

Carnegie
Mellon
University



MultiComp
Lab

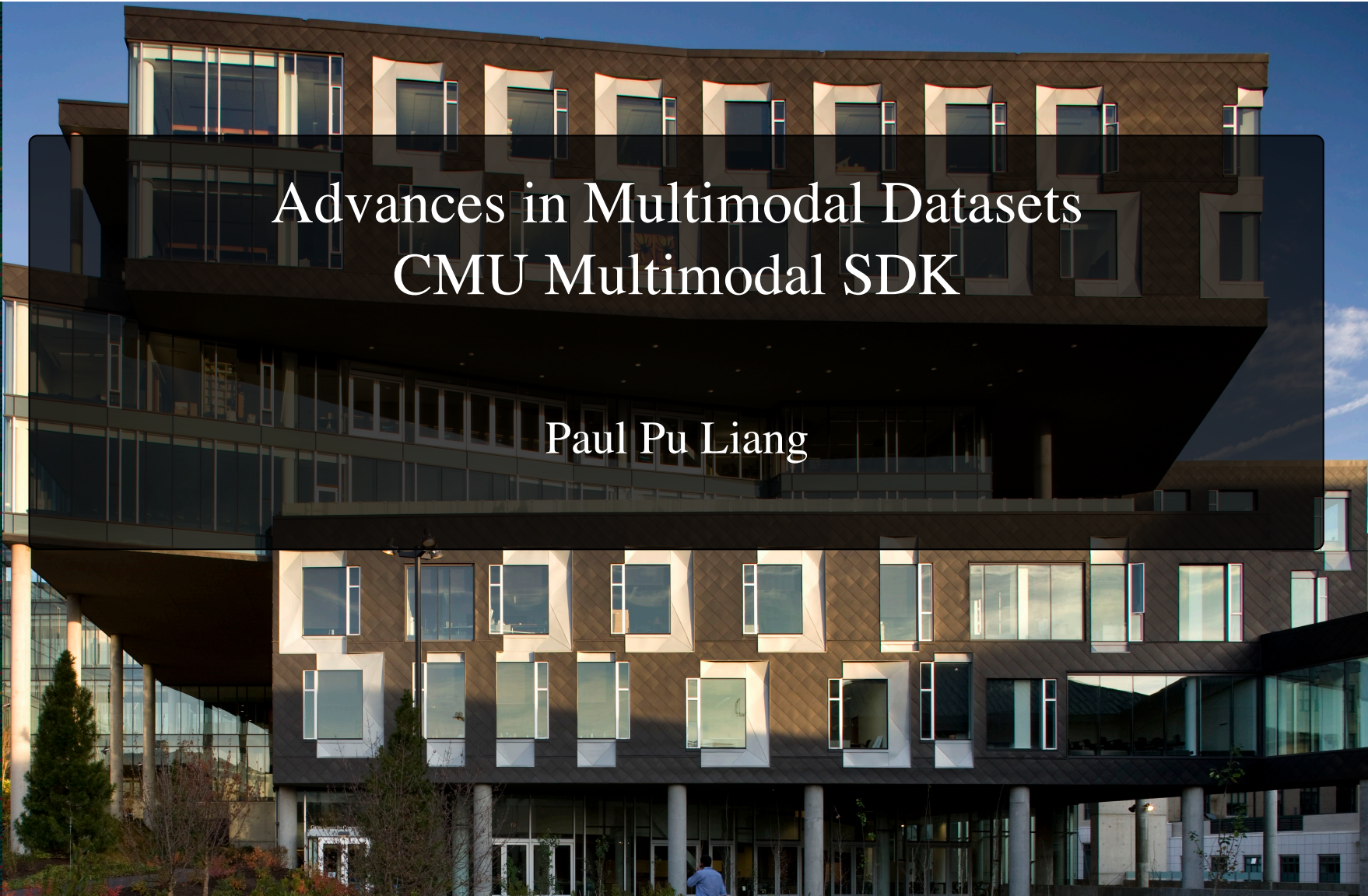
ML
MACHINE LEARNING
DEPARTMENT



USC University of
Southern California

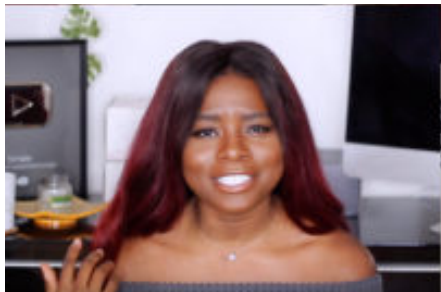
Advances in Multimodal Datasets CMU Multimodal SDK

Paul Pu Liang



Advancements in Modeling

- Human multimodal language has seen a surge of interest in fine-grained modeling.



I liked the movie I saw last night

Overall smile

Excited voice

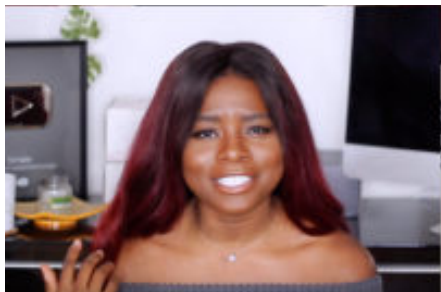
- Wouldn't account for changes or temporal co-occurrences.

Advancements in Modeling

- New approaches rely on word-level alignments to build a co-occurrence model (Gu et al. ACL 2018, Zadeh et al. AAAI 2018, Chen et al. ICMI 2017)

Advancements in Modeling

- New approaches rely on word-level alignments to build a co-occurrence model (Gu et al. ACL 2018, Zadeh et al. AAAI 2018, Chen et al. ICMI 2017)
- These can account for changes in each modality.



I liked the movie I saw last night

Smile

Neutral

Frown

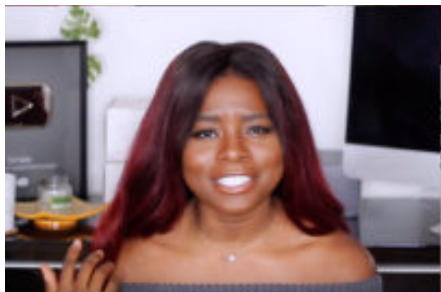
Excited

Neutral

Excited

Advancements in Modeling

- New approaches rely on word-level alignments to build a co-occurrence model (Gu et al. ACL 2018, Zadeh et al. AAAI 2018, Chen et al. ICMI 2017)
- These can account for changes in each modality.



I liked the movie I saw last night

Smile

Neutral

Frown

Excited

Neutral

Excited

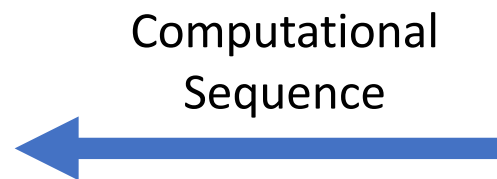
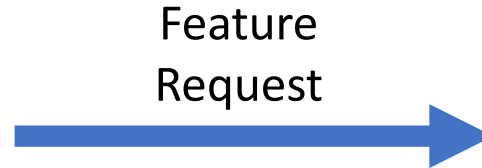
- These approaches complicate datasets and data processing.

CMU Multimodal SDK

- Publicly available SDK for loading and training multimodal temporal data.
- Call Sequence:



Multimodal
Scientist

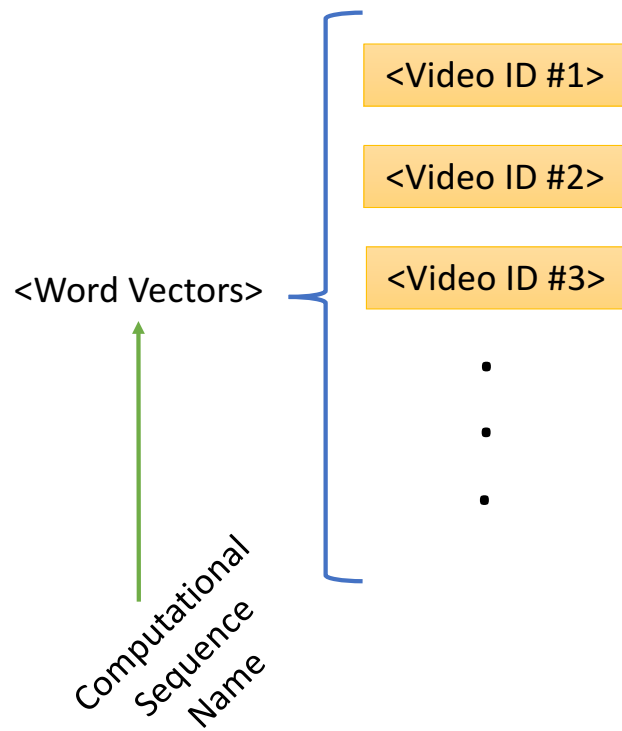


Validations
Parameters

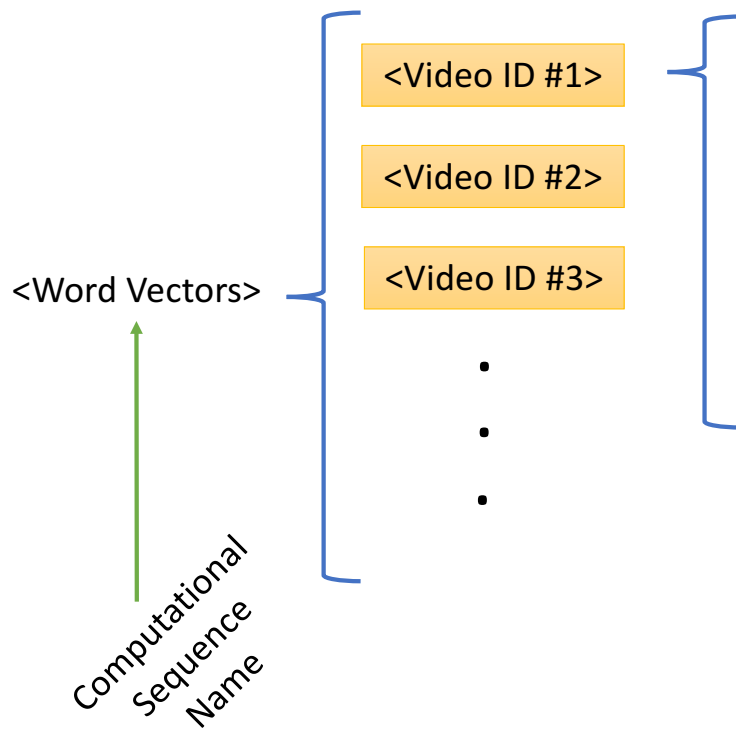


SDK Server

Computational Sequence



Computational Sequence



<Intervals>

S(0)	E(0)
S(1)	E(1)
S(2)	E(2)
.	.
.	.
.	.
S(T)	E(T)

<features>

f(0,0)	f(0,1)	f(0,2)	f(0,3)	f(0,4)
f(1,0)	f(1,1)	f(1,2)	f(1,3)	f(1,4)
f(2,0)	f(2,1)	f(2,2)	f(2,3)	f(2,4)
.
.
.
f(T,0)	f(T,1)	f(T,2)	f(T,3)	f(T,4)

Multimodal Computational Descriptors
in Hierarchical Format

Computational Sequence

- Suitable for temporal data from multiple modalities.
- Compatible with hdf5 (hierarchical data format) protocol.
- Stored using binary values and validated using trust server using sha256 hash – allowing for feature sharing among community.
- Storage using lossless floating point and string compression.

Computational Sequence

- Community can share their extracted features across network.



Multimodal
Scientist

Deploy Request
Set Private Key



Acknowledgment
Using UUID and SHA256



SDK Server

Current Datasets

- CMU-MOSEI – 23453 samples, sentence level sentiment and emotions
- CMU-MOSI – 2199 samples, sentence level sentiment

Current Datasets

- CMU-MOSEI – 23453 samples, sentence level sentiment and emotions
- CMU-MOSI – 2199 samples, sentence level sentiment
- ICT-MMMO – 340 samples, video level sentiment

Current Datasets

- CMU-MOSEI – 23453 samples, sentence level sentiment and emotions
- CMU-MOSI – 2199 samples, sentence level sentiment
- ICT-MMMO – 340 samples, video level sentiment
- MOUD – 386 samples, videos in Spanish, sentence level sentiment

Current Datasets

- CMU-MOSEI – 23453 samples, sentence level sentiment and emotions
- CMU-MOSI – 2199 samples, sentence level sentiment
- ICT-MMMO – 340 samples, video level sentiment
- MOUD – 386 samples, videos in Spanish, sentence level sentiment
- POM – 903 samples, video level personality traits

Current Computational Sequences

- Language
 - Glove word embeddings
- Vision
 - FACET descriptors
 - OpenFace descriptors
- Acoustic
 - Covarep
 - OpenSmile

Future Computational Sequences

- Vision

- Face Embeddings
- VGG-Face

- Acoustic

- Low-level features
- Phoneme Embeddings

Multisequence Alignment

Expected cross-sequence Alignment

Visual



Audio

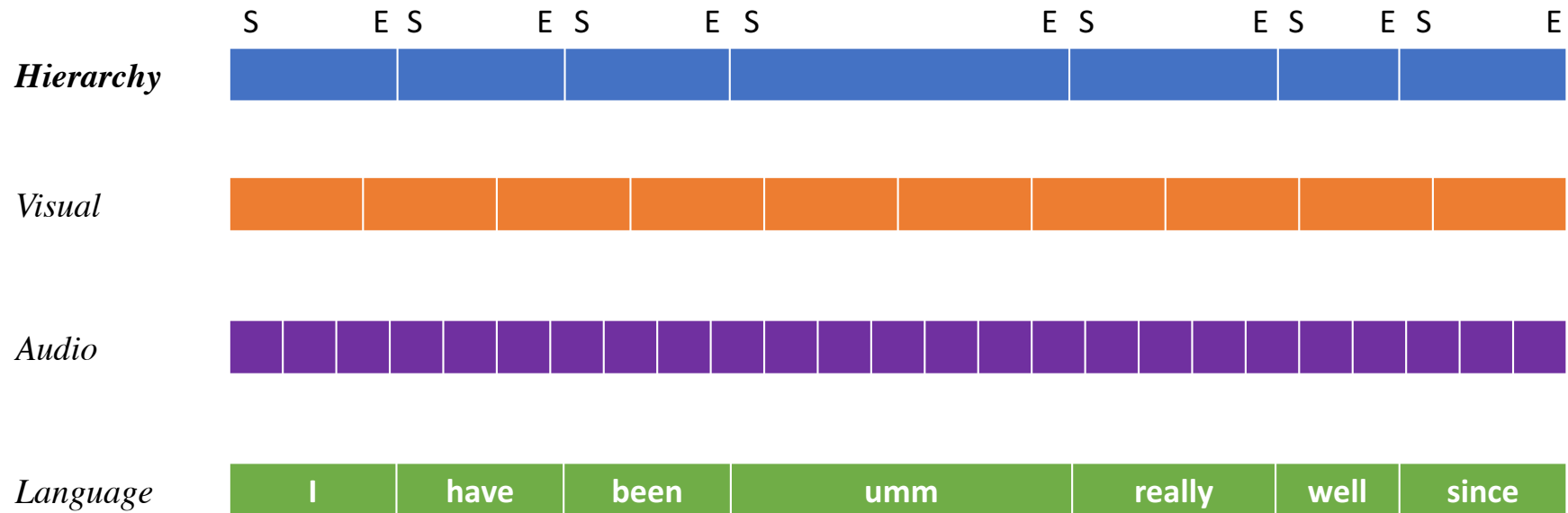


Language



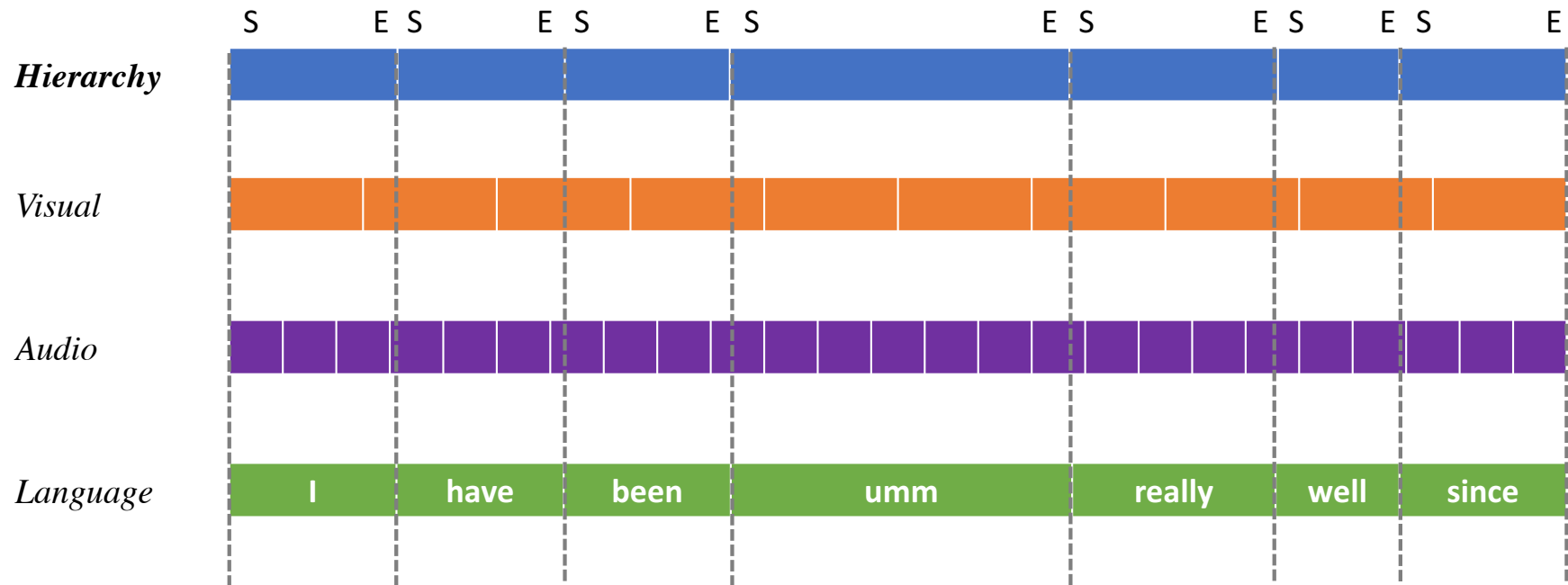
Multisequence Alignment

Expected cross-sequence Alignment



Multisequence Alignment

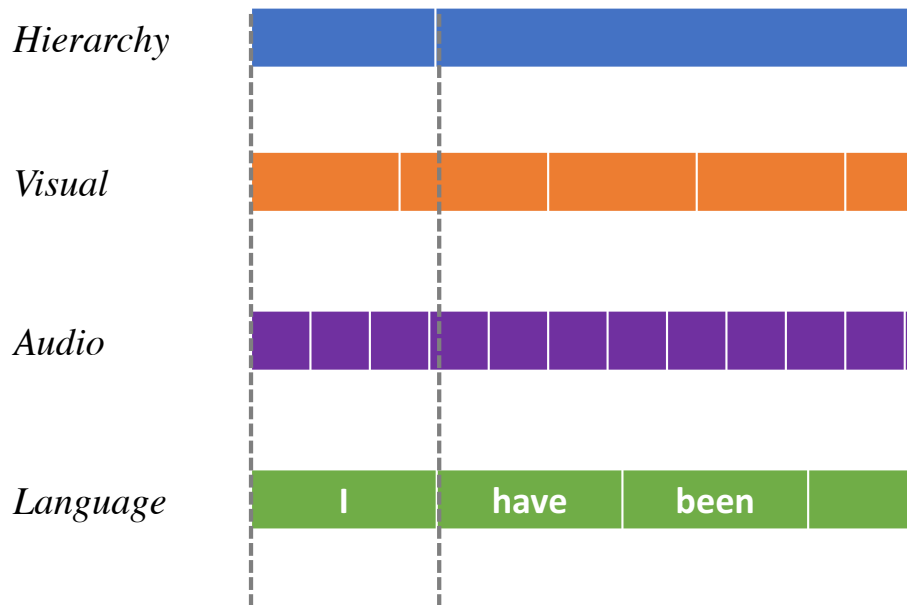
Cross-sequence Alignment



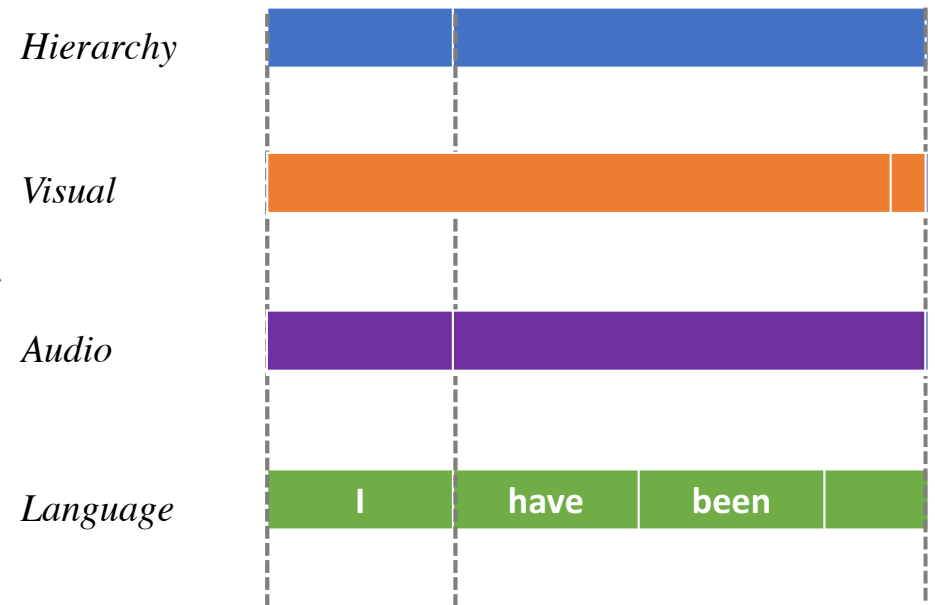
Multisequence Alignment

Cross-sequence Alignment

Non-flatten



Flatten



Multisequence Alignment

Cross-sequence Alignment

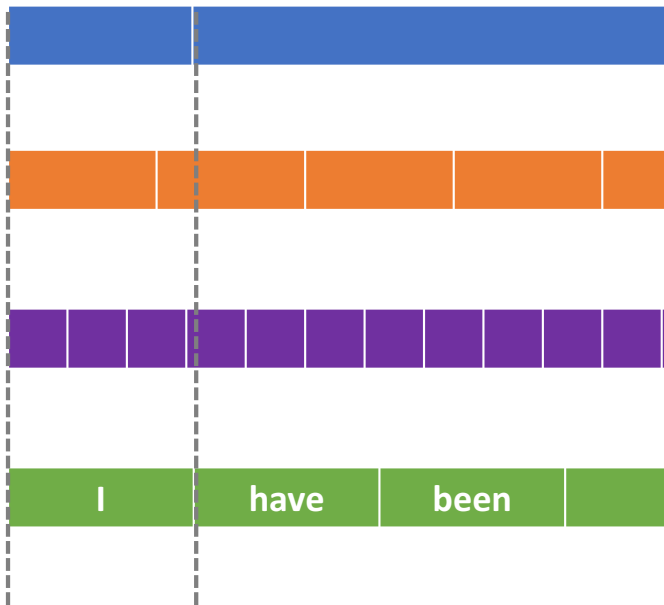
Non-flatten

Hierarchy

Visual

Audio

Language



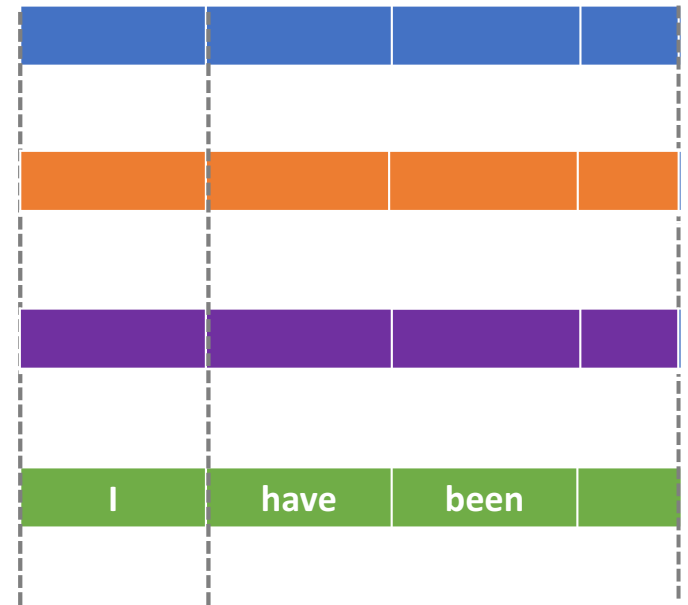
Flatten

Hierarchy

Visual

Audio

Language



Future Work

- Automated evaluation of submitted models
- Public leaderboard on the CMU-MultimodalSDK

The End!

Data: <https://github.com/A2Zadeh/CMU-MultimodalSDK>

Website: www.cs.cmu.edu/~pliang, www.amir-zadeh.com

Email: pliang@cs.cmu.edu, abagherz@cs.cmu.edu

Twitter: [@pliang279](https://twitter.com/pliang279)