

Heterogenous Graph Neural Networks for Action Prediction on Egocentric Action Scene Graphs

Olmo Ceriotti
Mat 886140

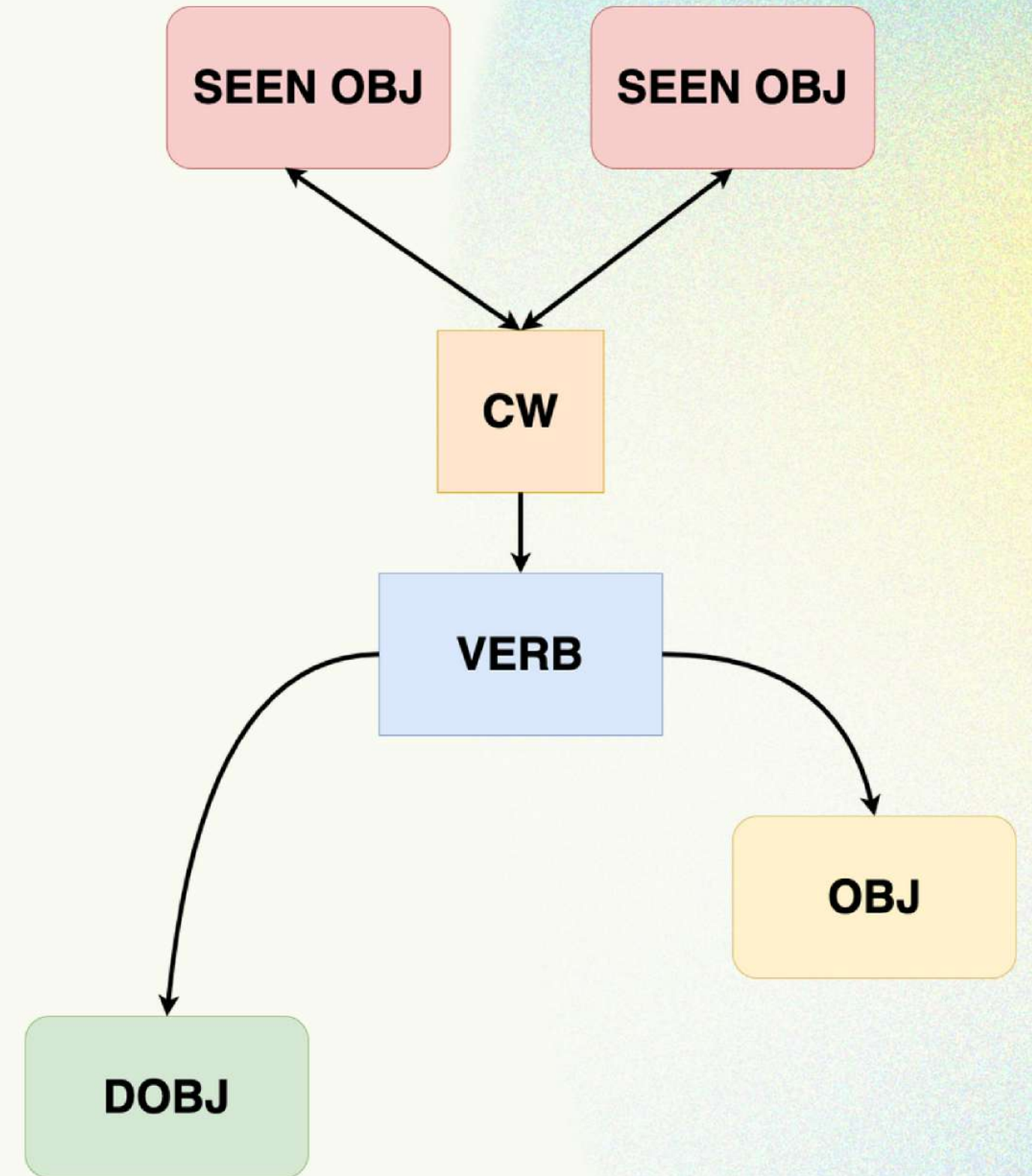
Relatore: Dott. Sorrenti Domenico
Correlatore: Prof. Ognibene Dimitri

Obiettivi

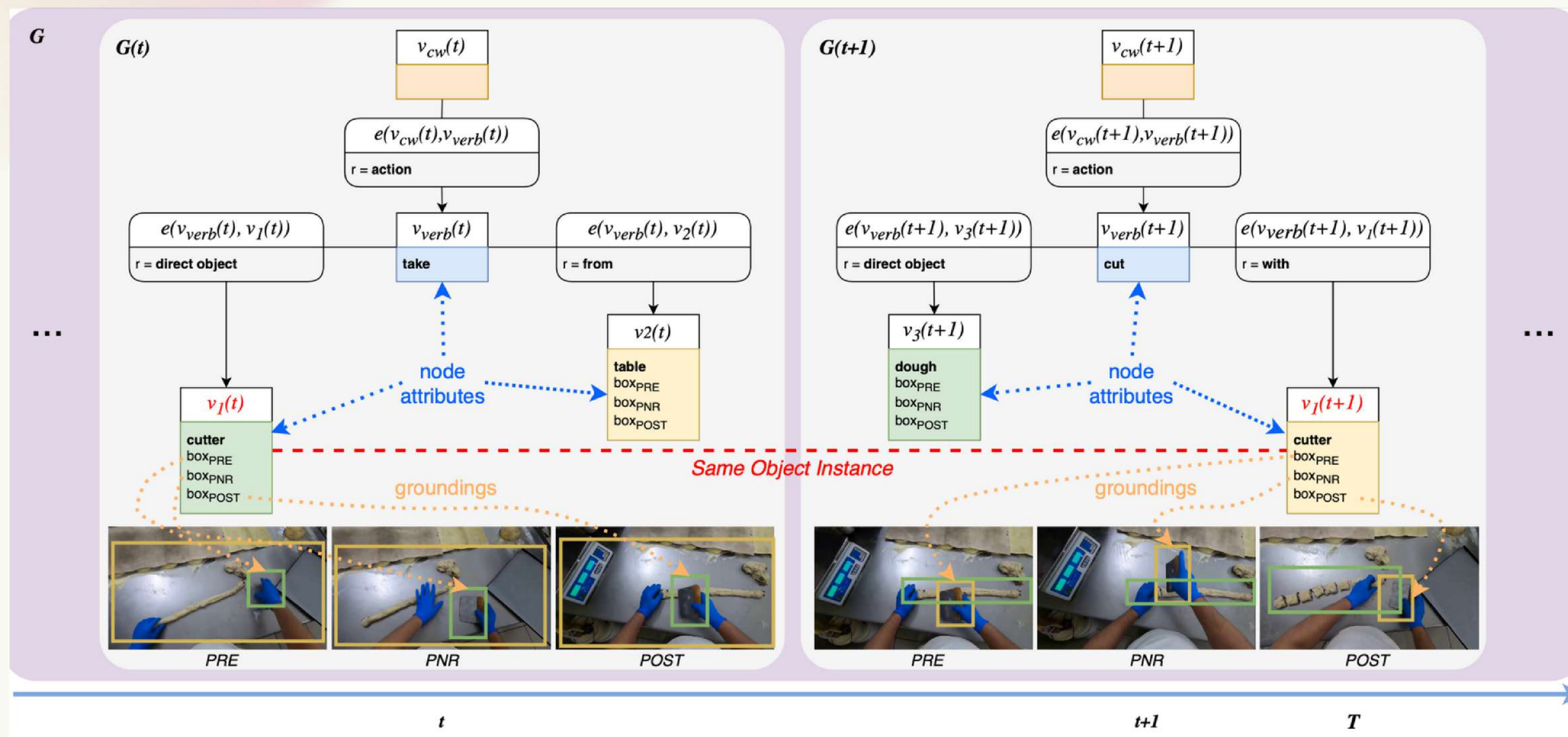
Dataset	Heterogeneous GNNs
<ul style="list-style-type: none">• Adattamento del dataset• Valorizzazione e modifica dei dati• Test dei cambiamenti effettuati	<ul style="list-style-type: none">• Definizione di 3 modelli• Training e inferenza su dataset

EASG: il Dataset

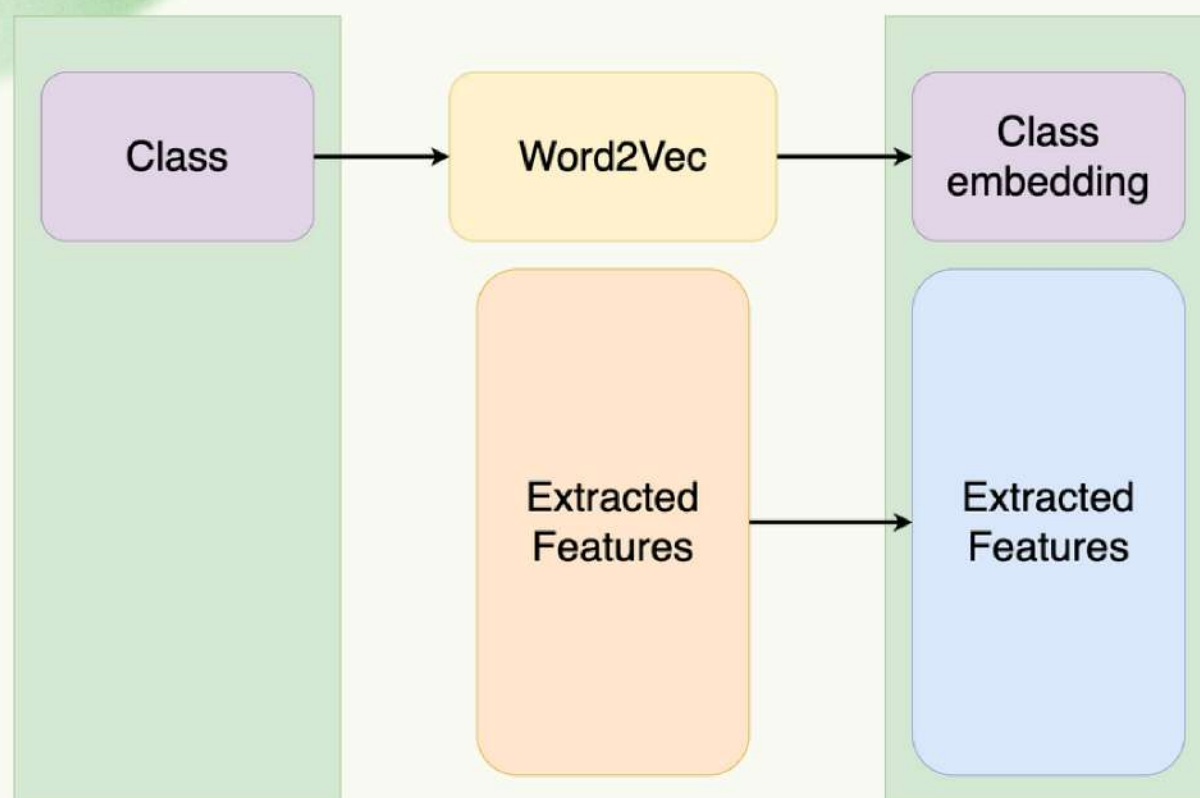
- Creato da Intel Labs e dall'Università di Catania
- Basato su Ego-4D
- Diviso in sequenze, ognuna contenente un insieme di grafi



EASG: Struttura

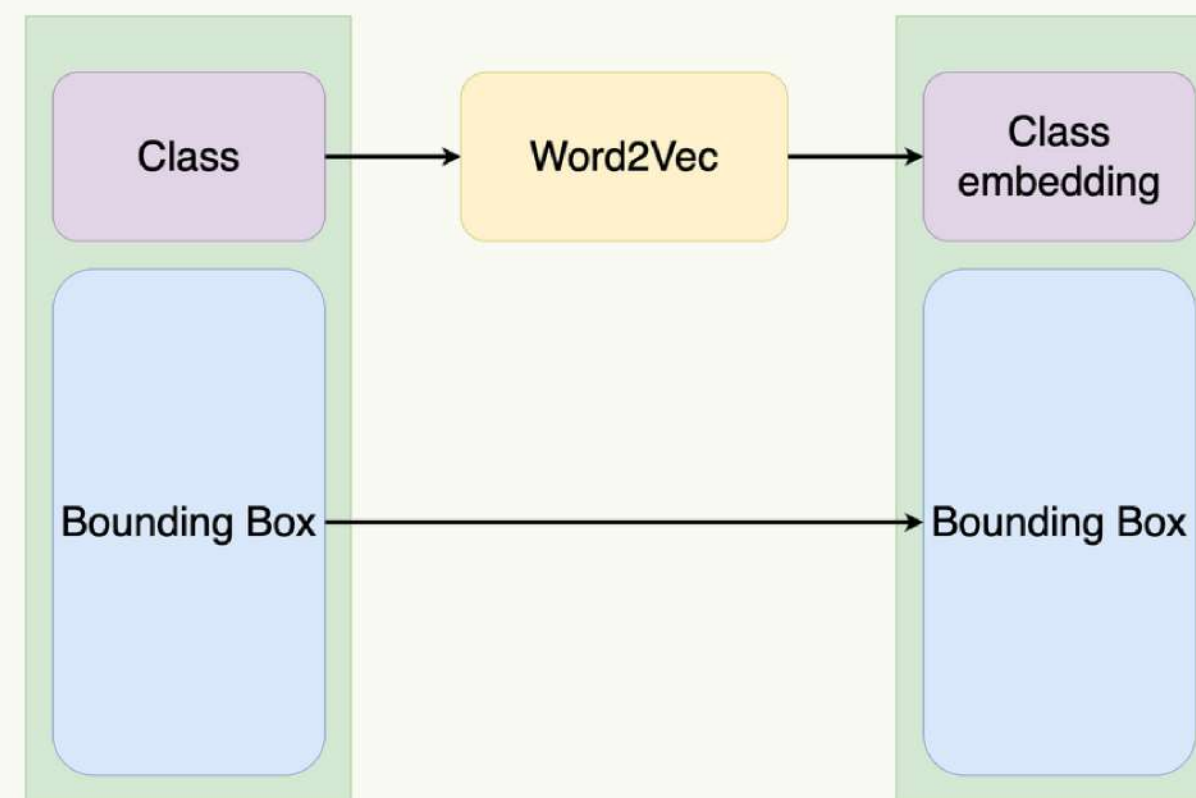


EASG: modifiche



Aggiunta features pre-estratte

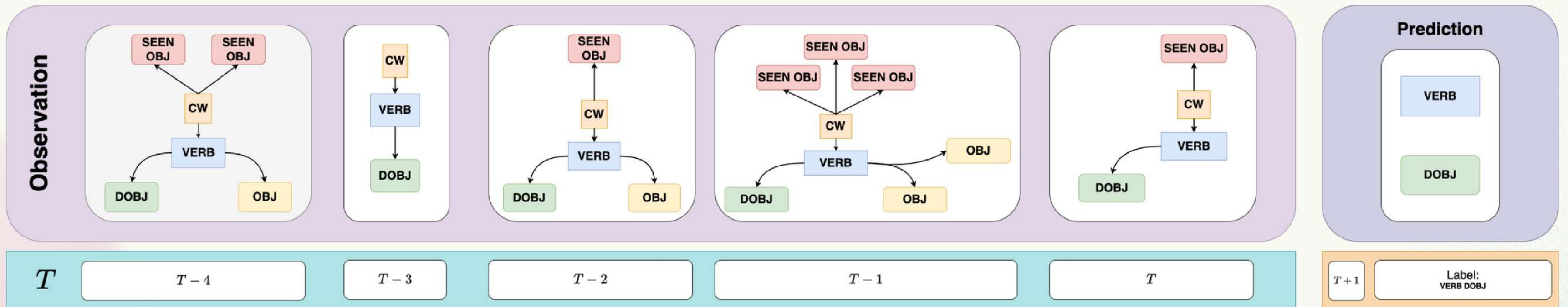
Le features estratte sono state utilizzate per la generazione dei nodi verbo nel dataset originale.



Aggiunta embedding Word2Vec

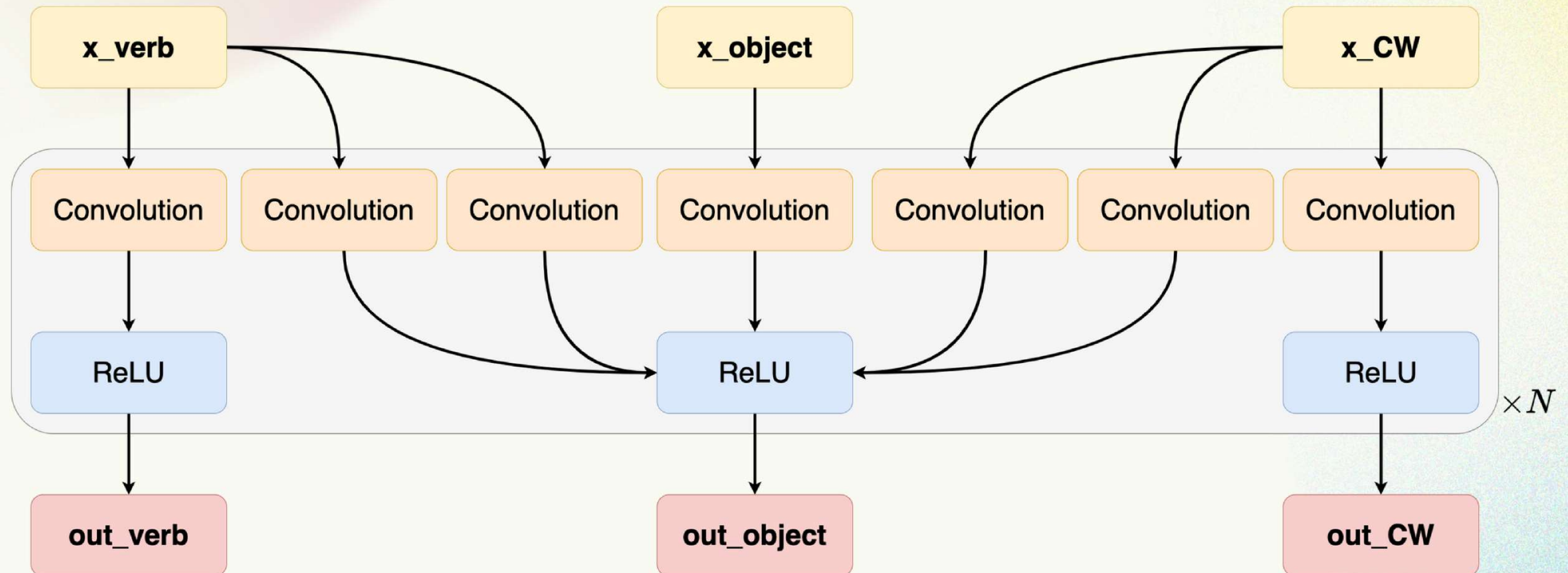
Il modello Word2Vec crea un embedding per ogni etichetta cercando di rappresentare al meglio la semantica dell'oggetto o del verbo in questione.

EASG: Task



Heterogenous GNNs

- Basate su GNNs
- Supportano computazione su grafi eterogenei
- Un layer per ogni tipo di arco



Modelli proposti

SAGE

- Seleziona casualmente nodi dal vicinato di un nodo.
- Aggrega le features calcolandone la media.
- Impara l'embedding del nodo partendo da questa selezione.

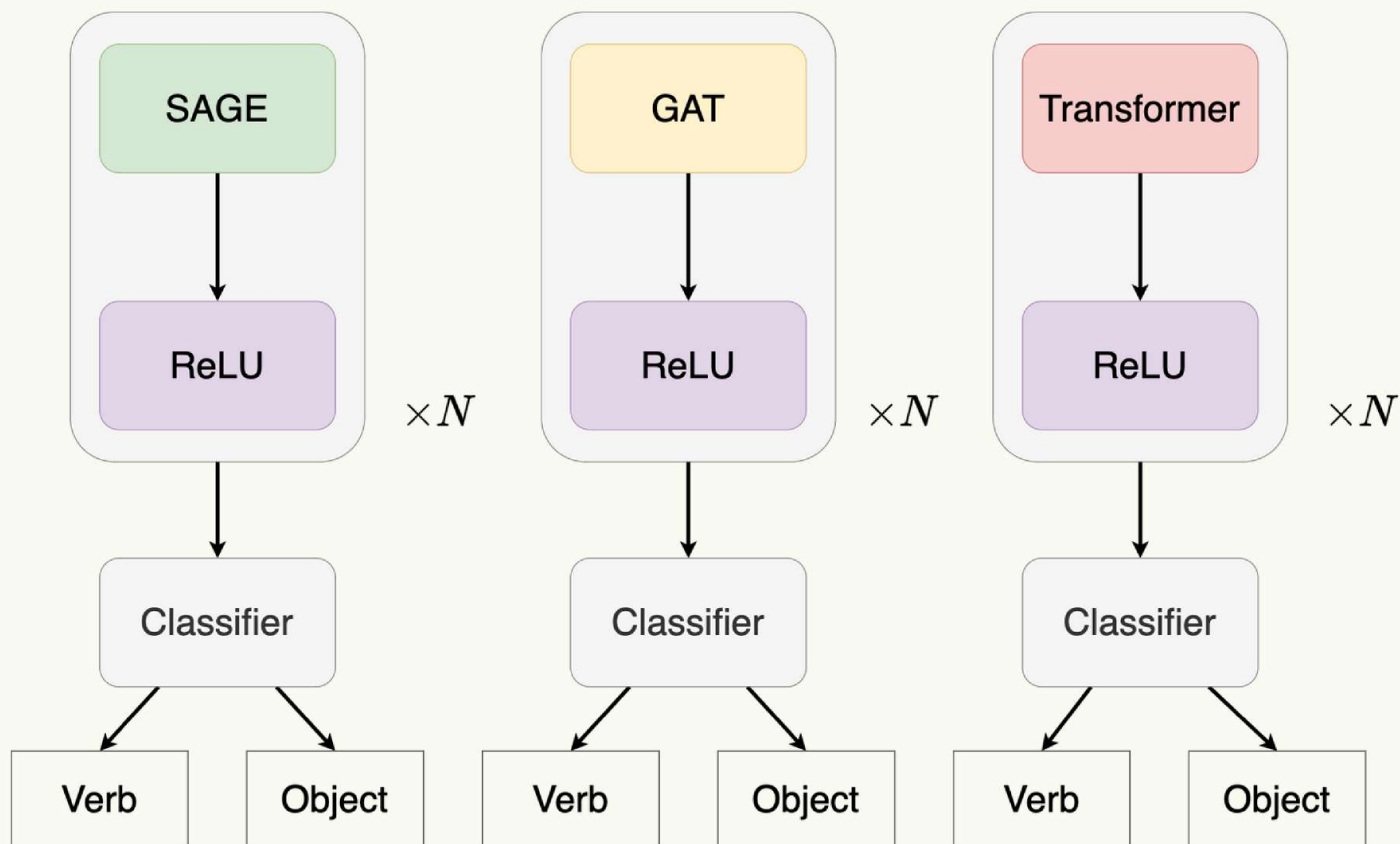
GAT

- Utilizza un meccanismo di attenzione.
- Ogni arco considerato diversamente dagli altri.
- Se presenti più attention heads i loro risultati sono concatenati,

Transformer

- Simula la struttura del classico Transformer.
- Calcola Query, Key e Value per ogni nodo.
- Cattura relazioni anche tra nodi molto distanti nel grafo.

Modelli proposti



Ambiente di training

Strumenti	Loss	Learning rate
<ul style="list-style-type: none">• Colab• WandB	<ul style="list-style-type: none">• Cross Entropy Loss• Due parametri allenati per quantificare l'importanza delle due classi verb e object	<ul style="list-style-type: none">• Massimo learning rate 0.01• Uso di uno scheduler con picco alla decima epoch• Andamento lineare in warm-up e decay

Risultati

Risultati del training dei modelli sul dataset con applicate le modifiche

	Verb		Object		Action	
	Top-1 accuracy	Top-5 accuracy	Top-1 accuracy	Top-5 accuracy	Top-1 accuracy	Top-5 accuracy
GPT	5.94	14.97	47.36	67.26	3.40	9.24
SAGE	14.67	35.33	10.89	10.92	2.22	5.33
GAT	13.78	40.67	20.67	21.11	1.56	7.77
Transformer	14.89	31.33	0.18	0.22	0	0

Risultati dataset

Risultati del training del modello
GAT su varie versioni del dataset

	Verb		Object		Action	
	Top-1 accuracy	Top-5 accuracy	Top-1 accuracy	Top-5 accuracy	Top-1 accuracy	Top-5 accuracy
Full	13.78	40.67	20.67	21.11	1.56	7.77
No pre- extracted features	15.11	32.22	18.22	18.89	2.22	4.66
No Word2Vec	15.33	38.89	20.02	20.22	2.66	8.44

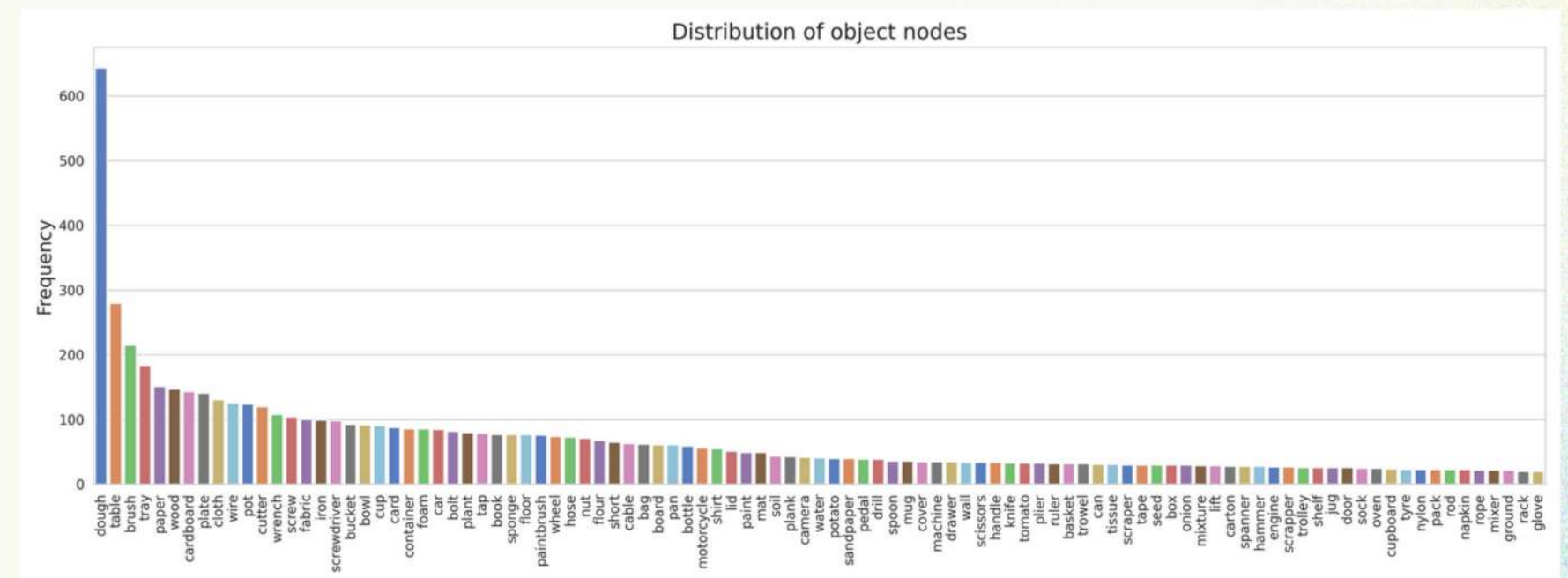
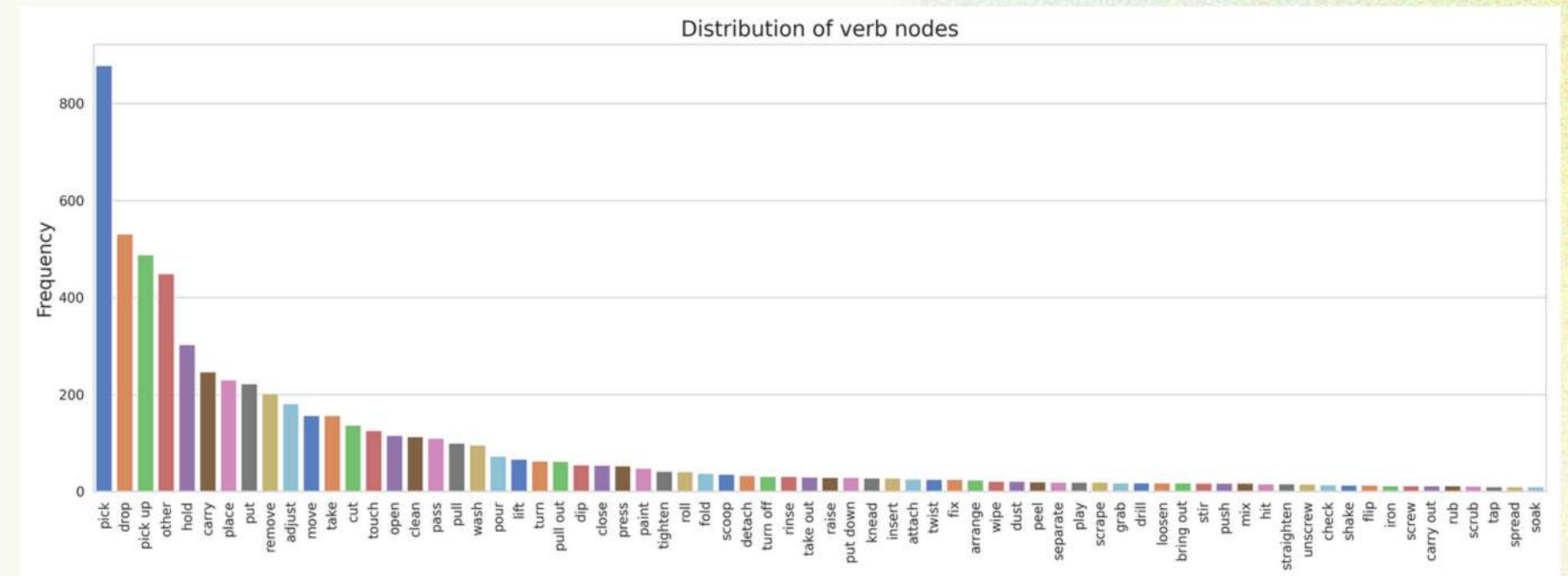
Analisi risultati

Generali:

- Risultati comparabili allo stato dell'arte
- Distribuzione labels
- Baseline GPT

Dataset:

- Importanza features pre-estrate
- Contributo embedding Word2vec



Conclusioni



Grazie!