

Abstract

A napjainkban tapasztalható robbanásszerű technológiai fejlődés és az egyre növekvő digitalizáció jelentős hatással van a társadalomra és a mindennapi életünkre. A modern világban az információtechnológia, a kommunikáció és az adatelemzés kulcsfontosságú szerepet játszanak az üzleti világban, a tudományos kutatásokban, az egészségügyben és számos más területen. Az adatok hatalmas mennyisége, amelyeket ma generálunk és gyűjtünk új lehetőségeket teremt az adatok elemzése és az összefüggések feltárása.

Ebben a kontextusban a szakdolgozat célja egyrészt, hogy mélyebb betekintést nyújtson a kauzalitásvizsgálati, azaz ok-okozati tényezők feltárásában használatos alapfogalmakba, bemutasson egy erre alkalmas algoritmust, valamint a gépi tanulási módszereket is megpróbálja kihasználni ezen a területen. A szakdolgozat másik főbb tudományos célja egy Taiwani kutatóközpont zebrahalak mozgását speciális környezetben rögzítő adatsorának részletes elemzése, valamint a zebrahalak esetleges viselkedésbeli különbségeinek feltárása.

Az kauzalitásvizsgálat és gépi tanulás területe széles körű és sokoldalú alkalmazási lehetőségeket kínál. Az üzleti világban például a kauzalitásvizsgálati módszerek segíthetnek a gazdasági folyamatok célzott elemzésében és megértésében. Az egészségügyben a betegségek közötti ok-okozati kapcsolat feltárása is egy hasonlóan fontos felhasználási terület. A gépi tanulási módszerekről is számtalan felhasználási területet lehetne felsorolni. A közlekedés területén a gépi tanulás lehetővé teszi a közlekedési minták előrejelzését és a forgalom optimalizálását, valamint a jövő az önvezető, sokkal biztonságosabb autóké, amelyeket gépi tanulási módszerekkel tanítanak. A bármelyikünk zsebében ott lapuló telefon is rengeteg gépi tanulási algoritmust használ, hogy a lehető leghatékonyabb módon könnyítse meg az életünket.

A szakdolgozat a következő módon épül fel. Az első részben áttekintést adunk az kauzalitásvizsgálat alapvető matematikai fogalmairól és egy erre alkalmas algoritmus, az NSRPS alapú kompressziós ráta módszer bemutatásával betekintést nyújtunk egy lehetséges kauzalitás-elemző módszer részleteibe. A második fő részben részletesen ismertetjük az LSTM neurális hálózatok elméletét és működését. Az utolsó részben pedig bemutatjuk az elért eredményeket és részletesen kielemezzük a halak viselkedését információelméleti (transzfer-entrópia és időkélesltetéses közös információ), valamint a korábban bemutatott gépi tanulási és kauzalitásvizsgálati módszerekkel.