


Computer-assisted Classification at China National Intellectual Property Administration



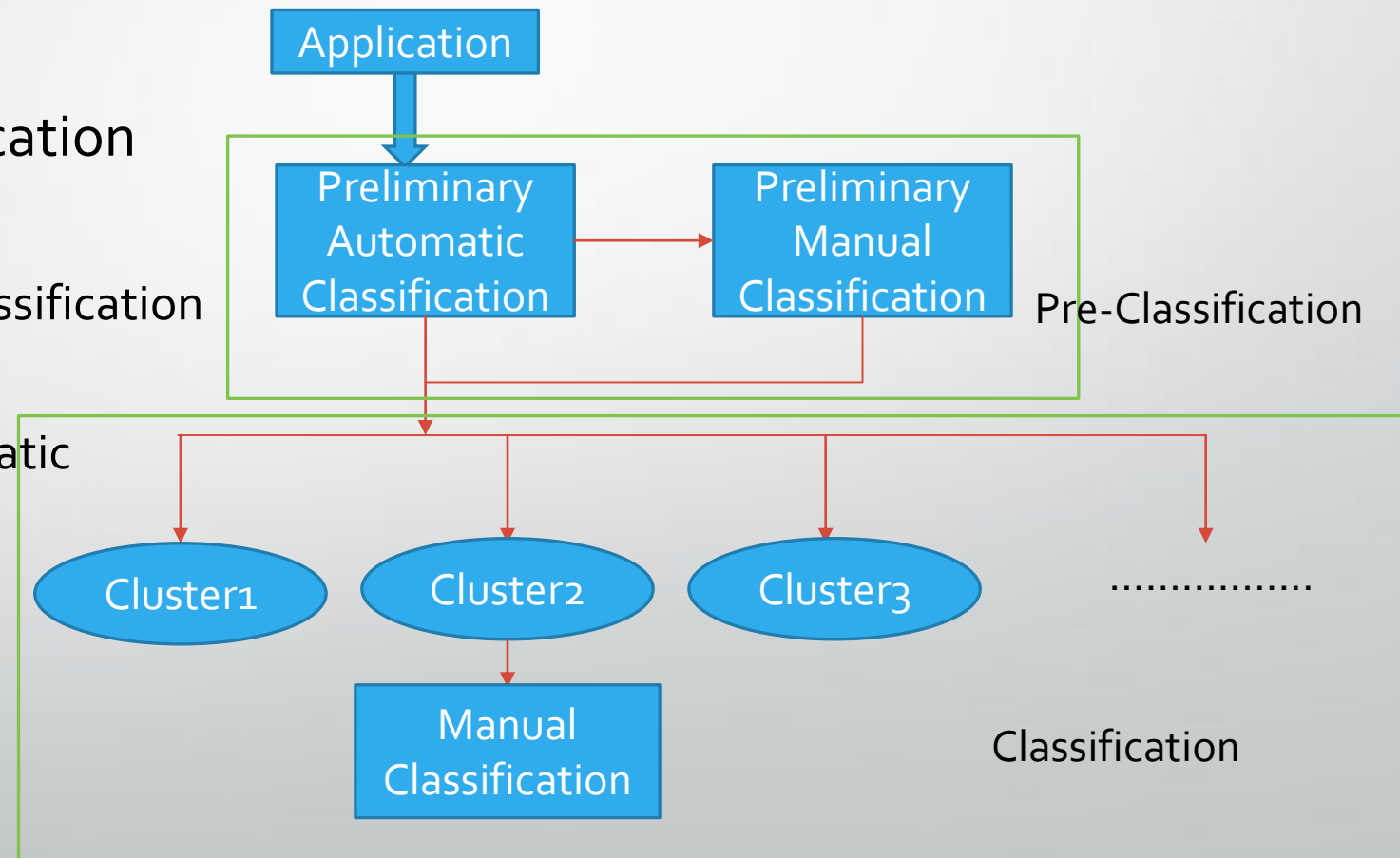
Content

- Computer-assisted Classification Practices at CNIPA
- Computer-assisted Classification Research at CNIPA
- Research Orientation in the Future

- 
- The classification of patent applications and the development of related technologies are mainly undertaken by China Patent Technology Development Company(CPDC).
 - CPDC is affiliated to China National Intellectual Property Administration(CNIPA).

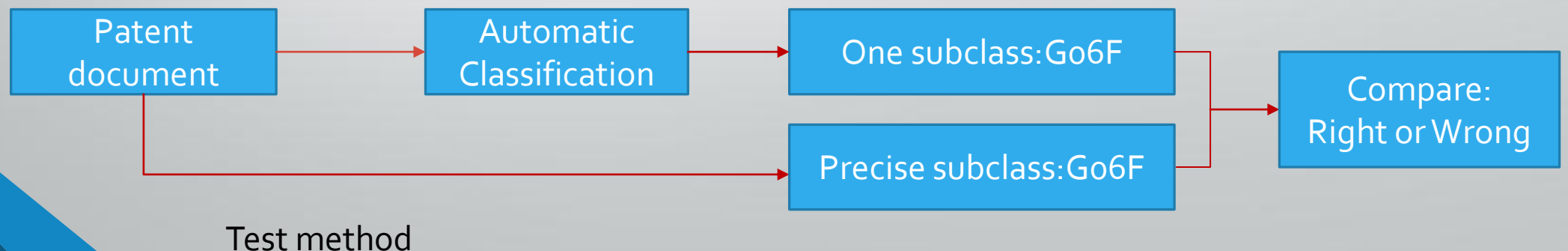
Computer-assisted Classification Practices

- Computer-assisted Classification Process at CNIPA
 - Automatic classification: classification to subclasses
 - E system and its dual automatic classification system



Computer-assisted Classification Practices

- Computer-assisted Classification Research and Development
 - Automatic classification application by using deep learning technology
 - Preliminary classification accuracy rate is 74.3% ; test data size is about 1.07million
 - 1 millisecond time consumed per patent document, 22 minutes 35 seconds consumed totally for 1.07 million patent documents



Computer-assisted Classification Research

- Natural Language Processing for Patent Documents
 - Preprocessing
 - Filter interference information
 - Chinese Word Segment
 - Difficulties in CWS
 - Offer training corpus for pre-training linguistic models
 - Less ambiguity word units
 - Semantic Role Labeling
 - More sentence level information

南京市长江大桥批准了铁路修建工程。

After chinese word segmenting:

南京/市长/江大桥/批准/了/铁路修建工程。

南京市/长江大桥/批准/了/铁路/修建工程。

The Mayor of Nanjing, Jiang Daqiao, approved the railway construction project.

The Nanjing Yangtze River Bridge approved the railway construction project.

Machine Translation

Computer-assisted Classification Research

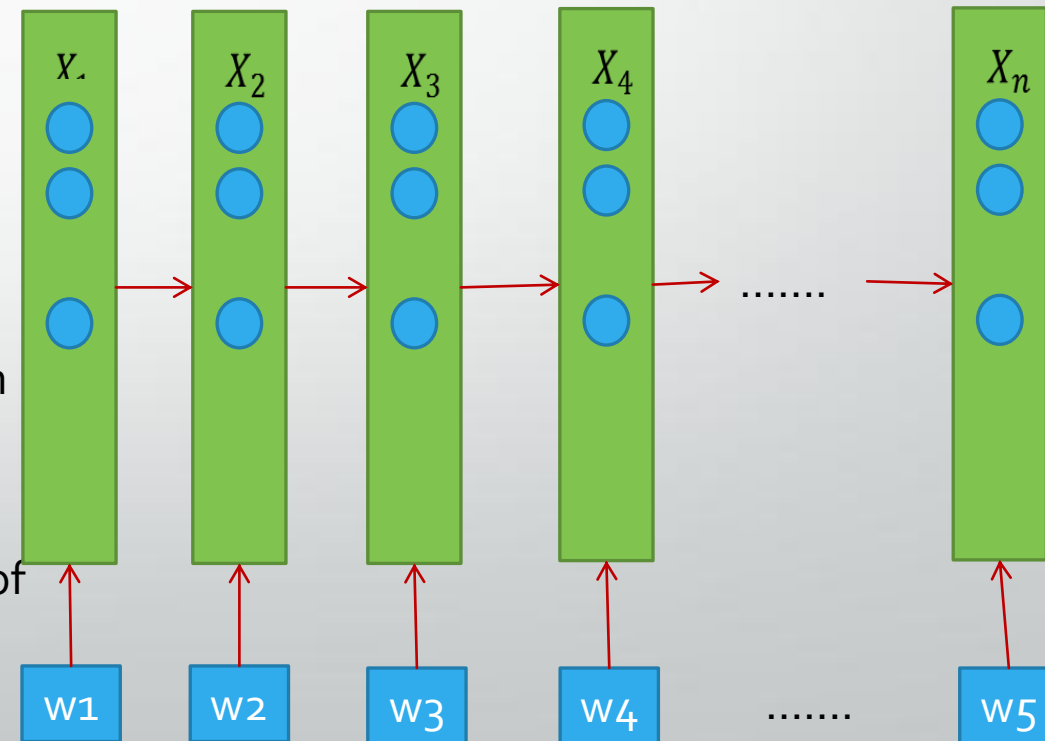
- Patent Classification Modeling
 - Deep Learning Model modules
 - Word Embedding
 - Model word level information with linguistic model, using continuous word vectors to represent discrete words
 - LSTM、GRU、Bidirectional LSTM、Bidirectional GRU
 - Model the sequence information and structure, combine long distance context information
 - Attention mechanism
 - Model the importance of different parts in a sequence, for example, a sentence contains different word units, the words have different syntactic and semantic functions, the importance of them is different
 - Transformers and Bert

Computer-assisted Classification Research

- Patent Classification Modeling

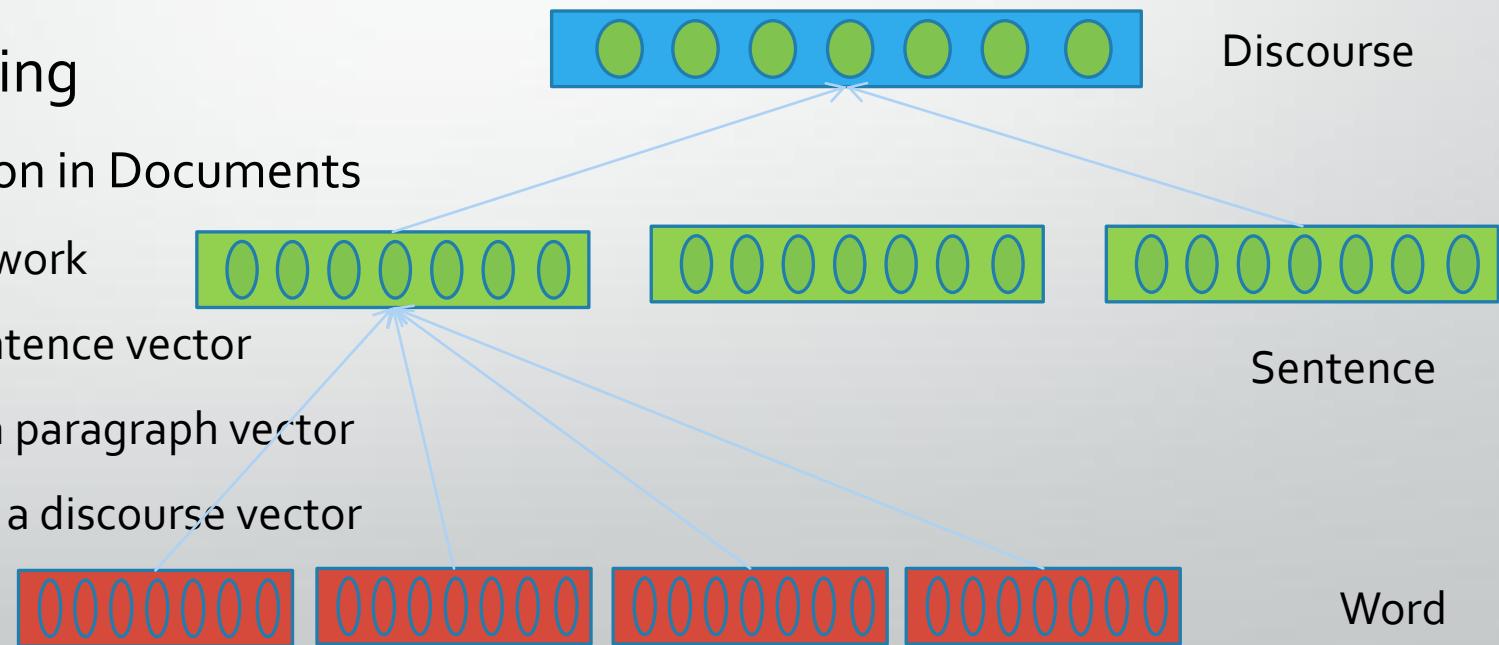
- Using Sequence Semantic Information

- Words in a sentence, sentences in a paragraph contain structural semantic information
 - Using LSTM、Bidirectional LSTM、GRU、Bidirectional GRU etc. to model words sequence in sentences
 - Using Attention Mechanism to model attention process and distinguish the semantic importance of different parts of the sentences



Computer-assisted Classification Research

- Patent Classification Modeling
 - Using Hierarchical Information in Documents
 - Hierarchical Attention Network
 - Form word vectors to a sentence vector
 - Form sentence vectors to a paragraph vector
 - Form paragraph vectors to a discourse vector



Computer-assisted Classification Research

- Automatic Classification of Non-patent Documents
 - Classifying scientific documents
- Automatic Industrial Classification of Patent Documents
 - Application of automatic classification methods to other classification system, such as 'Industrial classification for national economic activities' which is similar to 'International Standard Industrial Classification'

Research Orientation in the Future

- Improve the precision and efficiency of preliminary automatic classification
- Develop methods to automatically classify the applications to group level
- Develop methods to automatically classify/reclassify the backfiles or the new-filings before their publication



Thank you