A Appendix

Theorem 1: *The marginal instance contribution from an Additive* MIL *model,* $g(x_i)$ *is proportional to the Shapley value of that instance,* ϕ_i.

$$g(x_i) \propto \phi_i(V, x) = \sum_{S \subseteq F \setminus i} \frac{|S!|(|F| - |S| - 1|!)}{|F!|} V_{S \cup i}(x_{S \cup i}) - V_S(x_S) \tag{7}$$

Proof: The interpretation for the value function V_S is taken from [28] where it's defined as the expected value of the model given a specific input set x^*_S.

$$V_S(x_S) = \mathbb{E}[g(x)|x_S = x^*_S] \tag{8}$$

Since the conditional expectation is for the case where only the set S is known, rewriting the equation in the form of integrals and breaking it down by set S and its complement $\bar{S}$ gives:

$$V_S(x_S) = \int g(x) p(x_{\bar{S}}|x_S = x^*_S) dx_{\bar{S}} \tag{9}$$

$$= \int \left[(\sum_{j \in S} g(x^*_j)) + (\sum_{j \in \bar{S}} g(x_j)) \right] p(x_{\bar{S}}|x_S = x^*_S) dx_{\bar{S}} \tag{10}$$

$$= \int \sum_{j \in S} g(x^*_j)\, p(x_{\bar{S}}|x_S = x^*_S) dx_{\bar{S}} + \int (\sum_{j \in \bar{S}} g(x_j)\, p(x_{\bar{S}}|x_S = x^*_S) dx_{\bar{S}} \tag{11}$$

$$= \sum_{j \in S} g(x^*_j) \int p(x_{\bar{S}}|x_S = x^*_S) dx_{\bar{S}} + \sum_{j \in \bar{S}} \mathbb{E}[g(x_j)] \tag{12}$$

$$= \sum_{j \in S} g(x^*_j) + \sum_{j \in \bar{S}} \mathbb{E}[g(x_j)] \tag{13}$$

Equation 10 uses the model definition from equation 5 to express the function g into its linearly additive components over all instances which are either in set S or in $\bar{S}$. Similarly, we can write the value function when the i^{th} index is included in S by removing it from set $\bar{S}$ and adding it to S:

$$V_{S \cup i}(x_{S \cup i}) = V_S(x_S) + g(x^*_i) - \mathbb{E}[g(x_i)] \tag{14}$$

$$V_{S \cup i}(x_{S \cup i}) - V_S(x_S) = g(x^*_i) - \mathbb{E}[g(x_i)] \tag{15}$$

Since the second term here is the expected value of the model output, we can put this back in equation 7 to get an equivalence between the Shapley value and the instance contribution from an Additive MIL model.

$$\phi_i(V, x) \propto g(x_i)$$