

Exercise 9: Sound and music description

With this exercise you will learn to describe sounds with simple machine learning methods. You will learn to use the Freesound API to load pre-computed sound descriptors from Freesound and to perform sound clustering and classification with them. You will work with instrumental sounds, thus learning what audio features are useful for characterizing them.

There are four parts in this exercise: 1) Download sounds and descriptors from Freesound, 2) Select two descriptors for a good sound clustering, 3) Cluster sounds using k-means, and 4) Classify sounds using k-NN.

We provide the code for each task, thus no need to get involved in much programming.

Relevant Concepts

Freesound API

With the Freesound API you can browse, search, and retrieve information from Freesound, such as automatically extracted features from audio files. You can also perform advanced queries combining content analysis features and other metadata (tags, etc...). With the API you can do text searches similar to what you can do from the advanced searches in the website <http://freesound.org/search/>, but implementing the queries in software. If you are interested in knowing more about the Freesound API, you can see examples of using it with python in: <https://github.com/MTG/freesound-python/blob/master/examples.py> and you can read the API documentation, <http://www.freesound.org/docs/api/>

Sound descriptors

In this exercise, you will use sound descriptors that have been pre-computed with Essentia, <https://essentia.upf.edu> and are stored in the Freesound database together with the corresponding sounds. Many sound descriptors can be extracted using Essentia (http://essentia.upf.edu/documentation/algorithms_reference.html) and in Freesound, there is specific information of the descriptors available in Freesound, https://freesound.org/docs/api/analysis_index.html.

Euclidean distance

The Euclidean distance is the straight-line distance between two points in an n-dimensional space, thus the distance between points p and q is the length of the line segment connecting them. If $p = (p_1, p_2, \dots, p_n)$ and $q = (q_1, q_2, \dots, q_n)$ are two points in Euclidean n-space, then the distance, d , from p to q , or from q to p is given by the Pythagorean formula:

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

K-means clustering (k-means)

K-means clustering is a method of vector quantization that is popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. The problem is computationally difficult (NP-hard), however, efficient heuristic algorithms converge quickly to a local optimum.

Given a set of observations $\{x_1, x_2, \dots, x_n\}$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k ($\leq n$) sets $S = S_1, S_2, \dots, S_k$ so as to minimize the within-cluster sum of squares (WCSS) (i.e. variance). Formally, the objective is to find:

$$\arg \min_{S} \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg \min_{S} \sum_{i=1}^k |S_i| \text{Var } S_i, \text{ where } \mu_i \text{ is the mean of points in } S_i.$$

K-nearest neighbours classifier (k-NN)

K-nearest neighbours classification (k-NN) is a non-parametric method in which the input consists of the k closest training examples in the feature space. The output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor.

Part 1: Download sounds and descriptors from Freesound

Download a collection of instrumental sounds and their descriptors from Freesound using the Freesound API.

First get a Freesound API key from <http://www.freesound.org/api2/apply/> and create a directory in the workspace with the name `testDownload` to store the sounds and descriptors. You also need to install the pyjnius client for the freesound API. Do this by cloning the repository <https://github.com/MTG/freesound-python>, into a folder and installing it following the instructions. You will have to import the freesound module in the code.

You will be calling the function `download_sounds_freesound()` whose input parameters are:

```
1. queryText (string): A single word or a string of words without spaces (use hyphens), typically the name of the instrument. e.g. (e.g. "violin", "trumpet", "cello", "bassoon", etc).
2. tag (string): tag to be used for filtering the searched sounds (e.g., "multisample", "single-note", "velocity", "tenuto", etc).
3. duration (2 floating point numbers): min and max duration (seconds) of the sound to filter, e.g., (0,2).
4. APIKey (string): your API key.
5. outputDir (string): path to the directory where you want to store the sounds and their descriptors.
6. topResults (integer): number of results (sounds) that you want to download.
7. featureExt (file extension): file extension for storing sound descriptor (.json, typically).
```

In the call to `download_sounds_freesound()` you will have to choose the appropriate queryText, tag, and duration, to return single notes-strokes of instrumental sounds. The first twenty results of the query should be "good". Note that the tag can be empty. Example of a query to obtain single notes of violin could be:

```
download_sounds_freesound(queryText="violin", APIKey="your key", outputPath="testDownload", topResults=20, featureExt=".json")
```

This returns 20 single notes of violin sounds and the script stores them in the `testDownload` directory (the duration has to be created beforehand).

Download and store twenty sound examples of each instrument using the `download_sounds_freesound()` function given below. The examples need to be representative of the instrument and coherent, single notes (melodic instruments) or single strokes (percussion instruments), and shorter than 10 seconds. Specify a good query text, tag, and duration to query for the chosen instruments. Refine your search parameters until you get twenty adequate samples for each instrument. Select three instruments to be used out of this set: violin, guitar, bassoon, trumpet, clarinet, cello, naobo (cymbals used in Beijing Opera). Listen to the sounds downloaded and look at the descriptor json files.

Before using the API to download the sounds, we recommend to do the same query using the Freesound website and checking that the top 20 results are good.

```
In [1]:  
import os, sys  
import json  
import freesound as fs  
  
descriptors = [ 'lowlevel.spectral_centroid.mean',  
    'lowlevel.spectral_contrast.mean',  
    'lowlevel.dissonance.mean',  
    'lowlevel.hfc.mean',  
    'lowlevel.ihc.mean',  
    'sfz.longattackmean',  
    'sfz.longattacktime.mean',  
    'sfz.inharmonicity.mean']  
  
In [2]: def download_sounds_freesound(queryText = "", tag=None, duration=None, APIKey = "", outputPath = "", topResults = 5, featureExt = '.json'):  
    """  
    This function downloads sounds and their descriptors from freesound using the queryText and the tag specified in the input. Additionally, you can also specify the duration range to filter sounds based on duration.  
  
    Inputs:  
        (Input parameters marked with a * are optional)  
        queryText (string): query text for the sounds (eg. "violin", "trumpet", "cello", "bassoon" etc.)  
        tag* (string): tag used for filtering the searched sounds. (eg. "multisample", "single-note", "velocity", "tenuto", etc.)  
        duration* (tuple): min and the max duration (seconds) of the sound to filter, eg. (0,2,15)  
        APIKey* (string): your api key, which you can obtain from : www.freesound.org/api2/apply/  
        outputPath (string): path to the directory where you want to store the sounds and their descriptors  
        topResults (integer): number of results(sounds) that you want to download  
        featureExt (string): file extension for storing sound descriptors  
    Output:  
        This function downloads sounds and descriptors, and then stores them in outputPath. In this directory outputPath it creates a directory of the same name as that of the queryText. In this directory the directory as the sound id, contains all the sound files and links to the sound and freesound links for all the downloaded sounds in the outputPath.  
        NOTE: If the directory outputPath/queryText exists, it deletes the existing contents and stores new sounds from the current query.  
    ....  
  
    # Checking for the compulsory input parameters  
    if queryText == "":  
        print("\n")  
        print("Provide a query text to search for sounds")  
        return -1  
  
    if APIKey == "":  
        print("\n")  
        print("You need a valid freesound API key to be able to download sounds.")  
        print("Please apply for one here: www.freesound.org/api2/apply/")  
        print("\n")  
        return -1  
  
    if outputPath == "" or not os.path.exists(outputPath):  
        print("\n")  
        print("Please provide a valid output directory. This will be the root directory for storing sounds and descriptors")  
        return -1  
  
    # Setting up the Freesound client and the authentication key  
    fsCln = fs.FreesoundClient()  
    fsCln.set_token(APIKey,"token")  
  
    # Creating a duration filter string that the Freesound API understands  
    if duration and type(duration) == tuple:  
        flt_dur = "duration:" + str(duration[0]) + " TO " + str(duration[1]) + "]"  
    else:  
        flt_dur = ""  
  
    if tag and type(tag) == str:  
        flt_tag = "tag:" + tag  
    else:  
        flt_tag = ""  
  
    # Querying Freesound  
    page_size = 30  
    if not flt_tag + flt_dur == "":  
        qRes = fsCln.text_search(query=queryText ,filter = flt_dur + flt_tag,sort="score", fields="id,name,previews,username,url,analysis", descriptors="").join(descriptors), page_size=page_size  
    else:  
        qRes = fsCln.text_search(query=queryText ,sort="score",fields="id,name,previews,username,url,analysis", descriptors="").join(descriptors), page_size=page_size  
  
    outputPath = os.path.join(outputPath, queryText)  
    if os.path.exists(outputPath): # If the directory exists, it deletes it and starts fresh  
        os.system("rm -r " + outputPath)  
    os.makedirs(outputPath)  
  
    pageNo = 1  
    sndCnt = 0  
    indCnt = 0  
    totalSnds = min(qRes.count,200) # System quits after trying to download after 200 times  
  
    # Creating directories to store output and downloading sounds and their descriptors  
    downloadedSounds = []  
    while 1:  
        if indCnt >= totalSnds:  
            print("Not able to download required number of sounds. Either there are not enough search results on freesound for your search query and filtering")  
            break  
        sound = qRes[sndCnt - (pageNo-1)*page_size]  
        print("Resound-id: " + str(sound.id))  
        outDir = os.path.join(outputPath, queryText, str(sound.id))  
        if os.path.exists(outDir):  
            os.system("rm -r " + outDir)  
        os.makedirs(outDir)  
  
        pageNo += 1  
        sndCnt += 1  
        indCnt += 1  
        totalSnds = min(qRes.count,200) # System quits after trying to download after 200 times  
  
    # Creating directories to store output and downloading sounds and their descriptors  
    downloadedSounds = [  
    while 1:  
        if indCnt >= totalSnds:  
            print("Not able to download required number of sounds. Either there are not enough search results on freesound for your search query and filtering")  
            break  
        sound = qRes[sndCnt - (pageNo-1)*page_size]  
        print("Resound-id: " + str(sound.id))  
        outDir = os.path.join(outputPath, queryText, str(sound.id))  
        if os.path.exists(outDir):  
            os.system("rm -r " + outDir)  
        os.makedirs(outDir)  
  
        pageNo += 1  
        sndCnt += 1  
        indCnt += 1  
        totalSnds = min(qRes.count,200) # System quits after trying to download after 200 times  
  
    # Creating the list of files and Freesound Links  
    # Dumps the list of files and Freesound_Links  
    f = open(os.path.join(outputPath, "SoundList.txt"), 'w')  
    for elem in downloadedSounds:  
        fid.write(elem['name'] + "\n")  
    fid.close()  
  
In [1]: # 1.1: call download_sounds_freesound for 3 instruments with parameters to obtain adequate sounds  
## your code here  
  
### explain the coherence of the sound collections obtained  
....  
  
....
```

Part 2: Select two descriptors for a good sound clustering

Select two of the sound descriptors obtained from Task 1 in order to obtain a good clustering of the sounds of the three instruments in a two dimensional space. By visualizing the descriptor values of the sounds in a 2D plot you can choose the features that can help to better cluster these instruments.

You take as inputs the downloaded sounds folder (`targetDir`) and the descriptor pair indices (`descInput`) (see mapping) to create a 2-D scatter plot of the descriptor pair. The data points, sounds, from different instruments are shown with different colors. In addition, you can also plot the Freesound ID of the sounds with the points. Only plot the sounds of the 3 instruments chosen. Make sure that in `targetDir` you only have the 3 instruments chosen.

Choose a good pair of descriptors for the sounds of the 3 instruments you downloaded in Part 1. A good pair of descriptors leads to a point distribution where all the sounds of an instrument cluster together, with a good separation from the other instrument clusters. Try out different combinations of descriptor pairs. Write a short paragraph on the descriptor pairs you tried out, justifying your choices for selecting those particular descriptors. Based on the spectral and temporal features of the instruments and sounds, give an explanation of why (or why not) a good clustering is (or is not) achieved with the chosen pairs of descriptors.

From the code given you can generate a 2-D scatter plot of all sounds for the chosen descriptor pairs.

```
In [ ]:  
import os, sys  
import numpy as np  
import matplotlib.pyplot as plt  
from matplotlib.lines import Line2D  
from scipy.cluster.vq import vq, kmeans, whiten  
  
# Mapping of descriptors  
descriptorMapping = { 0: 'lowlevel.spectral_centroid.mean',  
    1: 'lowlevel.dissonance.mean',  
    2: 'lowlevel.hfc.mean',  
    3: 'sfz.longattackmean',  
    4: 'sfz.inharmonicity.mean',  
    5: 'lowlevel.spectral_contrast.mean',  
    6: 'lowlevel.spectral_contrast.mean',  
    7: 'lowlevel.spectral_contrast.mean',  
    8: 'lowlevel.spectral_contrast.mean',  
    9: 'lowlevel.spectral_contrast.mean',  
    10: 'lowlevel.spectral_contrast.mean',  
    11: 'lowlevel.hfrz.mean',  
    12: 'lowlevel.mfcc.mean',  
    13: 'lowlevel.mfcc.mean',  
    14: 'lowlevel.mfcc.mean',  
    15: 'lowlevel.mfcc.mean',  
    16: 'lowlevel.mfcc.mean'  
    }  
  
In [ ]:  
def convFtrDict2List(ftrDict):  
    """  
    This function converts descriptor dictionary to an np.array. The order in the numpy array (indices) are same as those mentioned in descriptorMapping dictionary.  
    """  
  
    Input:  
        ftrDict (dict): dictionary containing descriptors downloaded from the freesound  
    Output:  
        ftr (np.ndarray): Numpy array containing the descriptors for processing later on  
    ....  
    ftr = []  
    for key in range(len(descriptorMapping.keys())):  
        try:  
            ftrName, ind = key.split('.')[0].split(':')[1][:-1], int(descriptorMapping[key].split('.')[1][:-1])  
            ftr.append(ftrDict[ftrName][0][ind])  
        except:  
            ftr.append(ftrDict[key][0])  
    return np.array(ftr)  
  
def fetchDataDetails(inputDir, descExt = '.json'):  
    """  
    This function is used by other functions to obtain the information regarding the directory structure and the location of descriptor files for each sound  
    """  
    dataDetails = {}  
    for path, dname, fnames in os.walk(inputDir):  
        for fname in fnames:  
            if descExt in fname:  
                remPath, rname, sname = path.split('/')[1:-3], path.split('/')[-3], path.split('/')[-2]  
                if fname.endswith(descExt):  
                    fDict = json.load(open(os.path.join('/',remPath,rname,sname,fname), 'r'))  
                    dataDetails[sname] = {fname: fDict}  
    return dataDetails  
  
# 2.2: Select the descriptors to plot of the three instruments chosen  
inputDir = "testDownload/"  
  
## this is the main line to modify, select two descriptors, change the XX by a number from 0 to 16  
descInput = (XX,XX)  
  
# no need to change the code from here  
....  
  
....
```

Part 3: Cluster sounds using k-means

After visualizing the sound descriptors, you will now cluster the sounds using more than two descriptors. You can use as many descriptors as you need for the best clustering, the same set of sounds obtain in Task 1, starting from the descriptors that you found were good in Part 2, and then adding other descriptors that you feel can improve the k-means clustering of sounds. The function `cluster_sounds()` (see mapping) to create a 2-D scatter plot of the descriptor pair. The data points, sounds, from different instruments are shown with different colors.

You take as inputs the downloaded sounds folder (`targetDir`) and the descriptor pair indices (`descInput`) (see mapping) to create a 2-D scatter plot of the descriptor pair. The data points, sounds, from different instruments are shown with different colors. In addition, you can also plot the Freesound ID of the sounds with the points. Only plot the sounds of the 3 instruments chosen. Make sure that in `targetDir` you only have the 3 instruments chosen.

For this part, you can use as many descriptors as you need to achieve good clustering and classification performance. However it is best to use as few descriptors as possible in order to make it easier to explain the contribution of each descriptor. Choose the number of clusters to be the same as the number of instruments (i.e., 3). Ideally in such a case, all the sounds of an instrument cluster together, with a good separation from the other instrument clusters. Try out different combinations of descriptor pairs. Write a short paragraph on the descriptor pairs you tried out, justifying your choices for selecting those particular descriptors. Based on the spectral and temporal features of the instruments and sounds, give an explanation of why (or why not) a good clustering is (or is not) achieved with the chosen pairs of descriptors.

From the code given you can generate a 2-D scatter plot of all sounds for the chosen descriptor pairs.

```
In [ ]:  
import os, sys  
import numpy as np  
import matplotlib.pyplot as plt  
from matplotlib.lines import Line2D  
from scipy.cluster.vq import vq, kmeans, whiten  
  
# Mapping of descriptors  
descriptorMapping = { 0: 'lowlevel.spectral_centroid.mean',  
    1: 'lowlevel.dissonance.mean',  
    2: 'lowlevel.hfc.mean',  
    3: 'sfz.longattackmean',  
    4: 'sfz.inharmonicity.mean',  
    5: 'lowlevel.spectral_contrast.mean',  
    6: 'lowlevel.spectral_contrast.mean',  
    7: 'lowlevel.spectral_contrast.mean',  
    8: 'lowlevel.spectral_contrast.mean',  
    9: 'lowlevel.spectral_contrast.mean',  
    10: 'lowlevel.spectral_contrast.mean',  
    11: 'lowlevel.hfrz.mean',  
    12: 'lowlevel.mfcc.mean',  
    13: 'lowlevel.mfcc.mean',  
    14: 'lowlevel.mfcc.mean',  
    15: 'lowlevel.mfcc.mean',  
    16: 'lowlevel.mfcc.mean'  
    }  
  
In [ ]:  
def cluster_sounds(targetDir, nCluster = -1, descInput=[]):  
    """  
    This function clusters all the sounds in targetDir using kmeans clustering.  
    Input:  
        targetDir (string): Directory where sound descriptors are stored (all the sounds in this directory)  
        nCluster (int): Number of clusters to be used for kmeans clustering.  
        descInput (list): list of indices of the descriptors to be used for similarity/distance computation  
    Output:  
        Prints the class of each cluster (computed by a majority vote), number of sounds in each cluster and information (sound_id, sound-class and classification decision) of the sounds in each cluster.  
        descInput (list): list of indices of the descriptors assigned to each cluster.  
    ....  
    dataDetails = fetchDataDetails(targetDir, descExt = '.json')  
  
    This function is used by other functions to obtain the information regarding the directory structure and the location of descriptor files for each sound  
    """  
    dataDetails = {}  
    for path, dname, fnames in os.walk(targetDir):  
        for fname in fnames:  
            if descExt in fname:  
                remPath, rname, sname = path.split('/')[1:-3], path.split('/')[-3], path.split('/')[-2]  
                if fname.endswith(descExt):  
                    fDict = json.load(open(os.path.join('/',remPath,rname,sname,fname), 'r'))  
                    dataDetails[sname] = {fname: fDict}  
    return dataDetails  
  
# 2.2: Select the descriptors to plot of the three instruments chosen  
inputDir = "testDownload/"  
  
## this is the main line to modify, select two descriptors, change the XX by a number from 0 to 16  
descInput = (XX,XX)  
  
# no need to change the code from here  
....  
  
....
```

Explain the results

Assign a sound different from the sounds of the 3 instruments chosen to one of the 3 instrumental classes you chose in Part 1, using the k-nearest neighbours classifier (k-NN).

Given a new sound (query sound) and its descriptors, use the function `classify_sound_kNN()` for doing a k-NN classification. It uses a distance measure based on Euclidian distance which is implemented in `compute_similar_sounds()`. `classify_sound_kNN()` returns the instrument class that the query sound is classified into.

The goal of the exercise is to experiment with the k-NN classifier and be able to understand the result by being able to explain why a particular query sound, that is not from any of the defined classes, is actually classified to one of those.

4.1 Get query sounds from Freesound. To get query sounds and their descriptors, you can use `download_sounds_freesound()` function using different query texts (as you did in Part 1). Get sounds that are not from the 3 instruments you chose in Part 1, or at least that is none of the sounds you used to define the classes. If you use a sound from one of the three instruments make sure that in `targetDir` you only have the 3 instruments chosen.

4.2 Perform 5 classifications. You can use as many descriptors as you need (the fewer you use, the easier it will be, to explain the result). k is usually chosen to be an odd positive integer. Try out with different query sounds, different subsets of descriptors, and different values of k . Explain the reason for choosing the descriptors you used and the value of k you selected. Include cases where you think the classification is incorrect, and cases with a query sound of an instrument different from the starting classes. Based on the spectral and temporal features of the instruments and sounds, give an explanation of why (or why not) a good classification is (or is not) achieved with a sound from the chosen instrument class.</p