11-877 Advanced Multimodal Machine Learning **Spring 2022**

Week 1: Dimensions of Heterogeneity

Instructors: Louis-Philippe Morency, Amir Zadeh, Paul Liang *Synopsis Lead: Paul Liang*

Summary: Multimodal machine learning is the study of computer algorithms that learn and improve through the use and experience of multimodal data. It brings unique challenges for both computational and theoretical research given the heterogeneity of various data sources.

In week 1's discussion session, the class brainstormed about the various dimensions of heterogeneity commonly encountered in multimodal ML research. There were no assigned papers, but the following introductory resources may be helpful:

1. Survey paper: Multimodal Machine Learning: A Survey and Taxonomy [Baltrušaitis et al., 2018]
2. Lecture slides: https://cmu-multicomp-lab.github.io/adv-mmml-course/spring2022/schedule/lecture1-Introduction.pdf
3. 11-777 multimodal ML course: https://cmu-multicomp-lab.github.io/mmml-course/fall2020/
4. Reading list: https://github.com/pliang279/awesome-multimodal-ml

We summarize several main takeaway messages below:

1. Data collection: Heterogeneity can come in the form of different specialized sensors used to capture raw modalities, such as different sensor equipment, collection environments, sampling rates, time-scales, how raw data is stored and retrieved from files, and data storage formats.
2. Vocabulary/atoms: Heterogeneity in the set of basic atoms (vocabulary) comprising a modality, which can be discrete or continuous and come from different base distributions.
3. Structure: Heterogeneity in how basic atoms are composed to form global information, which can span spatial, temporal/sequential, hierarchical, graphical, and set-based compositions. Even within the same type of structure, there can be heterogeneity in information content (e.g., differing entropy of atoms and compositions), information density (e.g., low vs high-frequency sequential data), and information range (short vs long-term temporal relationships).
4. Invariances: When composed at a global level, there lie different invariant transformations that preserve meaning, such as spatial invariance for images and permutation invariance for sets and graphs.
5. Modeling considerations: In addition to reasoning about the data, there can also be heterogeneity in the models suitable for processing each modality. These dimensions of heterogeneity include inductive biases and invariances designed into models to capture unimodal representations, VC dimension and other model complexity measures, and optimization challenges for each type of model.
6. Task-dependent considerations: Finally, dimensions of heterogeneity may also depend on the task at hand. For example, heterogeneity may depend on conditional independence assumptions with respect to the label (e.g., the extent to which both modalities have overlapping information given the label), how dominant a modality is for a given multimodal problem (which captures heterogeneity with respect to data, modeling, optimization, and tasks), and task-dependent priors (e.g., spatial invariance for image classification but spatial equivariance for image segmentation).

References

Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2):423–443, 2018.