Supplementary Material: Multivariate Cooperative Game for Image-Report Pairs: Hierarchical Semantic Alignment for Medical Report Generation

1 Training and Implementation Details

Parameter	Value
Image Encoder	ViT [3]
Report Encoder	SciBERT [2]
Optimization	AdamW [4] with a learning rate of $2e-5$
Normalized Dimension	[64, **128**, 256, 512]
λ	[0.2, 0.4, 0.6, **0.8**, 1.0]
Device	1 Nvidia V100

Table 1. Details on hyper-parameters.

2 Axiomatic Properties of Game-theoretic Interaction

Axioms 1 *Given a set $\mathcal{P} = \{1, 2, ..., p\}$ of players, a characteristic function $\phi : 2^p \to \mathbb{R}$, and a coalition $\mathcal{C} = \{i, j\} \subseteq \mathcal{P}$, following properties are met for the Banzhaf Interaction score $\mathcal{B}(\mathcal{C})$. (a)* ***Symmetry:*** *If $\forall \mathcal{S} \subseteq \mathcal{P},\ \phi(\mathcal{S} \cup \mathcal{C}) = \phi(\mathcal{S} \cup \mathcal{C}^{'}), \sum_{i \in \mathcal{C}} \phi(\mathcal{S} \cup \{i\}) = \sum_{i' \in \mathcal{C}'} \phi(\mathcal{S} \cup \{i^{'}\})$, then $\mathcal{B}(\mathcal{C}) = \mathcal{B}(\mathcal{C}^{'})$; (b)* ***Dummy:*** *If $\forall \mathcal{S} \subseteq \mathcal{P},\ \phi(\mathcal{S} \cup \mathcal{C}) = \phi(\mathcal{S}), \sum_{i \in \mathcal{C}} \phi(\mathcal{S} \cup \{i\}) = 0$, then $\mathcal{B}(\mathcal{C}) = 0$; (c)* ***Additivity:*** *If $\phi(*)$ and $\phi^{'}(*)$ have $\mathcal{B}(\mathcal{C})$ and $\mathcal{B}^{'}(\mathcal{C})$ respectively, then $\phi(*) + \phi^{'}(*)$ is $\mathcal{B}(\mathcal{C}) + \mathcal{B}^{'}(\mathcal{C})$; (d)* ***Recursivity:*** *let $\mathcal{V}(*)$ denote the Banzhaf Value [1], then $\mathcal{V}(\mathcal{C}|\mathcal{N} \setminus \{i\}, \{j\} \cup \mathcal{C}) = \mathcal{V}(i|\mathcal{N} \setminus \{j\}) + \mathcal{V}(j|\mathcal{N} \setminus \{i\}) + \mathcal{B}(\mathcal{C})$.*

References

1. Banzhaf III, J.F.: Weighted voting doesn't work: A mathematical analysis. Rutgers Law Review (1964)
2. Beltagy, I., Lo, K., Cohan, A.: Scibert: A pretrained language model for scientific text. In: EMNLP (2019)

3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: ICLR (2021)
4. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: ICLR (2019)