

Тема: **Издавање знања из великих података добијених као резултат рачунарских симулација**

Наставник: **Вишња Симић**

Издавање знања из великих података данас игра веома важну улогу будући да се у свим сферама подаци непрекидно прикупљају и чувају информације које чекају да буду откривене. Један од извора великих података представљају и рачунарске симулације многобројних сценарија рада различитих реалних система, попут водопривредних система. У оваквом систему могуће је дефинисати бројне сценарије који обухватају различита стања оперативних компоненти као и узрочнике тих стања. Дејство сваког сценарија на систем зависи од много фактора, те се додатно може вршити и Монте-Карло варијација појединих улаза у симулације, чиме се обезбеђује потпуно сагледавање опсега понашања система у датом сценарију, али се уједно добијају резултати чија се количина мери у милијардама записа.

Откривање интерпретабилног знања из масивних података је захтеван поступак који подразумева проналажење потенцијално корисних патерна у подацима. Количина података са којим треба радити је несавладива за човека и једноставну употребу алата за табеларну обраду података, те се отуда мора прибећи делимичној аутоматизацији анализе података.

Студент пре свега треба да направи преглед и систематизацију постојећих комерцијалних алата за визуелизацију, анализу и разумевање података, као и постојећих библиотека отвореног кода за развој оваквих софтверских решења. Након увида у постојећа решења, потребно је дефинисати захтеве за развој прототипова софтверских модула за анализу, визуелизацију и издавање знања из података, којима би се надоместили недостаци постојећих решења нарочито када се има у виду примена на податке добијене симулацијом сценарија водопривредних система.

Потребно је у првом кораку развити прототип помоћног алата који омогућава интуитивнији поглед на велике податке - визуелни приказ података груписаних по одређеном критеријуму, графикони зависности између појединих величина. Анализа резултата симулације се често мора вршити интерактивно, вођена стручњаком. Због тога је интеракција један од кључних аспеката који се мора узети у обзир приликом дизајнирања овог алата. Нарочито задатак тражења непознатих, корисних релација између карактеристика у скупу података и њихово издавање захтева врло флексибилан и интуитиван интерфејс. Даље, треба обезбедити груписање података према различитим критеријумима сличности. Кластеризација се може вршити техникама ненадгледаног машинског учења, али је могуће применити фази кластеризацију. Следећи ниво је формирање предиктивног модела коришћењем техника надгледаног обучавања. Модел би дао могућност да за сваки сценарио који наступи у систему може направити процену његовог дејства на стање система и вероватноћу да ће такав сценарио довести до пада система.

Литература

1. Mohamed, Azlinah, et al. "The state of the art and taxonomy of big data analytics: view from new big data framework." *Artificial Intelligence Review* 53.2 (2020): 989-1037.
2. Cuzzocrea, Alfredo, et al. "An innovative framework for supporting big atmospheric data analytics via clustering-based spatio-temporal analysis." *Journal of Ambient Intelligence and Humanized Computing* 10.9 (2019): 3383-3398.
3. King, L. M., S. P. Simonovic, and D. N. D. Hartford. "Using system dynamics simulation for assessment of hydropower system safety." *Water Resources Research* 53.8 (2017): 7148-7174.