt tart nyilván az irodalom, a leghatékonyabb megközelítések gyakran az úgynevezett mély tanulás (deep learning) területhez kapcsolódnak. A mesterséges neurális hálózatok tanítása minden esetben egy felügyelt regressziós problémára kerül visszavezetésre, de kivitelezhető osztályozás

és nem felügyelt tanítás is. A hálózatok működésében két fázist különíthetünk el: tanítási fázisban a ismert bemeneti paraméterek és várt kimenetek ismeretében a súlyokat változtatjuk úgy, hogy egy veszteségfüggvény értékét (például az átlagos négyzetes hibát) minimalizáljuk ezzel. A feltanított neurális hálózat a predikciós fázisban ezután ismeretlen bemenet átadásakor kimenetet képez, mely lehet például egy kategóriába való tartozás valószínűsége.

Neurális hálózatokat a gyakorlatban előszeretettel alkalmazzák a különböző aktivitások előrejelzésére. A tevékenységnapló adatai alapján a műveletsori szekvenciák azonosíthatók, valamint az aktuális tevékenységsor pedig illeszthető a meglévő mintákra. A folyamatfeltárás neurális hálózat alapú megközelítése azonban számos problémát vet fel. Semmiféleképpen sem nevezhető tipikus ANN-feladatnak, mert a probléma megoldása egy nem felügyelt tanulást kíván meg, a neurális hálózatok pedig főként felügyelt tanulásra alkalmasak. Ezen a területen számos kutatási irány/probléma fog előtérbe kerülni, amelyek arra a kérdésre fogják keresni a választ, hogy miként lehet megalkotni egy olyan hálózatot, amely képes az említett problémát megoldani. A jövőben új, dinamikus struktúrájú neurális háló-modell kidolgozása fogják a kutatások alapját képezni.

2.4.1. Tanítóminta

A neurális hálózatok tanulási folyamatához tanítómintára van szükség. A korábban említésre került inputadatok hiánya/nem megfelelő minősége alapjaiban befolyásolhatja a tanulás folyamatát. Az inputadathalmaznak kellően komplexnek kell lennie – tartalmazva az összes előforduló folyamatot, azok aktivitásait – ahhoz, hogy a különböző megközelítések, hipotézisek és modellek maradéktalanul kipróbálhatók legyenek a modellalkotás során. Tovább kimondható, hogy egyetlen mintaadathalmaz önmagában szintén nem elegendő, különben a kidolgozott modell nem tesztelhető hatékonyan a valóságban előforduló sokszínűség kezelésében.

2.4.2. Erőforrás-szükséglet

Az utóbbi években egyre hangsúlyosabb szerepet kapnak az úgynevezett LSTM típusú neurális hálózatok, melyek számtalan helyen (Siri, Cortana, Google Voice Assistant, Alexa) bizonyították hatékonyságukat. Bár hatékonyak, célszerű odafigyelni a hátrányukra is:

Az LSTM-alapú hálózatok tanítási folyamata nehezebb, mert memória-sávszélességhez megfelelően igazított számítást igényelnek, ami a hardvertervezők szá-

mára egy külön fejtörés, valamint korlátozza a hálózat gyakorlati alkalmazhatóságát. Az LSTM-nek cellánként 4 lineáris rétegre (MLP-réteg) van szüksége ahhoz, hogy az egyes szekvencia-idő-lépésekben és az egyes szakaszokban fusson. A lineáris rétegek nagy mennyiségű memória sávszélességet igényelnek, ebből kifolyólag gyakran nem használhatnak sok számítási egységet, mert a rendszer nem rendelkezik elegendő memória-sávszélességgel a számítási egységek “táplálásához”. További számítási egységeket hozzáadni könnyű, azonban memória-sávszélességét növelni sokkal nehezebb. Ennek eredményeként az RNN / LSTM és a variánsok nem gyorsíthatók hardveresen megfelelően.

Tehát LSTM alapú hálózatokkal való dolgozás mindenféleképpen nagyobb erőforrás-szükségletet eredményez, amely főként a tanulási fázist érinti. Ráadásul a projekt kapcsán kezelt, illetve várt inputként szolgáló adathalmaz is nagy. Ki kell dolgozni tehát egy olyan környezetet az kutatók számára, ahol osztott módon képesek kihasználni a rendelkezésre álló szerverkapacitásokat. A mintarendszer rendelkezzen elegendő memória-, számítási és tárkapacitással. Így az oktatók nem a saját (általában alacsony teljesítményű) gépüket használják a komplexebb tanítási folyamatok tesztelésére, hanem egy sokkal gyorsabban eredményt szolgáltató környezetet.

3. Összegzés

A modern vállalatok ügyviteli folyamataik egyre komplexebbek, melyeket egyre összetettebb rendszerekkel lehet támogatni. Természetesen, mint minden más területhez hasonlóan, itt is felmerül az automatizálhatóság kérdése, amely segítségével a bonyolult folyamatok részben egyszerűsíthetők. Az RPA éppen egy ilyen kezdeményezés, amely során a felhasználói aktivitás naplózással és figyélésével felépíthető egy olyan automatizált döntéstámogató rendszer, amely képes a komplex folyamat bizonyos lépéseinek megoldására, vagy akár előrejelzésére. A terület meglehetősen modern volta miatt még nem kiforrott, nem állnak rendelkezésre dobozos termékek a megvalósításra. Jelen publikációban összefoglaltuk a terület legfontosabb követelményeit és jellegzetes problémáit, melyek jó alapot nyújtanak egy ilyen jellegű rendszer megvalósításához.