

The background features a light gray, swirling pattern on the left side, resembling stylized vines or water. Scattered throughout the background are several small, light gray butterfly silhouettes in various orientations.

Application of Sum-Product Network on Speech Recognition

BY Peidong Wang

Contents



motivation



former work



my solution



results



difficulties



future work

Motivation

- We can regard the output of a classical DNN as a linear polynomial. Like

$$y = a_1x_1 + a_2x_2 + \cdots + a_Nx_N$$

- If we add product nodes/layers to the network, the output will become a higher order polynomial. For example,

$$y = a_{1,1}x_1^2 + a_{1,2}x_1x_2 + \cdots + a_{N,N-1}x_{N-1}x_N + a_{N,N}x_N^2$$



Motivation

- Benefits:
 - 1. The number of the variables in the output is greater.
 - 2. Higher order polynomial is more expressive, compared with linear polynomial.



Former Work

- **Image Completion:**
- *Sum-product networks: A new deep architecture*
 - Hoifung Poon and Pedro Domingos
 - Proc. 12th Conf. on Uncertainty in Artificial Intelligence
- **Contribution:**
 - Proposed a new architecture called sum-product network.
 - Theoretical definitions and inferences provided.



Former Work

- **Image Classification:**
- *Discriminative Learning of Sum-Product Networks*
 - Robert Gens and Pedro Domingos
 - NIPS 2012 best student paper award
- **Contribution:**
 - Proposed a discriminative sum-product network structure.
 - Broadened the use of sum-product network to the area of image classification.



Former Work

- **Language Modeling:**

- *Language Modeling with Sum-Product Networks*

- Wei-Chen Cheng and Stanley Kok

- **Contribution:**

- Proposed an architecture based on discriminative sum-product network to train language models.

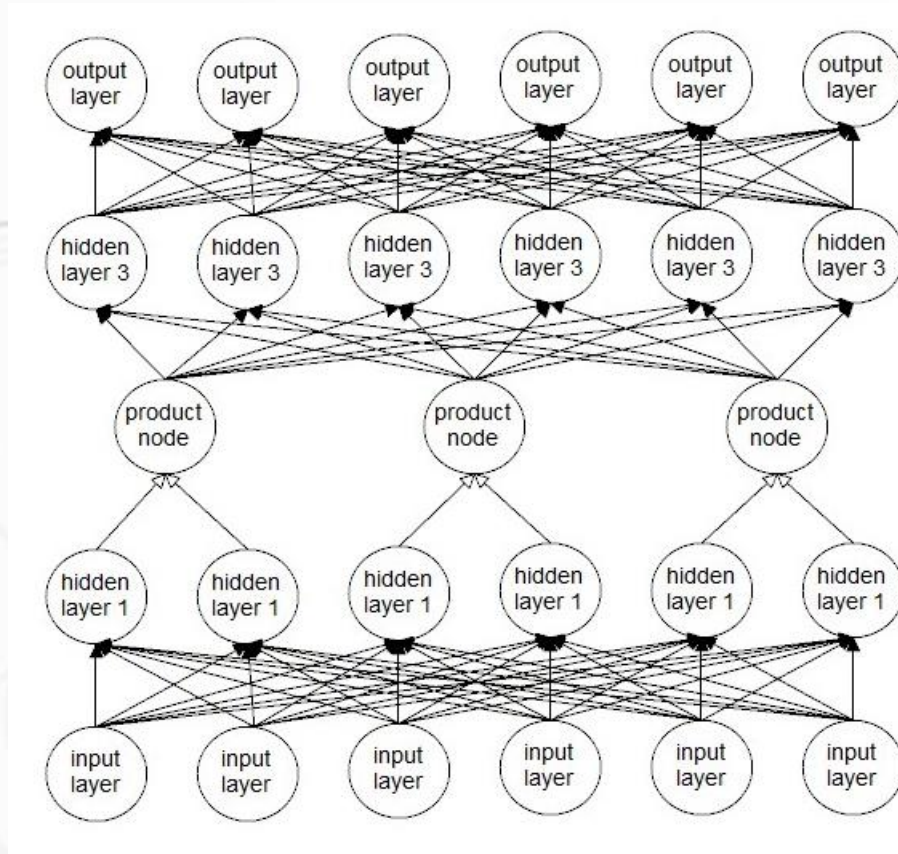


My Solution

- Substitute one classical layer with one product layer in a five-layer network.
- The transfer function of the product layer is simply $y=x$, rather than sigmoid function.
- A corresponding modification of Back Propagation algorithm.

My Solution

- Network Structure:

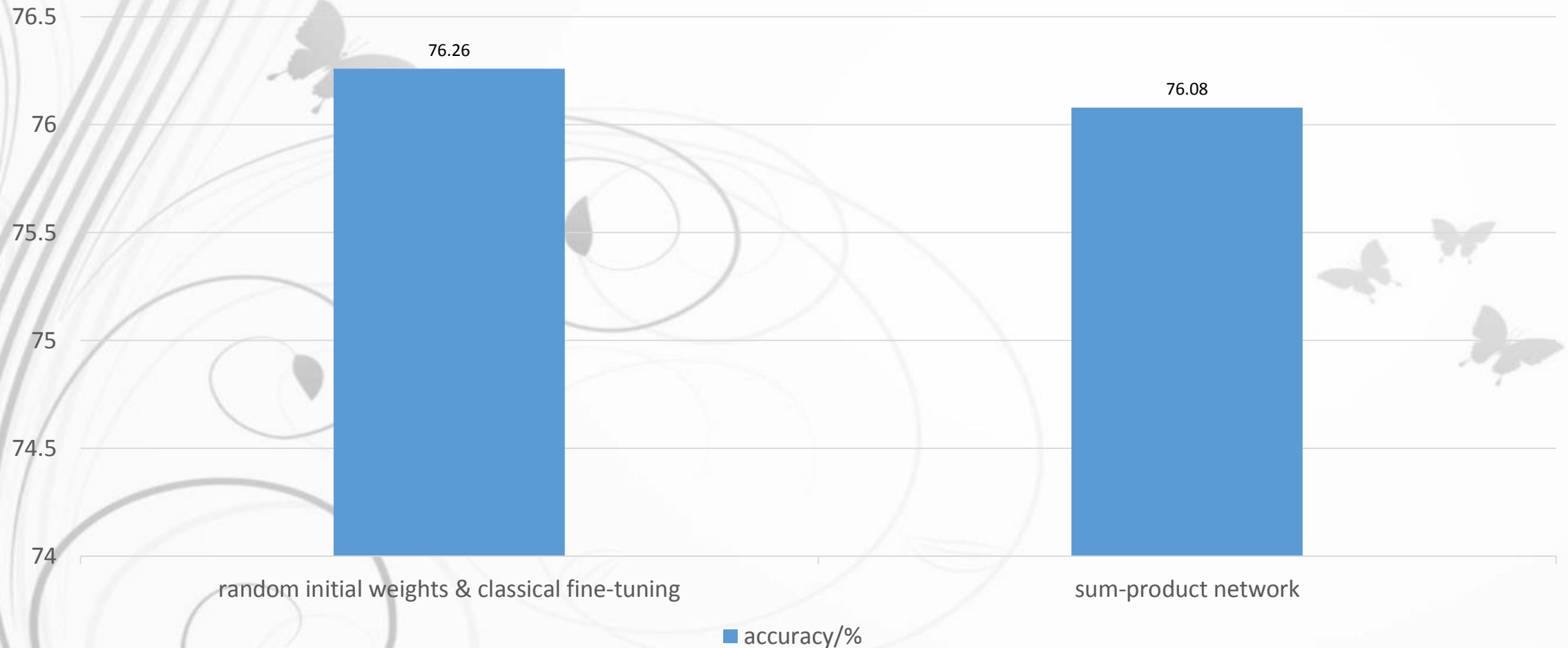


Results

- corpus:
 - TIMIT
- layer number (total):
 - five
- layer sizes:
 - 429 2048 1024 2048 183
- contrast group:
 - 429 2048 2048 183
 - random initial weights & classical fine-tuning

Results

comparison between classical network and SPN



Results

- look at accuracy:
 - comparable
 - common fluctuation on a small corpus
- look at complexity:
 - reduces half the complexity
 - each weight under product node is constantly one, no need to update

Results (Relevant Experiments)

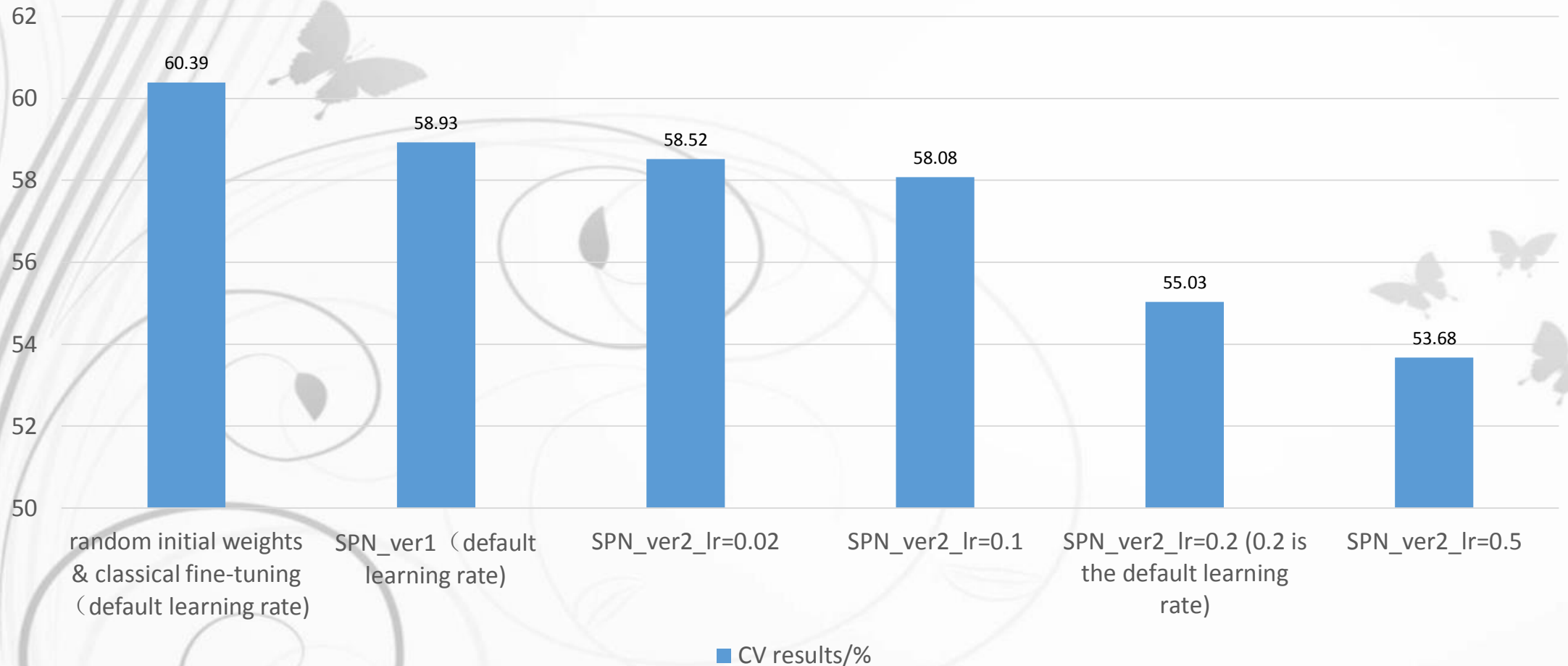
- Some relevant experiments were conducted in order to find the most suitable setting.
- These modifications mainly focused on changing the transfer functions and learning rates.
- The results are CV results, rather than formal results yield by testing.

Results (Relevant Experiments)

- Second version of SPN:
 - Changed the transfer function of the second layer (bottom-top order, the layer under the product layer) to $y=x$.
 - Change the transfer function of the product layer to sigmoid function.

Results (Relevant Experiments)

comparison between SPN version two and former networks

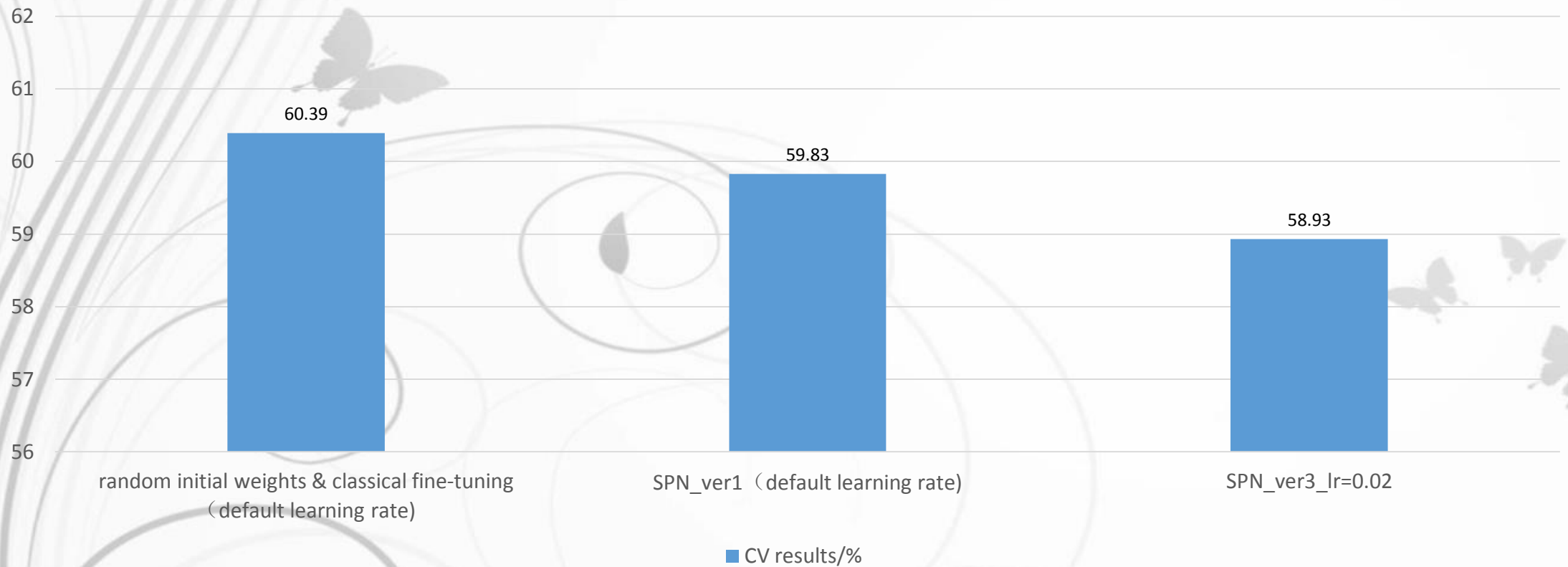


Results (Relevant Experiments)

- Third version of SPN:
 - Changed the transfer function of the second layer (bottom-top order, the layer under the product layer) to $y=x$, compared with SPN version one.
 - The learning rate was set 0.02, based on the knowledge acquired from the second version of SPN.

Results (Relevant Experiments)

comparison between SPN version three and former networks



Results (Summary)

- We can learn from the experiments above that the first version of SPN yields the best result of all of the three versions.
- Although the accuracies of SPNs are not persuasive at the first glance, SPNs have an undeniable reduction of complexity.
- If we regard the combination of one product layer and the nether sum layer as one layer, as is often used, SPNs seem to have a promising future.

Difficulties

- No former use of this architecture in the area of speech recognition./A lack of relevant papers and experiments. (Or we can regard it as a chance)
- Difficulties in regularizing the range of the outputs of product layers.
- Difficulties in deducing corresponding BP algorithm, accompanied by the addition of product layers.



Future Work

- Conduct experiments on a larger corpus.
 - Currently using PSC, which is a 65h Chinese corpus.
- Add layers, including product layers, to yield a better accuracy.
- Find ways to deal with the range of the outputs of product layers. Perhaps use ReLU function as the transfer function of the product layers, while using linear function on the nether sum layers.

The background features a light gray, swirling pattern on the left side, transitioning into a white background with faint, scattered butterfly silhouettes. The text 'Thank you' is centered in a blue, 3D-style font.

Thank you