

Feladatterv

Adattömörítési eljárások fejlesztése az inverz kaszkád modell felhasználásával

Jelen dokumentum az Adattömörítési algoritmusok vizsgálata című kutatási projekthez kapcsolódó kutatási feladatterv ismertetése. A dokumentum célja a kutatás hátterének bemutatása, a kutatási feladatok ismertetése, az elvárt eredmények megadása.

A kutatás háttere

A nagyintenzitású, ultrarövid lézerimpulzusok alkalmazása egyre szélesebb körben terjed el a különböző tudományágak között, az orvostudományoktól kezdve a fúziós kísérletekig számos helyen hasznosíthatóak. Az ilyen impulzusokat előállító lézerrendszerek paramétereinek diagnosztikájához igen jelentős adatmennyiség gyors tárolására és feldolgozására van szükség. A kísérletek szempontjából döntő fontosságúak az impulzusról impulzusra változó paraméterek rögzítése, mint például a csúcsintenzitás, nyaláb kétdimenziós térbeli profilja, a spektrum alakja, valamint az impulzus időbeli lefutása. A kísérletek során képi információ formájában nagymennyiségű (akár peta nagyságrendű) adat keletkezik, melynek feldolgozása a kísérleti helytől távol szuperszámítógépes környezetben történhet.

A kutatási projekt célja, hogy milyen új algoritmikus módszerekkel és technológiákkal lehet nagymennyiségű képből álló képsorozatot minél rövidebb idő alatt hálózaton továbbítani. A hálózatok véges sáv szélessége miatt a képeket az adatátvitel előtt tömöríteni kell. Jelen kutatási feladat célja a hatékony adatfeldolgozás kapcsán korábbi kutatások során kifejlesztett úgynevezett inverz kaszkád eljárások adattömörítés céljából történő felhasználásának részletes vizsgálata.

Kutatási feladat

Az adattömörítési problémák esetében központi szerepet játszik a kódoláshoz használt adatmodell kérdése. A diszkrét struktúrák kapcsán az adatok egymáshoz kapcsolódó relációjában a gráf struktúra adja az egyik lehetséges irányvonalat. Jelen kutatási feladatban egy olyan irányvonal feltárása a cél, amely a gráfokon értelmezett úgynevezett diffúziós vagy fertőzési folyamatok vizsgálatára épül. A diffúziós folyamatok esetében minden időpillanatban a csúcspontok egy adott részhalmazra adja a pillanatnyilag fertőzött csúcsok halmazát vagy sztochasztikus esetben ezen csúcsok egy fertőzési valószínűséget kapnak. A

fertőzés terjedése a csúcsok között a kapcsolódó éleken definiált fertőzési valószínűségek által adott. Jelen kutatáshoz az úgynevezett Domingos-Richardson kaszkád fertőzési modellt használjuk fel. Korábbi kutatások során ezen modell prototípus szintjén kifejlesztésre került, melynek adatbányászati felhasználhatóságát is megvizsgáltuk. A módszertanra ismertetése az [1] publikációban megtalálható. Ezen eljáráscsalád kifejlesztése hatékony módon egy összetett feladat, alapvetően csak közelítő, heurisztikus eljárások segítségével valósítható meg. Erre vonatkozó módszertan a [2] publikációban található. Az eljárások nagyobb gráfokon való működéséhez többprocesszoros megoldás kifejlesztése szükséges.

A kaszkád eljáráshoz azonban feltételezhető, hogy ismertek az éleken a fertőzési valószínűségek. A gyakorlatban ez általában nem teljesül, ezért mintaadatbázisok segítségével megvalósított tanuló eljárások alapján közelíthetjük az élfertőzési valószínűségeket. Erre vonatkozó prototípus került kifejlesztésre korábbi kutatások során. A cél ezen algoritmus és szoftver prototípus továbbfejlesztése, amely alkalmas lehet az adattömörítés hatékony megvalósításához. A korábbi megoldás hátránya a jelentős futásidő és a teszteredmények alapján tapasztalt pontosság hiányossága. Jelen kutatás keretében a kifejlesztendő új eljáráscsalád és szoftver prototípus esetében ezt kell javítani, tesztelésekkel igazolni.

Kutatás elvárt eredményei

A kutatás elvárt eredménye a korábban kifejlesztett inverz kaszkád algoritmus adattömörítési célra való adaptálhatóságának vizsgálatához egy szoftver prototípus kifejlesztése, amely többprocesszoros környezetben képes működni, az erőforrások hatékony kihasználásával. Ezen szoftver a korábban kifejlesztett „grid search” és gradiens alapú keresést megvalósító szoftver prototípust kell alapként felhasználnia (a módszertan a [3] publikációban elérhető), ennek továbbfejlesztését kell megvalósítani metaheurisztikus tanulási eljárások alkalmazásával, elsősorban a [4] publikációban ismertetett eljáráscsalád felhasználásával. A szükséges módszertan alapjai az [1], [2] és [3] publikációkban elérhetőek, a kutatás megvalósítása során ezek rendelkezésre fognak állni.

Kutatás szakmai környezete

A kutatást az SZTE JGYPK Informatika Alkalmazásai Tanszékével együttműködve kell megvalósítani. A kifejlesztendő eljárásokhoz rendszeres szakmai konzultáció, a tesztelésben való részvétel és a kutatási infrastruktúra használata biztosított.