

VÁLASZ OPPONENSI VÉLEMÉNYRE

OPPONENS:

Dr. Borosy András

Associate Principal Scientist, DSM-Firmenich, 1242 Satigny,

Switzerland

DOKTORI DOLGOZAT CÍME:

Integrált módszertanok az adatvezérelt szoftver szenzorok fejlesztésére

SZERZŐ:

Hanzelik Pál Péter

Köszönöm Opponensemnek, hogy elvállalta doktori dolgozatom bírálatát, örülök a támogató és elismerő véleményének. A kérdéseire a bírálatban való előfordulásuk sorrendjében adok választ.

1. kérdés:

How would the proposed framework perform in settings where system hierarchies are not well-defined or where data quality is poor?

Hogyan működne a javasolt keretrendszer olyan környezetben, ahol a rendszerhierarchiák nincsenek jól definiálva, vagy ahol az adatminőség gyenge?

Válasz:

A keretrendszer alkalmazható azokban az esetekben is, ahol a rendszerhierarchiák nincsenek definiálva, ezekben az esetekben a data reconciliation (DR) lépés nem alkalmazható, így a

modellek futtatása egymástól elszigetelten működne. Ilyen esetekben egy operátor például tudja könnyedén felügyelni a modellek peridációs eredményét és teljesítményét.

Rendszerhierarchiák nincsenek jól definiálva:

A DR ott alkalmazható, ahol vannak hierarchiák a kovariancia mátrixban, a mérések bizonytalansága figyelembe van véve, és ha az adat minősége egyes esetben gyengébb, akkor azt be lehet építeni a mérleghiba kiegyenlítésbe.

Gyenge adatminőség:

Adatminőség tesztelésére a Proof of Concept (PoC) mini projektek indítása lehet a megoldás, amelyek során feltárjuk a tisztaságát és a minőségét az adatoknak.

Ami az adatok minőségét illeti, külön kategóriákat javaslok:

- Ha hiányosak az adatok, akkor adatgenerálás alkalmazása megfontolandó.
- Ha az adatok nagy szórással állnak rendelkezésre, akkor a bizonytalansági faktort kell beépíteni a modellekbe.

Fontos, hogy mind a rendszerhierarchiák definiálásának hiányát, mind az adatminőség gyengeségének kérdését a fejlesztés során validálni kell.

2. kérdés:

What measures can be taken to ensure the reproducibility and accessibility of the framework for practitioners without advanced machine learning expertise?

Milyen intézkedések tehetők a keretrendszer reprodukálhatóságának és hozzáférhetőségének biztosítására a gépi tanulási tudással nem rendelkező szakemberek számára?

Válasz:

A keretrendszer hosszú távú felhasználásához és reprodukálásához fontos a felhasználók számára a trainingek megszervezése, amik megmutatják a keretrendszer lehetőségét, ha például operátorok, vagy kezdő mérnökök szeretnék elsajátítani az adattudománnyal kapcsolatos ismereteket.

Továbbá kiemelt fontosságú olyan folyamatok és felületek létrehozása, mely lehetőséget kínál interaktív dashboardok, vagy online real-time web alkalmazások használatához, akár „Low-code”, „No-code” környezetek létrehozásával.

Felhasználókkal közösen fejlesztett ellenőrző és kezelő felületek fejlesztése is gyakori eljárás. Például a Web alkalmazás során olyan plusz felületeket készíthetünk, ahol akár a real-time adatok klikkelésekkel történő modell újra tanítása és futtatása végrehajtható. Továbbá kiemelt fontosságú a kulcs felhasználók kurzusokon történő tréningezése annak érdekében, hogy pontosan megismerjék a különböző funkciók használatát, a keretrendszer korlátait. Ha lehetőség van rá, akkor az alkalmazás fejlesztésénél is érdemes bevonnani a felhasználókat és az észrevételeiket beépíteni a fejlesztési fázisokba.

Ezekkel a megoldásokkal könnyíthető és gyorsítható a keretrendszer reprodukálhatósága és hozzáférhetőségének biztosítása a gépi tanulási tudással nem rendelkező szakemberek számára.

3. kérdés:

Have you considered the long-term impact of artificial data generation on model interpretability and robustness in production environments?

Figyelembe vették már a mesterséges adatgenerálás hosszú távú hatását a modell értelmezhetőségére és robusztusságára éles környezetben?

Válasz:

Egyelőre nem, de tervben van, hogy alkalmazva lesz, még hozzá neurális hálózatok esetében, például a közetek savoldhatóságának becslésére használt CNN algoritmus teljesítményének az összehasonlítása a mesterséges spektrumokat nem és a mesterséges spektrumokat tartalmazó spektrumok adatkészletén.

Továbbá tervben van a mesterséges spektrumgenerálás alkalmazása spektrum könyvtár bővítésekre. Például fázisdiagramok esetében kolonna lepárlás esetén csak bizonyos pontokban állnak rendelkezésre mért spektrumok, ezekhez generálnánk a teljes fázisdiagramot lefedő spektrumokat.

4. kérdés:

Are there specific industries or data types where you anticipate the framework would not be applicable or would require significant adaptation?

Vannak-e olyan konkrét iparágak vagy adattípusok, ahol a keretrendszer várhatóan nem lenne alkalmazható, vagy jelentős módosítást igényelne?

Válasz:

A keretrendszer önmagában alkalmazható minden iparágban, ahol fontos szerepe van a minőségbiztosításnak, vagy a laboratóriumi méréseknek és az azokra épülő folyamatoknak. A disszertációban bemutatott FT- MIR- és Raman mérések használatosak gyógyszeriparban, élelmiszeriparban, autóiparban és petrokémiában is. Ezekben az iparágakban nagy jelentőséggel bír a mérleghiba kiegyenlítés és a gépi tanulási algoritmusok alkalmazása is.

Továbbá a mesterséges spektrumok, például az hulladék égető erőműveket segíti abban, hogy pontosan diagnosztizálják az égetendő anyagok halmazát, amik később bekerülnek a kemencébe.

A keretrendszer használatát nehezítheti a fokozott hatósági ellenőrzés, törzskönyvezés például a gyógyszeriparban, vagy az élelmiszeriparban előírt szabványok.

A javasolt módszertan alkalmassá tehető a gyógyszeriparban alkalmazott GMP (Good Manufacturing Practice), GLP (Good Laboratory Practice) és GAMP 5 (Good Automated Manufacturing Practice) irányelvek szerinti működésre. Bár a jelen értekezés fókuszán kívül esett ezen szabályozási követelmények gyakorlati implementációja, a fejlesztett architektúra elvi szinten támogatja az adatintegritás biztosítását és a modellek teljes életciklus-validációját.

Budapest, 2026. április 24.



Hanzelik Pál Péter

doktorjelölt