

Abstrakt: Rozpoznávání cover verzí písní je oblast problematiky získávání informací z hudby, která se zabývá úkolem rozpoznat, zda dvě odlišné audio nahrávky obsahují různé verze téže písně. Jelikož cover verze se mohou lišit v tempu, tónině, instrumentaci a dalších vlastnostech, bylo během uplynulých let vytvořeno mnoho důmyslných příznaků vhodných pro tento účel. Provedli jsme důkladnou analýzu 32 příznaků použitých v pracích zabývajících se touto problematikou, přičemž rozlišujeme příznaky přesné a škálovatelné. Přesné příznaky jsou založeny na sekvencích harmonických deskriptorů (typicky jsou to tzv. „chroma“ vektory) a vedou k lepším výsledkům, avšak za cenu vyšší výpočetní náročnosti. Škálovatelné příznaky mají malou konstantní velikost a zachycují pouze obecné rysy dané audio nahrávky, díky čemuž je jejich výpočet rychlý a hodí se tak pro použití s velkými data-sety. Vybrali jsme 7 škálovatelných a 3 přesné příznaky, které jsme použili pro konstrukci našeho dvouúrovňového systému pro rozpoznávání cover verzí, přičemž škálovatelné příznaky jsou použity na první úrovni k prořezání data-setu a přesné na druhé úrovni pro zpřesnění výsledků. Dva různé modely strojového učení jsou použity pro zkombinování škálovatelných resp. přesných příznaků.

Pro analýzu příznaků a ohodnocení našeho systému jsme použili Million Song Dataset. Experimenty ukázaly, že škálovatelné příznaky dosahují lepších výsledků než ty přesné, což nás vedlo k rozhodnutí použít v našem systému pouze těchto 7 škálovatelných příznaků. Výsledky takto upraveného systému jsou srovnatelné s nejvýkonnějšími metodami testovanými na tomto data-setu. Následuje analýza překvapivě špatných výsledků přesných příznaků, na jejímž základě jsme dospěli k závěru, že na vině je pravděpodobně nižší kvalita použitého datasetu a že škálovatelné příznaky vykazují pozoruhodnou robustnost vůči nízké kvalitě dat.