Propuesta de Proyecto Final: Estimación de nota musical

Saul Ivan Rivas Vega

Aprendizaje Automatizado

3 de marzo de 2020

1. Descripción y delimitación del problema

Basado en atributos en el dominio de la frecuencia como son los coeficientes por ventana de muestreo de la transformada **constante-Q** estimar la nota musical de guitarra presente en un archivo de audio de una pieza musical monofónica con notas simples.

2. Objetivos

- Extraer los atributos frecuenciales por ventana del conjunto de datos.
- Entrenar un modelo predictivo para estimar la nota musical presente en una ventana dados sus atributos frecuenciales.
- Realizar un transcript de alguna melodía pre-grabada.

3. Justificación

Se han realizado múltiples estudios en la estimación de frecuencia fundamental o de tono [1–4]. En aplicaciones el ser en tiempo real como en [1] sería de gran utilidad, sin embargo los métodos con mayor precisión son los que analizan archivos pre-grabados como es el caso de [2,3], ambos superando al método estándar YIN [4] que se encuentra en múltiples bibliotecas de extracción de información musical como ESSENTIA [5].
El presente trabajo tiene como justificación el poder utilizar las propiedades fundamentales para la estimación del tono como en [4, 6, 7] y a su vez beneficiarse de los métodos de aprendizaje automatizado para ofrecer una opción para la estimación de notas musicales balanceando la eficiencia en la cantidad de atributos requeridos y la precisión de la estimación, lo cual podría dejar una base para un sistema posterior para análisis en tiempo real.

4. Base de datos a utilizar o estrategia para recopilarla

La base de datos es NSynth del grupo de investigación musical **Magenta** en Google [8], el dataset contiene 305,979 notas musicales, cada una con un distinto tono, timbre, y envoltura, obtenidos de 1,006 instrumentos grabando clips de monofónicos con una taza de muestreo de 16kHz de 4 segundos con anotaciones de nota musical en el rango del formato MIDI (21-108) con 5 velocidades (25, 50, 75, 100, 127). La nota se mantuvo por 3 segundos dejando el último segundo para el decaimiento.

5. Análisis exploratorio de los datos

Para obtener un subconjunto del dataset se seleccionaron los clips de audio que cumplieran:

- Su nota estuviera entre Do1 (C1) y Do4 (C4), de esta forma la resolución frecuencial podrá extraer valores de los armónicos, por ejemplo Do8 (C8) esta al borde y todos sus armónicos posteriores se pierden en nuestra resolución.
- Fueran de guitarra acústica, esto permitiría pruebas en entornos reales mas fácilmente.
- Su velocidad fuera de 127, es decir que la nota se tocó con la mayor potencia posible.

Lo cual da un total de 814 clips de audio, provenientes de 22 instrumentos tocando cada uno las 37 notas del rango a velocidad 127.

De los clips de audio seleccionados se obtuvieron los valores de la transformada **Constante-Q** utilizando la biblioteca Librosa [9] para python.

Como el ejemplo de la figura 1, en el cual se muestran los componentes frecuenciales de un clip de audio correspondiente a la nota Do4 (C4) de una guitarra acústica.

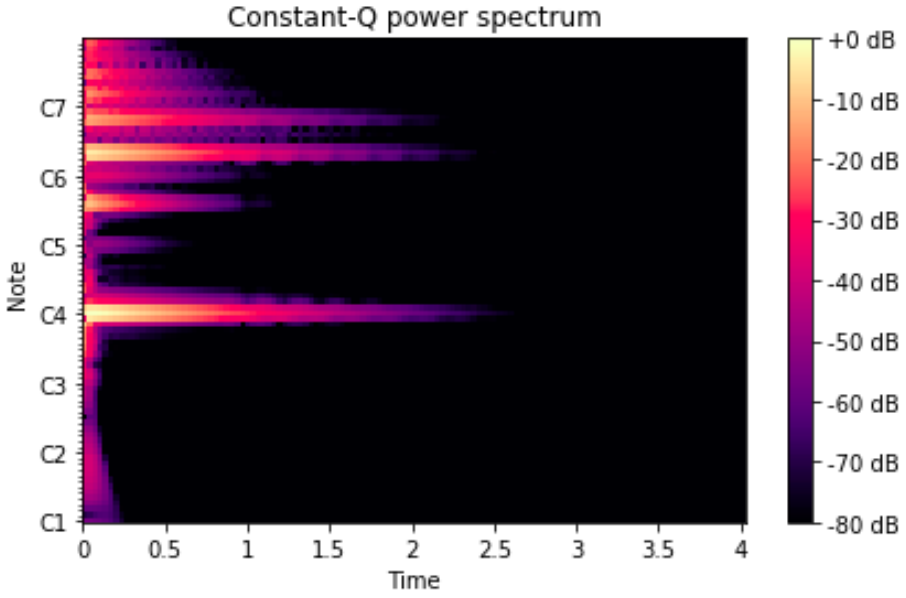

Figura 1: Transformada **Constante-Q**.

Se organizaron un Data Frame los atributos frecuenciales de las primeras 4 ventanas de muestreo (aproximadamente 1 segundo), añadiendo al final la clase a la que pertenece (la nota MIDI).

El Data Frame terminó con 3256 registros 4 por cada clip, con sus respectivas componentes frecuenciales de la transformada las cuales van de la nota 24 (Do1, C1) hasta la nota 107 (Do8, C8) y finalmente mostrando en la última columna su clase, que en este caso es la nota en el rango MIDI de 24 a 60 (Do1-Do4, C1-C4).

	segment_name	note_24	note_25	note_26	note_27	note_28	note_29
0	guitar_acoustic_001-042-127_seg_0	0.1548396758272465	0.21162235488528283	0.27339401081193865	0.33918030901241114	0.4105982238979489	0.4860930198251672
1	guitar_acoustic_001-042-127_seg_1	0.14865352420076297	0.20248682732713844	0.25893947386066524	0.31955088081352306	0.3837413824757683	0.4510814140314256
2	guitar_acoustic_001-042-127_seg_2	0.13246161845622467	0.17783167815307035	0.22213563008240203	0.2694108793899601	0.3167880061470374	0.36207314734165963
3	guitar_acoustic_001-042-127_seg_3	0.10867733895496602	0.14073855298161575	0.17048595547463752	0.19929073737870456	0.22416698824840192	0.24261158137301447
4	guitar_acoustic_004-033-127_seg_0	0.2665652028886773	0.43013401153111186	0.5818432409547265	0.699358685484074	0.7593690342389483	0.7292998568142924
...	...	...	...	...	...	...	...
3251	guitar_acoustic_026-060-127_seg_3	0.3128918375472502	0.31833682526523954	0.31756775407281007	0.3135946669912808	0.3042061507873016	0.28850208200586736
3252	guitar_acoustic_002-042-127_seg_0	0.45726032931360555	0.4228486327314009	0.30577141242939543	0.1026331800263919	0.18385927643938918	0.5254607895038081
3253	guitar_acoustic_002-042-127_seg_1	0.4387334557786869	0.4036912168580886	0.29341380217306556	0.11491179704115223	0.18970652164073423	0.49636591989079254
3254	guitar_acoustic_002-042-127_seg_2	0.38990515722765556	0.35416741550726255	0.2604998314719337	0.13367541309159833	0.19685114090432898	0.42202188422619125
3255	guitar_acoustic_002-042-127_seg_3	0.3178505825897622	0.2831098900920131	0.2123089908501331	0.1425680006007804	0.19041034343059968	0.317695888845671

3256 rows × 86 columns

Figura 2: DataFrame con los datos.

note_102	note_103	note_104	note_105	note_106	note_107	NOTE_CLASS
1.832943404286949e-05	8.316579176575785e-06	3.9076648754583695e-05	3.5763742118685775e-05	5.105543487994791e-05	2.11414371213532e-05	42
0.014533650297154168	0.008226192868327858	0.010360438782171614	0.014441596849455026	0.008667981941052722	0.003155229680306545	42
0.0018141528729064146	0.004119008154421155	0.0044064551799480 6	0.0009024598898683947	0.0006787049657676877	0.0013428408854527668	42
0.00309586712346008	0.0008150099748880614	0.002419585980409444	0.0024946556969354544	0.0015128041767926825	0.0002903539822933042	42
8.970850564369581e-07	6.791063943939393e-07	1.4770395968244498e-06	1.9648396650626448e-07	7.979831352343817e-07	1.6047887803798125e-07	33
...	...	...	...	...	...	...
0.011131912336763196	0.0331234307951551	0.02373260817132681	0.018034044540355773	0.03701983165047271	0.04718326573072677	60
1.2072932703553653e-06	4.276874305732596e-06	2.8458301007481548e-05	1.2967388577397697e-05	3.389439542710119e-06	2.0961770858611154e-06	42
0.03531958849508476	0.005206877502471014	0.015915537402639034	0.03260285235128136	0.06745135639540611	0.05517935759730043	42
0.021916981556870798	0.007041478630295643	0.010519602186439847	0.02244587464419417	0.02173963034875114	0.01034854306201443	42
0.007705038901932117	0.0038751564084174915	0.0031746565204511155	0.00447291856527498	0.00335268698271044	0.0025471661892740916	42

Figura 3: DataFrame con los datos, mostrando la clase en la última columna.

Referencias

[1] O. Das, J. O. S. Iii, and C. Chafe, "Real-time Pitch Tracking in Audio Signals with the Extended Complex Kalman Filter," p. 7, 2017.

[2] J. W. Kim, J. Salamon, P. Li, and J. P. Bello, "CREPE: A Convolutional Representation for Pitch Estimation," *arXiv:1802.06182 [cs, eess, stat]*, Feb. 2018, arXiv: 1802.06182. [Online]. Available: http://arxiv.org/abs/1802.06182

[3] M. Mauch and S. Dixon, "PYIN: A fundamental frequency estimator using probabilistic threshold distributions," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 659–663, iSSN: 2379-190X.

[4] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, Apr. 2002, publisher: Acoustical Society of America. [Online]. Available: https://asa.scitation.org/doi/10.1121/1.1458024

[5] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra, "ESSENTIA: an Audio Analysis Library for Music Information Retrieval," Nov. 2013.

[6] J. C. Brown, "Calculation of a constant Q spectral transform," *The Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 425–434, Jan. 1991. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.400476

[7] J. Brown and M. Puckette, "An efficient algorithm for the calculation of a constant Q transform," *Journal of the Acoustical Society of America*, vol. 92, p. 2698, Nov. 1992.

[8] J. Engel, C. Resnick, A. Roberts, S. Dieleman, D. Eck, K. Simonyan, and M. Norouzi, "Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders," *arXiv:1704.01279 [cs]*, Apr. 2017, arXiv: 1704.01279. [Online]. Available: http://arxiv.org/abs/1704.01279

[9] B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and Music Signal Analysis in Python," Jan. 2015. [Online]. Available: https://scinapse.io/papers/2191779130