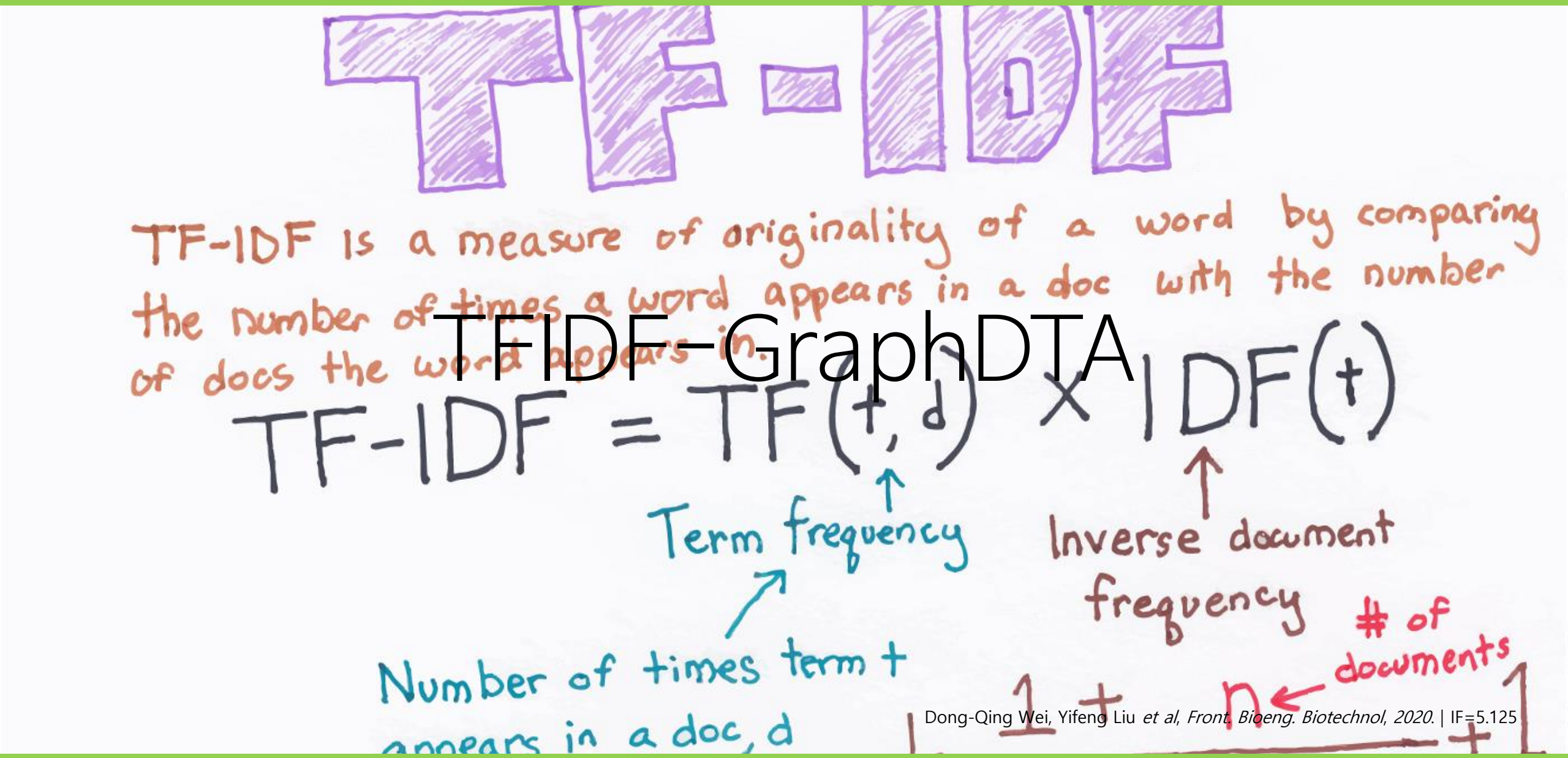


TFIDF-GraphDTA

Dong-Qing Wei, Yifeng Liu *et al*, *Front. Bioeng. Biotechnol*, *2020*. | IF=5.125

frontiers in Bioengineering and Biotechnology

METHODS
published: 03 April 2020
doi: 10.3389/fbioe.2020.00267

Check for updates

Dipeptide Frequency of Word Frequency and Graph Convolutional Networks for DTA Prediction

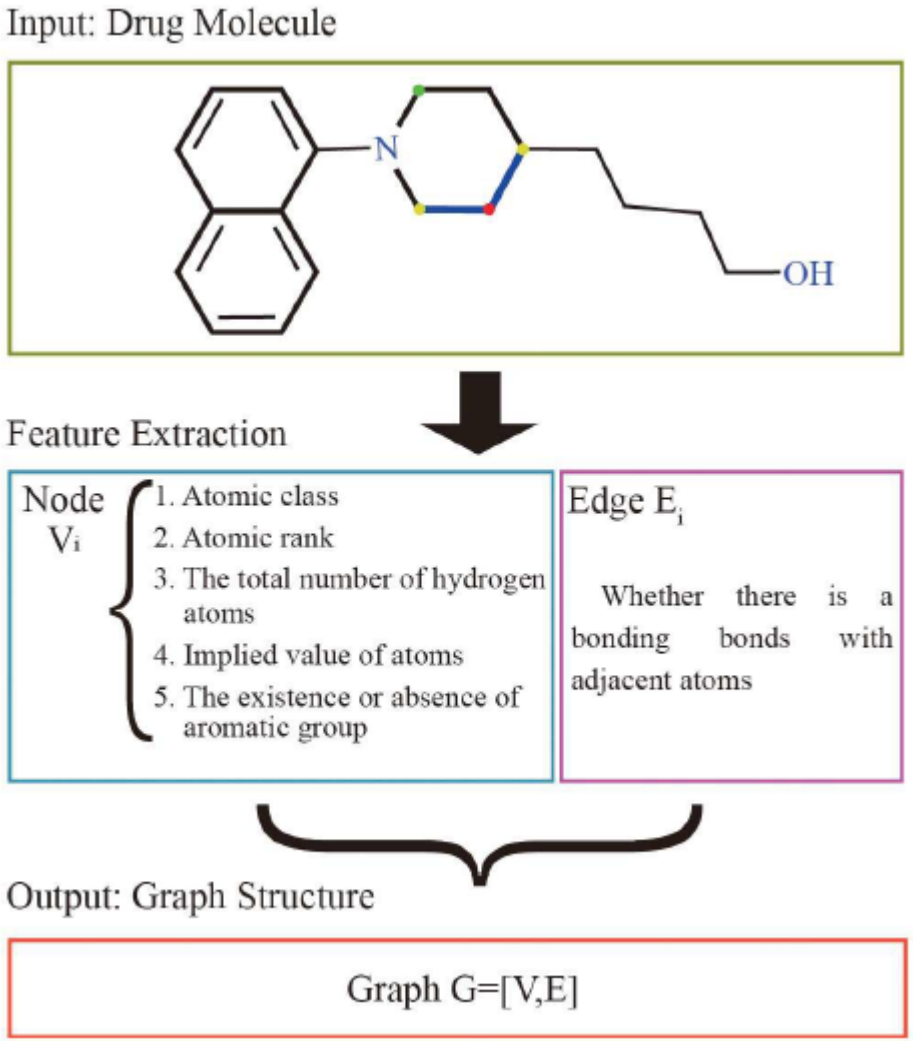
Xianfang Wang[1,2*], *Yifeng Liu*[2], *Fan Lu*[2], *Hongfei Li*[2], *Peng Gao*[2] *and Dongqing Wei*[3]

[1] School of Computer Science and Technology, Henan Institute of Technology, Xinxiang, China, [2] School of Computer and Information Engineering, Henan Normal University, Xinxiang, China, [3] School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

Outline

- Drug Molecular Feature Extraction
- Protein Sequence Feature Extraction
- Network Model Construction
- Results and Discussion

Drug Molecular Feature Extraction

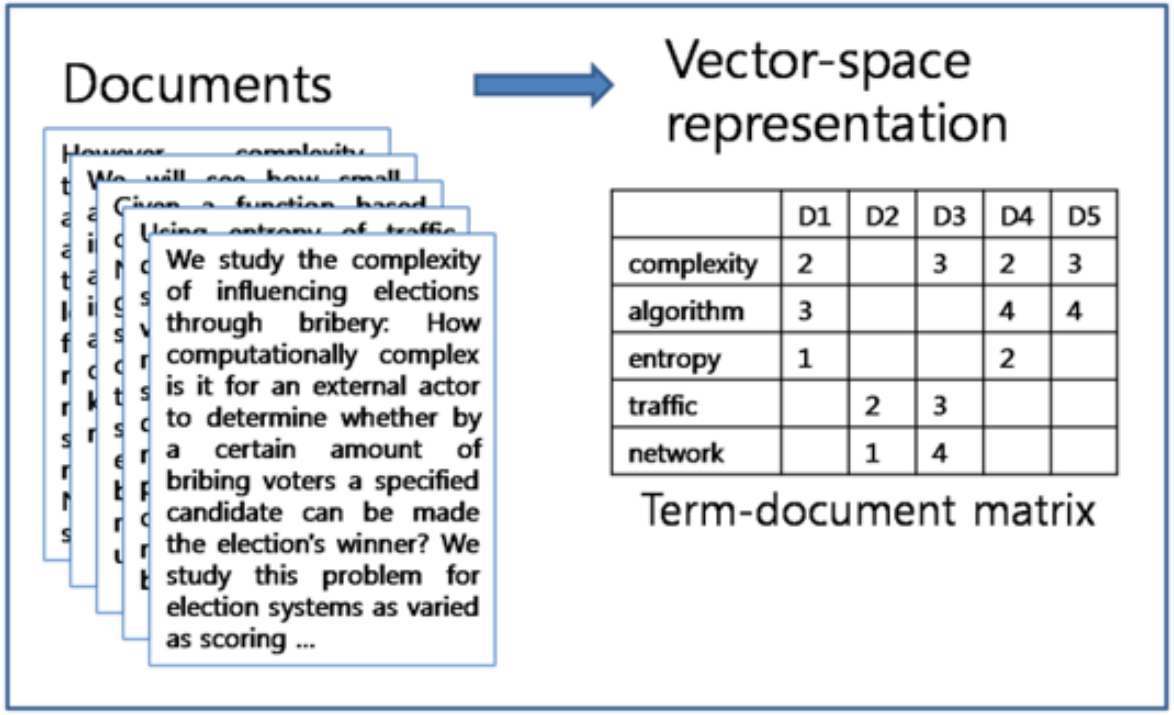


$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (8, 8)$$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots \\ x_{81} & x_{82} & x_{83} \end{bmatrix} \quad (8, 3)$$

Protein Sequence Feature Extraction

- term frequency - inverse document frequency (TF-IDF) algorithm is employed.
- This algorithm plays an important role in natural language process (NLP)
- polypeptide frequency is similar to the calculation process of TF in bioinformatics.



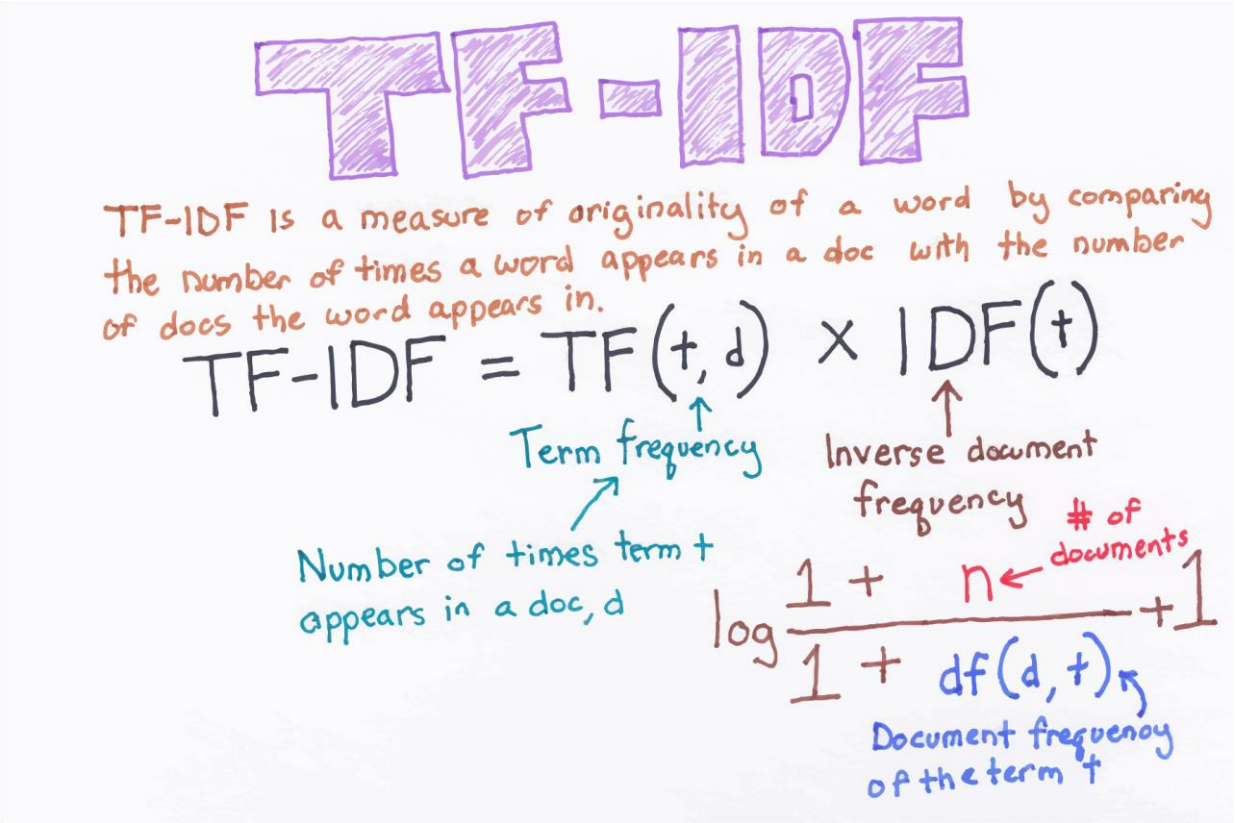
	D1	D2	D3	D4	D5
complexity	2		3	2	3
algorithm	3			4	4
entropy	1			2	
traffic		2	3		
network		1	4		

term	df_t	idf_t
car	18,165	1.65
auto	6723	2.08
insurance	19,241	1.62
best	25,235	1.5

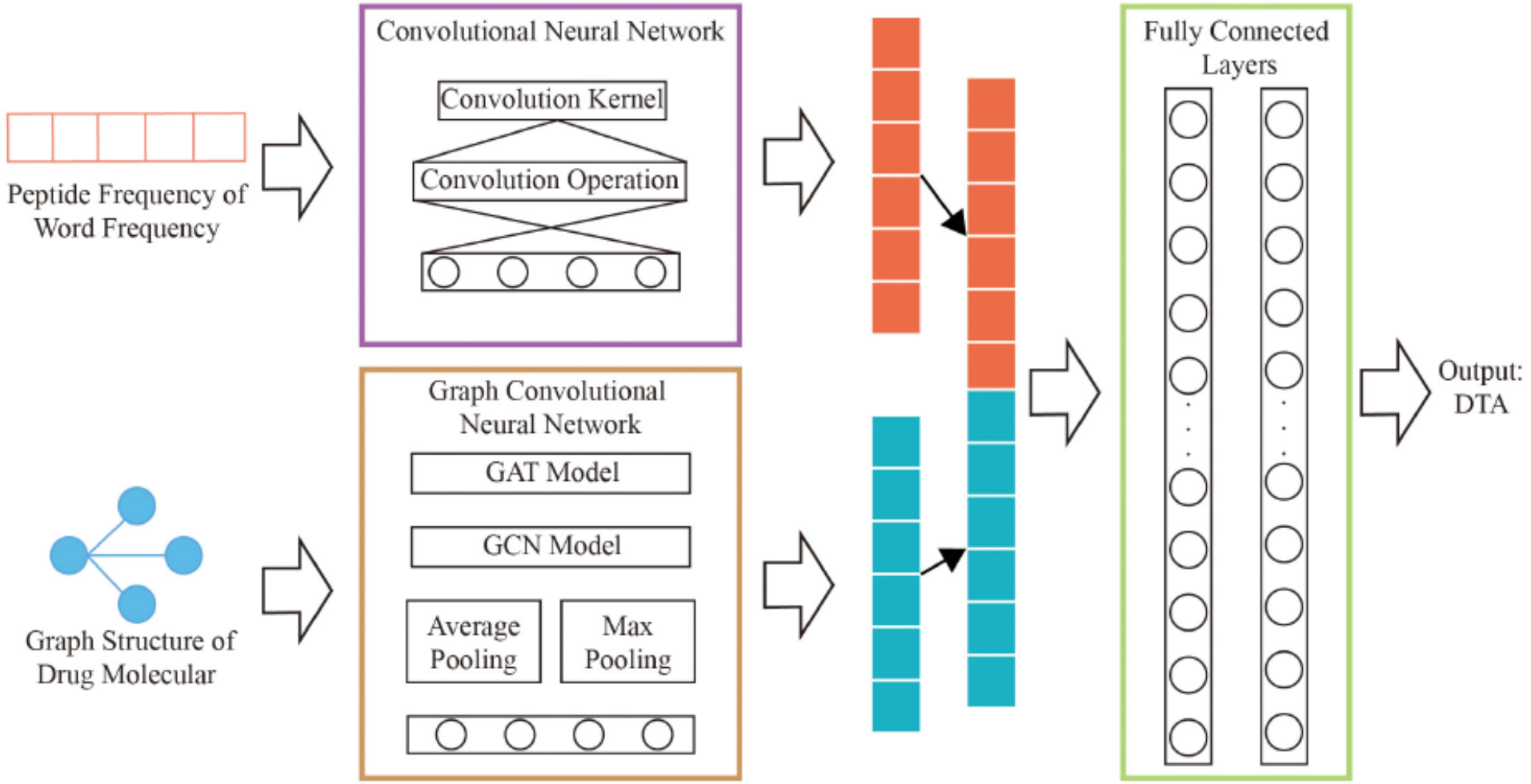
$$idf_j = log\left[\frac{n}{df_j}\right]$$

Protein Sequence Feature Extraction

- term frequency - inverse document frequency (TF-IDF) algorithm is employed.
- This algorithm plays an important role in natural language process (NLP)
- polypeptide frequency is similar to the calculation process of TF in bioinformatics.



Network Model Construction



Result and Discussion

Features	KIBA		Davis	
	MSE	CI	MSE	CI
DeepDTA	0.194	0.863	0.261	0.878
WipeDTA	0.179	0.875	0.262	0.886
This model	**0.126**	**0.901**	**0.220**	**0.899**