SYNAPSE: Summarization & Answering with Precision via Scalable Encoding

George R., Manas M. & Yajat R.

The Basic Idea

The central idea of this project is to develop **Synapse**, a novel hierarchical framework designed to overcome the computational limits of standard Transformer models when processing long-form documents.

The basic idea is to pass entire chunks of sentences as a single token to the model and thus increase the context size of the model.

We test this using summarization as the downstream task for the model.

Dataset

We use Narrative QA dataset by Google Deepmind.

It contains Movie Scripts, Reference Summaries and related QAs to the script.
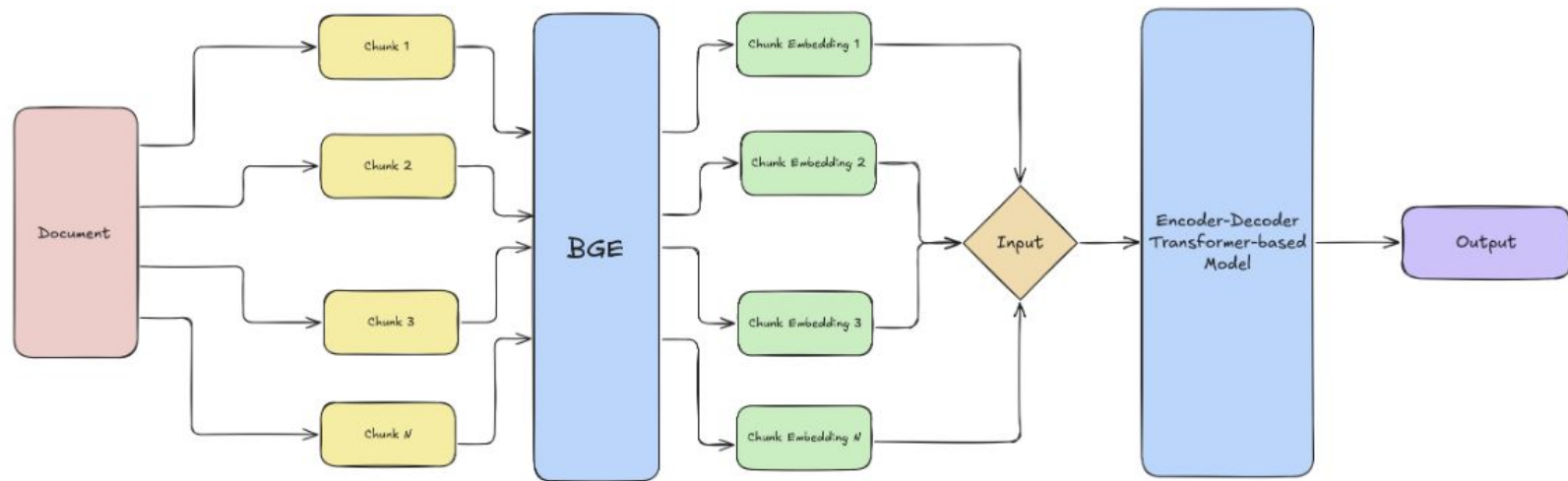
Architecture

Figure 1: Overview of our proposed pipeline. The system converts the document into semantic chunks which are compressed into chunk embeddings before passing them to the transformer for the specified task.

Chunking

We used Graph Based chunking method to semantically chunk the documents.

Algorithm 1 Graph-Based Semantic Chunking

Require: Document $D = \{s_1, \ldots, s_n\}$, similarity threshold τ

Compute cosine similarity for all sentence pairs to form a similarity matrix

Construct an undirected graph G where each sentence is a node

Add an edge between nodes s_i and s_j if $\text{CosSim}(s_i, s_j) \geq \tau$

Identify all connected components in G

Each connected component forms a chunk; add it to C

return C

Mini LM vs BGE

τ	Chunks	Avg	Med	Max
0.1	1	3847.000	3847	3847
0.2	1	3847.000	3847	3847
0.3	1	3847.000	3847	3847
0.4	126	30.532	1	3715
0.5	1447	2.659	1	2070
0.6	3012	1.277	1	200
0.7	3609	1.066	1	45
0.8	3775	1.019	1	11
0.9	3818	1.008	1	11

τ	Chunks	Avg	Med	Max	Outlier%
0.1	1	4647.000	4647	4647	0.0
0.2	1	4647.000	4647	4647	0.0
0.3	1	4647.000	4647	4647	0.0
0.4	1	4647.000	4647	4647	0.0
0.5	1	4647.000	4647	4647	0.0
0.52	1	4647.000	4647	4647	0.0
0.54	4	1161.750	1	4644	0.0
0.56	26	178.731	1	4622	0.0
0.566	38	122.289	1	4609	2.6
0.57	45	103.267	1	4602	2.2
0.573	52	89.365	1	4595	1.9
0.576	61	76.180	1	4586	1.6
0.58	75	61.960	1	4570	1.3
0.6	214	21.715	1	4427	0.5
0.63	651	7.138	1	3968	0.3
0.67	1836	2.531	1	2709	0.2
0.7	2948	1.576	1	1354	0.6
0.73	3778	1.230	1	372	0.8
0.76	4253	1.093	1	87	0.4
0.8	4496	1.034	1	34	1.5
0.9	4612	1.008	1	10	0.3

Results - MiniLM + AE

Metric	BART (Full)			T5 (Full)			LLAMA (Adapter)			T5 (Adapter)		
	FT	Base	Diff	FT	Base	Diff	Adapt	Base	Diff	Adapt	Base	Diff
R-1	0.202	0.073	+176%	0.101	0.055	+85%	0.003	0.172	-99%	0.099	0.054	+84%
R-2	0.005	0.010	-51%	0.004	0.008	-51%	0.000	0.026	-99%	0.002	0.008	-69%
R-L	0.085	0.051	+68%	0.057	0.039	+49%	0.002	0.078	-97%	0.054	0.039	+40%
BS-P	0.738	0.746	-1.1%	0.744	0.766	-2.8%	0.687	0.734	-6.4%	0.737	0.784	-5.9%
BS-R	0.759	0.770	-1.5%	0.758	0.769	-1.4%	0.760	0.781	-2.6%	0.757	0.764	-0.9%
BS-F1	0.748	0.758	-1.3%	0.751	0.767	-2.1%	0.722	0.757	-4.6%	0.747	0.774	-3.5%

Results - BGE

Training Strategy	Metric	T5-Base			BART-Large		
		Trained	**Base**	**Improv.**	**Trained**	**Base**	**Improv.**
Fully Finetuned (Enc+Dec)	R-1	0.190	0.022	+762.6%	0.213	0.093	+130.4%
	R-2	0.034	0.002	+1435.2%	0.041	0.012	+249.7%
	R-L	0.105	0.018	+497.8%	0.126	0.061	+106.8%
	BS-F1	0.813	0.760	+6.9%	0.810	0.767	+5.6%
Decoder Frozen (Enc only)	R-1	0.100	0.022	+355.2%	0.208	0.093	+125.3%
	R-2	0.016	0.002	+633.7%	0.031	0.012	+164.7%
	R-L	0.064	0.018	+267.0%	0.119	0.061	+95.5%
	BS-F1	0.776	0.760	+2.1%	0.803	0.767	+4.7%
Untrained (No pretrained)	R-1	0.097	0.021	+367.2%	0.239	0.092	+158.7%
	R-2	0.022	0.002	+1004.4%	0.052	0.012	+336.4%
	R-L	0.066	0.018	+271.2%	0.137	0.061	+126.5%
	BS-F1	0.797	0.762	+4.5%	0.785	0.767	+2.4%
Cleaned Dataset (Preprocessed)	R-1	0.199	0.021	+858%	0.177	0.039	+350%
	R-2	0.037	0.001	+2608%	0.031	0.003	+1096%
	R-L	0.111	0.017	+542%	0.109	0.023	+374%
	BS-F1	0.813	0.737	+10.3%	0.809	0.742	+9.0%

Samples

Sample 3 (T5 full finetune)

Reference: ... Under Armitage's guidance, the children adapt from an aristocratic lifestyle to that of simple foresters. After Armitage's death, Edward takes charge and the children develop and expand the farmstead, aided by the entrepreneurial spirit of the younger brother Humphrey. They are assisted by a gypsy boy, Pablo, who they rescue ...

Generated: The story is set in the early years of the 19th century and begins with a brief history of the family's life. It tells the story of a young man named John, who was born into slavery at the age of five to a wealthy widower-turned-poverty. He has been raised by his aunt, Mrs. Pearce, for many years as a child, but has never had the opportunity to marry him. At the beginning of the 20th century, he met his future wife, Mary, when she was only seven years old when she became pregnant. When John's father died, John returned to England to pursue a career in her youth. After a

Sample 9 (T5 full finetune)

Reference: ... During his stay in Paris, Yorick is informed that the police inquired for his passport at his hotel. Without a passport at a time when England is at war with France (Sterne traveled to Paris in January 1762, before the Seven Years' War ended), he risks imprisonment in the Bastille. Yorick decides to travel to Versailles where he visits the Count de B**** to acquire a passport. ... Yorick returns to Paris, and continues his voyage to Italy after staying in Paris for a few more days. Along the way he decides to visit ...

Generated: The novel is set in the early years of the 19th century. It tells the story of a young man, who has just returned from a trip to Paris with his wife and two children. The story begins when he discovers that he has been mistreated by his father for not having enough money to pay off his debts. He decides to take a job as a waitress at a small-town hotel. After a few days, however, he finds himself in a situation where he can't afford to pay his rents due to a low-paid position at the hotel. When he comes back, it becomes clear to him that if he

Sample 12 (T5 full finetune)

Reference: In 965 AD, Odin, king of Asgard, wages war against the Frost Giants of Jotunheim and their leader Laufey, to prevent them from conquering the nine realms, starting with Earth. The Asgardian warriors defeat the Frost Giants and seize the source of their power, the Casket of Ancient Winters. ...

Generated: In the early 20th century, a group of aliens from the distant past have been dispatched to destroy a planet. The aliens are being hunted and killed by a mysterious force that has destroyed the planet's solar system. As a result, the Earth's population has fallen to a low point, and the spacecraft is about to be dismantled. They are joined by an army of scientists, led by Captain James T. Kirk, who is in charge of the project. Kirk is assigned the task of destroying the planet on his way out of the galaxy, as well as completing a mission to find a meteor strike on the planet