Results on Different Text Similarity Metrics

Dataset

"A transformer is a deep learning model that adopts the mechanism of attention, differentially weighing the significance of each part of the input data. It is used primarily in the field of natural language processing (NLP) and in computer vision (CV).",

"Like recurrent neural networks (RNNs), transformers are designed to handle sequential input data, such as natural language, for tasks such as translation and text summarization. However, unlike RNNs, transformers do not necessarily process the data in order. Rather, the attention mechanism provides context for any position in the input sequence. For example, if the input data is a natural language sentence, the transformer does not need to process the beginning of the sentence before the end. Rather, it identifies the context that confers meaning to each word in the sentence. This feature allows for more parallelization than RNNs and therefore reduces training times.",

"Transformers are the model of choice for NLP problems, replacing RNN models such as long short-term memory (LSTM). The additional training parallelization allows training on larger datasets than was once possible. This led to the development of pretrained systems such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), which were trained with large language datasets, such as Wikipedia Corpus and Common Crawl, and can be fine-tuned for specific tasks.",

"Before transformers, most state-of-the-art NLP systems relied on gated RNNs, such as LSTM and gated recurrent units (GRUs), with added attention mechanisms. Transformers built on these attention technologies without using an RNN structure, highlighting the fact that attention mechanisms alone can match the performance of RNNs with attention.",

"Gated RNNs process tokens sequentially, maintaining a state vector that contains a representation of the data seen after every token. To process the n-th token, the model combines the state representing the sentence up to token n-1 with the information of the new token to create a new state, representing the sentence up to token n. Theoretically, the information from one token can propagate arbitrarily far down the sequence, if at every point the state continues to encode contextual information about the token. In practice this mechanism is flawed: the vanishing gradient problem leaves the model's state at the end of a long sentence without precise, extractable information about preceding tokens.",

"This problem was addressed by attention mechanisms. Attention mechanisms let a model draw from the state at any preceding point along the sequence. The attention layer can access all previous states and weighs them according to a learned measure of relevancy, providing relevant information about far-away tokens.",

"A clear example of the value of attention is in language translation, where context is essential to assigning the meaning of a word in a sentence. In an English-to-French translation system, the first word of the French output most probably depends heavily on the first few words of the English input. However, in a classic LSTM model, in order to produce the first word of the French output, the model is given only the state vector of the last English word. Theoretically, this vector can encode information about the whole English sentence, giving the model all necessary knowledge. In practice this information is often poorly preserved by the LSTM. An attention mechanism can be added to address this problem: the decoder is given access to the state vectors of every English input word, not just the last, and can learn attention weights that dictate how much to attend to each English input state vector.",

"When added to RNNs, attention mechanisms increase performance. The development of the Transformer architecture revealed that attention mechanisms were powerful in themselves, and that sequential recurrent processing of data was not necessary to achieve the performance gains of RNNs with attention. Transformers use an attention mechanism without an RNN, processing all tokens at the same time and calculating attention weights between them in successive layers.",

"Like earlier models, the transformer adopts an encoder-decoder architecture. The encoder consists of encoding layers that process the input iteratively one layer after another, while the decoder consists of decoding layers that do the same thing to the encoder's output.",

"The function of each encoder layer is to generate encodings that contain information about which parts of the inputs are relevant to each other. It passes its encodings to the next encoder layer as inputs. Each decoder layer does the opposite, taking all the encodings and using their incorporated contextual information to generate an output sequence. To achieve this, each encoder and decoder layer makes use of an attention mechanism.",

"For each input, attention weighs the relevance of every other input and draws from them to produce the output. Each decoder layer has an additional attention mechanism that draws information from the outputs of previous decoders, before the decoder layer draws information from the encodings."

"Both the encoder and decoder layers have a feed-forward neural network for additional processing of the outputs, and contain residual connections and layer normalization steps.",

"The transformer building blocks are scaled dot-product attention units. When a sentence is passed into a transformer model, attention weights are calculated between every token simultaneously. The attention unit produces embeddings for every token in context that contain information about the token itself along with a weighted combination of other relevant tokens each weighted by its attention weight."

Randomly Selected Key

[9] "Like earlier models, the transformer adopts an encoder-decoder architecture. The encoder consists of encoding layers that process the input iteratively one layer after another, while the decoder consists of decoding layers that do the same thing to the encoder's output."

N-Word Synonym Swapped Key

"UnLike previous models, the transformer adopts an encoder-decoder architecture. The encoder consists of decoding layers that process the output iteratively one layer after another, while the decoder consists of decoding layers that do the same thing to the encoder's output."

Part Sentence Swapped Key

"The transformer adopts an encoder-decoder architecture like earlier models. While the decoder consists of decoding layers that do the same thing to the encoder's output, the encoder consists of encoding layers that process the input iteratively one layer after another."

Paragraph Added Key

"For each input, attention weighs the relevance of every other input and draws from them to produce the output. Each decoder layer has an additional attention mechanism that draws information from the outputs of previous decoders, before the decoder layer draws information from the encodings. Like earlier models, the transformer adopts an encoder-decoder architecture. The encoder consists of encoding layers that process th e input iteratively one layer after another, while the decoder consists of de coding layers that do the same thing to the encoder's output."

Result

Metrics	Exact Copy Past	N-Word Synonym Swap	Sentence Swap	Part-Sentence Swap	Paragraph Addition
Jaccard	[0.1016949152542373, 0.09090909090909091, 0.04395604395604396, 0.05714285714285714, 0.07368421052631578, 0.07042253521126761, 0.058823529411764705, 0.09210526315789473, 1.0, 0.1267605633802817, 0.16176470588235295, 0.0684931506849315]	[0.0847457627118644, 0.07, 0.044444444444444446, 0.057971014492753624, 0.07446808510638298, 0.08695652173913043, 0.059322033898305086, 0.09333333333333334, 0.8, 0.14492753623188406, 0.16417910447761194, 0.06944444444444445]	[0.0847457627118644, 0.07, 0.044444444444444446, 0.057971014492753624, 0.07446808510638298, 0.08695652173913043, 0.059322033898305086, 0.09333333333333334, 0.8, 0.14492753623188406, 0.16417910447761194, 0.06944444444444445]	[0.1016949152542373, 0.09090909090909091, 0.04395604395604396, 0.05714285714285714, 0.07368421052631578, 0.07042253521126761, 0.058823529411764705, 0.09210526315789473, 1.0, 0.1267605633802817, 0.16176470588235295, 0.0684931506849315]	[0.1016949152542373, 0.08, 0.04395604395604396, 0.05714285714285714, 0.07368421052631578, 0.07042253521126761, 0.06779661016949153, 0.10666666666666667, 0.6842105263157895, 0.1267605633802817, 0.14492753623188406, 0.0684931506849315]
Dice	[0.2033898305084746, 0.18181818181818182, 0.08791208791208792, 0.11428571428571428, 0.14736842105263157, 0.14084507042253522, 0.11764705882352941, 0.18421052631578946, 2.0, 0.2535211267605634, 0.3235294117647059, 0.136986301369863]	[0.1694915254237288, 0.14, 0.08888888888888889, 0.11594202898550725, 0.14893617021276595, 0.17391304347826086, 0.11864406779661017, 0.18666666666666668, 1.6, 0.2898550724637681, 0.3283582089552239, 0.1388888888888889]	[0.2033898305084746, 0.18181818181818182, 0.08791208791208792, 0.11428571428571428, 0.14736842105263157, 0.14084507042253522, 0.11764705882352941, 0.18421052631578946, 2.0, 0.2535211267605634, 0.3235294117647059, 0.136986301369863]	[0.2033898305084746, 0.16, 0.08791208791208792, 0.11428571428571428, 0.14736842105263157, 0.14084507042253522, 0.13559322033898305, 0.21333333333333335, 1.368421052631579, 0.2535211267605634, 0.2898550724637681, 0.136986301369863]	[0.22784810126582278, 0.25862068965517243, 0.12612612612612611, 0.13186813186813187, 0.19298245614035087, 0.27586206896551724, 0.2074074074074074, 0.23157894736842105, 1.1636363636363636, 0.3146067415730337, 0.9565217391304348, 0.21978021978021978]
Cosine	[0.48462375, 0.49867198, 0.28365431, 0.2941991 , 0.54750979, 0.30450009, 0.55427621, 0.41909854, 1., 0.52742292, 0.57073542, 0.27429019]	[0.48462375, 0.49867198, 0.28365431, 0.2941991 , 0.54750979, 0.30450009, 0.55427621, 0.41909854, 1. , 0.52742292, 0.57073542, 0.27429019]	[0.48462375, 0.49867198, 0.28365431, 0.2941991 , 0.54750979, 0.30450009, 0.55427621, 0.41909854, 1. , 0.52742292, 0.57073542, 0.27429019]	[0.48462375, 0.49867198, 0.28365431, 0.2941991 , 0.54750979, 0.30450009, 0.55427621, 0.41909854, 1. , 0.52742292, 0.57073542, 0.27429019]	[0.48462375, 0.49867198, 0.28365431, 0.2941991 , 0.54750979, 0.30450009, 0.55427621, 0.41909854, 1. , 0.52742292, 0.57073542, 0.27429019]
Pearson	[0.44739511460027376, 0.45256725972001105, 0.2056563001329873, 0.23398899799714282, 0.5132604967020087, 0.24165846581721243, 0.5163903235084463, 0.36759597796625937, 0.9999999999999996, 0.4887802772901767, 0.5361621917899391, 0.20817593335327225]	[0.44739511460027376, 0.45256725972001105, 0.2056563001329873, 0.23398899799714282, 0.5132604967020087, 0.24165846581721243, 0.5163903235084463, 0.36759597796625937, 0.9999999999999996, 0.4887802772901767, 0.5361621917899391, 0.20817593335327225]	[0.44739511460027376, 0.45256725972001105, 0.2056563001329873, 0.23398899799714282, 0.5132604967020087, 0.24165846581721243, 0.5163903235084463, 0.36759597796625937, 0.9999999999999996, 0.4887802772901767, 0.5361621917899391, 0.20817593335327225]	[0.44739511460027376, 0.45256725972001105, 0.2056563001329873, 0.23398899799714282, 0.5132604967020087, 0.24165846581721243, 0.5163903235084463, 0.36759597796625937, 0.9999999999999996, 0.4887802772901767, 0.5361621917899391, 0.20817593335327225]	[0.44739511460027376, 0.45256725972001105, 0.2056563001329873, 0.23398899799714282, 0.5132604967020087, 0.24165846581721243, 0.5163903235084463, 0.36759597796625937, 0.9999999999999996, 0.4887802772901767, 0.5361621917899391, 0.20817593335327225]
Spearman	[0.12541538025352447, 0.08406911001185906, 0.016827657540226866, 0.023719207843792074, 0.05373094179502448, 0.012476170320756853, 0.046027370148606486, 0.16097982464333196, 1.0, 0.18545188076044544, 0.23723372131192752, 0.006072156529226541 5]	[0.12541538025352447, 0.08406911001185906, 0.016827657540226866, 0.023719207843792074, 0.05373094179502448, 0.012476170320756853, 0.046027370148606486, 0.16097982464333196, 1.0, 0.18545188076044544, 0.23723372131192752, 0.0060721565292265415]	[0.12541538025352447, 0.08406911001185906, 0.016827657540226866, 0.023719207843792074, 0.05373094179502448, 0.012476170320756853, 0.046027370148606486, 0.16097982464333196, 1.0, 0.18545188076044544, 0.23723372131192752, 0.0060721565292265415]	[0.12541538025352447, 0.08406911001185906, 0.016827657540226866, 0.023719207843792074, 0.05373094179502448, 0.012476170320756853, 0.046027370148606486, 0.16097982464333196, 1.0, 0.18545188076044544, 0.23723372131192752, 0.0060721565292265415]	[0.12541538025352447, 0.08406911001185906, 0.016827657540226866, 0.023719207843792074, 0.05373094179502448, 0.012476170320756853, 0.046027370148606486, 0.16097982464333196, 1.0, 0.18545188076044544, 0.23723372131192752, 0.0060721565292265415]

Kendall Tau	[0.12384918044732406, 0.08144768146866946, 0.016341100744813625, 0.023159157811448856, 0.052123269583617945, 0.012258489539967334, 0.04413765501707866, 0.1577458254047948, 1.0, 0.18047438171911284, 0.2315026494914091, 0.005857038916733792]	[0.12384918044732406, 0.08144768146866946, 0.016341100744813625, 0.023159157811448856, 0.052123269583617945, 0.012258489539967334, 0.04413765501707866, 0.1577458254047948, 1.0, 0.18047438171911284, 0.2315026494914091, 0.005857038916733792]	[0.12384918044732406, 0.08144768146866946, 0.016341100744813625, 0.023159157811448856, 0.052123269583617945, 0.012258489539967334, 0.04413765501707866, 0.1577458254047948, 1.0, 0.18047438171911284, 0.2315026494914091, 0.005857038916733792]	[0.12384918044732406, 0.08144768146866946, 0.016341100744813625, 0.023159157811448856, 0.052123269583617945, 0.012258489539967334, 0.04413765501707866, 0.1577458254047948, 1.0, 0.18047438171911284, 0.2315026494914091, 0.005857038916733792]	[0.12384918044732406, 0.08144768146866946, 0.016341100744813625, 0.023159157811448856, 0.052123269583617945, 0.012258489539967334, 0.04413765501707866, 0.1577458254047948, 1.0, 0.18047438171911284, 0.2315026494914091, 0.005857038916733792]
Euclidean	[1.1538996284868939, 1.1467907481264772, 1.2823174355464255, 1.279956082293892, 1.1062040586090442, 1.275884519211488, 1.0958310218263925, 1.2016197851926989, 0.0, 1.104557961017176, 1.0745746731921395, 1.289844939061417]	[1.1723685186762434, 1.1713471378488962, 1.2866231218008393, 1.2841195072776608, 1.1167412908198493, 1.2724486130310735, 1.1086237481315944, 1.2088502469468412, 0.3880353772382839, 1.1066384926421984, 1.0855660114236956, 1.2936867081896344]	[1.1538996284868939, 1.1467907481264772, 1.2823174355464255, 1.279956082293892, 1.1062040586090442, 1.2758845192114878, 1.0958310218263922, 1.2016197851926986, 0.0, 1.104557961017176, 1.0745746731921397, 1.2898449390614168]	[1.1538996284868939, 1.1467907481264772, 1.2823174355464255, 1.279956082293892, 1.1062040586090442, 1.2758845192114878, 1.0958310218263922, 1.2016197851926989, 0.0, 1.104557961017176, 1.0745746731921395, 1.289844939061417]	[1.0992207363041158, 1.089409277885323, 1.2546994095694184, 1.2258987644067427, 1.0540072406203995, 1.1760777389776176, 1.018858351466245, 1.1362278992055295, 0.6211863535299019, 0.9699067554958443, 0.6420930145552705, 1.220189247568999]
Manhattan	[-0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651]	[-0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651]	[-0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651]	[-0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651]	[-0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651, -0.6532468086004651]
Hamming	[0.1935483870967742, 0.3709677419354839, 0.3193548387096774, 0.25161290322580643, 0.3870967741935484, 0.24193548387096775, 0.532258064516129, 0.25161290322580643, 0.0, 0.24838709677419354, 0.24193548387096775, 0.27419354838709675]	[0.1935483870967742, 0.3709677419354839, 0.3193548387096774, 0.25161290322580643, 0.3870967741935484, 0.24193548387096775, 0.532258064516129, 0.25161290322580643, 0.0, 0.24838709677419354, 0.24193548387096775, 0.27419354838709675]	[0.1935483870967742, 0.3709677419354839, 0.3193548387096774, 0.25161290322580643, 0.3870967741935484, 0.24193548387096775, 0.532258064516129, 0.25161290322580643, 0.0, 0.24838709677419354, 0.24193548387096775, 0.27419354838709675]	[0.1935483870967742, 0.3709677419354839, 0.3193548387096774, 0.25161290322580643, 0.3870967741935484, 0.24193548387096775, 0.532258064516129, 0.25161290322580643, 0.0, 0.24838709677419354, 0.24193548387096775, 0.27419354838709675]	[0.1935483870967742, 0.3709677419354839, 0.3193548387096774, 0.25161290322580643, 0.3870967741935484, 0.24193548387096775, 0.532258064516129, 0.25161290322580643, 0.0, 0.24838709677419354, 0.24193548387096775, 0.27419354838709675]
Minkowski	[60.0, 115.0, 99.0, 78.0, 120.0, 75.0, 165.0, 78.0, 0.0, 77.0, 75.0, 85.0]	[60.0, 115.0, 99.0, 78.0, 120.0, 75.0, 165.0, 78.0, 0.0, 77.0, 75.0, 85.0]	[60.0, 115.0, 99.0, 78.0, 120.0, 75.0, 165.0, 78.0, 0.0, 77.0, 75.0, 85.0]	[60.0, 115.0, 99.0, 78.0, 120.0, 75.0, 165.0, 78.0, 0.0, 77.0, 75.0, 85.0]	[60.0, 115.0, 99.0, 78.0, 120.0, 75.0, 165.0, 78.0, 0.0, 77.0, 75.0, 85.0]