12/6/2024

Dukát Nándor, YZV8QV

Gaszner Ádám, BIA823

Házi Feladat Dokumentáció

Karakterfelismerés

**Bevezetés**

Ez a dokumentáció egy mélytanulási osztályozó modell implementációját, tanítását, valamint a kapott eredmények értékelését ismerteti. Az osztályozási feladat során a karakterfelismerés problémájával foglalkoztunk, amelyhez kisméretű fekete-fehér képekből álló adathalmazt használtunk. A képekhez karakterazonosítók tartoznak, és a cél az ismeretlen tesztképek karakterének (0-9, a-z, A-Z) felismerése manuális beavatkozás nélkül, mélytanulási osztályozó segítségével. A megoldás során nagy hangsúlyt fektettünk az eredményességre és a modell teljesítményének objektív értékelésére.

A projekt elején egyszerűbb osztályozó algoritmusokkal kezdtünk, például a RandomForestClassifier használatával, amely előtt dimenziócsökkentést végeztünk az adatokon. Ez lehetővé tette a gyorsabb kezdeti kísérletezést, miközben alapvető működési tapasztalatokat szereztünk. Ahogy haladtunk előre a tananyagban, és mélyebb megértést szereztünk a mélytanulásról, áttértünk bonyolultabb modellek, például neurális hálózatok implementálására, hogy tovább javítsuk a modell teljesítményét.

A képek preprocesszálásához a cv2 könyvtárat használtuk, különös tekintettel az élkiemelési technikákra, amelyek segítették a modell számára releváns jellemzők kiemelését. Az adathalmaz feldolgozásakor a képeket az eredeti méretükről 32x32 pixeles méretre csökkentettük, mivel ez kiegyensúlyozta a számítási teljesítményt és a modell hatékonyságát. Megpróbáltuk nagyobb méretben is feldolgozni a képeket, azonban a számítási idő exponenciálisan megnőtt, és nem értünk el jobb eredményeket.

Az osztályozó modellünk mély neurális hálózatként több konvolúciós rérteget alkalmat, amelyet magas számú epoch mellett tanítottunk, miközben early stopping és dropout mechanizmust használtunk a túlilleszkedés elkerülése érdekében.

A projekt lépései közé tartozott a modell megtervezése és implementálása, a tanulási folyamat dokumentálása, a tesztadathalmazra történő predikciók előállítása. A dokumentáció kitér a tervezési döntésekre, amelyek meghatározták a végső megoldást, és részletezi a résztvevők közötti munkamegosztást.

**Feladat és adathalmaz ismertetése**

Az adathalmaz kisméretű fekete-fehér képekből áll, amelyek az angol ábécé kis- és nagybetűit, valamint a számjegyeket (0-9) reprezentálják. Minden karakterhez egy dedikált mappa tartozik, amelyben 900 tanítóelem található. Az összes képet egységesen 32x32 pixel méretre skáláztuk, majd a megfelelő címkékkel láttuk el a további feldolgozás érdekében.

A tanító adatok összeállításakor megfigyeltük, hogy a képek minden karakter esetében széles körű varianciát mutatnak, így biztosítva a modell számára a megfelelő generalizációs képességet. Ez lehetővé tette, hogy a különböző típusú bemeneti karakterek felismerése hatékonyan megtörténjen.

Az adathalmaz kizárólag fekete-fehér színskálát használ, amely jelentősen egyszerűsíti a preprocesszálási folyamatokat. Az adatok előfeldolgozásához a cv2 könyvtárat alkalmaztuk, különös hangsúlyt fektetve az élkiemelés technikáira. Ez az eljárás kiemelte a képek jellegzetességeit, ami a modell tanulási folyamatának hatékonyságát nagymértékben növelte.

Az adathalmaz megfelelően kiegyensúlyozott, az osztályok címkéinek eloszlása egyenletes, így az osztályozó modell tanítása során nem lépett fel osztálytorzítás. Az adatokat tanulási és tesztelési részhalmazokra osztottuk, biztosítva ezzel a modell fejlesztésének és objektív kiértékelésének alapfeltételeit.

**Megvalósítás lépései**