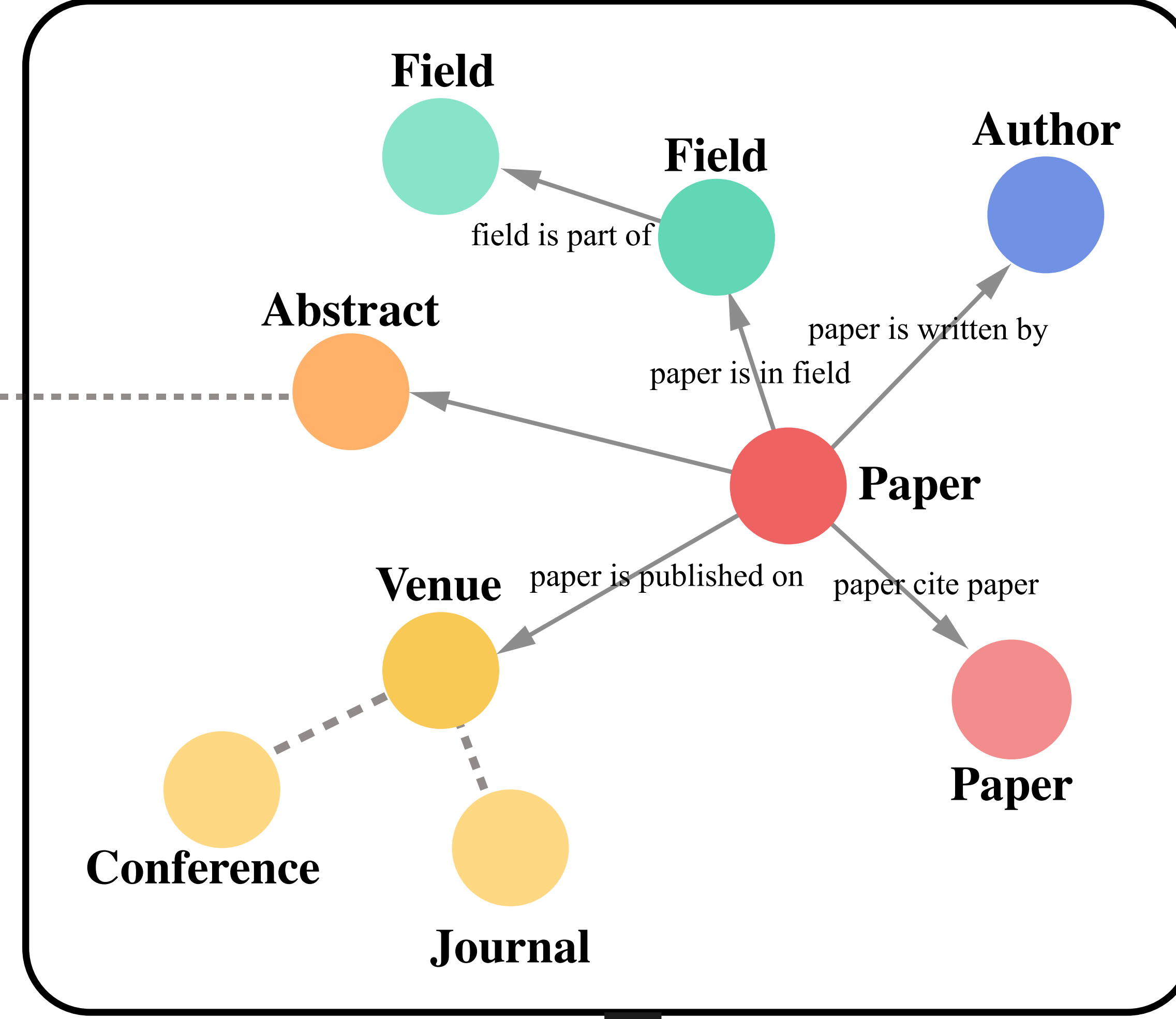


Text input

Multiple instance learning (MIL) is a variation of supervised learning where a single class label is assigned to a bag of instances. In this paper, we state the MIL problem as learning the Bernoulli distribution of the bag label where the bag label probability is fully parameterized by neural networks. Furthermore, we propose a **neural network-based permutation-invariant aggregation operator that corresponds to the attention mechanism**. Notably, an application of the proposed attention-based operator provides insight into the contribution of each instance to the bag label. We show empirically that our approach achieves comparable performance to the best MIL methods on benchmark MIL datasets and it outperforms other methods on **a MNIST-based MIL dataset and two real-life histopathology datasets without sacrificing interpretability**.

Graph input

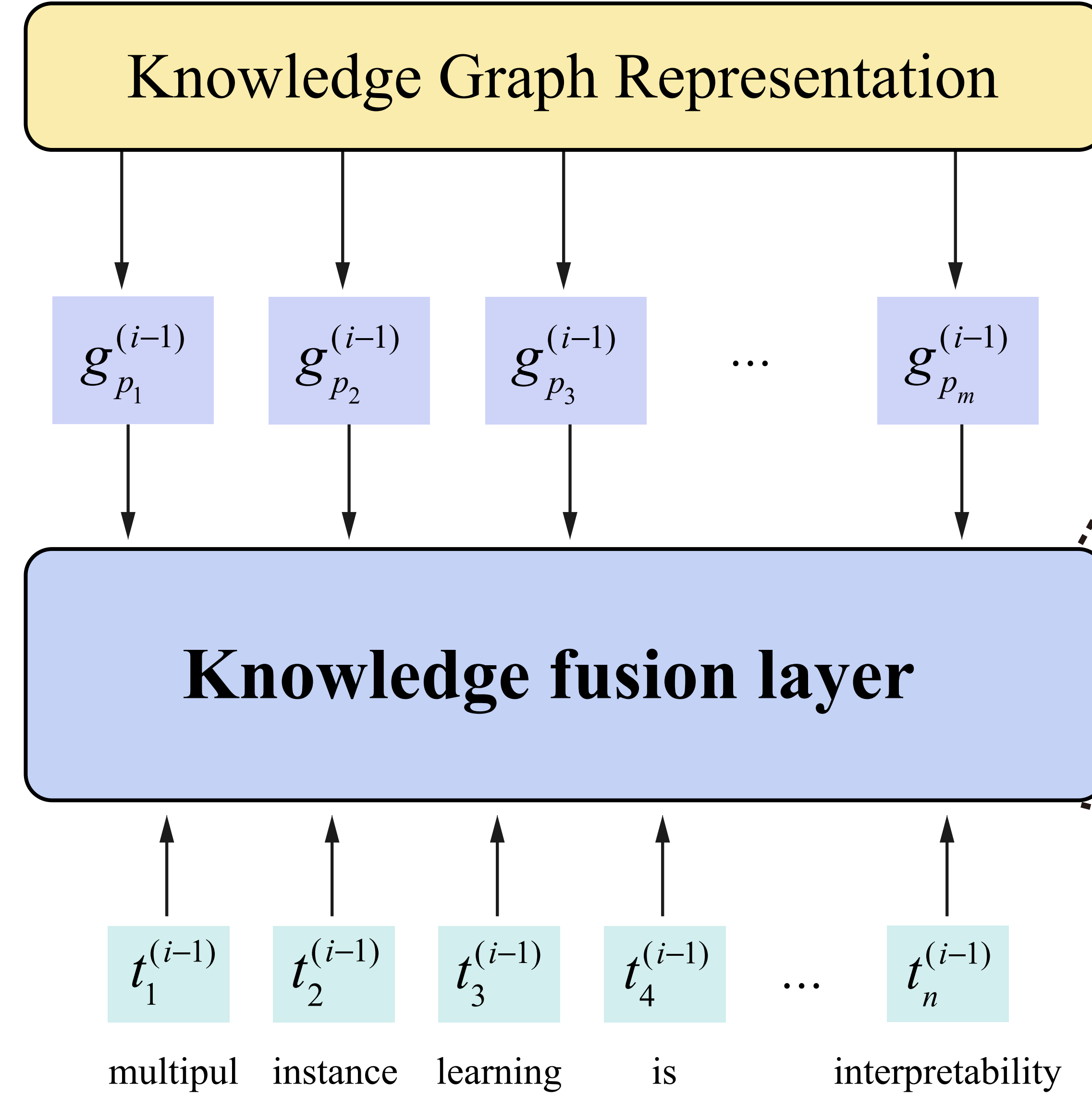


Representation Learning

Graph embedding

Knowledge fusion

Token input



Token input

$t_1^{(i-1)}$ $t_2^{(i-1)}$ $t_3^{(i-1)}$ $t_4^{(i-1)}$... $t_n^{(i-1)}$
multipul instance learning is interpretability

Graph node embedding

$g_{p_i}^{(i-1)}$

T-Encoder

Multi-head Attention

Feed Forword Network

$M \times$

$t'^{(i-1)}$

G-Encoder

Multi-head Attention

$t_1'^{(i-1)}$ $t_2'^{(i-1)}$ $t_3'^{(i-1)}$ $t_4'^{(i-1)}$... $t_n'^{(i-1)}$

$g_{p_i}^{(i-1)}$

$h_1^{(i-1)}$ $h_2^{(i-1)}$ $h_3^{(i-1)}$ $h_4^{(i-1)}$... $h_n^{(i-1)}$

Text&Graph Fusion

$N \times$

Token output

$t_1^{(i)}$ $t_2^{(i)}$ $t_3^{(i)}$ $t_4^{(i)}$... $t_n^{(i)}$

Graph output

$g_{p_i}^{(i)}$