

## Classify textual claims as *Supported*, *Refuted* or *NotEnoughInfo* using evidence retrieved from Wikipedia

### Frame-based Evidence Retrieval

Frame-matching was used to identify and retrieve sentences that provide evidence relevant to the claim, where a frame is a semantic schema describing a situation, event or relations and its participants

**Claim:** Maximum Overdrive is only a 1980 romance film  
(Frames: *Temporal\_Pattern*, *Origin*, *Behind\_the\_scene*, *Personal\_Relationship*)

#### Semantically Similar

**E1** – Maximum Overdrive is 1986 American science fiction horror dark comedy film written and directed by Stephen King  
(Frames: *Temporal\_Pattern*, *Origin*, *Behind\_the\_scene*)

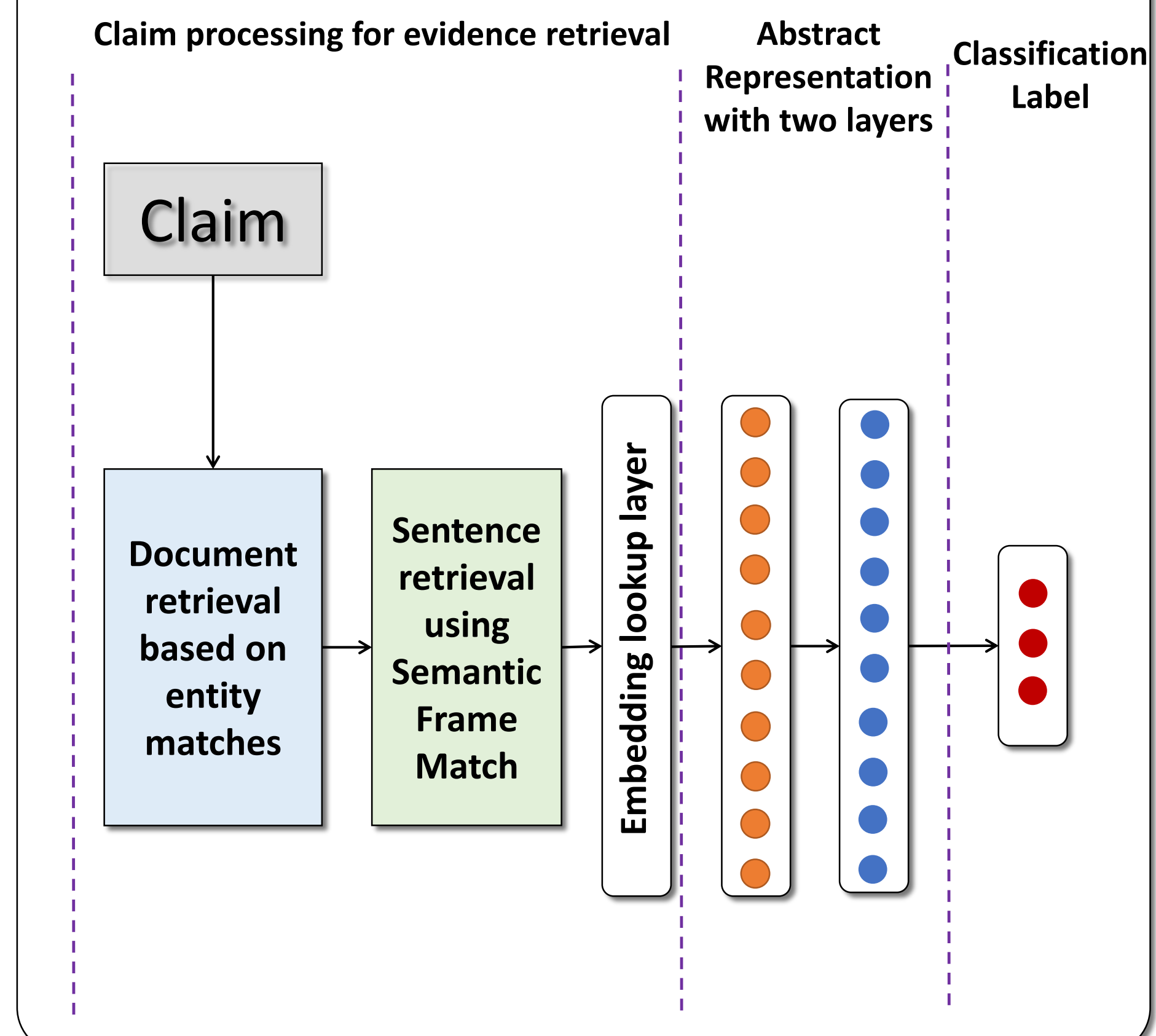
**E2** – In 1988, Maximum Overdrive was nominated for “Best Film” at the International Fantasy Film Awards.  
(Frames: *Temporal\_Pattern*, *Behind\_the\_scene*)

#### Semantically Dissimilar

**E3** – He considers the process a learning experience, after which he intended never to direct again.  
(Frames: *Education\_Teaching*, *Memorization*)

We used our KB of FrameNet annotations of Wikipedia article sentences (Ferraro et al., Concretely Annotated Corpora, AKBC Workshop, NIPS, 2014)

### Classification



Who	Classifier	Training Type	Classifier		Predicting Evidence		
			FEVER Score	ACC	Precision	Recall	F1
UMBC-1	MLP	(Entire Doc.) Frame Match	0.2572	0.4398	0.4868	0.3346	0.3966
UMBC-2	MLP	Frame Match U Intro. Sec.	0.2628	0.4479	0.4868	0.3346	0.3966
UMBC-3	MLP	Frame Match $\cap$ Intro. Section	0.2599	0.4069	0.4868	0.3346	0.3966
Baseline-1	MLP	Throne et al. 2018 Table 4	0.1942	0.4064	-	-	-
Baseline-2	DA	CodaLab Results	0.3127	0.5137	-	-	0.1718

Parameter	Value
Learning rate	0.01
Layers	2
Optimizer	SGD
Hidden Layer size	50
L2 regularize	1e-06
Epoch	2
Batch size	64
Dropout	0.5

MLP classifier parameter values

Performance on development dataset on difference settings; better evidence retrieval enables good classification performance with a simple classifier model

**Claim:** Last Man Standing does not star Tim Allen

**Predicted evidence:**

1. Timothy Allen Dick (born June 13, 1953), known professionally as Tim Allen, is an American actor, comedian and author
2. He is known for his role as Tim “The Toolman” Taylor in the ABC television show Home Improvement (1991) as well as for his starring roles in several films, including the role of Buzz Lightyear in the Toy Story franchise
3. From 2010 to 2017, he starred as Mike Baxter in the TV series Last Man Standing

**Predicted Label:** REFUTED (due to evidence (3)); **Gold standard Label:** REFUTED

✓ Relevant evidence is correctly retrieved and is classified correctly as refuting the claim

Human	Support	Refute	Neither
<b>Support</b>	4646	171	1849
<b>Refute</b>	3050	1198	1618
<b>Neither</b>	4123	391	2152

Dev confusion matrix for frame-based sentence retrieval only (NFC)

**Claim:** Rocky Mountain High is an Australian Song

**Predicted evidence (Correct):**

1. “Rocky Mountain High” is a folk rock song written by John Denver and Mike Taylor about Colorado, and is one of the two official state songs of Colorado
2. The song also made #3 on the Easy Listening chart, and was played by some country music stations
3. Denver told concert audiences in the mid-1970s that the song took him an unusually long nine months to write
4. Members of the Western Writers of America chose it as one of the Top 100 Western songs of all time

**Predicted Label:** SUPPORTED; **Gold standard Label:** REFUTED

✗ Relevant evidence correctly retrieved, but misclassified as supporting the claim

Human	Support	Refute	Neither
<b>Support</b>	4499	173	1994
<b>Refute</b>	2777	2125	1764
<b>Neither</b>	3968	365	2333

Dev confusion matrix for frame-based sentence retrieval only (NFUC)

Human	Support	Refute	Neither
<b>Support</b>	4370	87	2209
<b>Refute</b>	3122	1474	2070
<b>Neither</b>	4159	214	2293

Dev confusion matrix for frame-based sentence retrieval only (NFIC)

Error analysis examples of predicting evidence & classification. Frame-based retrieval returns high-quality evidence sentences, but performance is reduced by accuracy of classifier and automatic frame annotator

**Acknowledgement:** Supported by gifts from the IBM AI Horizons Network and Northrop Grumman and a grant from the National Science Foundation.