



DOCTORATS
INDUSTRIALS



EL PLA DE
DOCTORATS
INDUSTRIALS

PROJECTE DE DOCTORAT INDUSTRIAL EXPEDIENT 2016 DI 046

DADES DE L'EMPRESA I DE L'ENTORN ACADÈMIC

Títol del projecte

Music/Speech Detection in Broadcast Media Programs

Empresa

BMAT Licensing S.L.U

Responsable de l'empresa

Emilio Molina Martínez

Universitat

Universitat Pompeu Fabra

Director/a de tesi

Emilia Gómez Gutiérrez

Treballador/a de l'empresa i doctorand/a

Blai Meléndez Catalán

BREU DESCRIPCIÓ DEL PROJECTE DE RECERCA

La classificació de segments entre música o veu parlada en senyals d'àudio és una tasca del camp de Music Information Retrieval per a la que existeixen diversos algorismes que obtenen resultats que freguen el 90 % de precisió segons el Music Information Retrieval Evaluation eXchange (MIREX), en bases de dades controlades i generades en entorns de laboratori. En canvi, a BMAT s'han dut a terme experiments preliminars en entorns industrials on s'ha detectat que els mètodes actuals presenten debilitats quan els apliquem a casos de televisió i ràdio en que poden aparèixer una gran diversitat de sons difícils d'englobar en una d'aquestes dues classes i que, de fet, fan que la tasca pugui arribar a ser difícil fins i tot per un humà.

La nostra hipòtesi és que en aquest escenari és necessari un procés previ de reconeixement d'esdeveniments sonors que inclogui un ventall de classes més ampli que faciliti, en segona instància, la segmentació de l'àudio en blocs de música o veu parlada. La principal contribució del projecte serà el desenvolupament d'un algoritme propi capaç de dur a terme aquesta tasca amb una precisió més elevada que la dels algorismes actuals. A més, es proposaran estratègies de anotació manual i de generació de bases de dades anotades manualment per l'entrenament i avaluació.



Generalitat de Catalunya
Departament d'Empresa i Coneixement
Secretaria d'Universitats i Recerca



Agència
de Gestió
d'Ajuts
Universitaris
i de Recerca



EL PLA DE DOCTORATS INDUSTRIALS

Caldrà, per tant, estar familiaritzat amb l'estat de l'art dels mètodes de classificació de senyals sonors per poder-ne extreure idees i identificar els punts on es pugui contribuir durant el doctorat.

Ara per ara, un dels camps que ofereix major potencial és el de fer servir eines avançades d'aprenentatge automàtic (machine learning). Concretament, els mètodes que utilitzen aprenentatge profund (deep learning) estan obrint molts camins que val la pena explorar com demostren el resultats que s'estan obtenint en altres tasques i àmbits de recerca.

Un pas molt important per a aquest objectiu, i que en si es pot considerar una segona contribució, és la creació d'un dataset públic aprofitant que BMAT té accés a una gran quantitat de dades d'àudio de televisió i ràdio. Per a l'anotació d'aquest dataset es desenvoluparà una eina eficient que proporcioni una gran precisió, i s'utilitzarà una metodologia que en permeti el control i n'asseguri la qualitat incloent, per exemple, mètodes com el crowd sourcing.

