

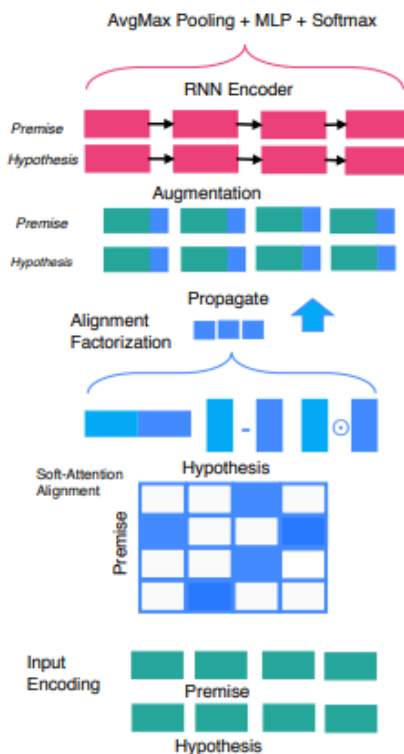
CAFE

C'est une nouvelle architecture d'apprentissage profond pour l'inférence en langage naturel (NLI). Dans un objectif précis: Étant donné une prémisse et une hypothèse, la NLI vise à détecter si la seconde implique ou contredit la première.

Architecture de modèle:

C'est une architecture dans laquelle le modèle prend en entrée une prémisse et une hypothèse sont comparées, compressées puis propagées vers des couches supérieures pour un apprentissage amélioré de la représentation. Ensuite, des couches de factorisation sont adaptées pour une compression efficace et expressive des vecteurs d'alignement en caractéristiques scalaires, qui sont ensuite utilisées pour augmenter les représentations des mots de base.

La figure ci-dessous représente l'architecture complète de notre modèle.



Pour tester l'efficacité de modèle et pour un objectif de comparaison de performance , trois ensembles de données de référence ont été utilisés: SNLI, MultiNLI, SciTail.

Les résultats obtenus montrent que le modèle CAFE est compétitif , avec une accuracy test qui varient selon les données(SNLI,MultiNLI, SciTail) et les différents paramètres de modèle choisis.

Les résultats sont montrés dans le tableaux suivants:

Tableau1 pour les données SNLI, Tableau 2: Comparaison MultiNLI, SciTail

Model	Params	Train	Test
Single Model (w/o Cross Sentence Attention)			
300D Gumbel TreeLSTM (Choi et al., 2017)	2.9M	91.2	85.6
300D DISAN (Shen et al., 2017)	2.4M	91.1	85.6
300D Residual Stacked Encoders (Nie and Bansal, 2017)	9.7M	89.8	85.7
600D Gumbel TreeLSTM (Choi et al., 2017)	10M	93.1	86.0
300D CAFE (w/o CA)	3.7M	87.3	<u>85.9</u>
Single Models			
100D LSTM with attention (Rocktäschel et al., 2015)	250K	85.3	83.5
300D mLSTM (Wang and Jiang, 2016b)	1.9M	92.0	86.1
450D LSTMN + deep att. fusion (Cheng et al., 2016)	3.4M	88.5	86.3
200D DecompAtt + Intra-Att (Parikh et al., 2016)	580K	90.5	86.8
300D NTI-SLSTM-LSTM (Yu and Munkhdalai, 2017b)	3.2M	88.5	87.3
300D re-read LSTM (Sha et al., 2016)	2.0M	90.7	87.5
BiMPM (Wang et al., 2017)	1.6M	90.9	87.5
448D DIIN (Gong et al., 2017)	4.4M	91.2	88.0
600D ESIM (Chen et al., 2017b)	4.3M	92.6	88.0
150D CAFE (SUM+2x200D MLP)	750K	88.2	87.7
200D CAFE (SUM+2x400D MLP)	1.4M	89.4	88.1
300D CAFE (SUM+2x600D MLP)	3.5M	89.2	<u>88.3</u>
300D CAFE (AVGMAX+300D HN)	4.7M	89.8	88.5
Ensemble Models			
600D ESIM + 300D Tree-LSTM (Chen et al., 2017b)	7.7M	93.5	88.6
BiMPM (Wang et al., 2017)	6.4M	93.2	88.8
448D DIIN (Gong et al., 2017)	17.0M	92.3	88.9
300D CAFE (Ensemble)	17.5M	92.5	89.3
External Resource Models			
BiAttentive Classification + CoVe + Char (McCann et al., 2017)	22M	88.5	88.1
KIM (Chen et al., 2017a)	4.3M	94.1	88.6
ESIM + ELMo (Peters et al., 2018)	8.0M	91.6	88.7
200D CAFE (AVGMAX + 200D MLP) + ELMo	1.4M	89.5	89.0

Table 1: Performance comparison of all published models on the SNLI benchmark.

Model	MultiNLI		SciTail
	Match	Mismatch	-
Majority	36.5	35.6	60.3
NGRAM [#]	-	-	70.6
CBOW ^b	65.2	64.8	-
BiLSTM ^b	69.8	69.4	-
ESIM ^{#,b}	72.4	72.1	70.6
DecompAtt [#] -	-	-	72.3
DGEM [#]	-	-	70.8
DGEM + Edge [#]	-	-	77.3
ESIM [†]	76.3	75.8	-
ESIM + Read [†]	77.8	77.0	-
CAFE	78.7	77.9	83.3
CAFE Ensemble	80.2	79.0	-

Table 2: Performance comparison (accuracy) on MultiNLI

