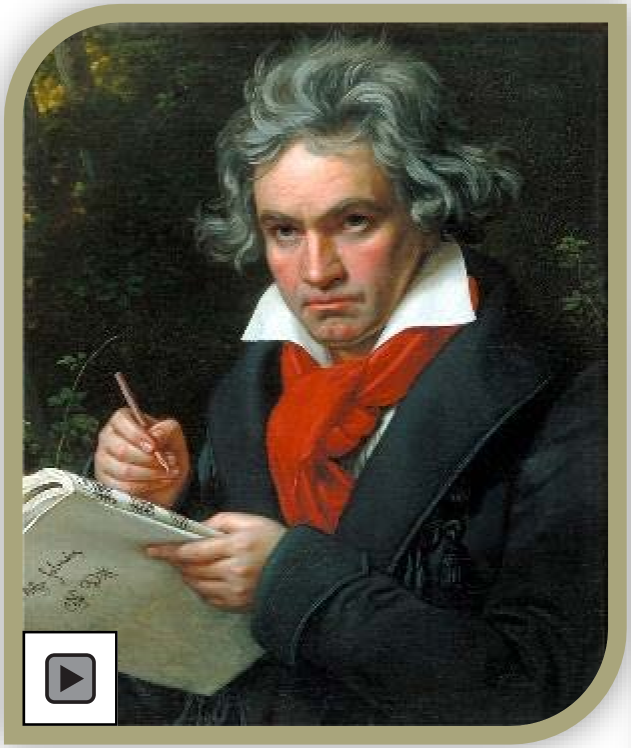
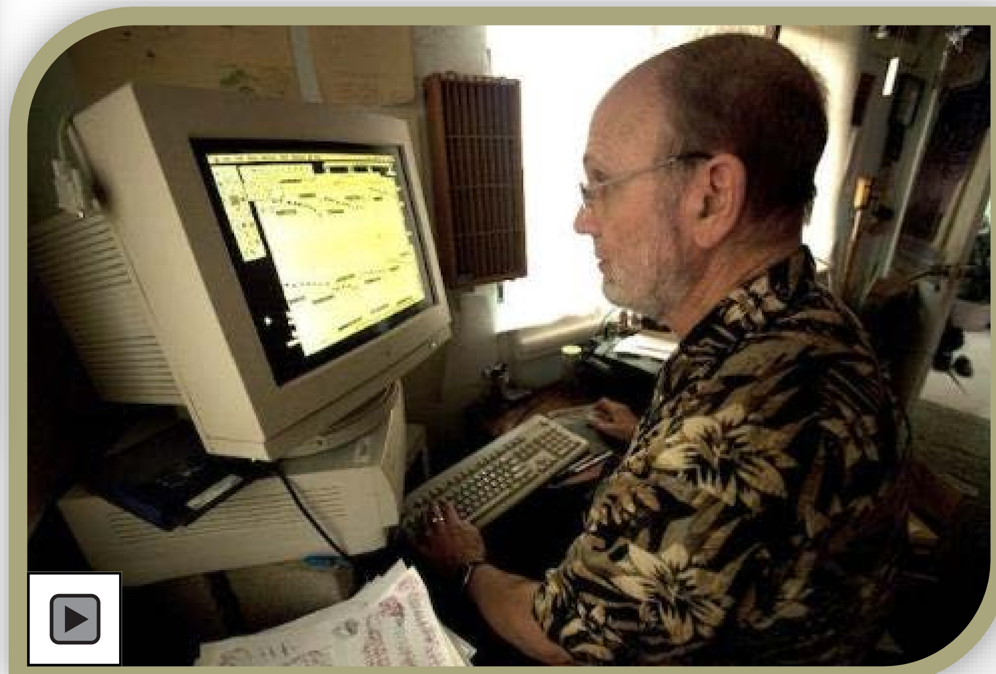
Musical Motif Discovery in Non-Musical Media

Daniel Johnson
Dan Ventura
Computer Science Department
Brigham Young University

Computational Composition Approaches



#1: Mimic

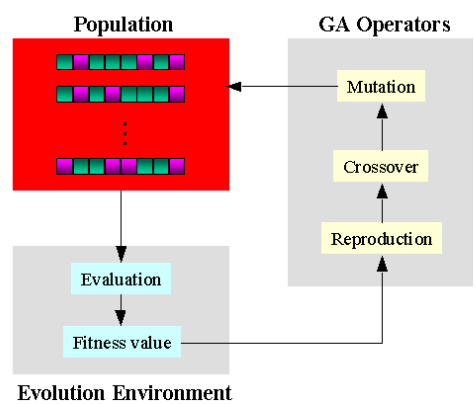


Computational Composition Approaches

#2: Innovate

Swarms

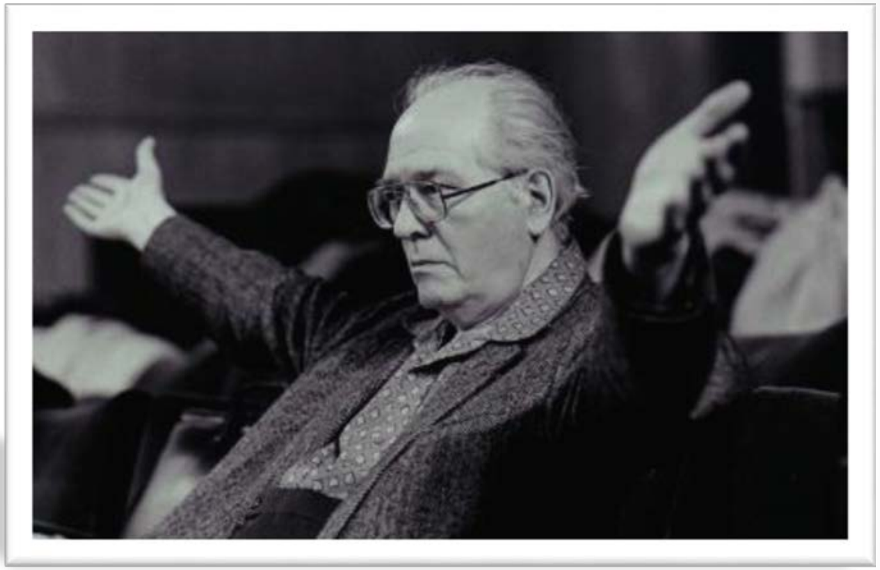
Genetic Algorithms



Genetic Algorithm Evolution Flow

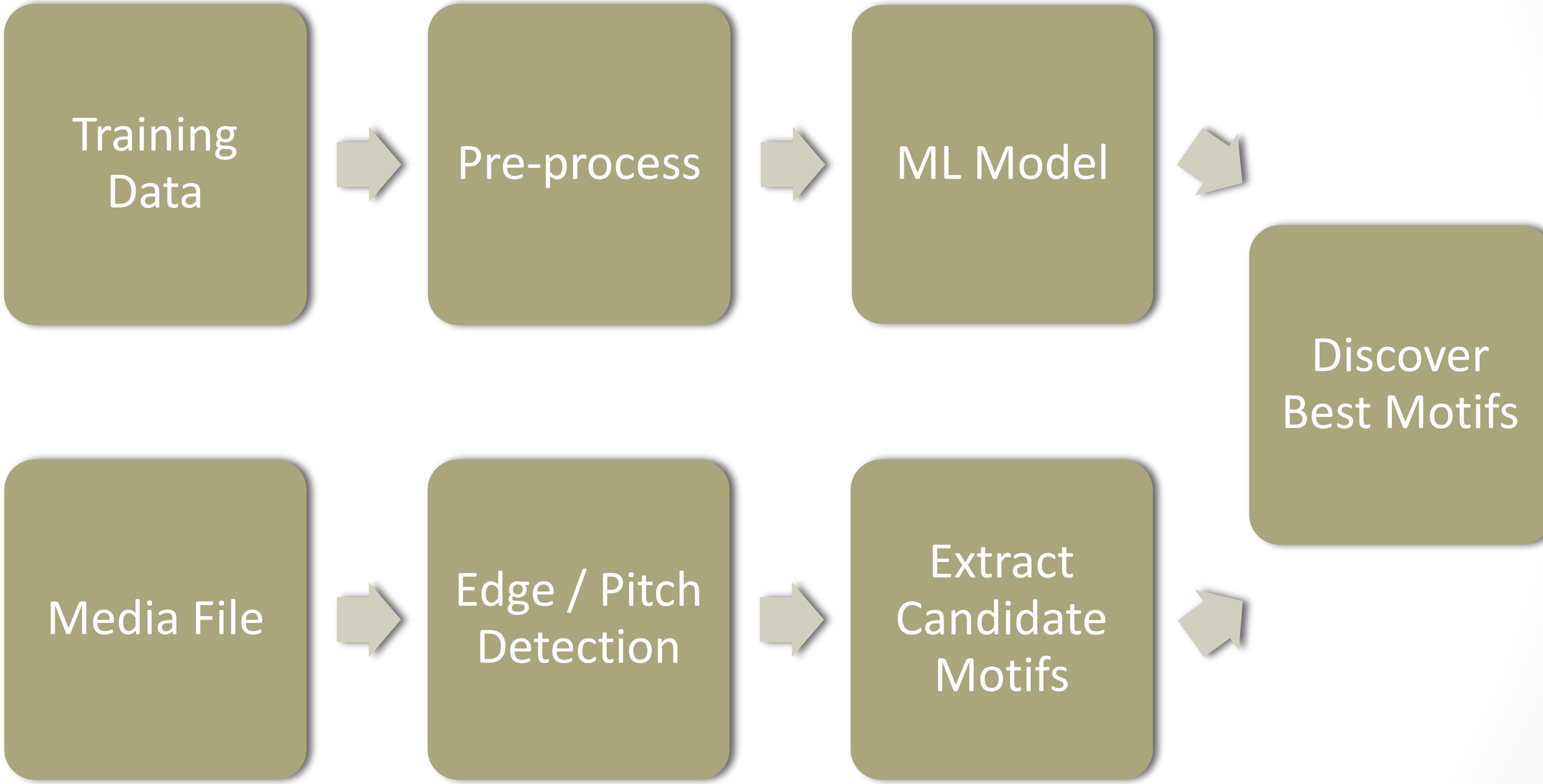
Human Composition Approaches

Mimic

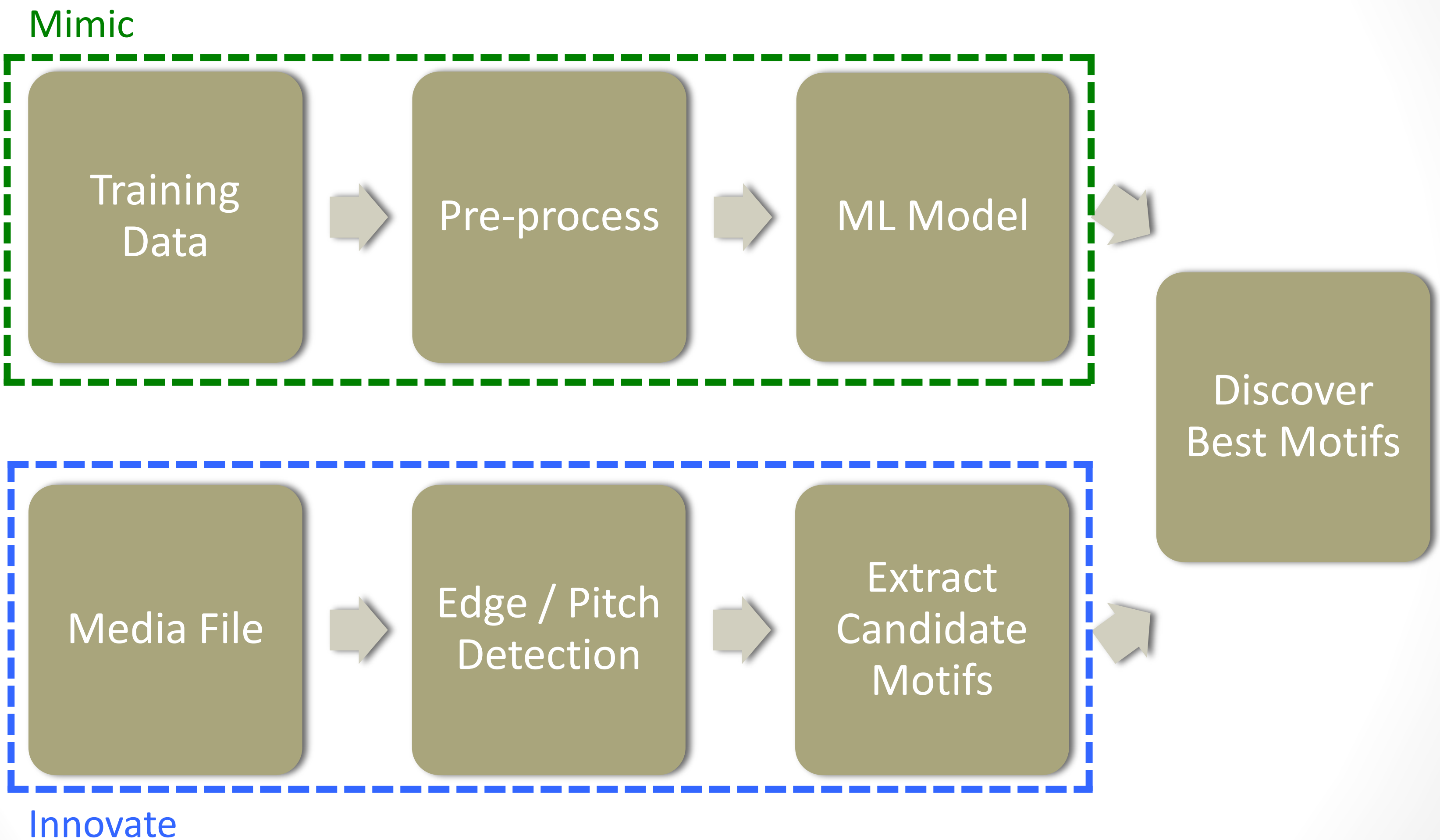


AND Innovate

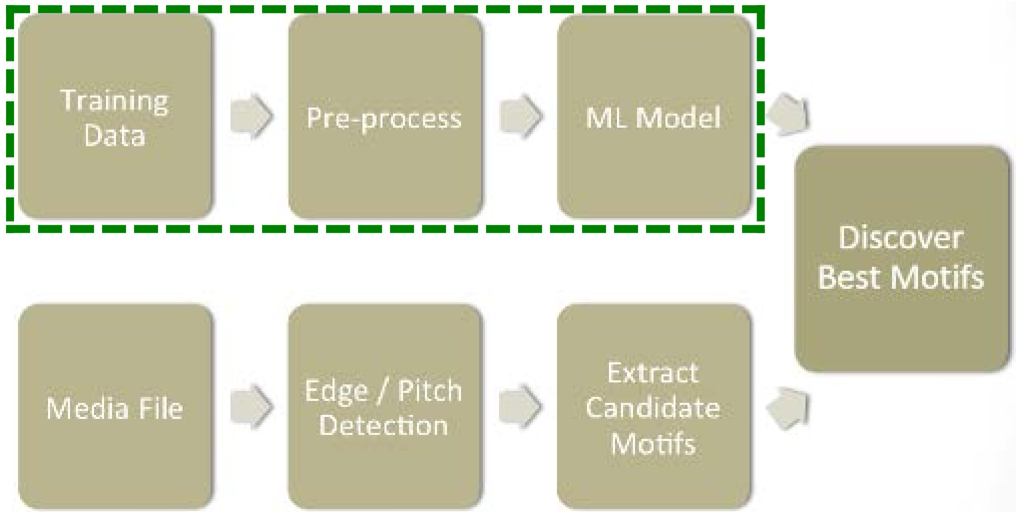
Project Description



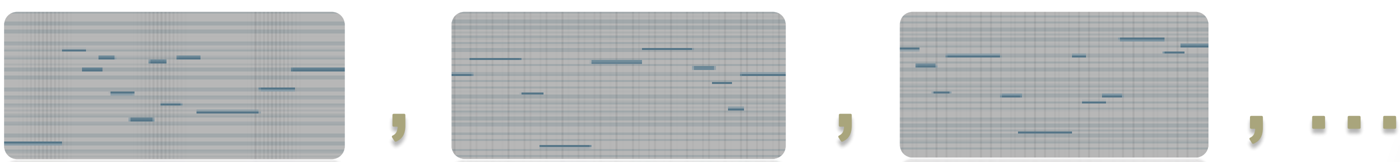
Project Description



Project Description

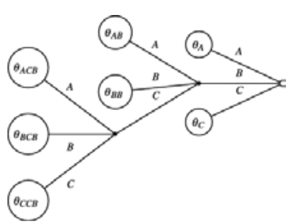


- Training data – 9383 monophonic MIDI themes

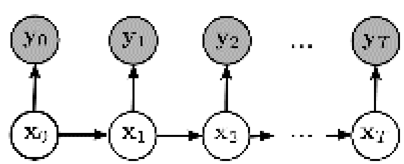


- Machine Learning Models

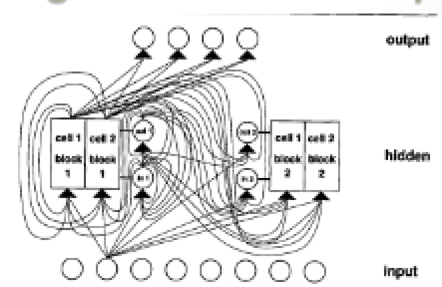
4 Variable Order Markov Models



Hidden Markov Model

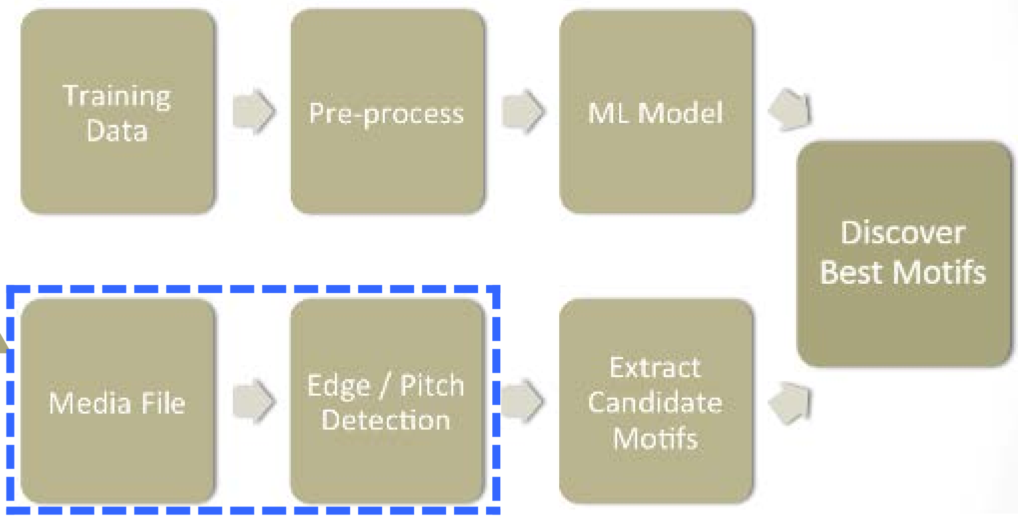


Long Short-Term Memory



Project Description

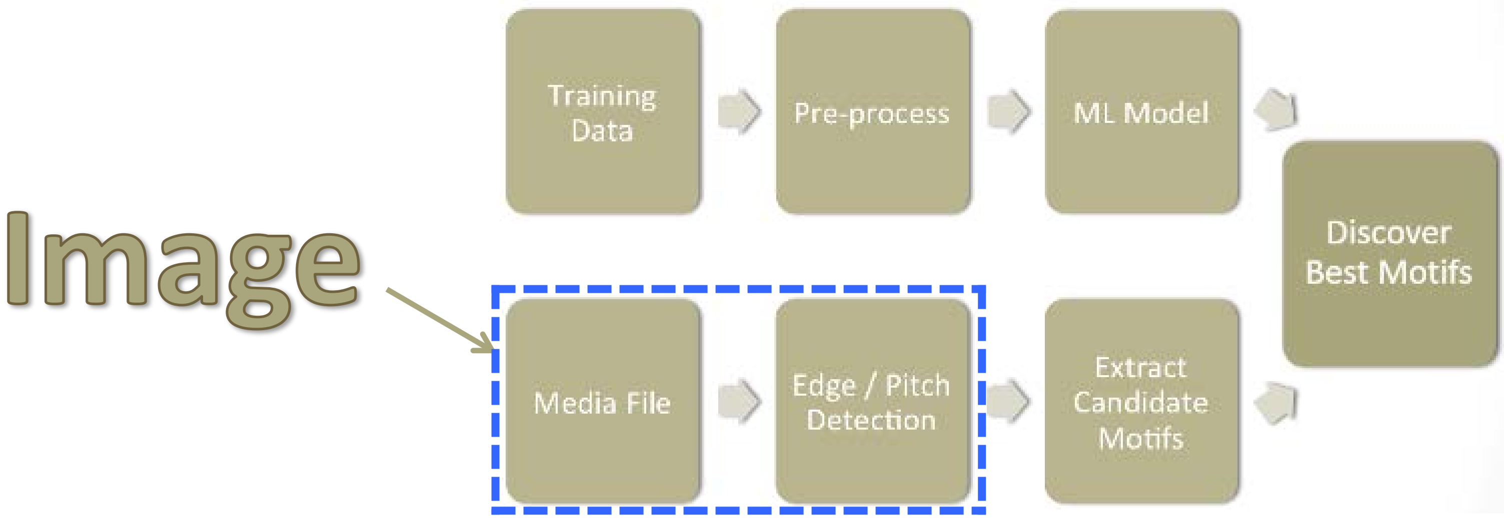
Audio

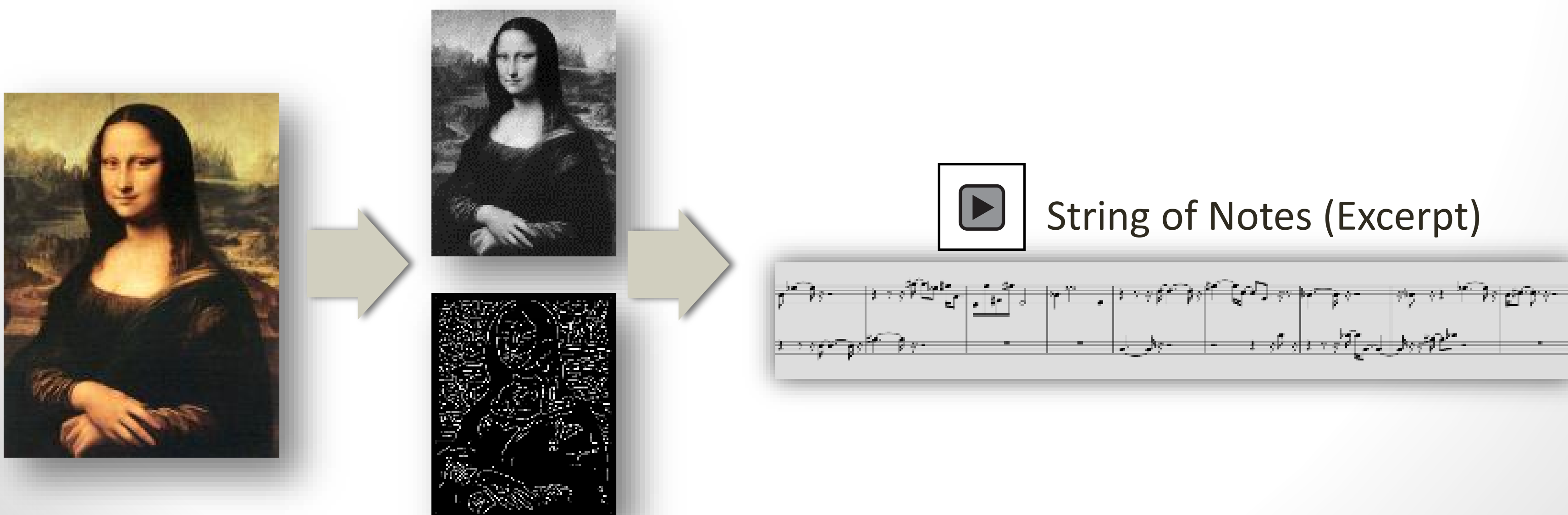


Pitch Detection

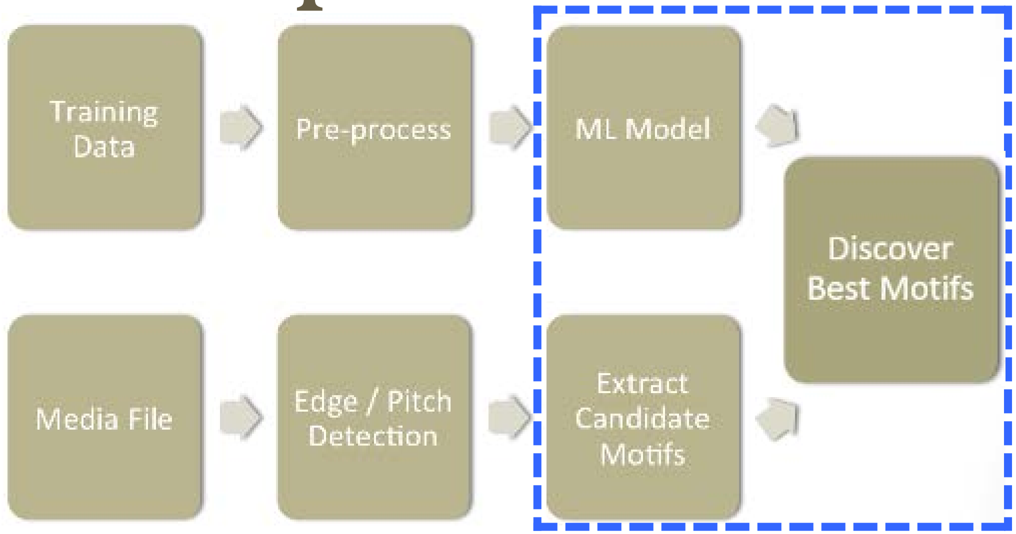
String of Notes

Project Description

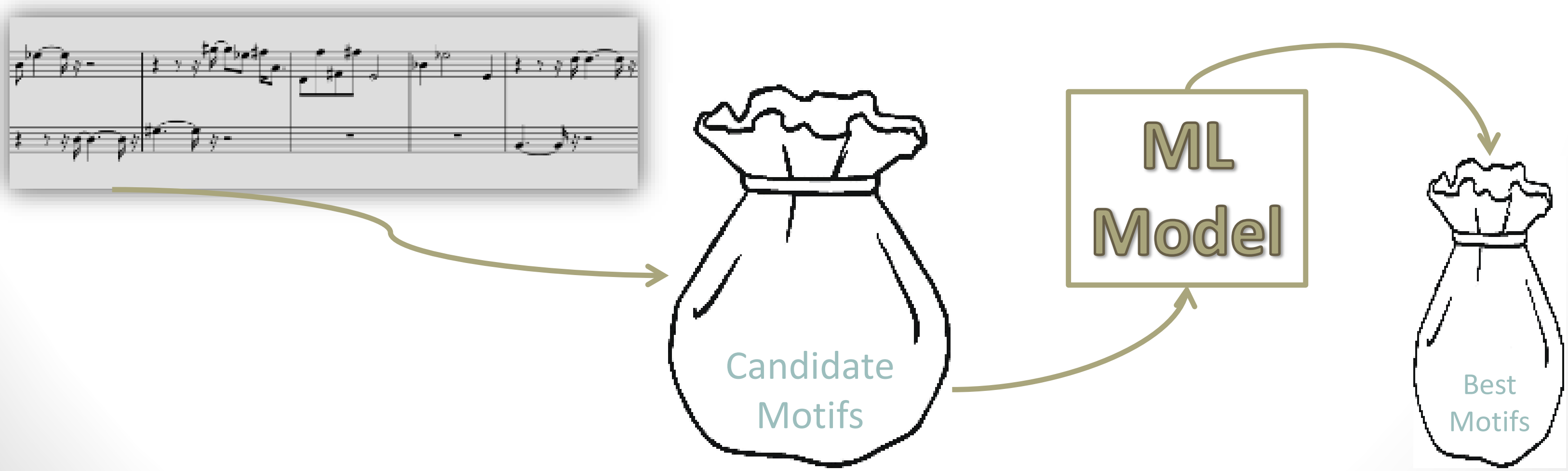




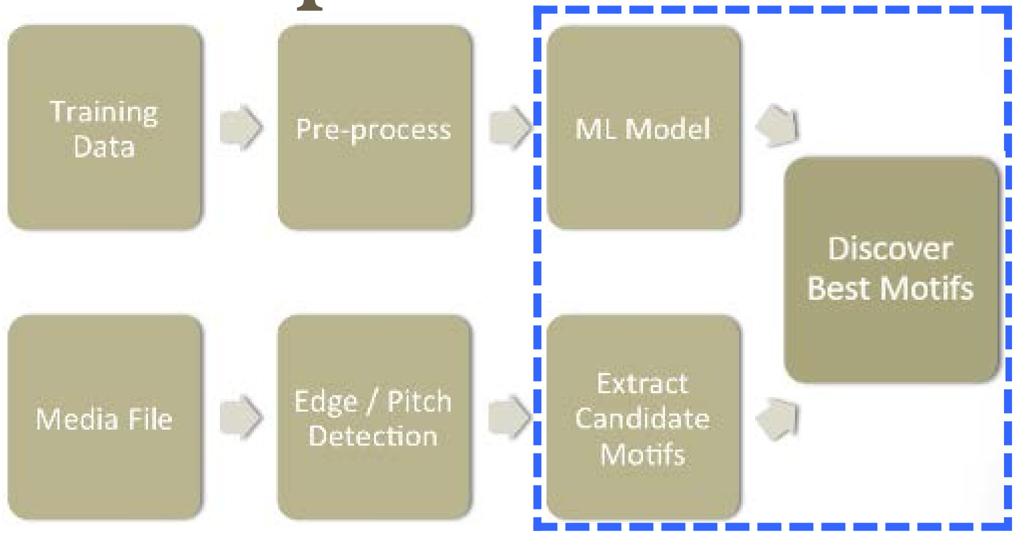
Project Description



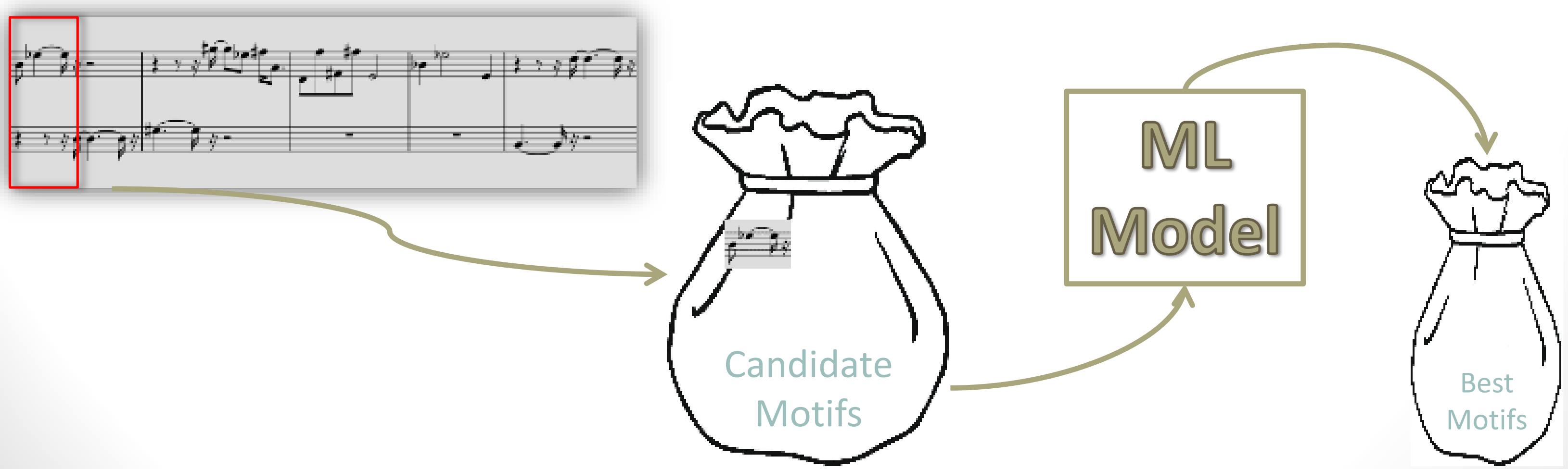
String of notes from edge/pitch detection



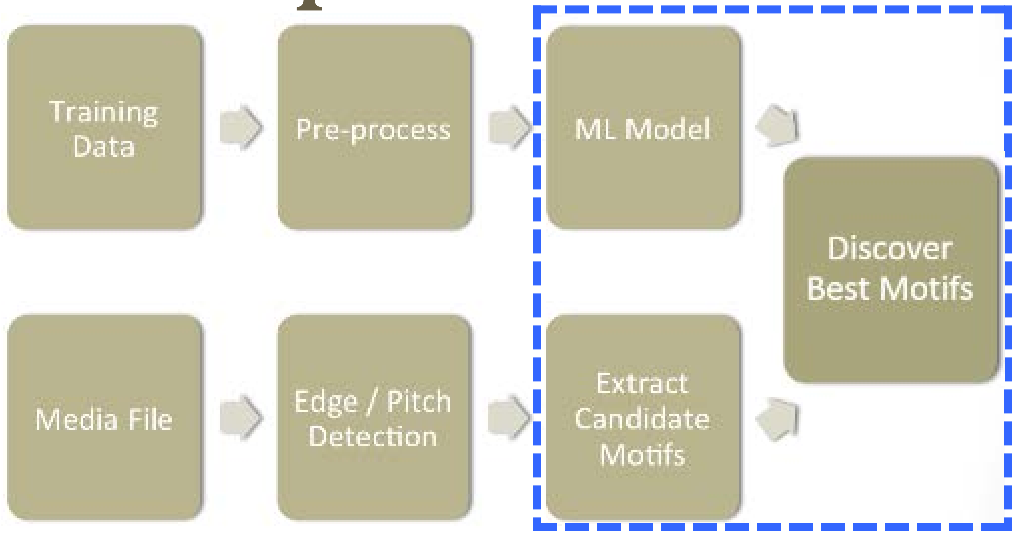
Project Description



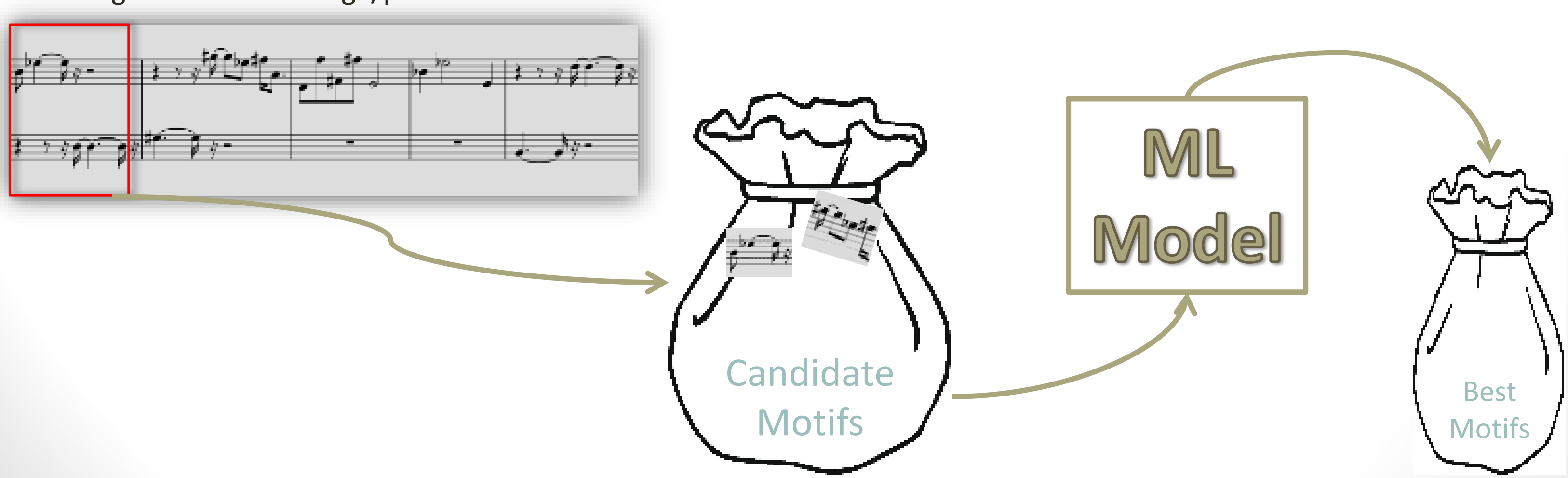
String of notes from edge/pitch detection



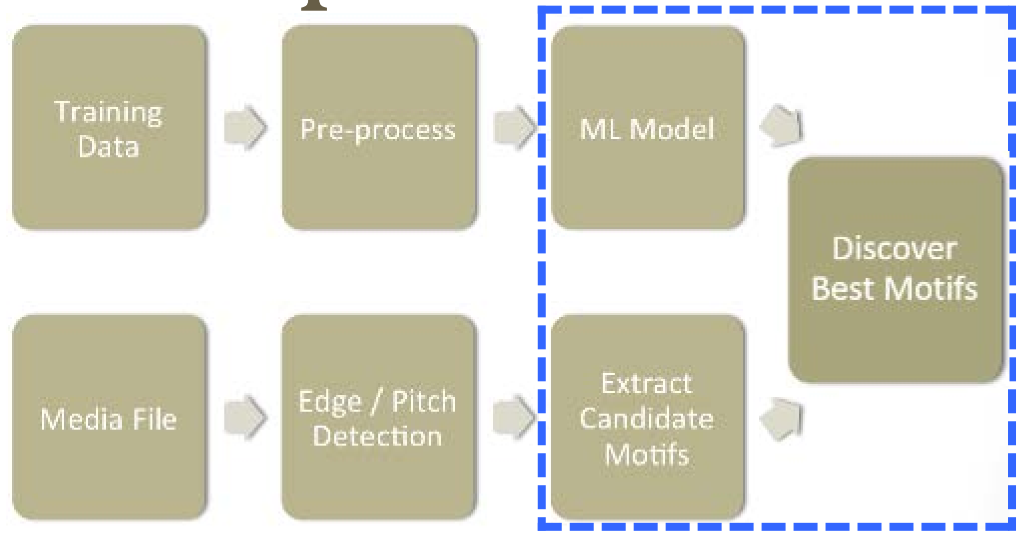
Project Description



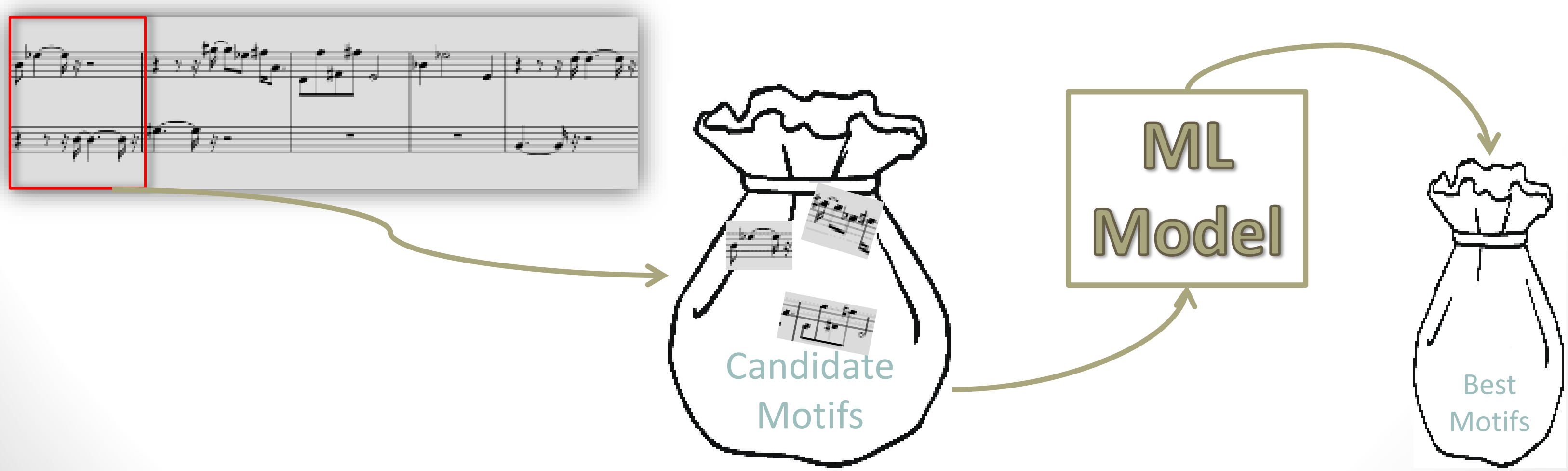
String of notes from edge/pitch detection



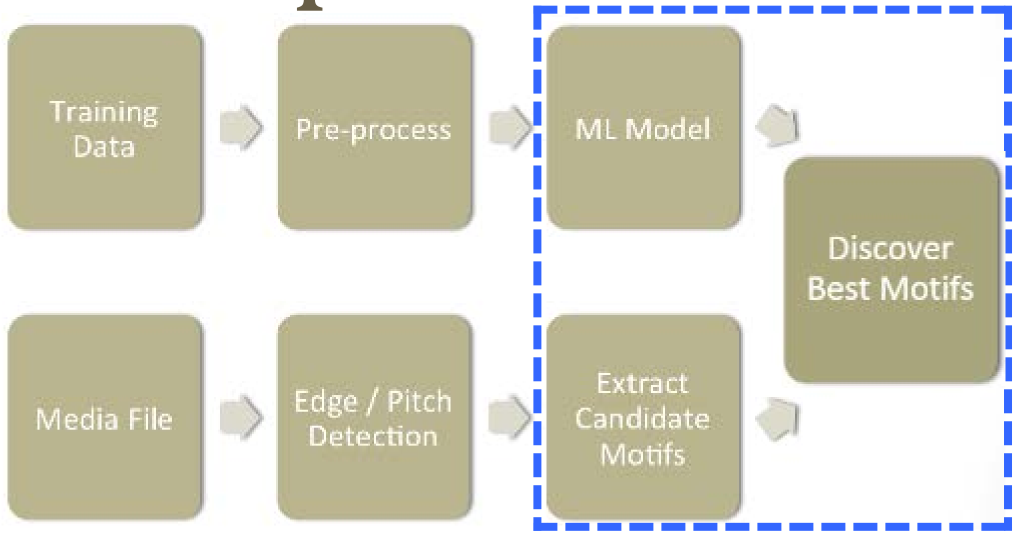
Project Description



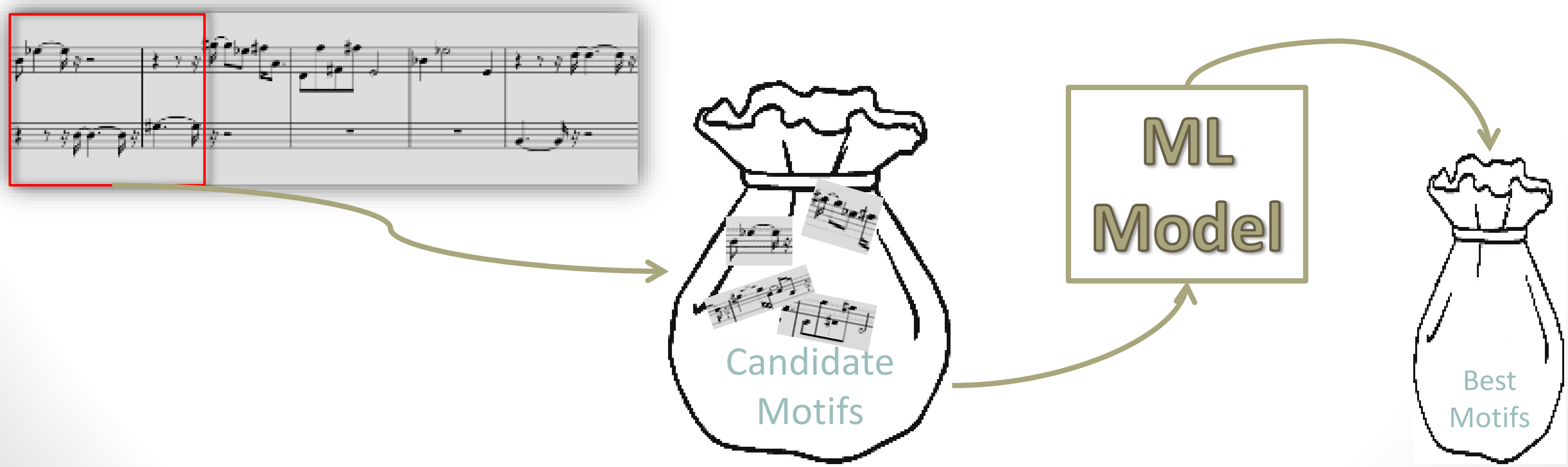
String of notes from edge/pitch detection



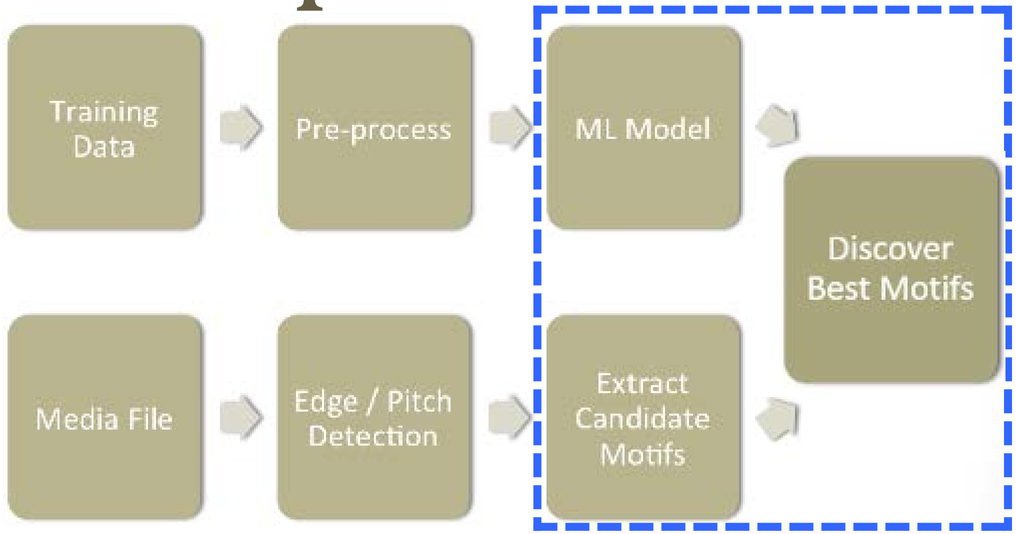
Project Description



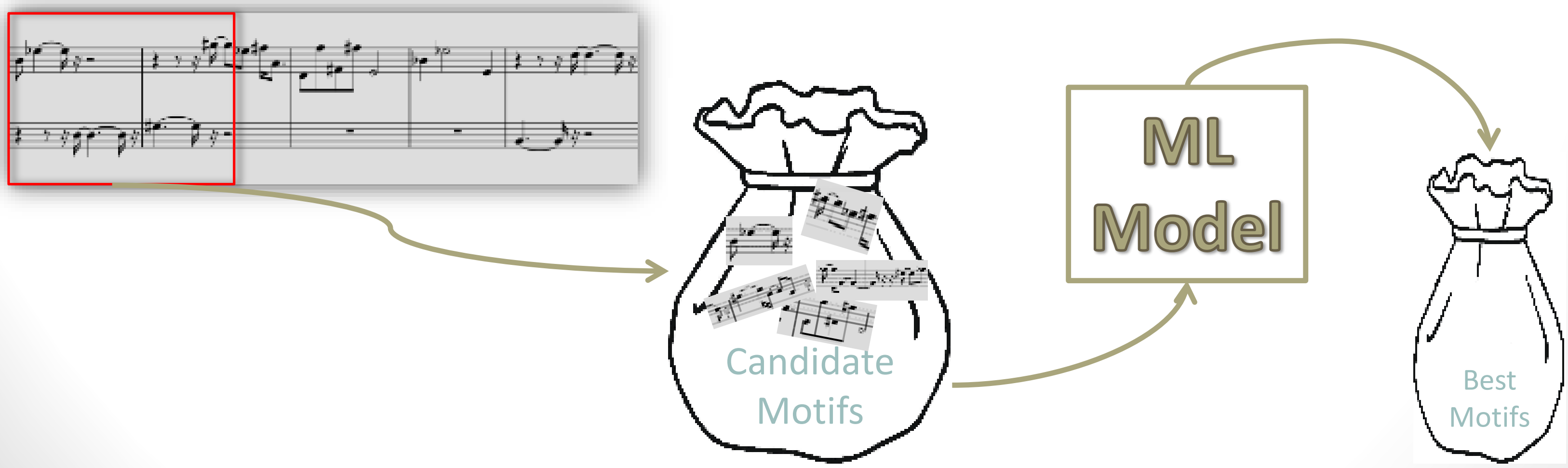
String of notes from edge/pitch detection



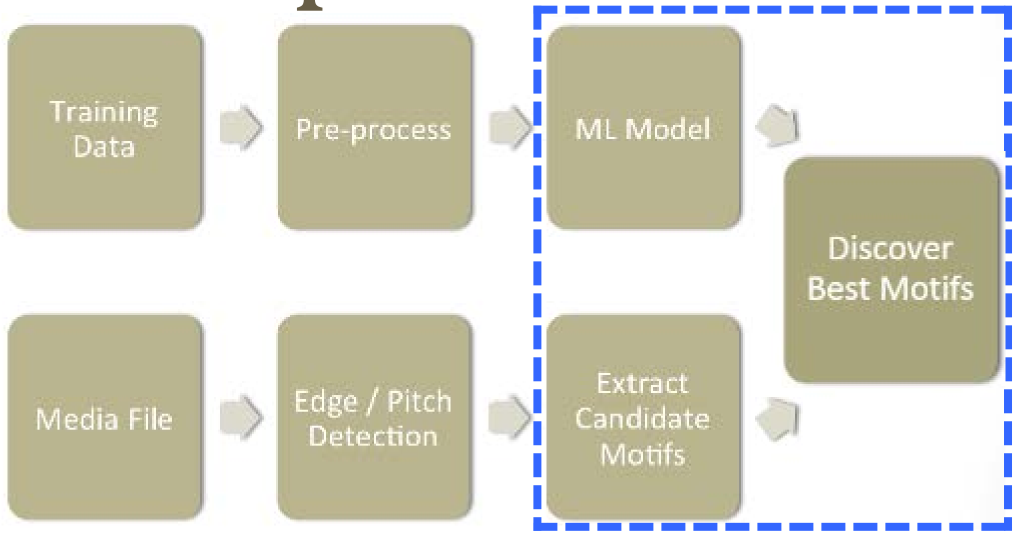
Project Description

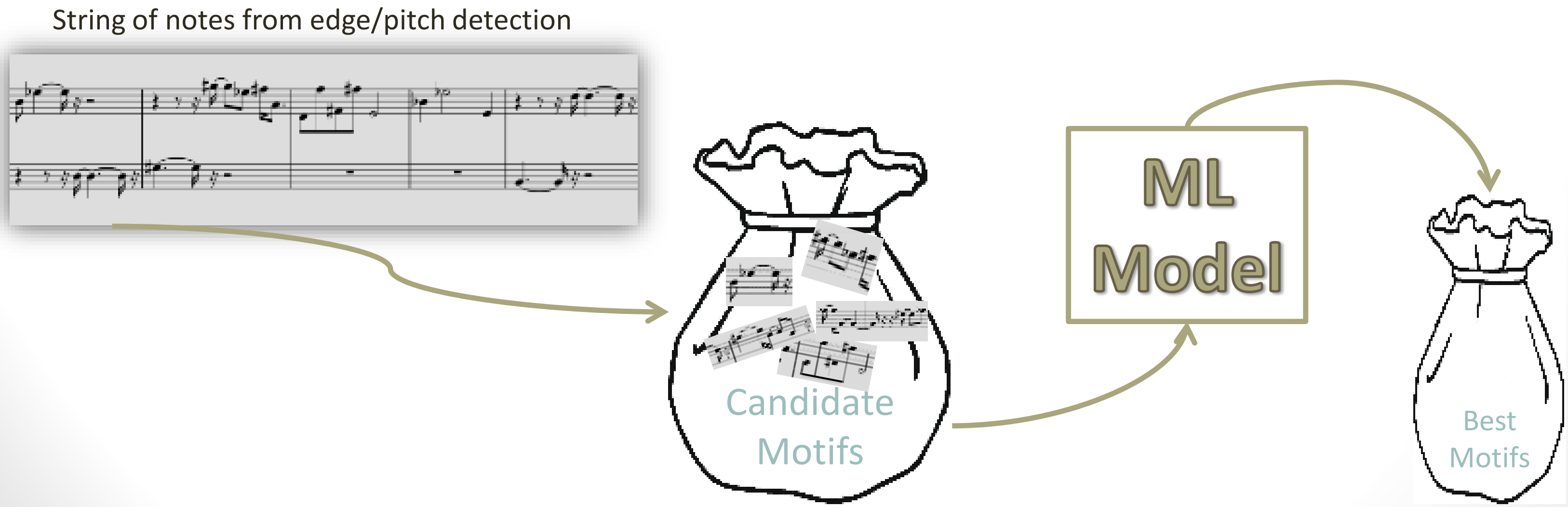


String of notes from edge/pitch detection

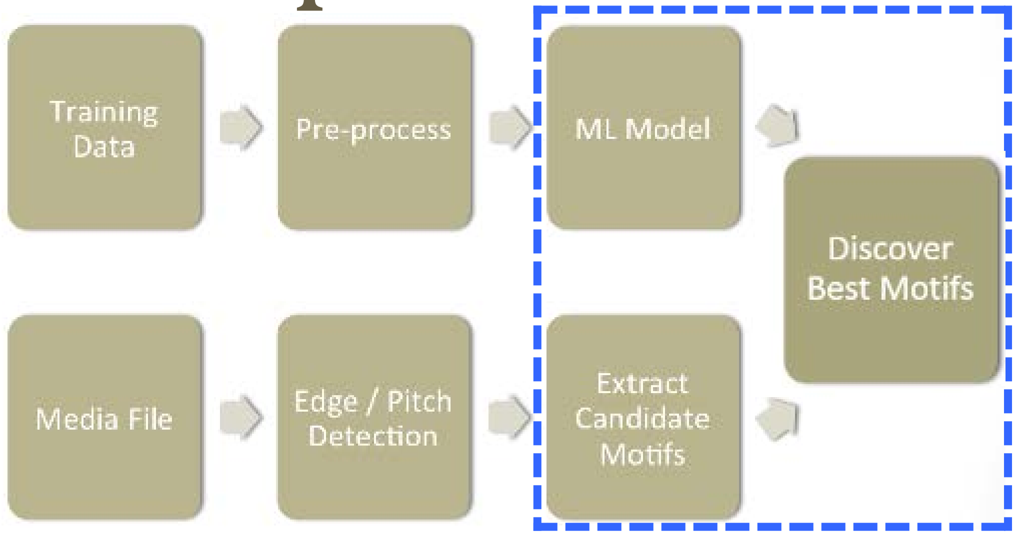


Project Description

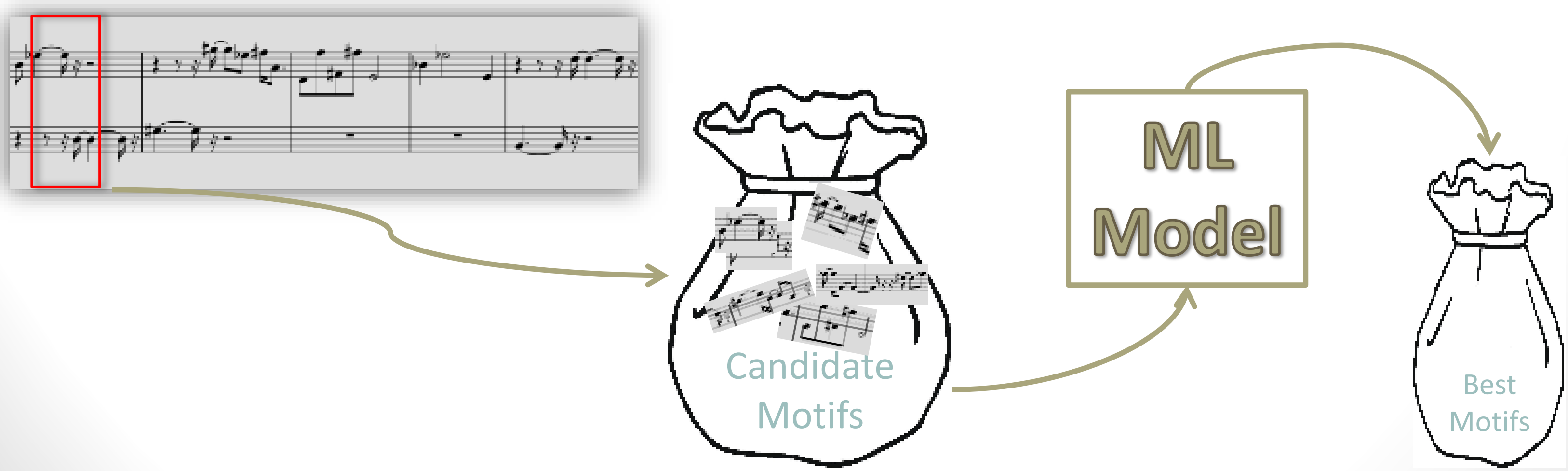




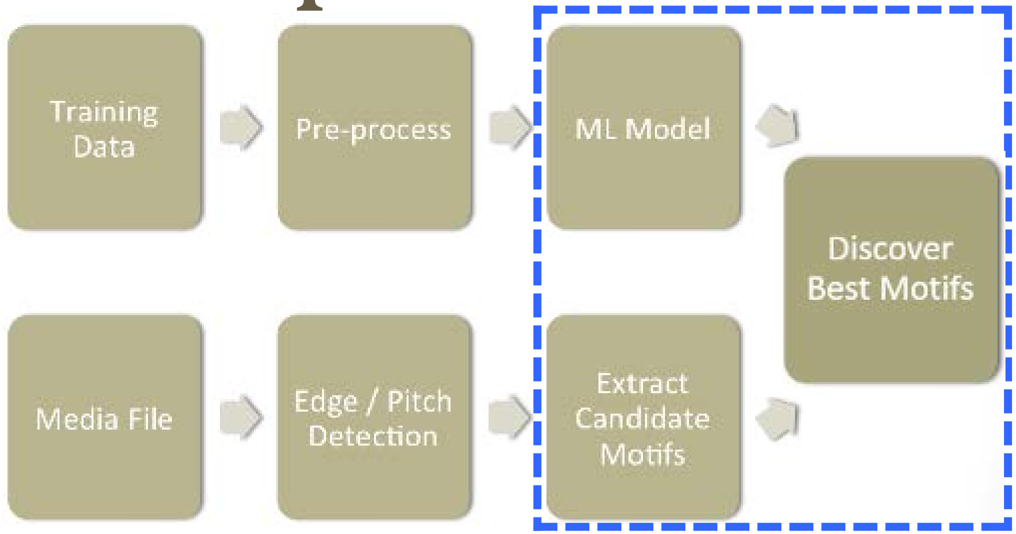
Project Description



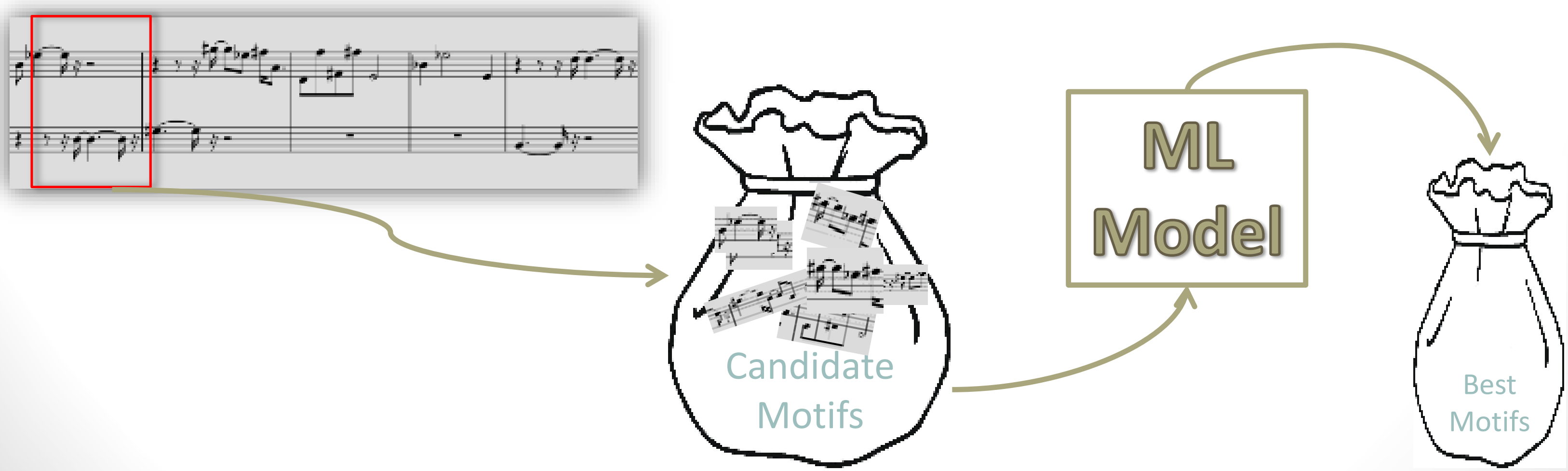
String of notes from edge/pitch detection



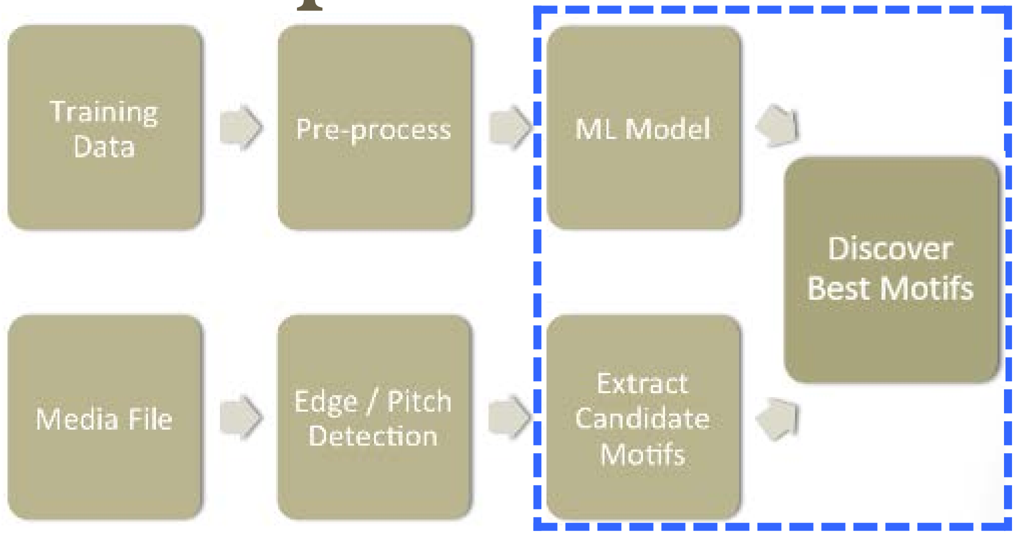
Project Description



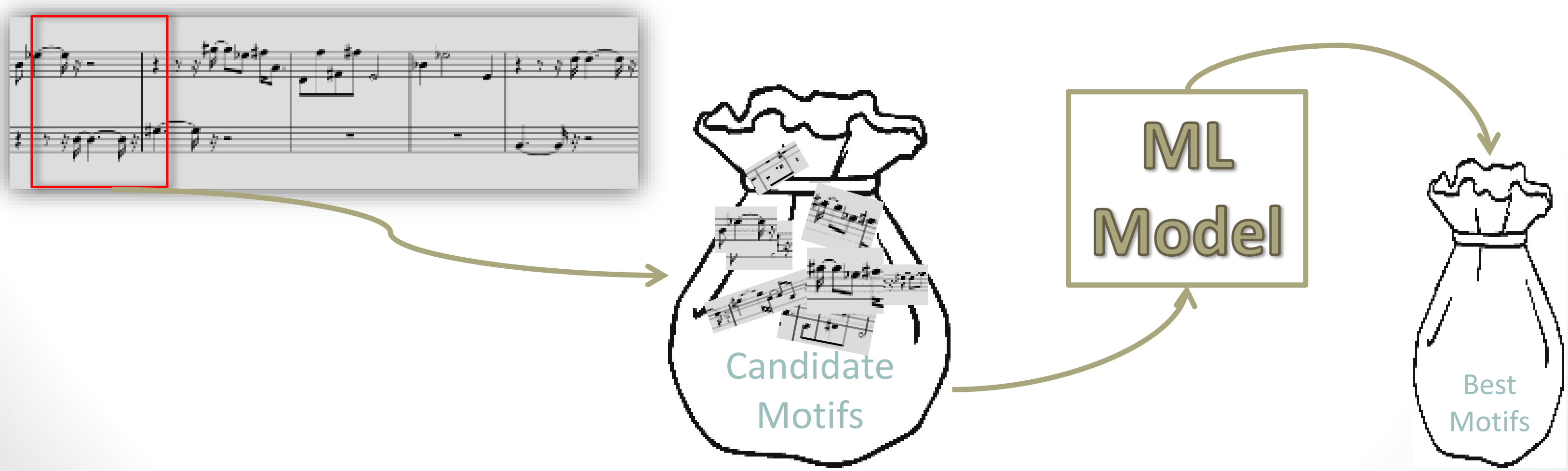
String of notes from edge/pitch detection



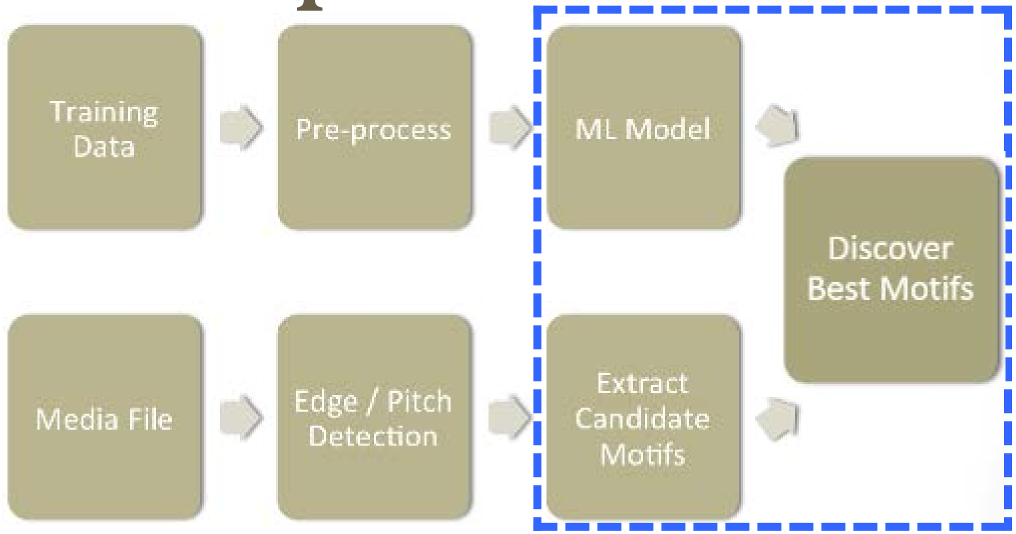
Project Description



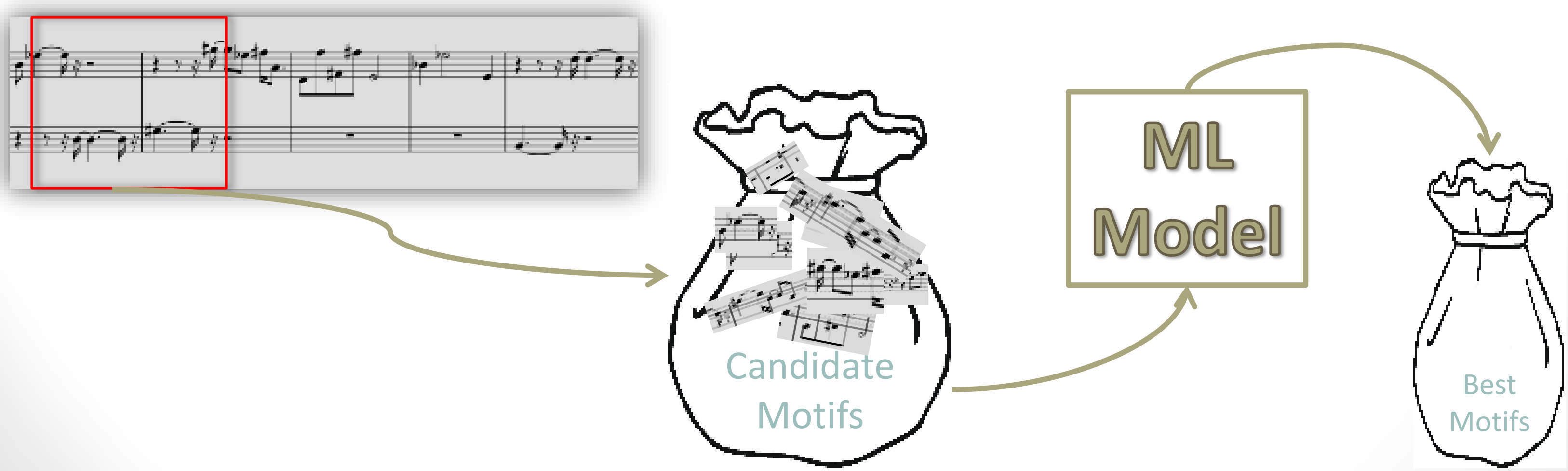
String of notes from edge/pitch detection



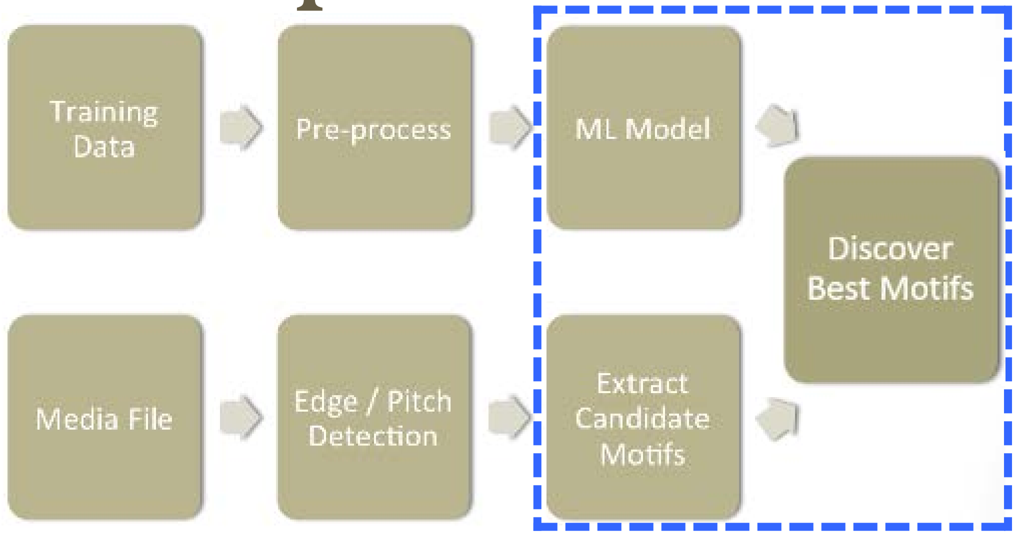
Project Description



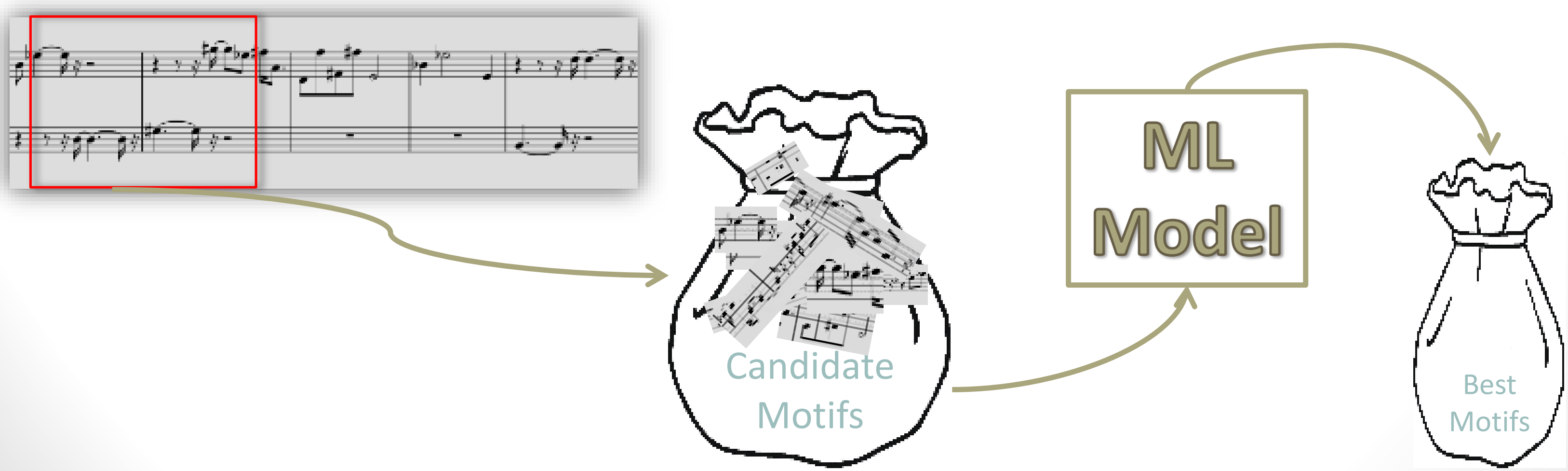
String of notes from edge/pitch detection



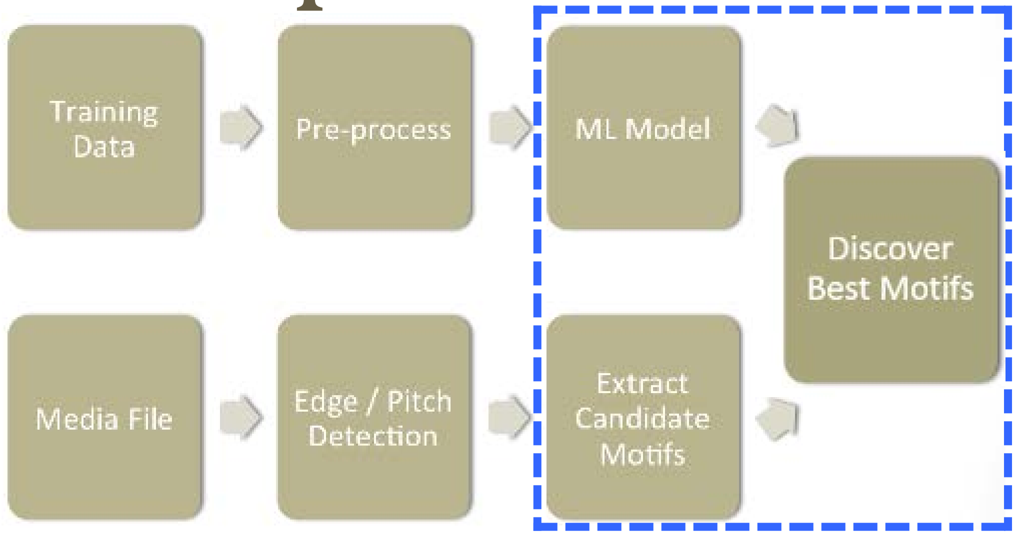
Project Description



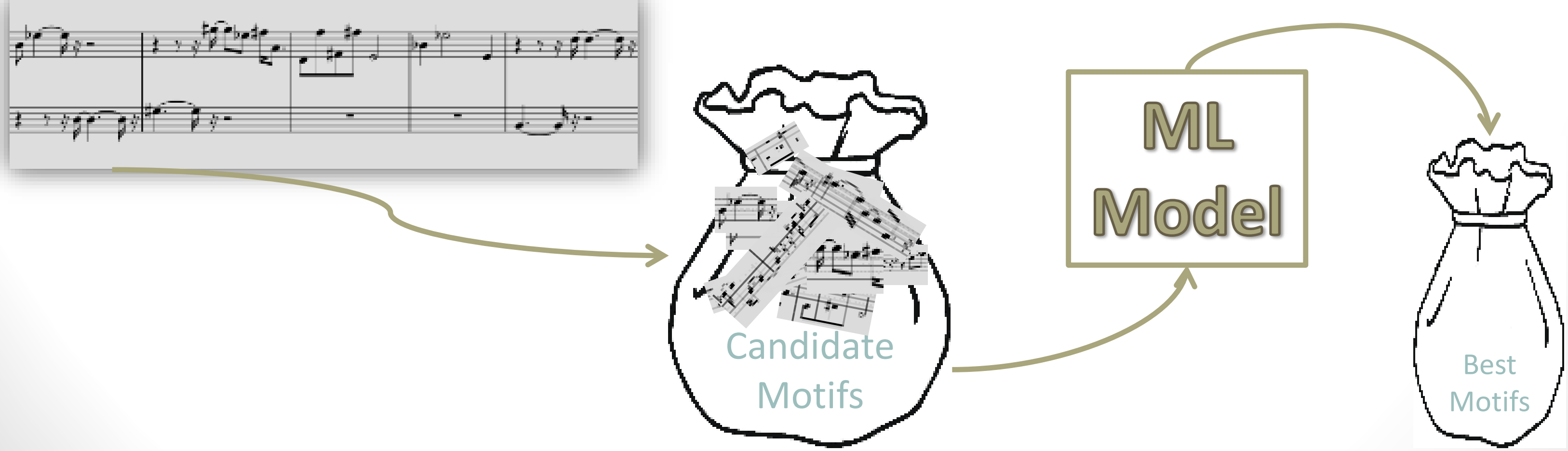
String of notes from edge/pitch detection



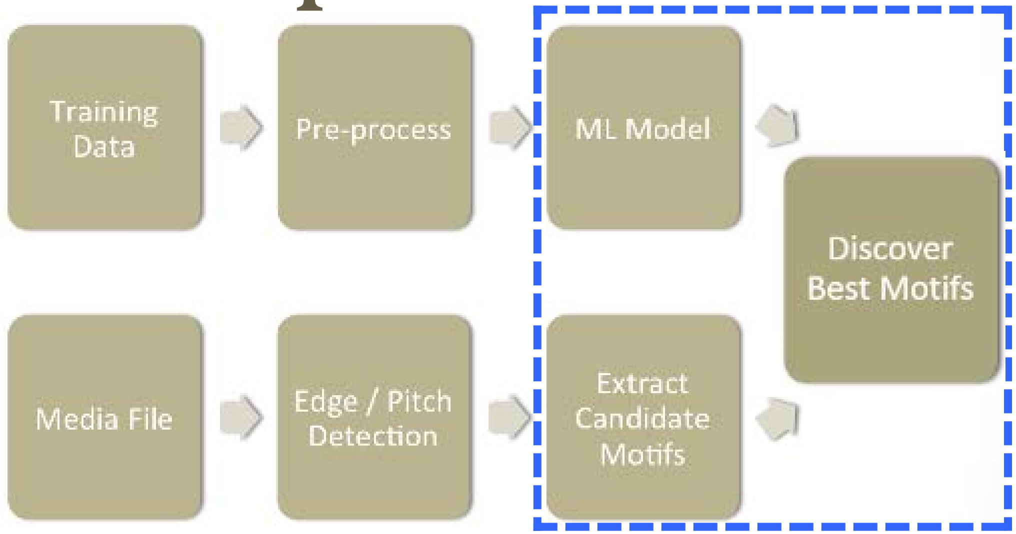
Project Description



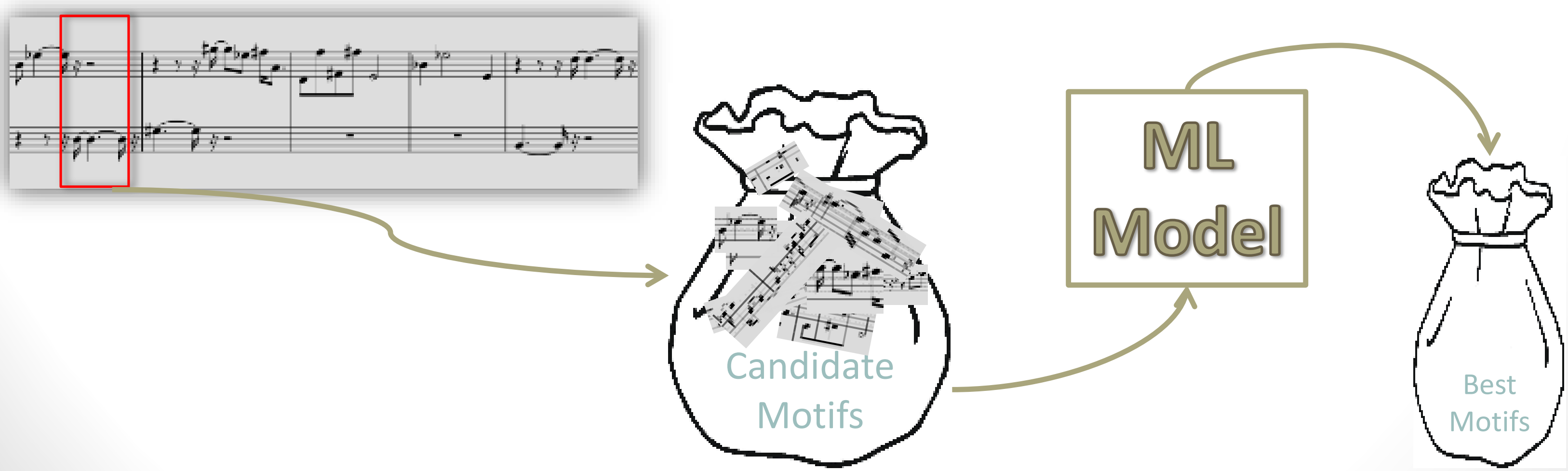
String of notes from edge/pitch detection



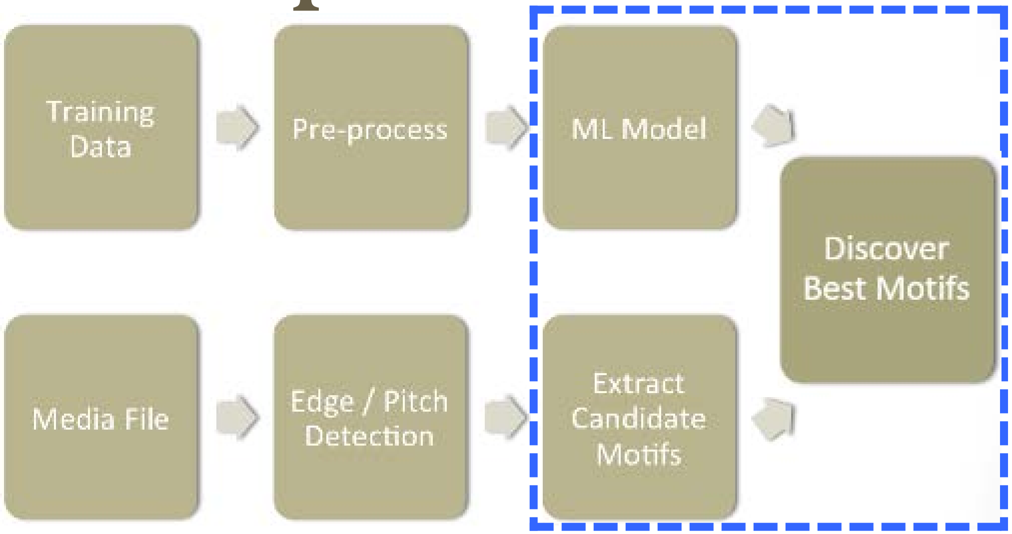
Project Description



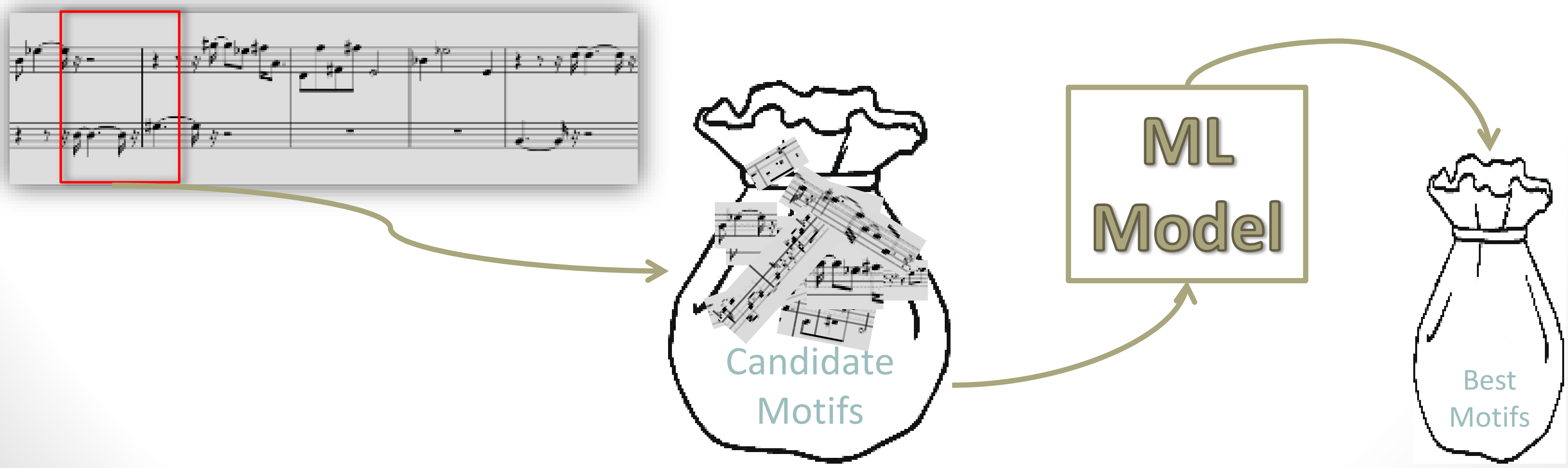
String of notes from edge/pitch detection



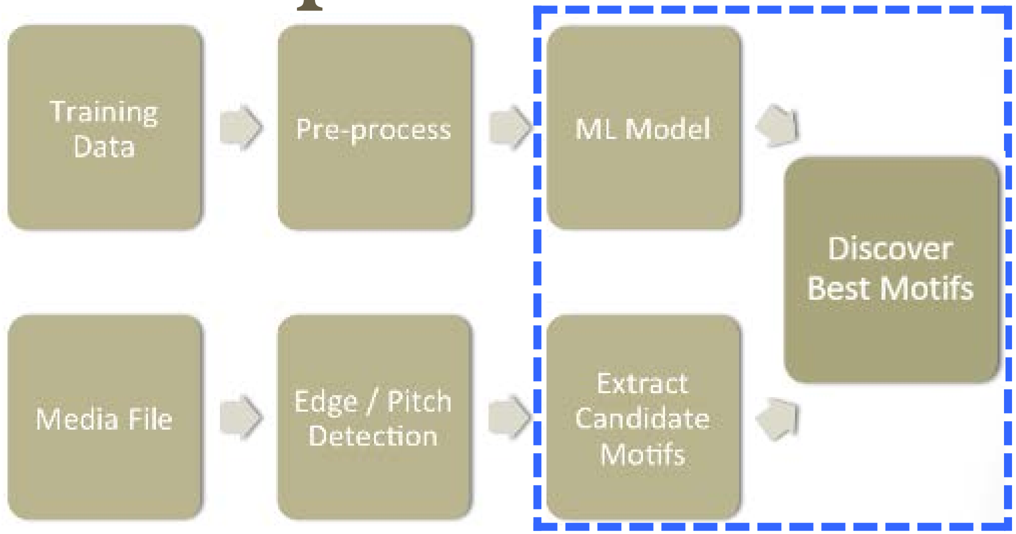
Project Description



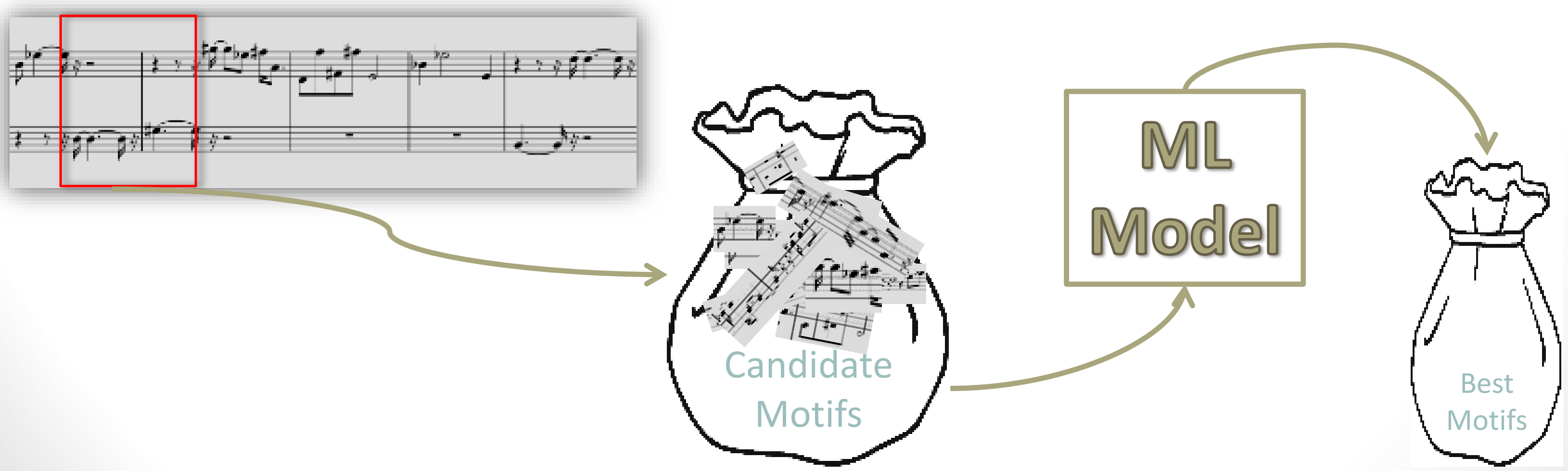
String of notes from edge/pitch detection



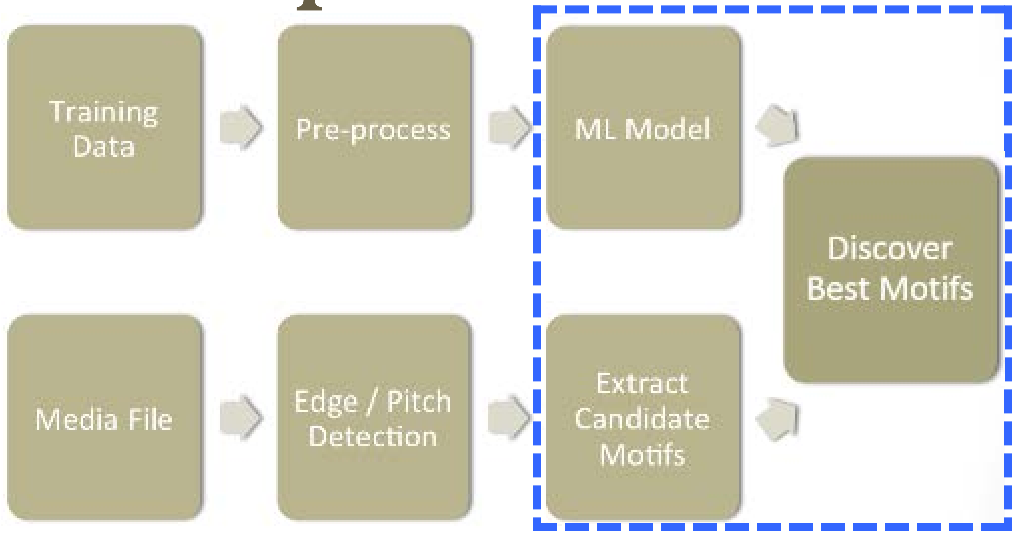
Project Description



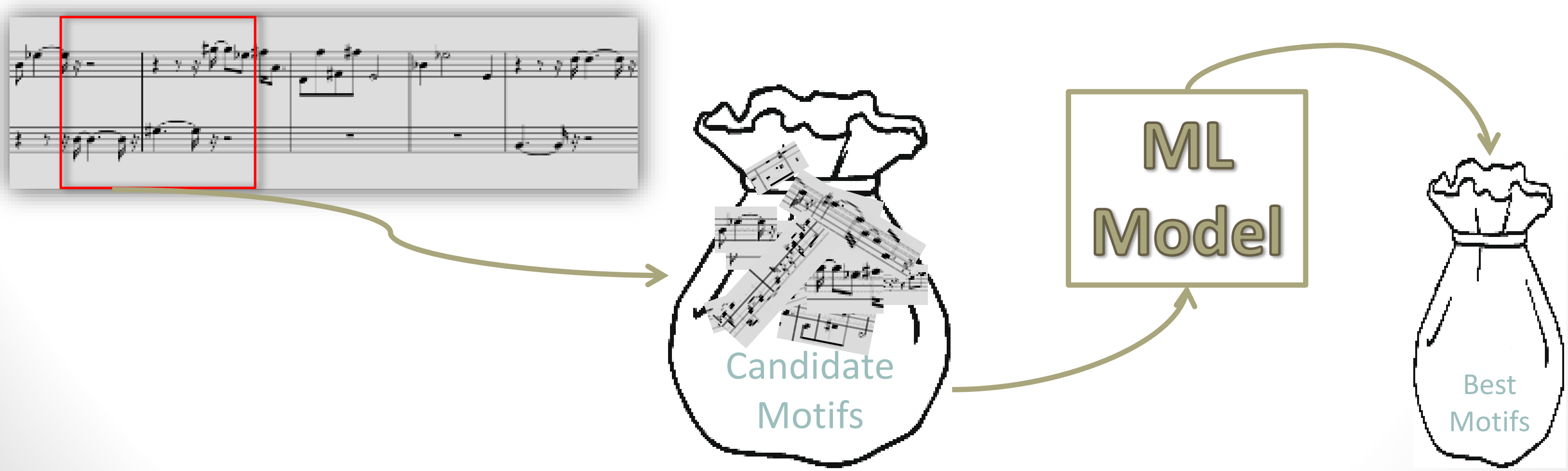
String of notes from edge/pitch detection



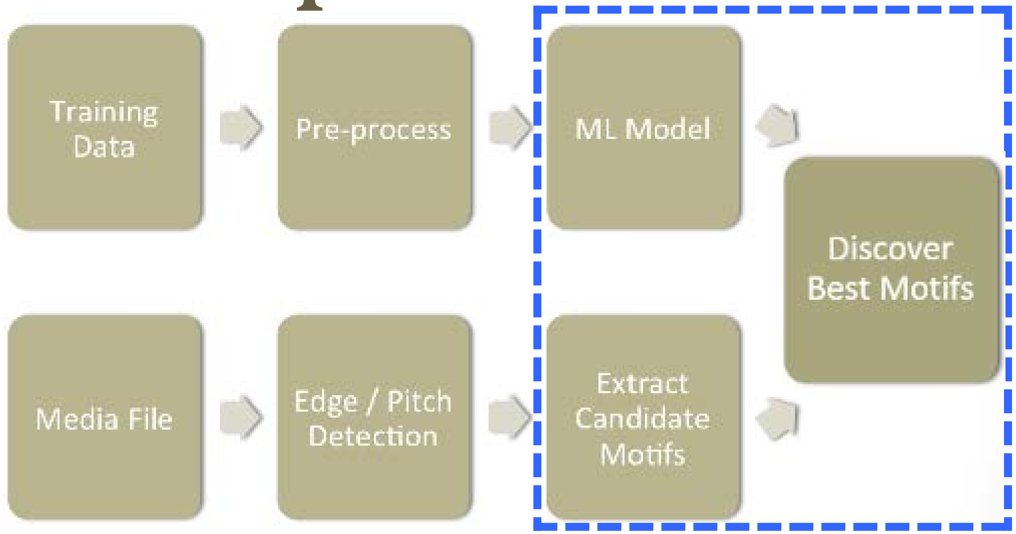
Project Description



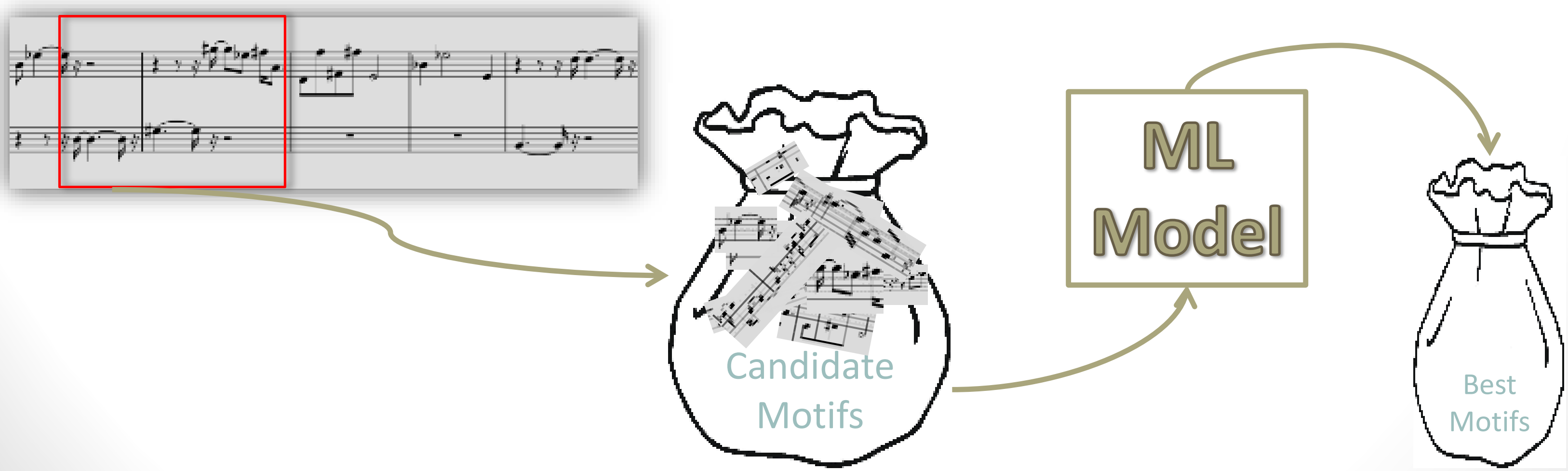
String of notes from edge/pitch detection



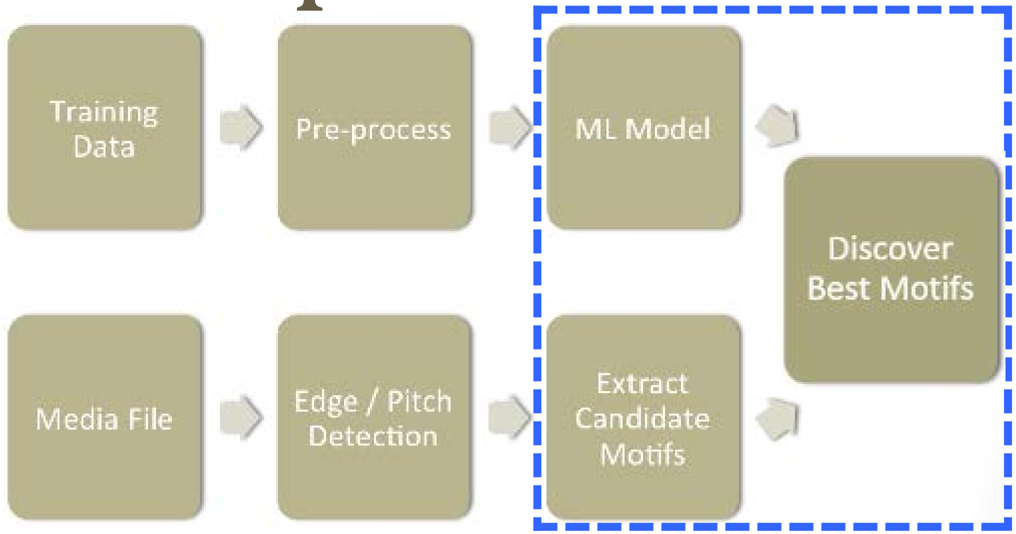
Project Description



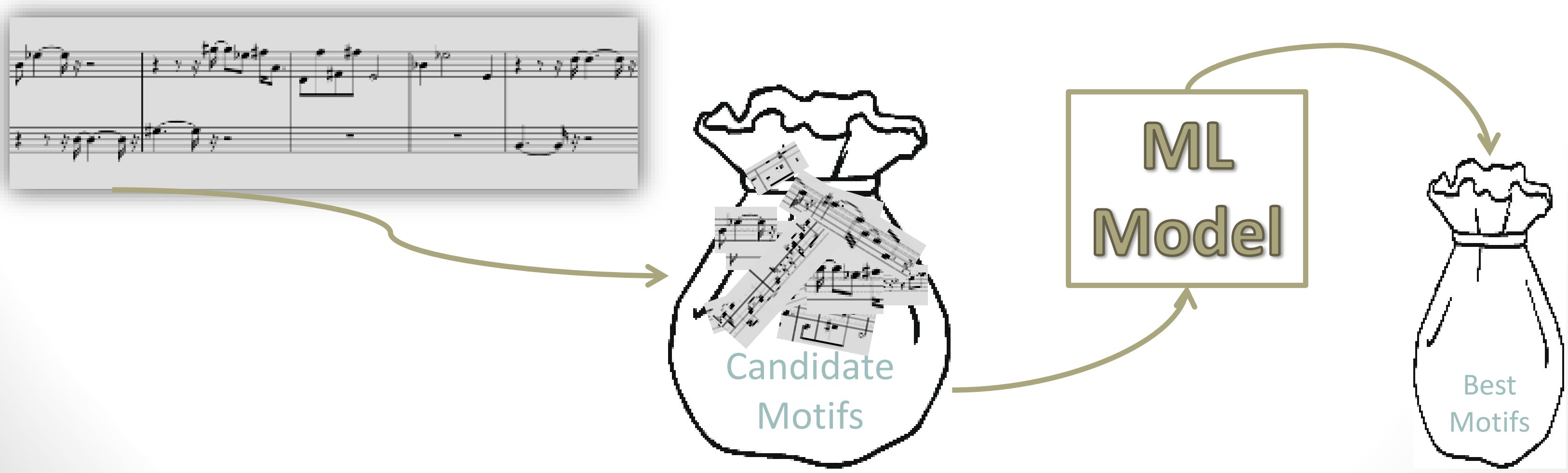
String of notes from edge/pitch detection



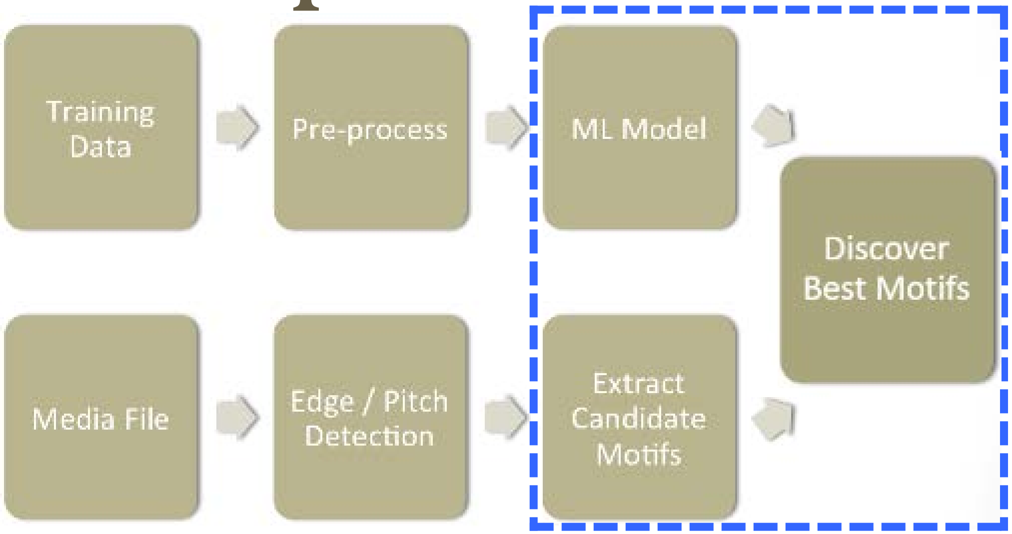
Project Description



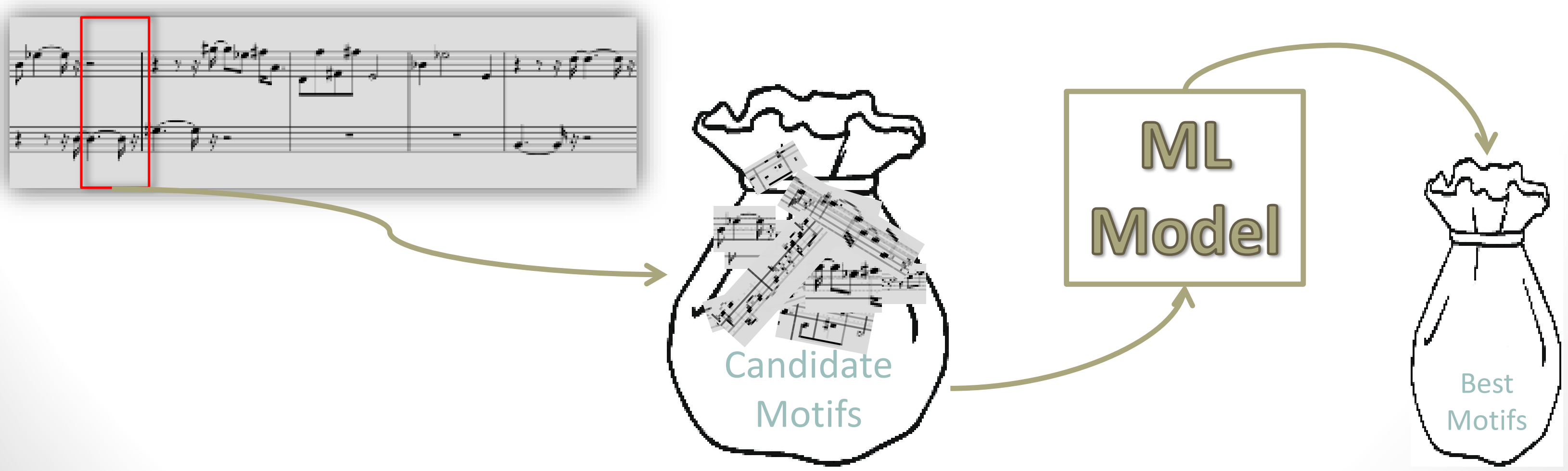
String of notes from edge/pitch detection



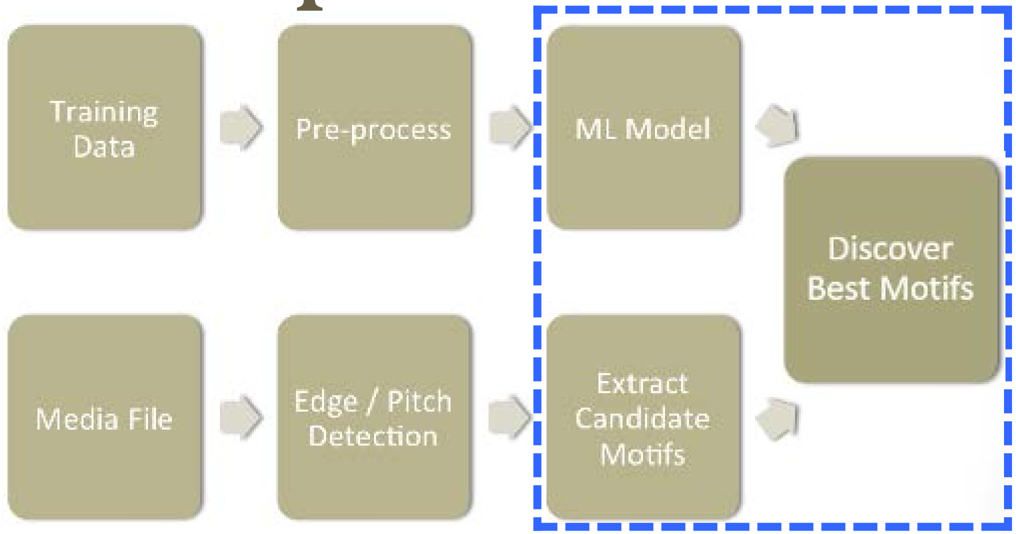
Project Description



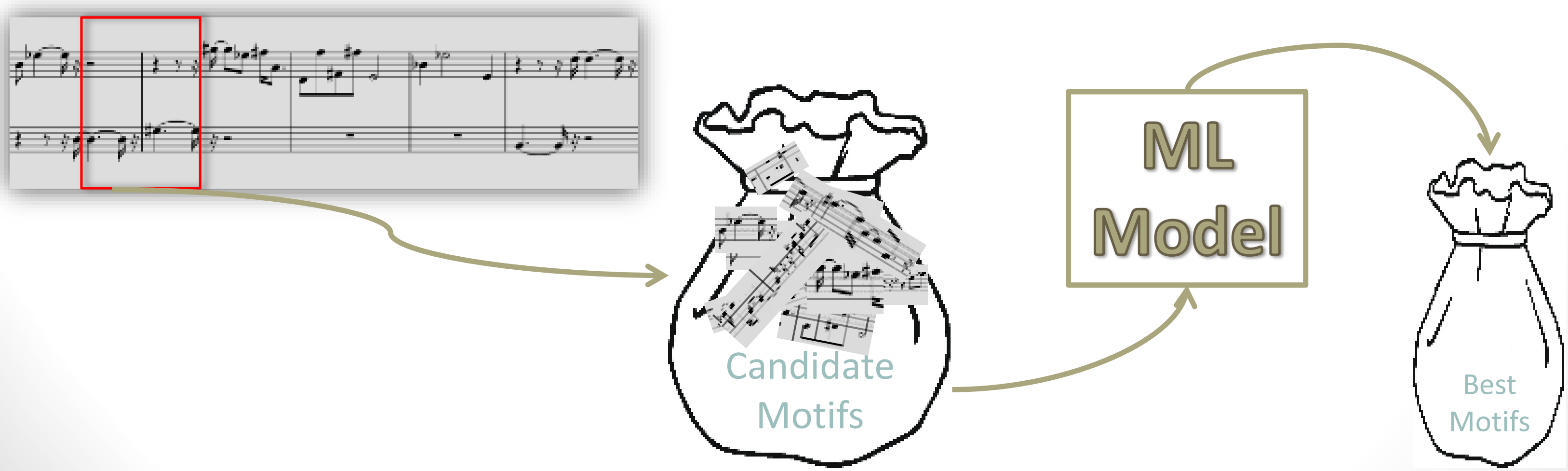
String of notes from edge/pitch detection



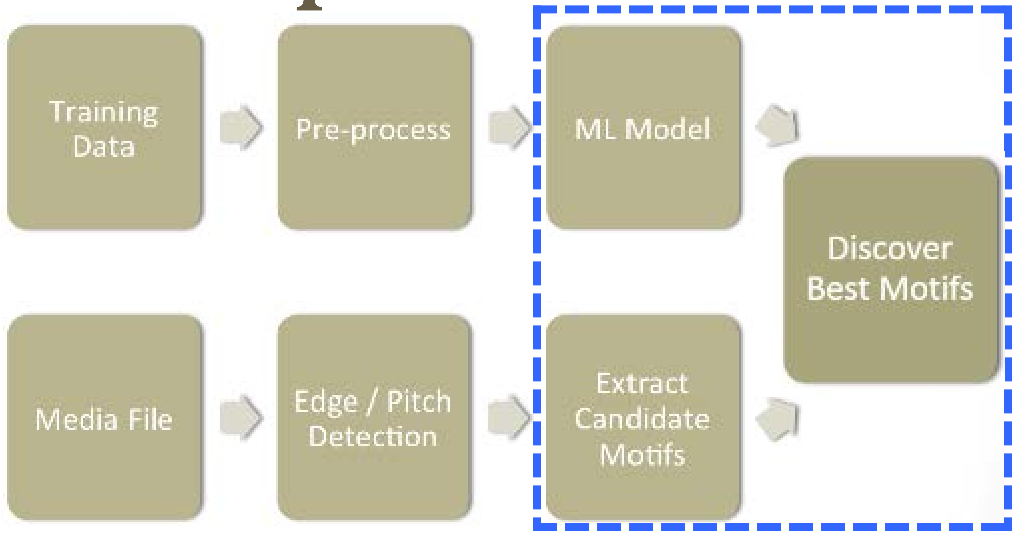
Project Description



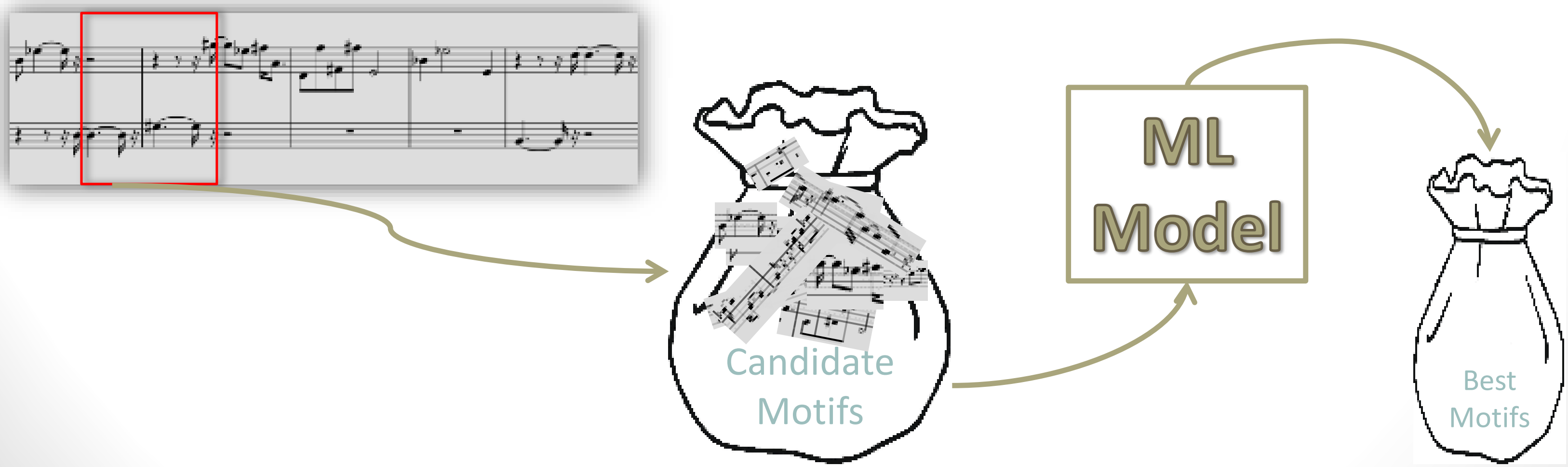
String of notes from edge/pitch detection



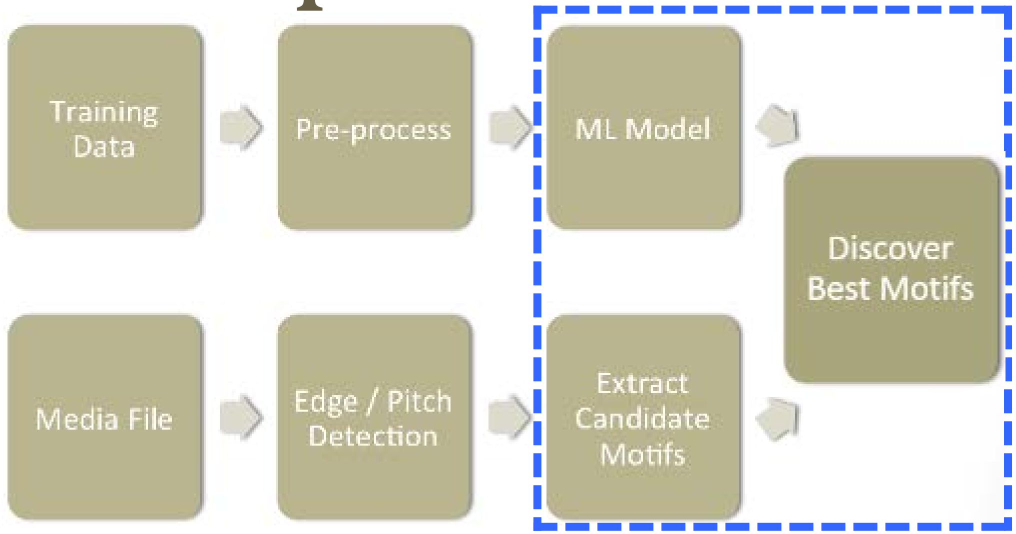
Project Description



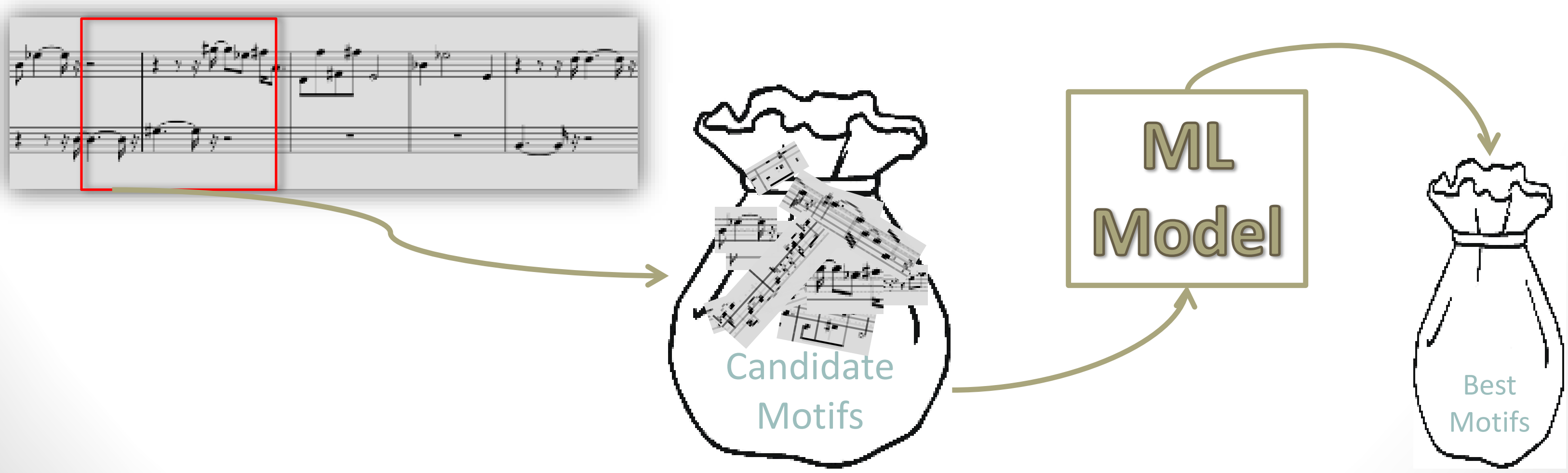
String of notes from edge/pitch detection



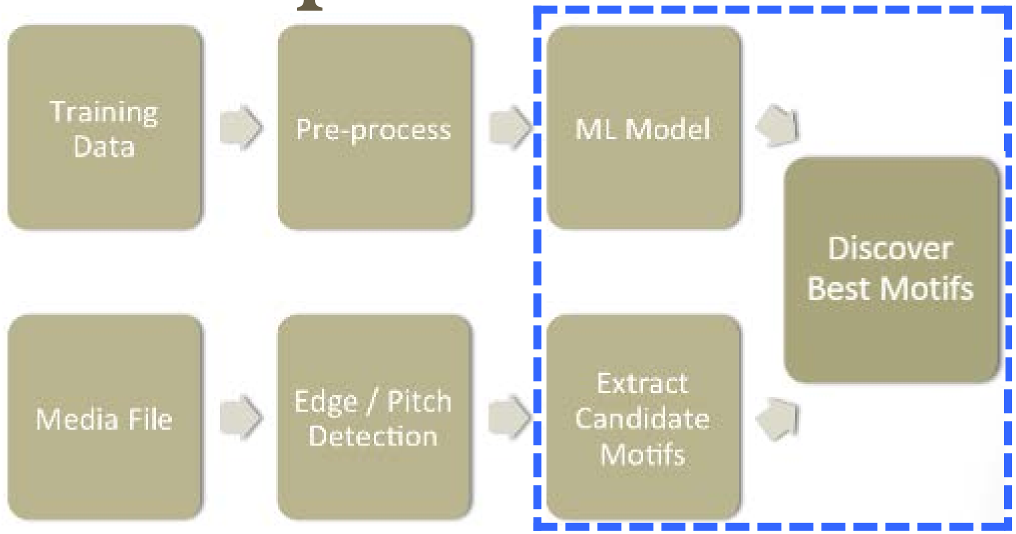
Project Description



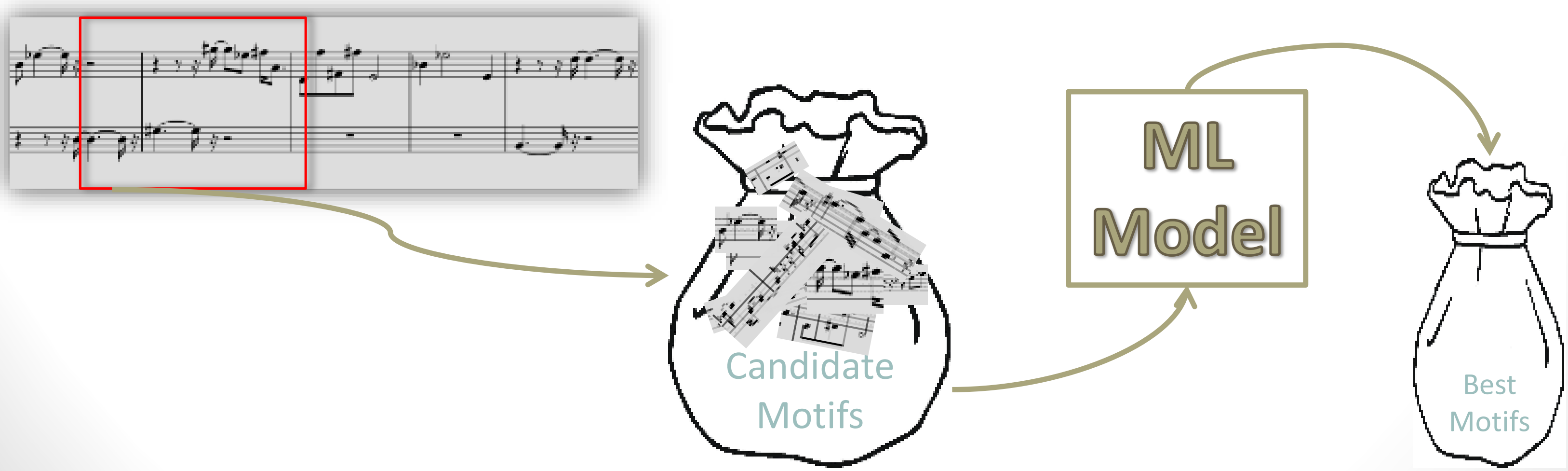
String of notes from edge/pitch detection



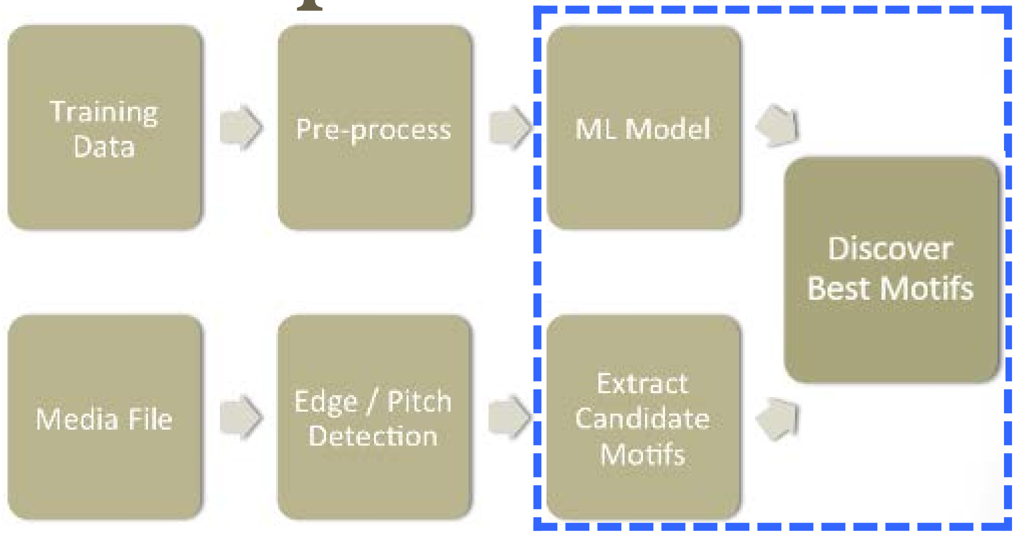
Project Description

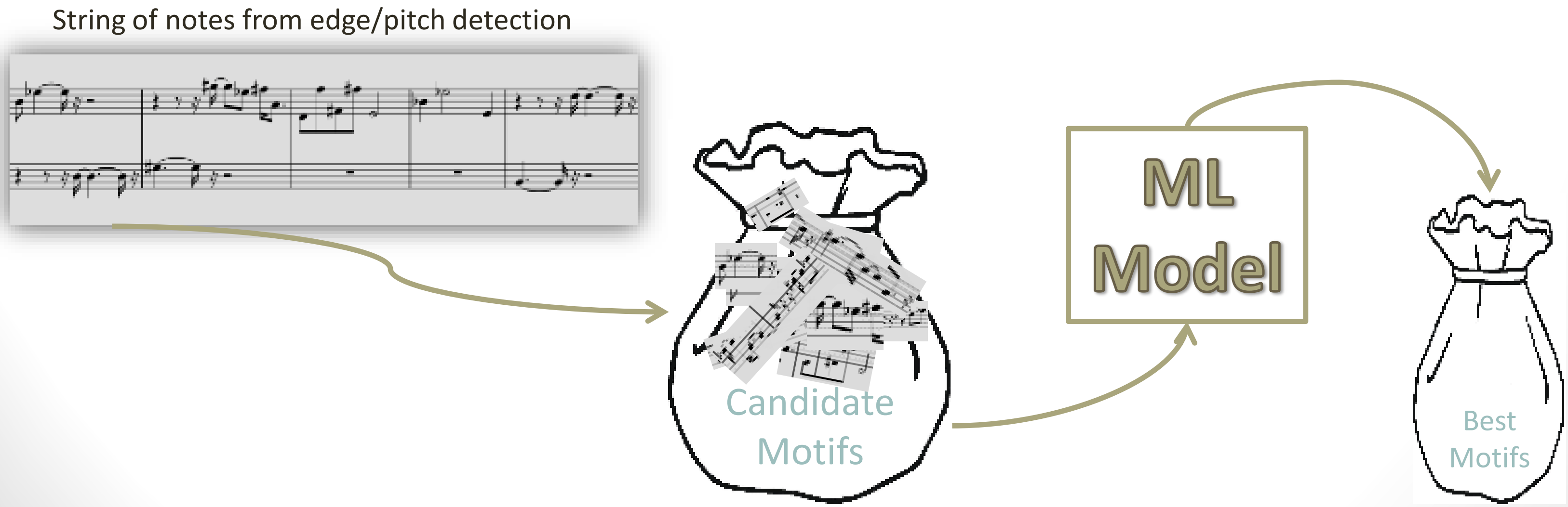


String of notes from edge/pitch detection

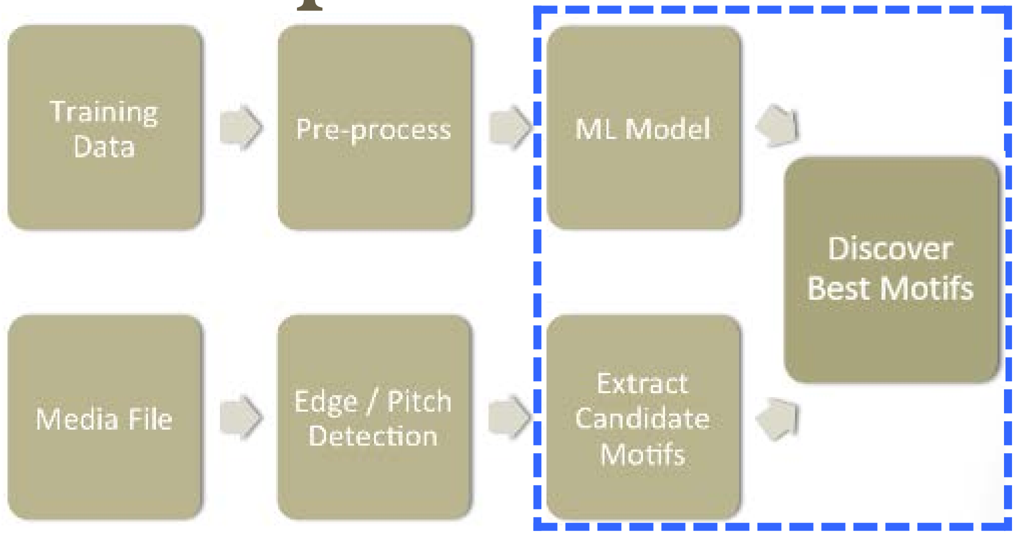


Project Description

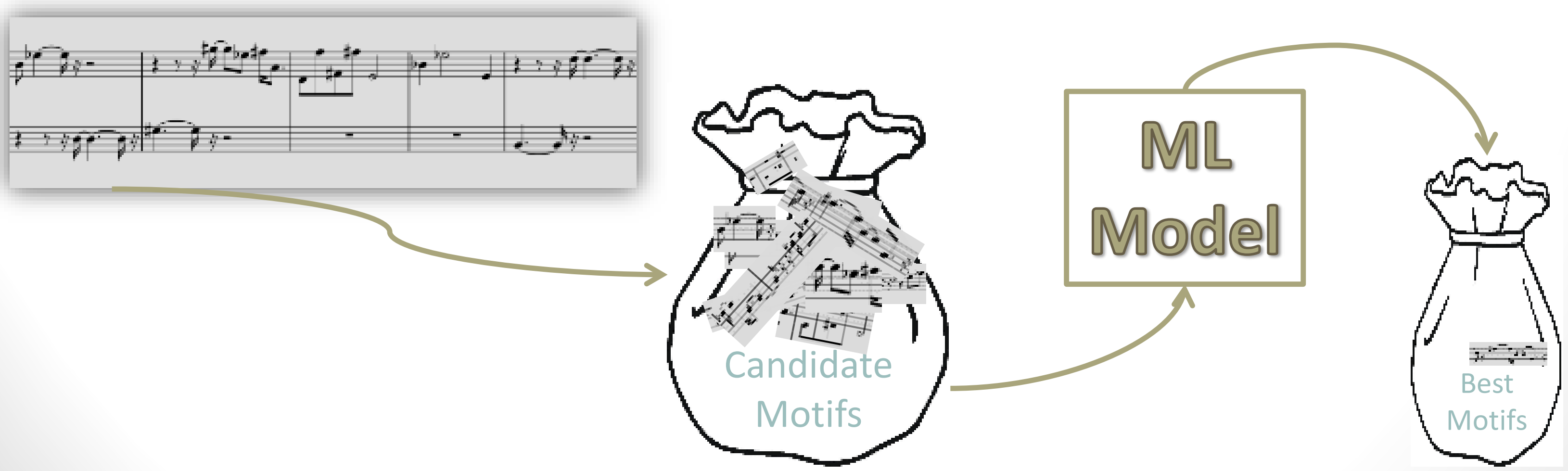




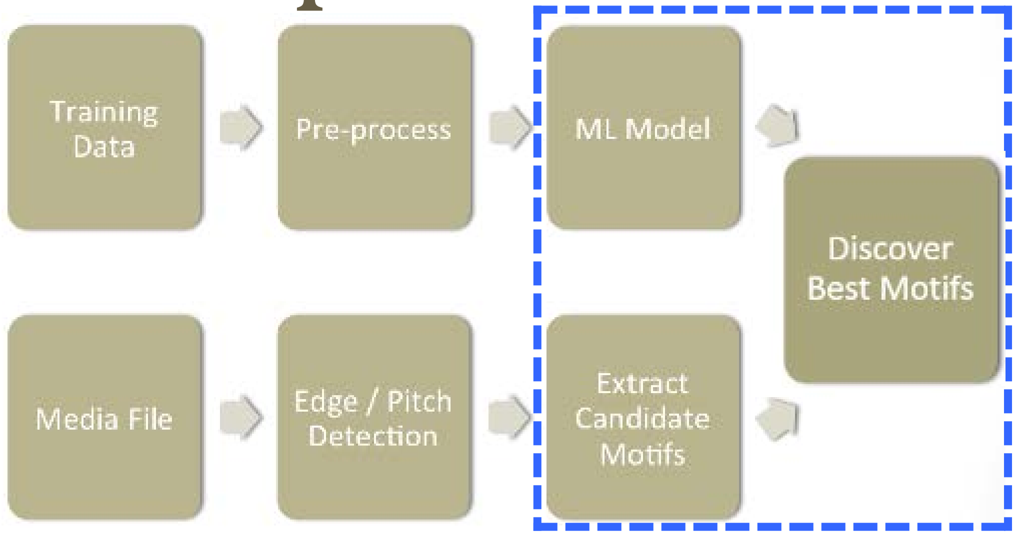
Project Description



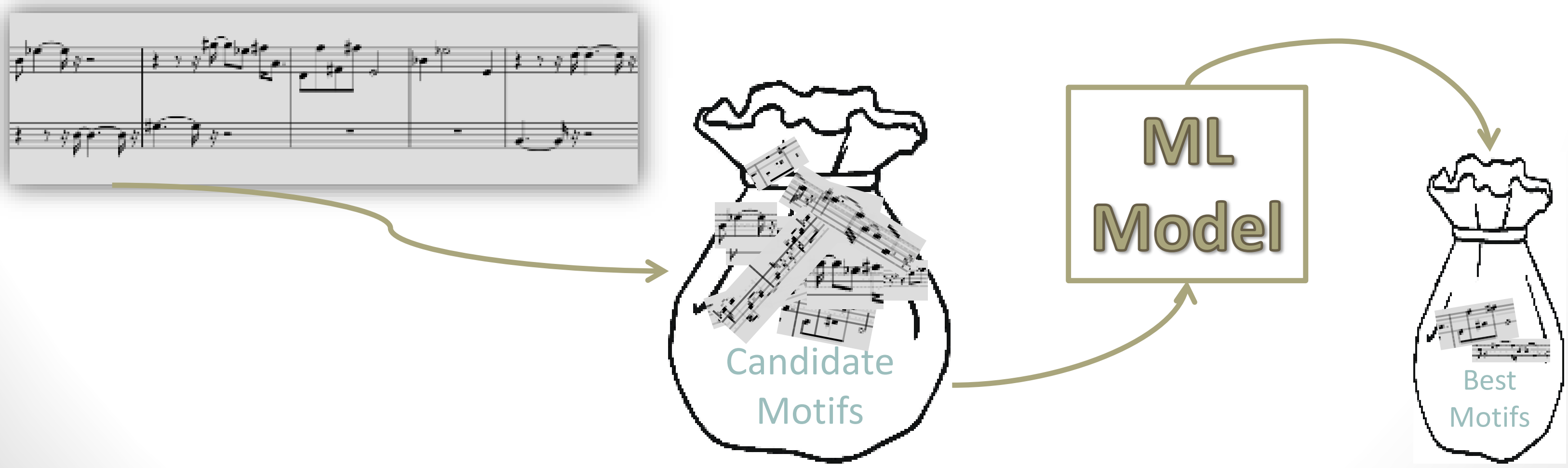
String of notes from edge/pitch detection



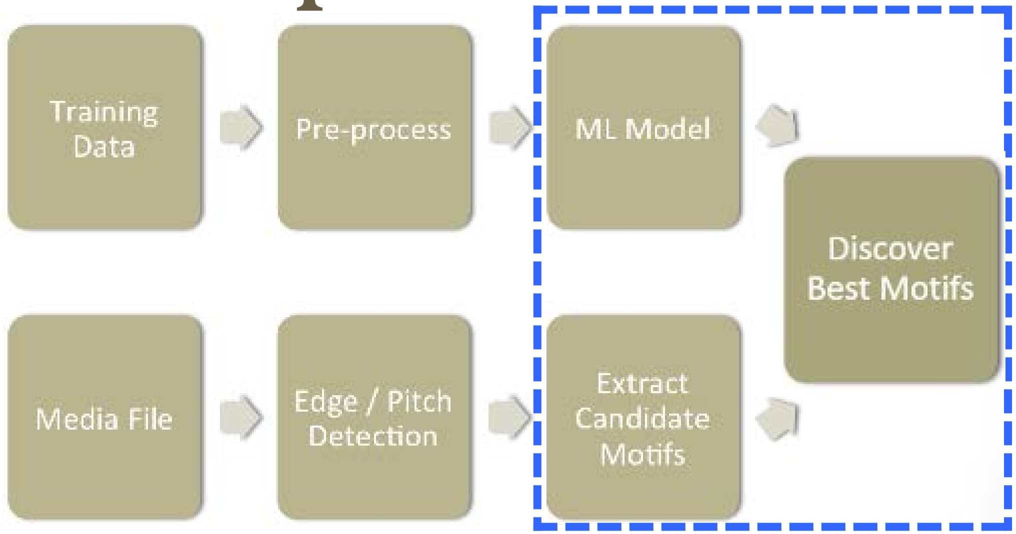
Project Description



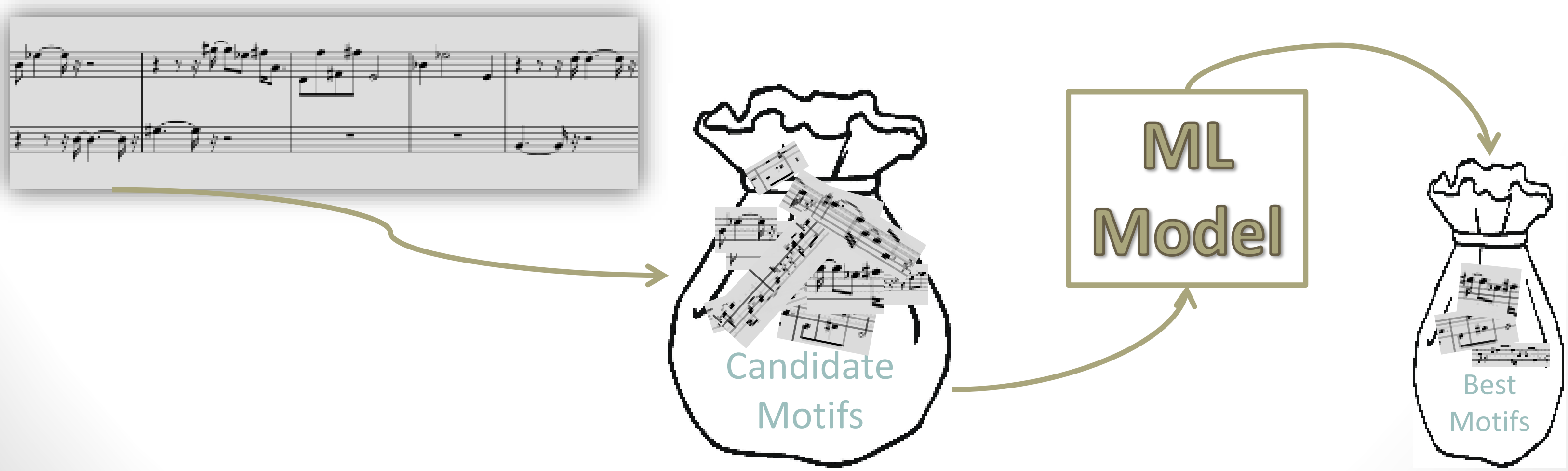
String of notes from edge/pitch detection



Project Description



String of notes from edge/pitch detection



Example Outputs

Input: Neverland.mp3

Discovered Motifs

LZMS

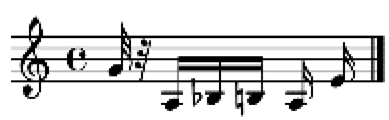
CTW

PPM

HMM

PST

LSTM



More examples available at axon.cs.byu.edu/motif-discovery

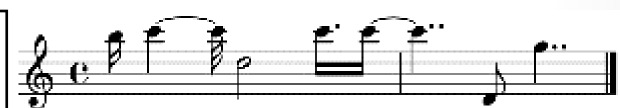
Example Outputs

Input: landscape.jpg

Discovered Motifs

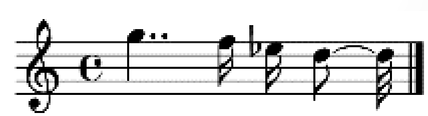
CTW

LZMS

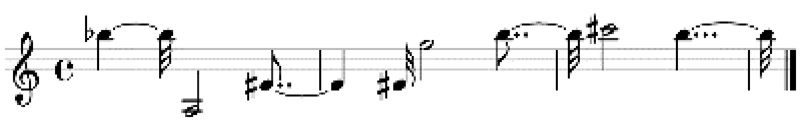


HMM

PPM

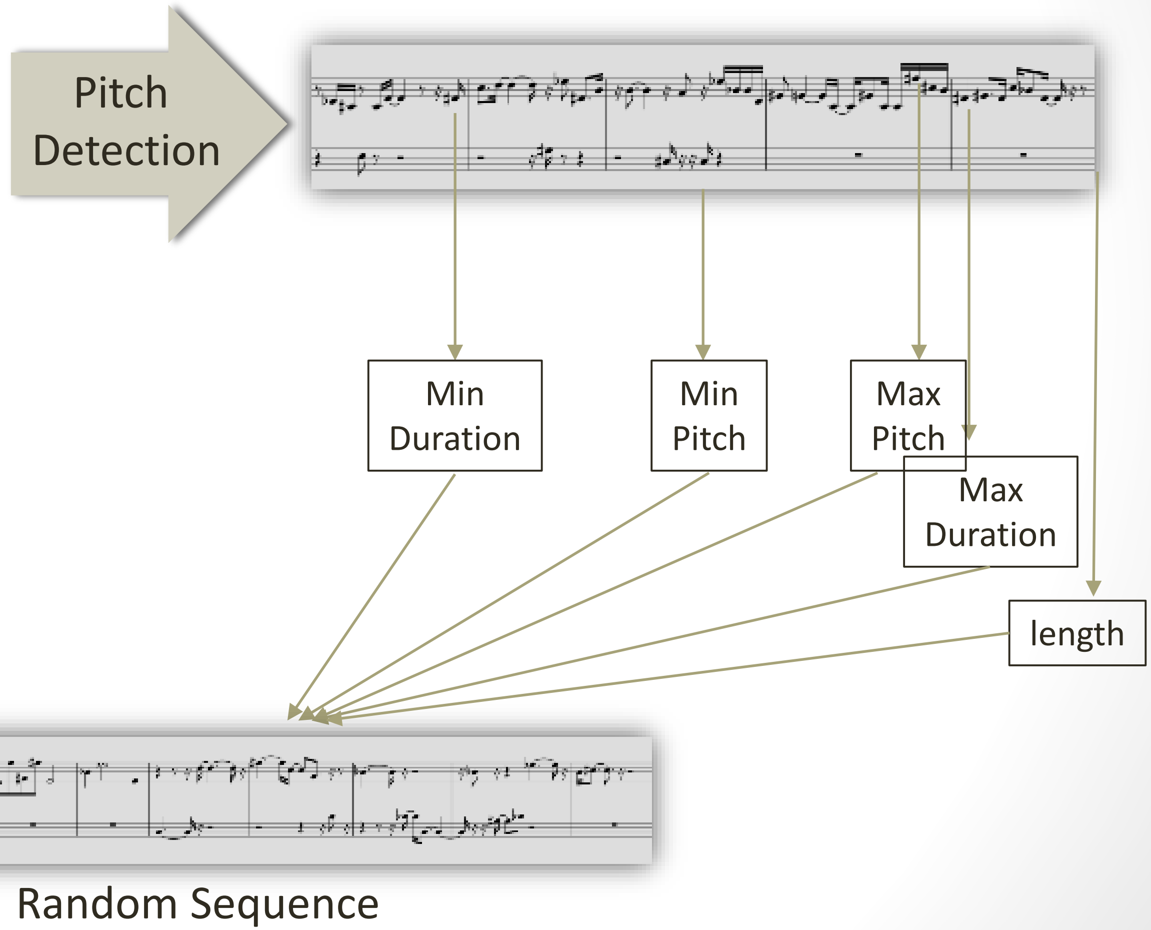


LSTM



PST

Preliminary Evaluation of Inspirational Sources



Results

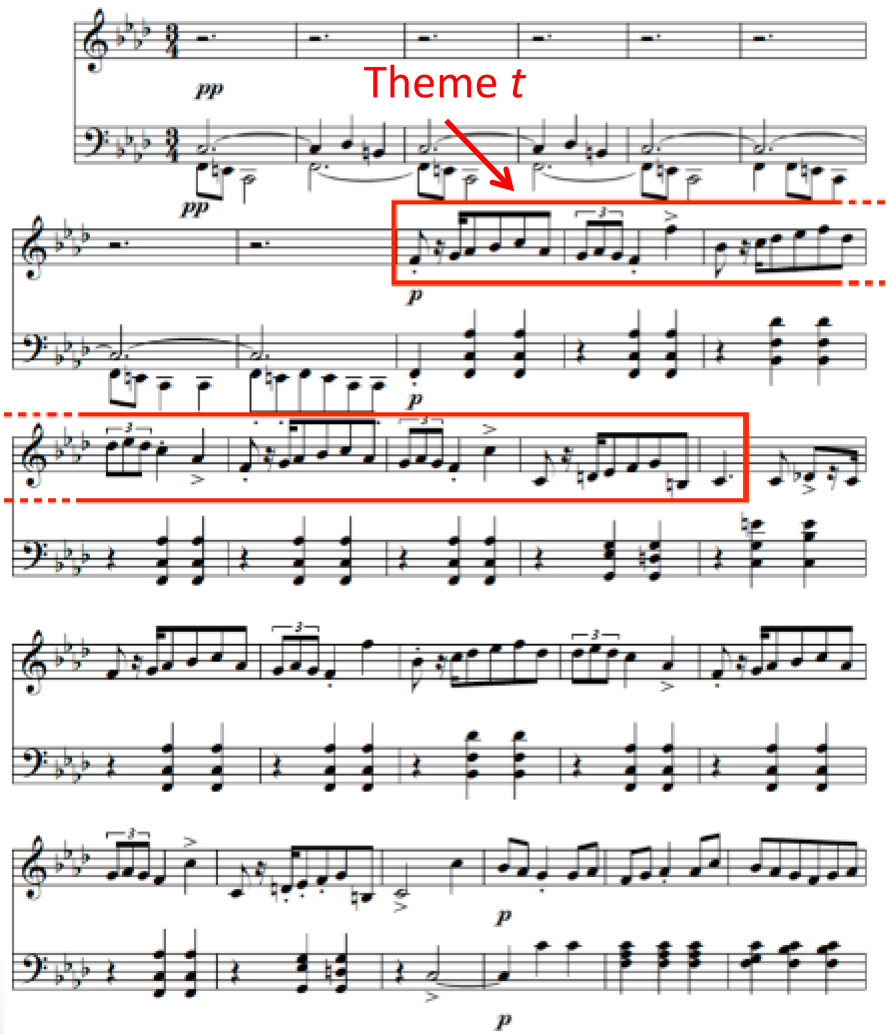
Inspirational Audio File Name	Pitch Entropy	Random Pitch Entropy	Rhythm Entropy	Random Rhythm Entropy
Reunion2005.wav	4.521	5.122	1.478	2.58
Neverland.wav	4.376	5.153	1.641	2.804
Birdsong.wav	4.835	5.156	3.317	5.154
ThunderAndRain.wav	4.465	5.152	3.196	6.283
SparklingWater.wav	5.002	5.151	0.54	2.321
TropicalRain.wav	4.994	5.164	2.485	4.083
PleasantBeach.wav	4.698	5.136	3.856	6.761
ChallengerDisasterAddress.wav	4.071	4.87	2.034	3.57
InauguralAddress.wav	4.865	5.162	1.914	5.037
MLKDream.wav	5.013	5.16	1.913	5.796
DarthVaderBreathing.wav	2.86	2.795	1.429	2.104
R2D2.wav	4.868	4.746	1.364	3.203
Lightsabers.wav	3.671	5.042	1.867	3.567
ChewbaccaRoar.wav	2.722	2.922	1.357	2.171
Blasters.wav	4.17	4.272	2.251	3.726
Average	4.342	4.734	2.043	3.944

Inspirational Image File Name	Pitch Entropy	Random Pitch Entropy	Rhythm Entropy	Random Rhythm Entropy
Motif.jpg	6.269	6.953	4.19	14.399
Fociz.jpg	6.451	6.999	4.095	15.437
Bioplazm2.jpg	6.743	6.988	4.201	15.369
LightPaintMix.jpg	5.989	6.869	4.922	14.487
Variation-Investigation.jpg	6.52	6.965	3.903	15.813
Pollock-Number5.jpg	6.099	6.79	3.75	12.737
Dali-ThePersistenceofMemory.jpg	6.115	6.684	4.634	13.662
Monet-ImpressionSunrise.jpg	5.073	6.583	4.486	13.813
DaVinci-MonaLisa.jpg	6.305	6.657	4.985	11.8
Vermeer-GirlWithaPearlEarring.jpg	6.465	6.869	4.844	14.156
Landscape.jpg	6.304	6.999	4.373	15.076
Stonehenge.jpg	5.739	6.374	4.851	14.787
River.jpg	6.252	6.869	5.057	14.994
Fish.jpg	5.59	6.882	4.547	15.104
Bird.jpg	5.837	6.227	5.655	14.012
Average	6.117	6.78	4.566	14.376

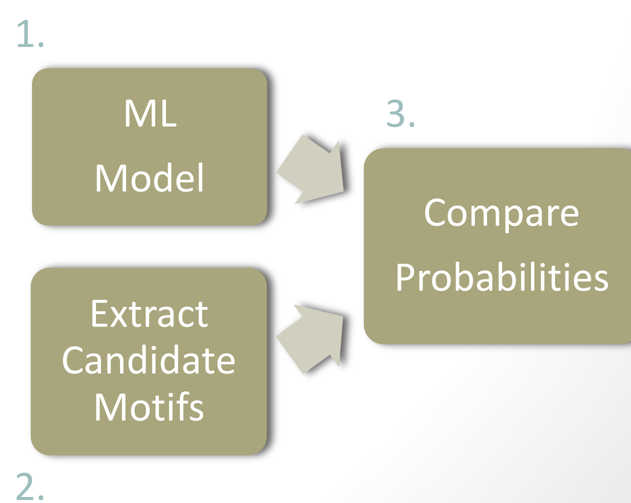
P-values: 0.175 0.0004 2.51 X 10^-5 1.36 X 10^-18

Changes to 0.003 when we remove the 3 shortest files

Evaluation of Motif Extraction Process

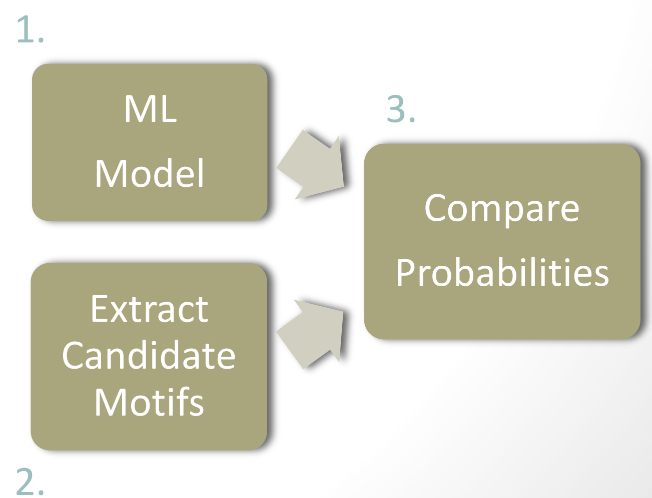


- For each theme *t* in validation set:
  1. Train ML model on all 9383 themes except *t*
  2. Extract all candidate motifs from *t*'s full score
  3. Mean probability of motifs inside of *t* should be higher than mean probability of motifs not in *t* (according to ML model)



Evaluation of Motif Extraction Process

- For each theme *t* in validation set:
  1. Train ML model on all 9383 themes except *t*
  2. Extract all candidate motifs from *t*'s full score
  3. Mean probability of motifs inside of *t* should be higher than mean probability of motifs not in *t* (according to ML model)



Results

$$Q(S|model) = \frac{\sum_{m \in S} norm(|m|)Pr(m|model)}{|S|}$$

$$U = \frac{Q(C_t|model) - Q(C_{-t}|model)}{\min\{Q(C_t|model), Q(C_{-t}|model)\}}$$

U values

Score File Name	CTW	HMM	LSTM	LZMS	PPM	PST	Average
BachBook1Fugue15	4.405	4.015	3.047	2.896	11.657	4.951	5.162
BachInvention12	-2.585	-5.609	26.699	1.078	0.534	13.191	5.551
BeethovenSonata13-2	1.065	-0.145	7.769	8.876	4.973	9.182	5.287
BeethovenSonata6-3	-0.715	-5.320	2.874	0.832	1.283	4.801	0.626
ChopinMazurka41-1	6.902	0.808	-7.690	3.057	18.965	-24.363	-0.387
Corelli5-8-2	-6.398	-1.270	-0.692	-2.395	-1.166	1.690	-1.705
Grieg43-2	2.366	1.991	-2.622	0.857	8.800	-7.740	0.609
Haydn33-3-4	14.370	2.370	1.189	6.155	8.475	0.841	5.567
Haydn64-6-2	1.266	2.560	-1.092	0.855	1.809	-0.133	0.878
LisztBallade2	-0.763	-0.610	-1.754	-0.046	1.226	0.895	-0.175
MozartK331-3	0.838	0.912	3.829	0.756	3.222	5.413	2.495
MozartK387-4	-4.227	-0.082	-91.960	-2.127	-3.453	-31.614	-22.244
SchubertImprGFlat	49.132	3.169	0.790	8.985	59.336	1.122	20.422
SchumannSymph3-4	0.666	2.825	-2.154	0.289	1.560	-6.830	-0.607
Vivaldi3-6-1	7.034	2.905	0.555	7.055	9.633	-0.367	4.469
Average	4.890	0.568	-4.081	2.475	8.457	-1.931	

Evaluation of Structural Quality of Motifs

1. Measure predictability

Actual Themes Entropy < Discovered Motifs Entropy < Candidate Motifs Entropy



2. Measure innovation

Actual Themes

ML Model

.145

>

Discovered Motifs

ML Model

.00274

Results

$$Q(S|model) = \frac{\sum_{m \in S} norm(|m|)Pr(m|model)}{|S|}$$

$$R = \frac{Q(A|model) - Q(E|model)}{\min\{Q(A|model), Q(E|model)\}}$$

Entropy and R Values

Bioplazm2.jpg	CTW	HMM	LSTM	LZMS	PPM	PST	Average
training motif pitches	1.894	1.979	1.818	1.816	1.711	1.536	1.793
discovered motif pitches	2.393	2.426	1.944	1.731	2.057	1.759	2.052
candidate motif pitches	2.217	2.328	2.097	2.104	1.958	1.784	2.081
training motif rhythms	1.009	1.051	0.976	0.970	0.927	0.822	0.959
discovered motif rhythms	2.110	2.295	1.789	2.212	0.684	1.515	1.767
candidate motif rhythms	2.387	2.466	2.310	2.309	2.132	1.934	2.256
R	7.567	13.296	20.667	4.603	-0.276	7.643	8.917

Landscape.jpg	CTW	HMM	LSTM	LZMS	PPM	PST	Average
training motif pitches	1.894	1.979	1.818	1.816	1.711	1.536	1.793
discovered motif pitches	1.974	2.074	2.143	1.833	2.027	1.675	1.954
candidate motif pitches	2.429	2.531	2.598	2.341	2.271	2.028	2.367
training motif rhythms	1.009	1.051	0.976	0.970	0.927	0.822	0.959
discovered motif rhythms	1.984	1.863	2.175	1.983	0.727	1.455	1.698
candidate motif rhythms	1.549	1.712	1.810	1.509	1.396	1.329	1.551
R	0.805	0.236	1.601	0.429	4.624	1.283	1.496

Results

$$Q(S|model) = \frac{\sum_{m \in S} norm(|m|)Pr(m|model)}{|S|}$$

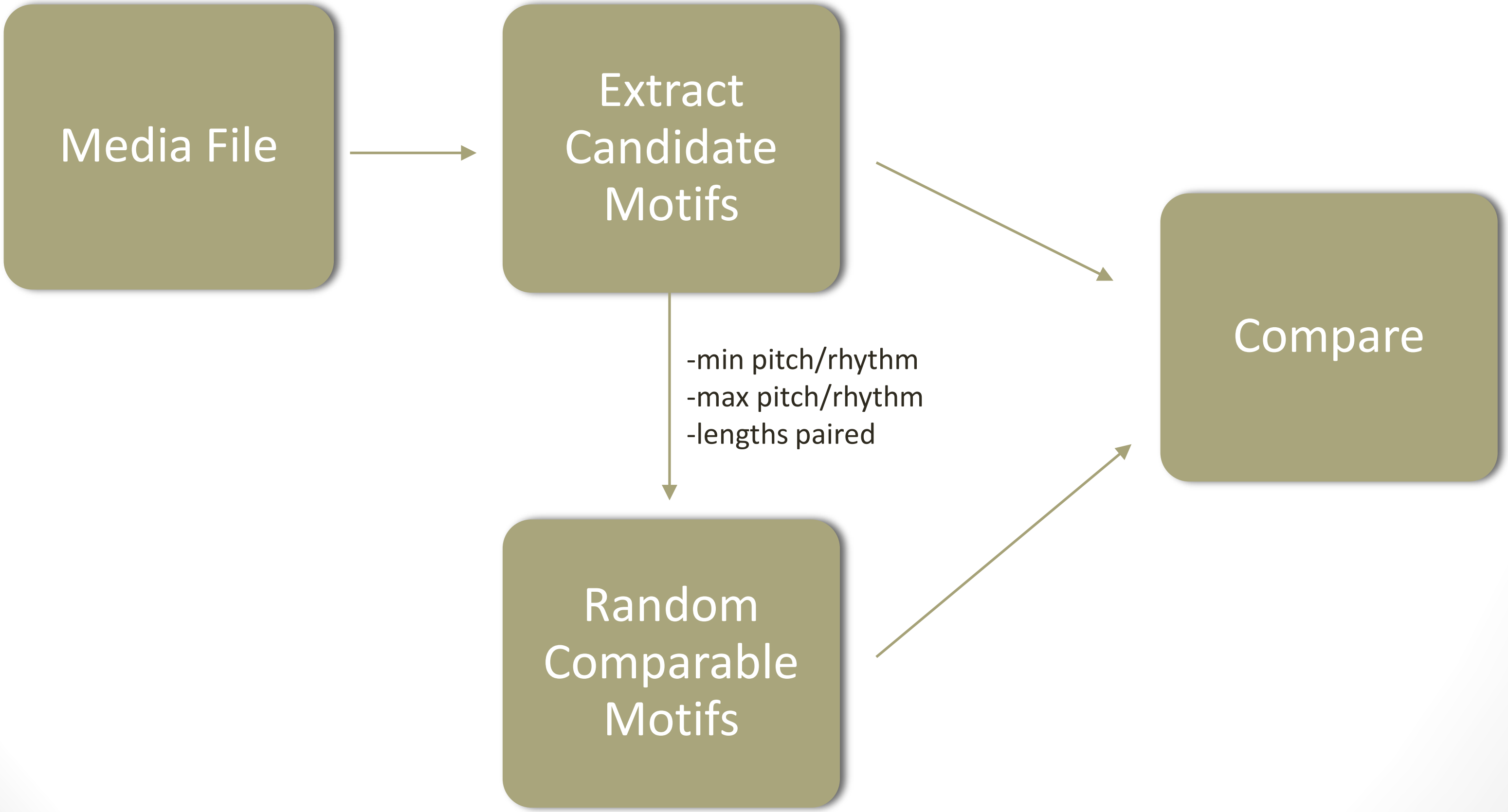
$$R = \frac{Q(A|model) - Q(E|model)}{\min\{Q(A|model), Q(E|model)\}}$$

Entropy and R Values

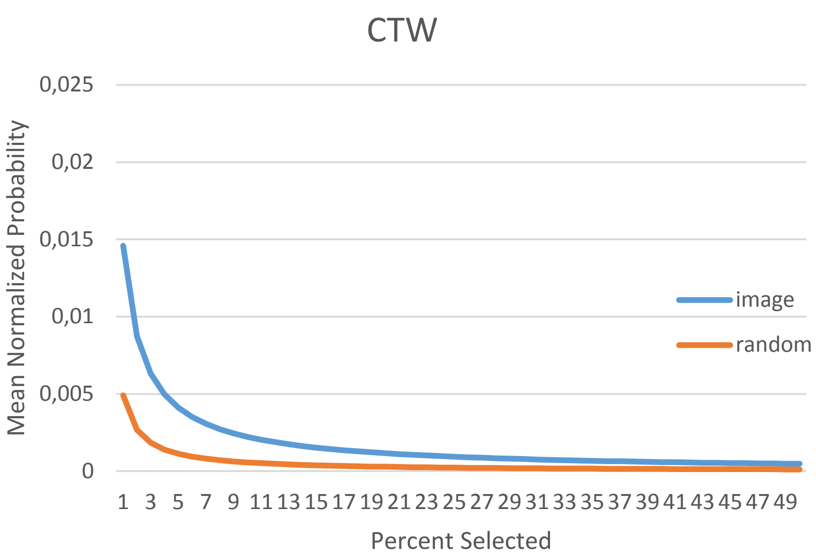
Lightsabers.wav	CTW	HMM	LSTM	LZMS	PPM	PST	Average
training motif pitches	1.894	1.979	1.818	1.816	1.711	1.536	1.793
discovered motif pitches	2.076	1.884	1.881	1.652	2.024	1.586	1.850
candidate motif pitches	2.225	2.097	2.217	1.876	2.115	1.755	2.048
training motif rhythms	1.009	1.051	0.976	0.970	0.927	0.822	0.959
discovered motif rhythms	1.534	1.309	2.024	1.623	0.860	1.225	1.429
candidate motif rhythms	1.540	1.524	1.541	1.502	1.548	1.276	1.489
R	5.637	0.793	27.227	4.812	6.768	7.540	8.796

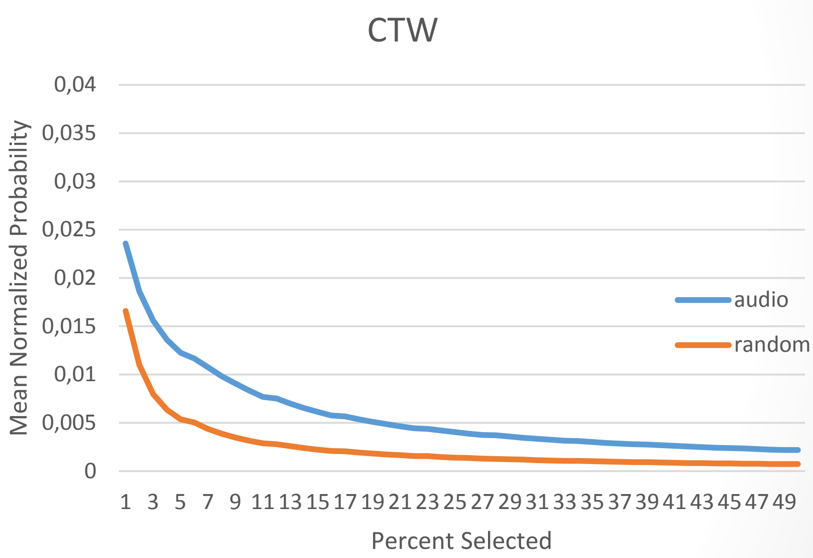
Neverland.wav	CTW	HMM	LSTM	LZMS	PPM	PST	Average
training motif pitches	1.894	1.979	1.818	1.816	1.711	1.536	1.793
discovered motif pitches	1.823	2.480	2.132	1.773	1.997	1.701	1.984
candidate motif pitches	2.153	2.248	2.250	2.141	2.242	1.839	2.146
training motif rhythms	1.009	1.051	0.976	0.970	0.927	0.822	0.959
discovered motif rhythms	1.550	1.587	1.560	1.779	0.289	1.128	1.315
candidate motif rhythms	1.472	1.469	1.471	1.477	1.469	1.226	1.431
R	1.520	10.163	24.968	4.283	0.257	6.865	8.010

Comparison of Media Inspiration and Random Inspiration



Comparison of Media Inspiration and Random Inspiration





Subset Training

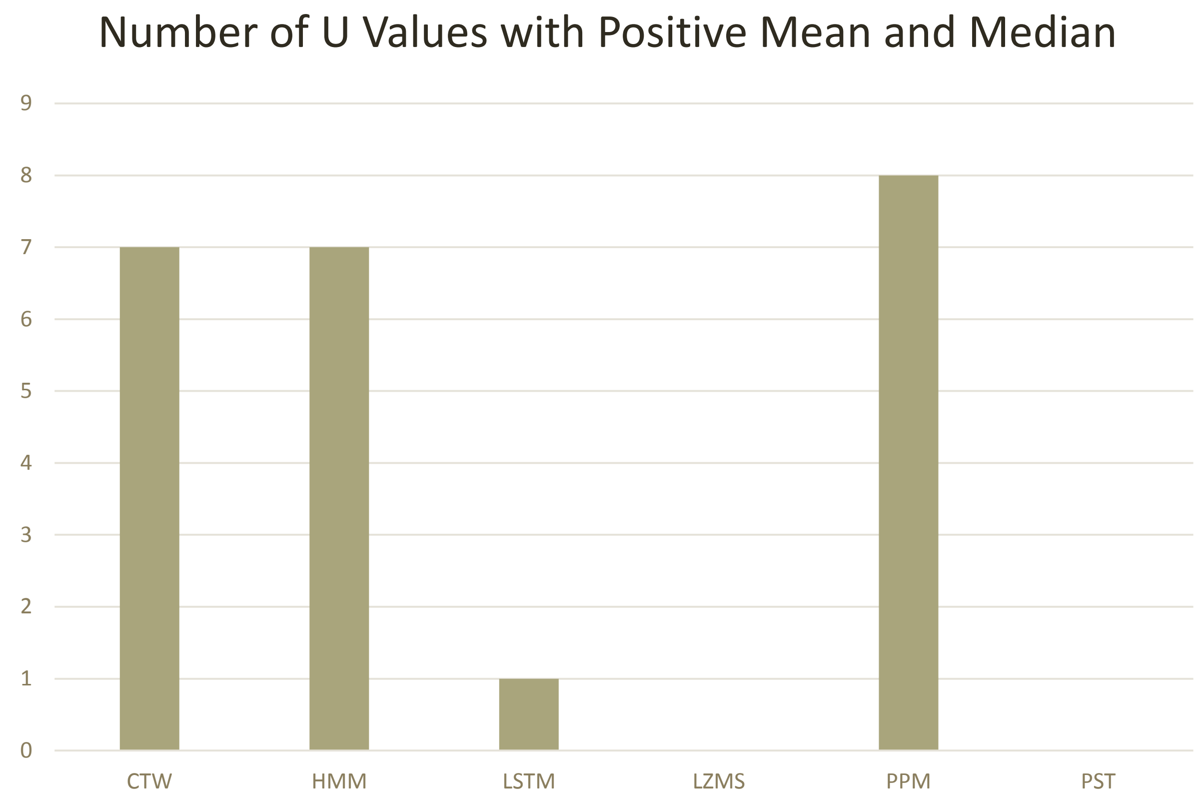
How does Bach relate to the others?

Measure U values after training on only a subset of the training data.

U values after training on only themes by Bach are included below.

Score File Name	CTW	HMM	LSTM	LZMS	PPM	PST	Mean	Median
BachBook1Fugue15	5.421	11.259	3.742	7.463	9.198	7.335	7.403	7.399
BachInvention12	0.375	-2.840	9.200	-1.859	0.626	25.351	5.142	0.500
BeethovenSonata13-2	13.057	0.278	2.190	4.217	2.511	2.490	4.124	2.500
BeethovenSonata6-3	-0.555	-1.706	5.588	-0.933	0.090	6.614	1.516	-0.233
ChopinMazurka41-1	28.394	5.482	-19.081	1.020	9.915	-377.693	-58.661	3.251
Corelli5-8-2	-40.103	-9.672	-0.018	-11.364	-17.721	5.819	-12.176	-10.518
Grieg43-2	3.399	7.232	-1.365	-0.385	8.187	-9.831	1.206	1.507
Haydn33-3-4	21.489	12.861	1.044	28.487	23.981	4.451	15.385	17.175
Haydn64-6-2	7.344	4.420	-1.303	1.864	7.226	-2.316	2.872	3.142
LisztBallade2	0.426	-0.414	-1.268	0.097	-0.234	-0.170	-0.261	-0.202
MozartK331-3	0.352	-0.445	7.057	-0.414	1.325	12.214	3.348	0.839
MozartK387-4	-3.223	-2.825	-48.039	-495.799	-20.821	-69.631	-106.723	-34.430
SchubertImprGFlat	-0.764	-0.146	7.800	5.716	3.255	9.671	4.255	4.486
SchumannSymph3-4	-14.501	-1.129	-7.549	-191.425	-23.026	-19.069	-42.783	-16.785
Vivaldi3-6-1	-0.013	-2.725	-5.421	-0.394	-0.072	-1.076	-1.617	-0.735
Mean	1.406	1.309	-3.161	-43.581	0.296	-27.056		
Median	0.375	-0.414	-0.018	-0.385	1.325	2.490		

Subset Training



Conclusion

- Innovation and value shown through *U*, *R*, and entropy
- CTW and PPM stand out as the best ML models for motif discovery
- Media inspiration is more efficient than random generation

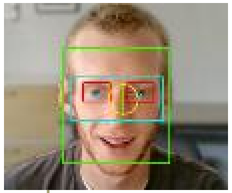
Future Work

- Motif discovery as starting point for full-length compositions
- Global structure
- Additional modes of inspiration (text, video, sleep data)
- Inspirational sources combined over time

CARL: Composer of Anomalous Reactive Lead Sheets

- Example using this motif discovery system as a starting point
- ABAB form, popular style

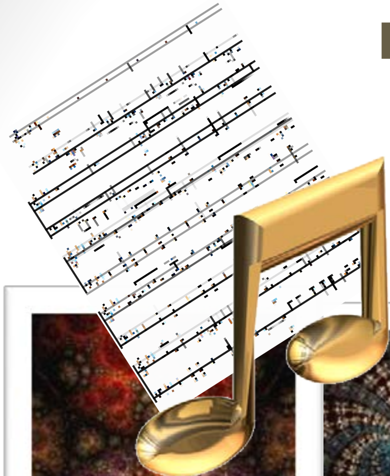
Face Detection



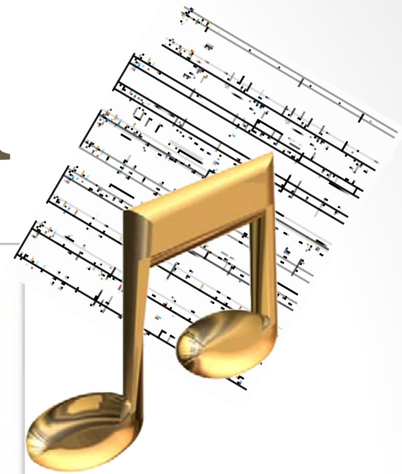
Motif Discovery

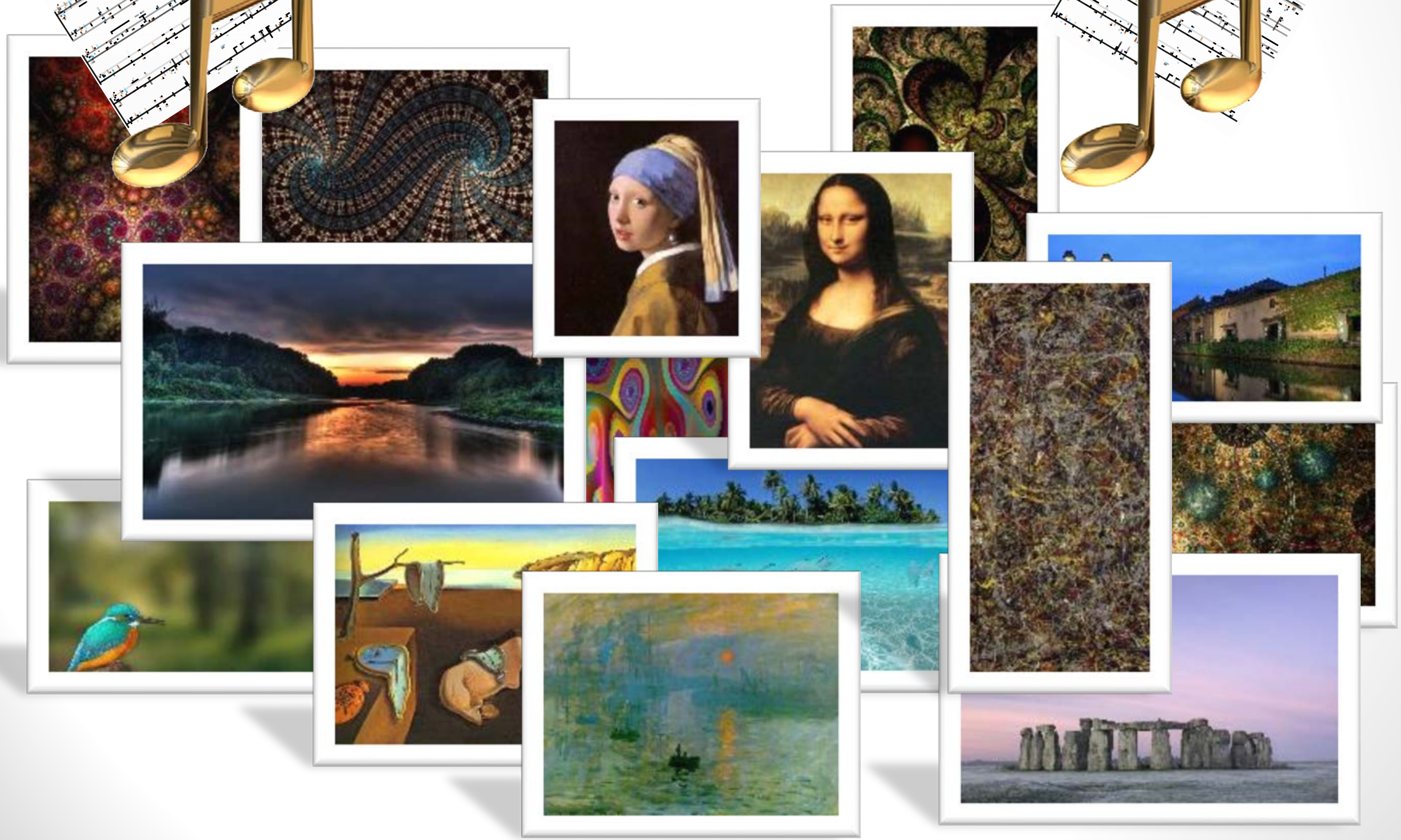
Melody Generation

Harmonization



Thank You





Human Composition Approaches

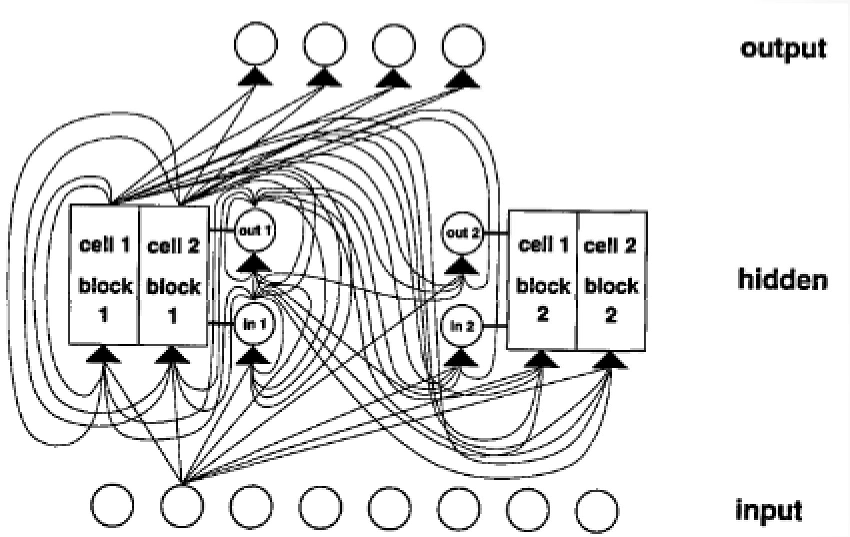
- Composers from period of common practice serve both goals
  - Beethoven – pushes classical to romantic
  - Wagner – bridges gap to atonality
  - Schoenberg – pushes atonality to a theoretical maximum
- Composers seek musical and non-musical inspiration
  - Olivier Messiaen – birdsong
  - Claude Debussy – nature, poetry, paintings
  - Franz Liszt – programme music

Related Work

- Markov models
  - Variable order
    - Dubnov et al. apply prediction suffix trees to music
    - Begleiter et al. compare multiple VMMs such as PST, CTW, PPM
  - Random fields
    - Lavrenko and Pickens apply to polyphonic music
  - Hierarchical/hidden
    - Weiland et al. use HHMMs to capture long term dependencies
- Neural networks
  - Long short-term memory (LSTM)
    - Solves vanishing gradient problem in RNNs
    - Eck and Schmidhuber apply to blues music
  - Adaptive Resonance Theory (ART)
    - Smith and Garnett capture medium-level structures using three hierarchical levels

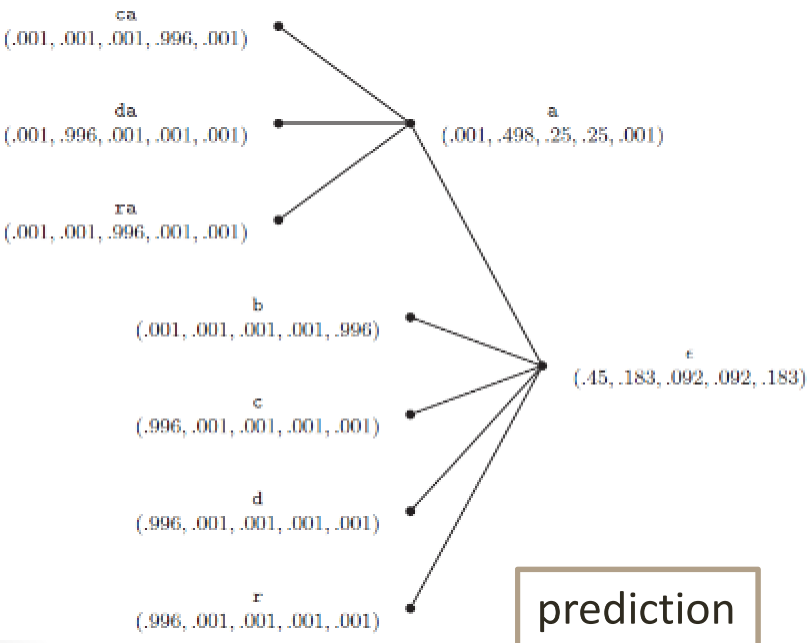
Related Work

Neural Networks



long
short-term
memory

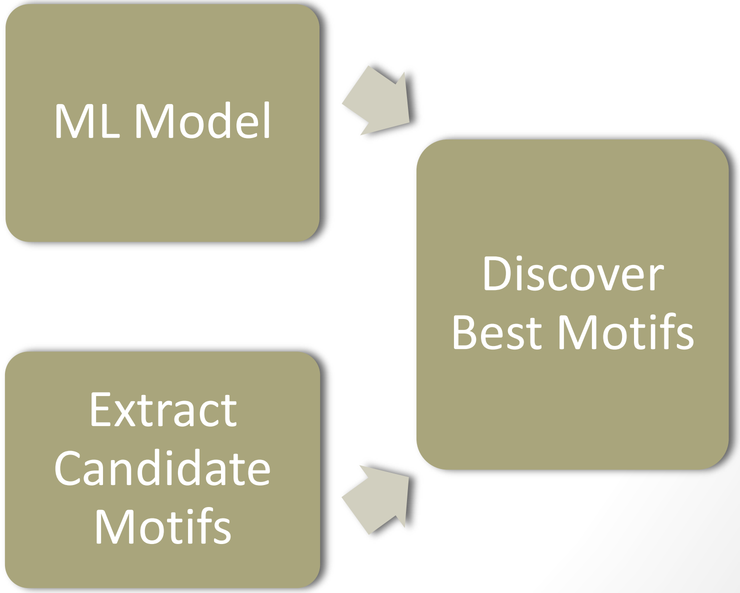
Markov Models



prediction
suffix
tree

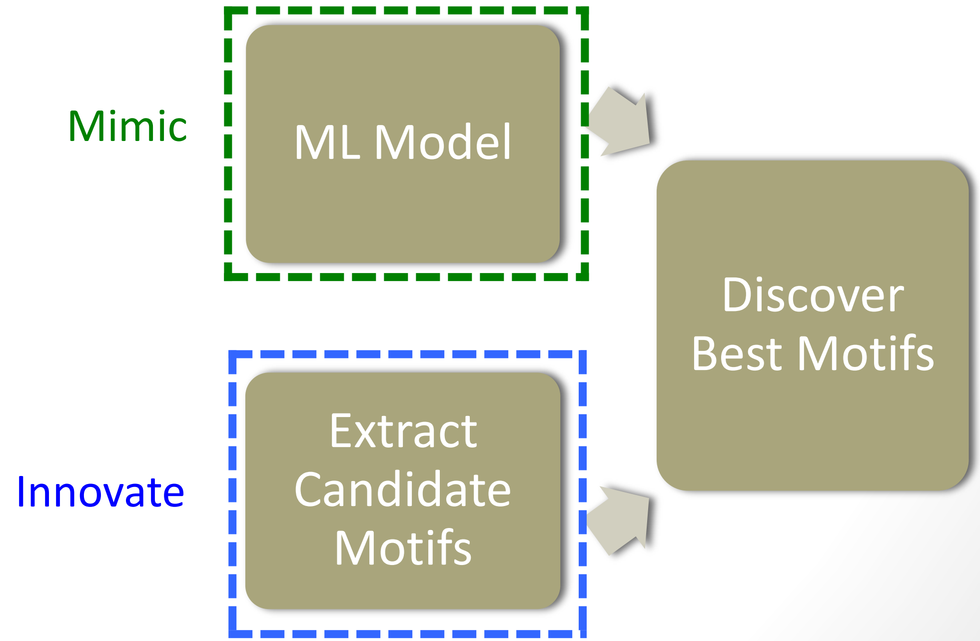
Our Approach

1. Train a machine learning (ML) model
2. Extract candidate motifs from an inspirational source
3. Use ML model to select motifs that most resemble the training data



Our Approach

1. Train a machine learning (ML) model
2. Extract candidate motifs from an inspirational source
3. Use ML model to select motifs that most resemble the training data



Length Normalization

Since shorter motifs are naturally more probable than longer motifs, an additional normalization step is taken in Algorithm 2. We would like each motif length to have equal probability:

$$P_{equal} = \frac{1}{(l\_max - l\_min + 1)} \tag{3.2}$$

Since the probability of a generative model emitting a candidate motif of length l is

$$P(l) = \sum_{m \in C, |m| = l} Pr(m|model) \tag{3.3}$$

we introduce a length-dependent normalization term that equalizes the probability of selecting motifs of various lengths.

$$norm(l) = \frac{P_{equal}}{P(l)} \tag{3.4}$$

Extract Candidate Motifs

String of notes from edge/pitch detection

Algorithm 1 *extract_candidate_motifs*

Input: $\boxed{notes}$, l_min, l_max

$candidate\_motifs \leftarrow \{\}$

for $l\_min \leq l \leq l\_max$ **do**

  for $0 \leq i \leq |notes| - l$ **do**

    $motif \leftarrow (notes_i, notes_{i+1}, \ldots, notes_{i+l-1})$

    $candidate\_motifs \leftarrow candidate\_motifs \cup motif$

return $candidate\_motifs$

Discover Best Motifs

Algorithm 2 *discover_best_motifs*

Input: $notes$, $model$, num_motifs, l_min, l_max
$C \leftarrow extract\_candidate\_motifs(notes, l\_min, l\_max)$
$best\_motifs \leftarrow \{\}$
while $|best\_motifs| < num\_motifs$ **do**
    $m^* \leftarrow \underset{m \in C}{\operatorname{argmax}}[norm(|m|)Pr(m|model)]$
    $best\_motifs \leftarrow best\_motifs \cup m^*$
    $C \leftarrow C - \{m^*\}$
return $best\_motifs$

Evaluation of Motif Extraction Process

- Test set = hand-picked scores and their matching themes
- Extract candidate motifs C from score, part by part, in a similar fashion performed on media files
- Split C into two disjoint sets
  - C_t = motifs that are subsets of the score's theme
  - C_-t = motifs that are not subsets of the score's theme

$$Q(S|model) = \frac{\sum_{m \in S} norm(|m|) Pr(m|model)}{|S|}$$

$$U = \frac{Q(C_t|model) - Q(C_{-t}|model)}{\min\{Q(C_t|model), Q(C_{-t}|model)\}}$$

Evaluation of Motifs

1. Compare entropy of actual music themes (A), extracted motifs (E), and candidate motifs (C)
   - Looking for entropy(A) < entropy(E) < entropy(C)
2. Measure level of innovation of extracted motifs (E) against actual music themes (A)

$$Q(S|model) = \frac{\sum_{m \in S} norm(|m|)Pr(m|model)}{|S|}$$

$$R = \frac{Q(A|model) - Q(E|model)}{\min\{Q(A|model), Q(E|model)\}}$$