

Válasz opponensi kérdésekre

DR. SZALAY ZSIGMOND GÁBOR

- 1. Jelen kutatás során kézi erővel készítette el a kiinduló adatbázist, ugyanakkor dolgozatában a mesterséges intelligencia tárgykörébe eső algoritmusokat mutat be, alkalmaz. Adja magát a kérdés, hogy a kiindulási adatbázis létrehozásában milyen automatikus, önjáró alkalmazásokat tud elképzelni, javasolni? Különös tekintettel arra, hogy egyfajta robot auditorként lehet tekinteni a végeredményre. Ebből a szempontból fokozott igény mutatkozna a teljes feladatsor „robot” általi megoldására.**

Válasz: Maga a szöveges adatok automatizált feldolgozása egy ismételt PhD kutatás keretein belül vizsgálható, és számos PhD dolgozat készíthető belőle, mivel legalább annyira komplex és szerteágazó témáról van szó, mint az, ami a dolgozatban bemutatott automatizált eljárásokra vonatkozik. Az alábbi, eljárások érdekes további kutatási alapot szolgálhatnak az adatok automatizált strukturálásához:

Alapvetően szöveges riportokról lévén szó, adja magát, hogy a természetes nyelvi feldolgozás által nyújtott eszköztárból válogasson a kutató, és mivel adat-strukturálásról van szó, ezért a feladat a felügyelet nélküli gépi tanulás kategóriájába tartozik. Egy lehetséges megoldás a szöveges adatok szemantikai feldolgozása (pl. Python NLTK segítségével) és klaszterezése a hasonló auditmegállapításoknak a szemantikai értékelést követően pl. neurális háló alapon önszervező térképekkel. A kialakult klaszterek és egy adott auditjelentéshez tartozó megállapítások két dimenzió mentén egy adattábla (OAM) automatizált feltöltését teszik lehetővé (sor az auditjelentés, oszlop a klasztertagság), mely alapul szolgálhat a dolgozatban bemutatott osztályozó eljárásnak.

Egy másik megoldás lehet a kulcsszavak alapján történő keresés, ahol a dolgozatban használt ISO szabvány egyes kontrollterületeinek kulcs szavait keresi az algoritmus az auditjelentésben, majd statisztikai alapon kategorizálja az egyes megállapításokat. Természetesen ez rendkívüli módon finomhangolható/bonyolítható pl. kereső eljárások alkalmazásával, ahol statisztikai eljárások alkalmazása helyett egy genetikai program keresi meg egy adott kategóriába illeszthető legjobb auditmegállapítást, tehát amelyik legjobban leírja az előre definiált attribútum tulajdonságait. Finomhangolható a klaszterezési eljárás is pl. fuzzy logikával, ahol az egyes auditmegállapítások több kategóriát érintve strukturálhatók. Érezhetően a megoldások tárháza végesen sok.

A legtriviálisabb megoldás természetesen szokásos OLAP-riportokká egyszerűsíti az OAM-képzést: az összes adatot el kell látni ezek keletkezése pillanatában előre eltervezett dimenzionálási leíró meta-adatokkal és ezek alapján pontosan úgy lehet OAM-ot, azaz táblázatot generálni, mint ahogy ma a minőségi statisztikai adatbázisok esetén az már szokás (pl. OECD, ENSZ, EUROSTAT).

- 2. (Az előző kérdést tovább fűzve...) Milyen kritériumok mentén lenne alkalmas, egyúttal milyen más üzleti, szakmai problémakörben lenne használható az Ön által készített algoritmus?**

Válasz: Mind a három célkitűzés mentén fejlesztett modellezési eljárás (hibrid ajánlórendszer/döntéstámogató rendszer, genetikai potenciál kereső eljárás, modell-preferencia levezető eljárás) szándékosan kontextus független megoldások, mely a dolgozatban kijelentésre is kerül és büszke

is erre a tulajdonságra a kutató. Így, ha a dolgozatban felhasznált adatok körét nem tekintjük a megoldások szerves részének, tulajdonképpen bármely olyan problémára illeszthető modellekről van szó, ahol (C1) osztályozási problémáról van szó, (C2) genetikai-potenciál értelmezése lehetséges pl. termőföld kihasználtság optimalizáló, (C3) versengő modell-alternatívák elérhetők. Ha az adatokat is a megoldások részének tekintjük, akkor pl. auditmegállapítások minőségellenőrzésén túl a megoldás kockázat-elemzésekben is felhasználható, ahol a cél a rendszerek sérülékenységének és kitettségének mérése, kockázati besorolása.