More Efficient Learning by Structuring, Classifying and Understanding Lectures in Online Courses

藉助線上課程之自動結構化、分類與理解以提升學習效率

SPEAKER：沈昇勳（SHENG-SYUN SHEN）
ADVISOR：李琳山（LIN-SHAN LEE）

Outline

- Introduction
- Structuring Lectures (課程內容結構化)
  - Within A Course
  - Between Courses
- Classifying Lectures (對課程小段做內容精確分類)
- Understanding Lectures (機器對課程小段內容做了解)
- Conclusion

Outline

- Introduction
- Structuring Lectures（課程內容結構化）
  - Within A Course
  - Between Courses
- Classifying Lectures（對課程小段做內容精確分類）
- Understanding Lectures（機器對課程小段內容做了解）
- Conclusion

Introduction

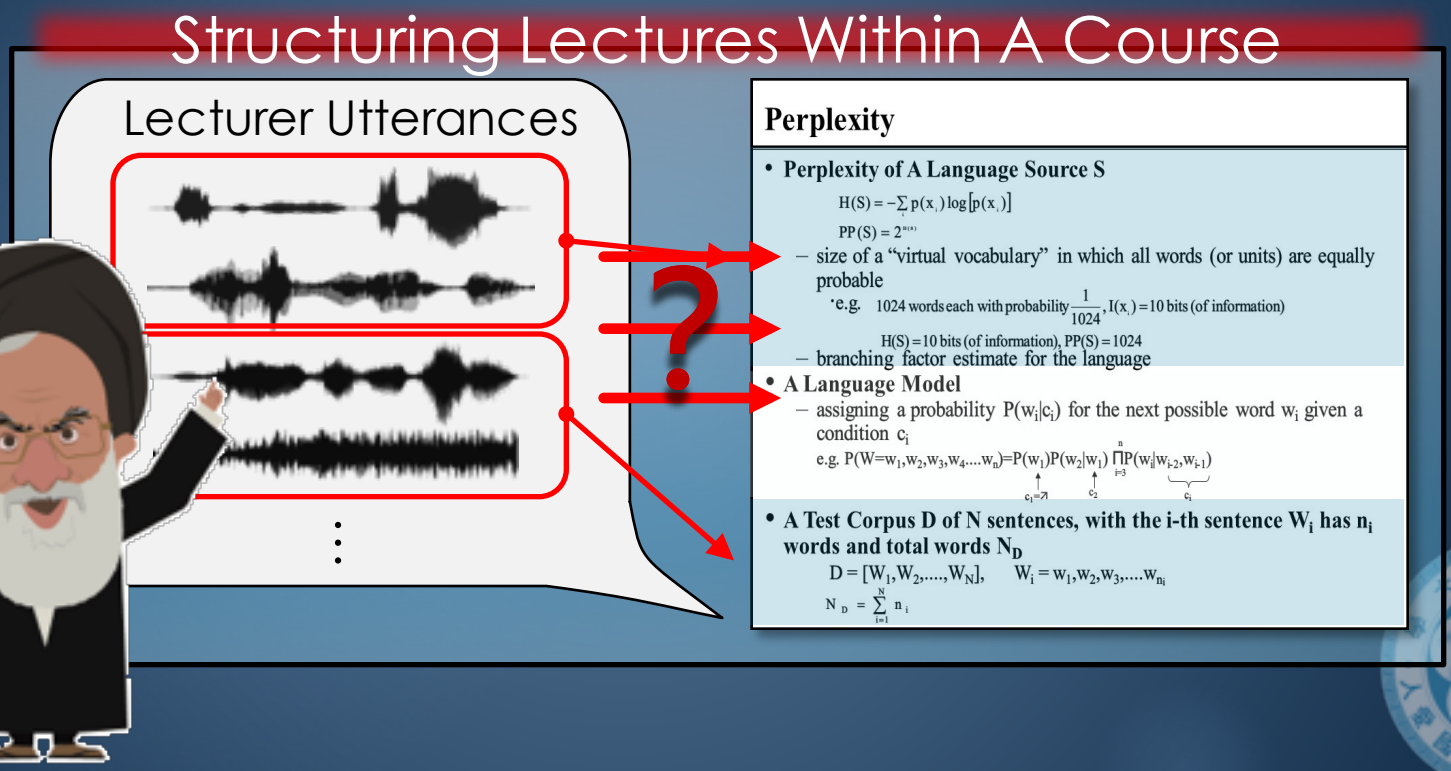
背景介紹

Motivation

- A huge number of online courses are available
  - On Massive Open Online Courses (MOOCs) platform
  - Coursera, edX, etc.

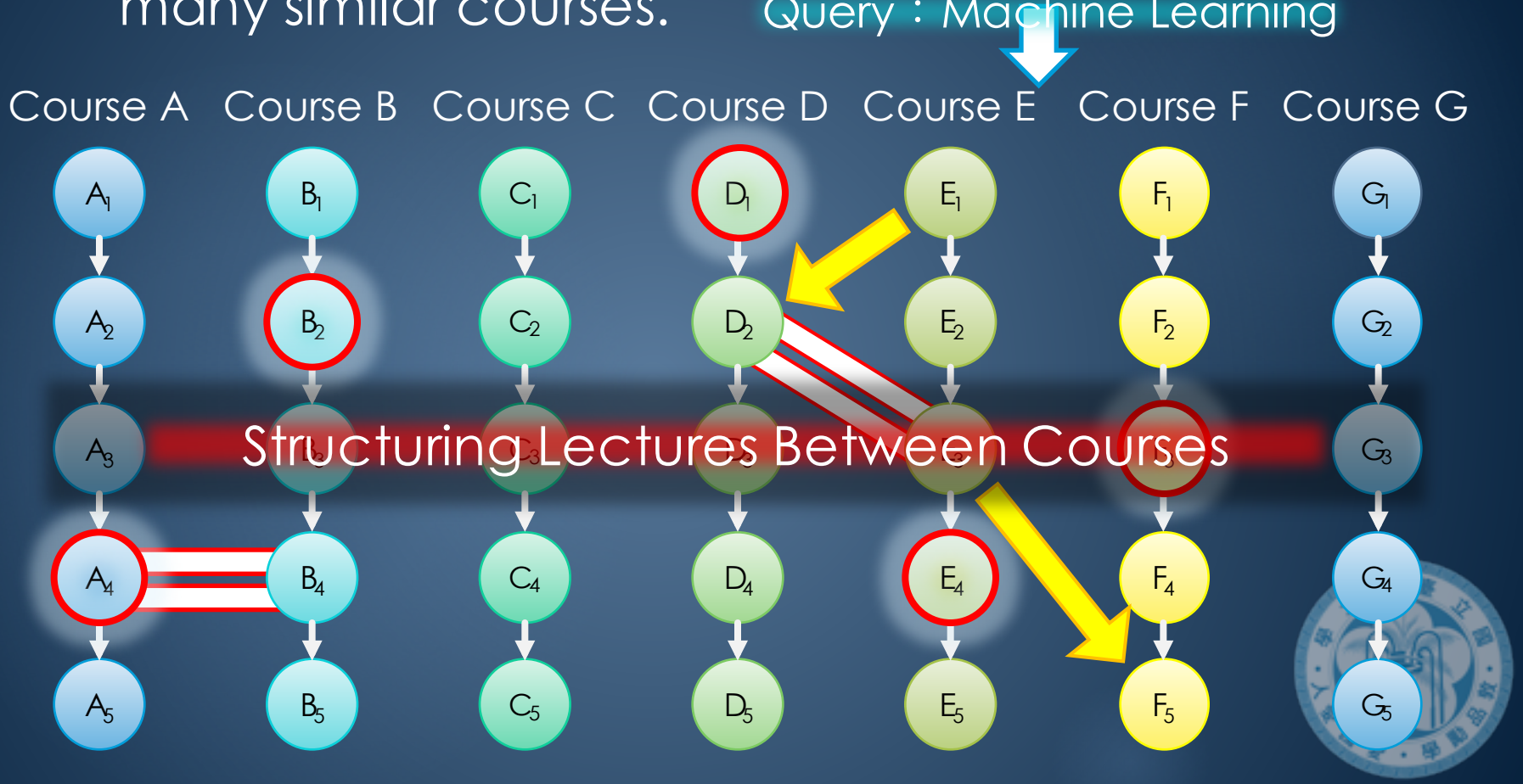
Motivation

- Not sure about which part of the slide is the lecturer talking about.



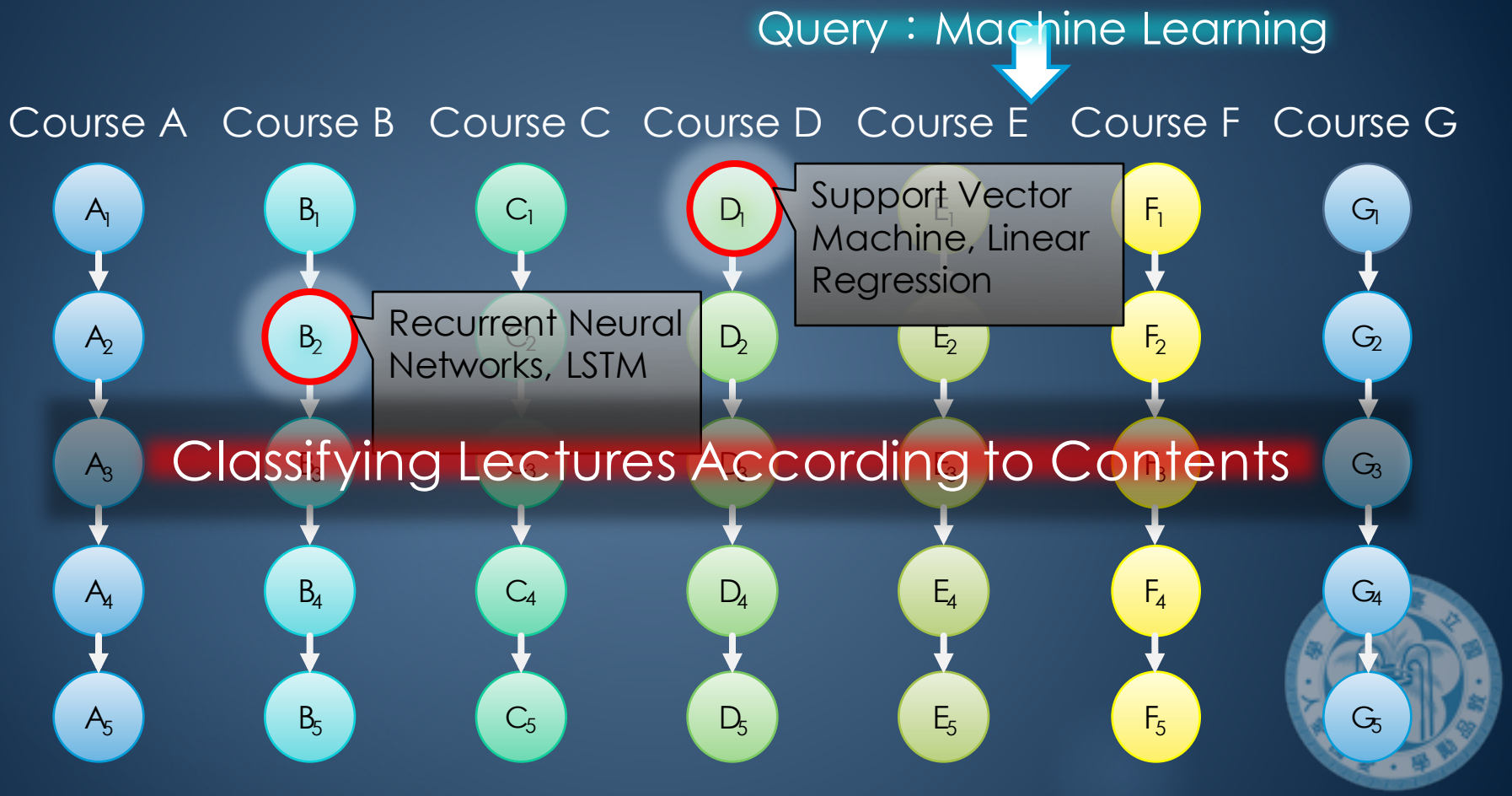
Motivation

- Not knowing which lecture to choose among many similar courses.



Motivation

- Not knowing the details about lectures.



Motivation

- Wondering whether machines could really understand the content of lectures.

Understanding Lecture Contents

Outline

- Introduction
- Structuring Lectures（課程內容結構化）
  - Within A Course
  - Between Courses
- Classifying Lectures（對課程小段做內容精確分類）
- Understanding Lectures（機器對課程小段內容做了解）
- Conclusion

Structuring Lectures

課程內容結構化

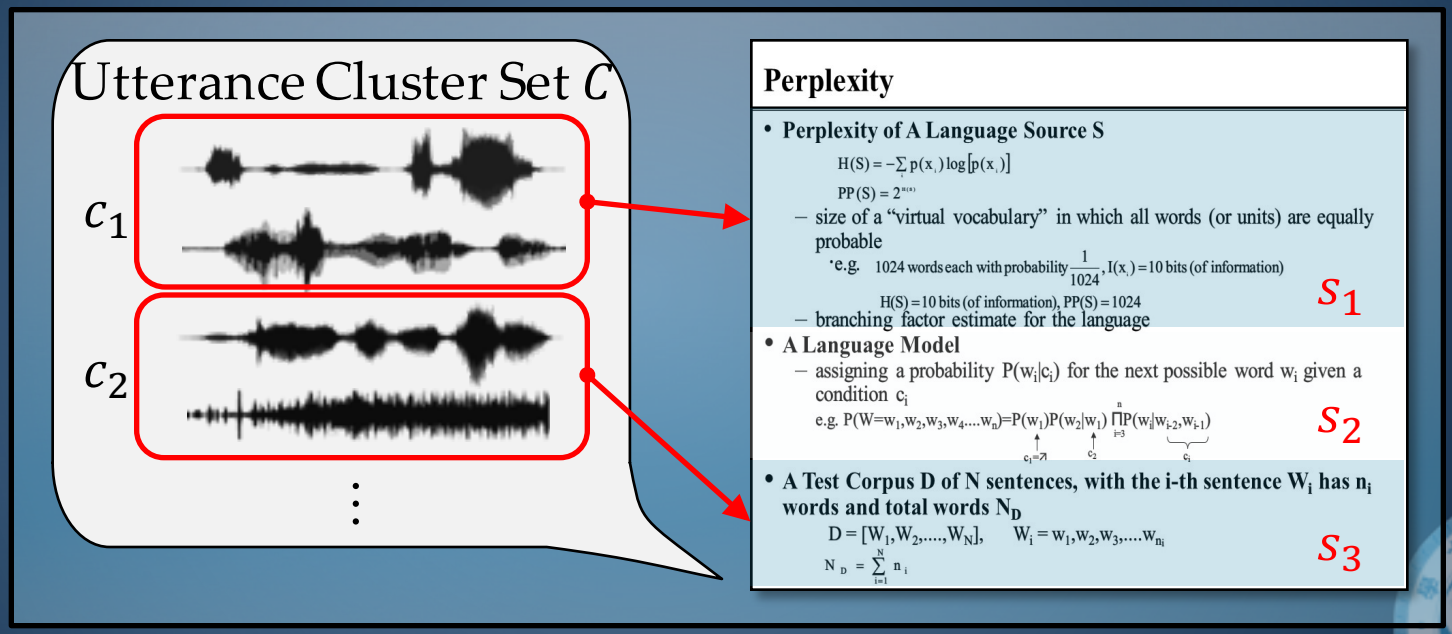
Structuring Lectures

WITHIN A COURSE

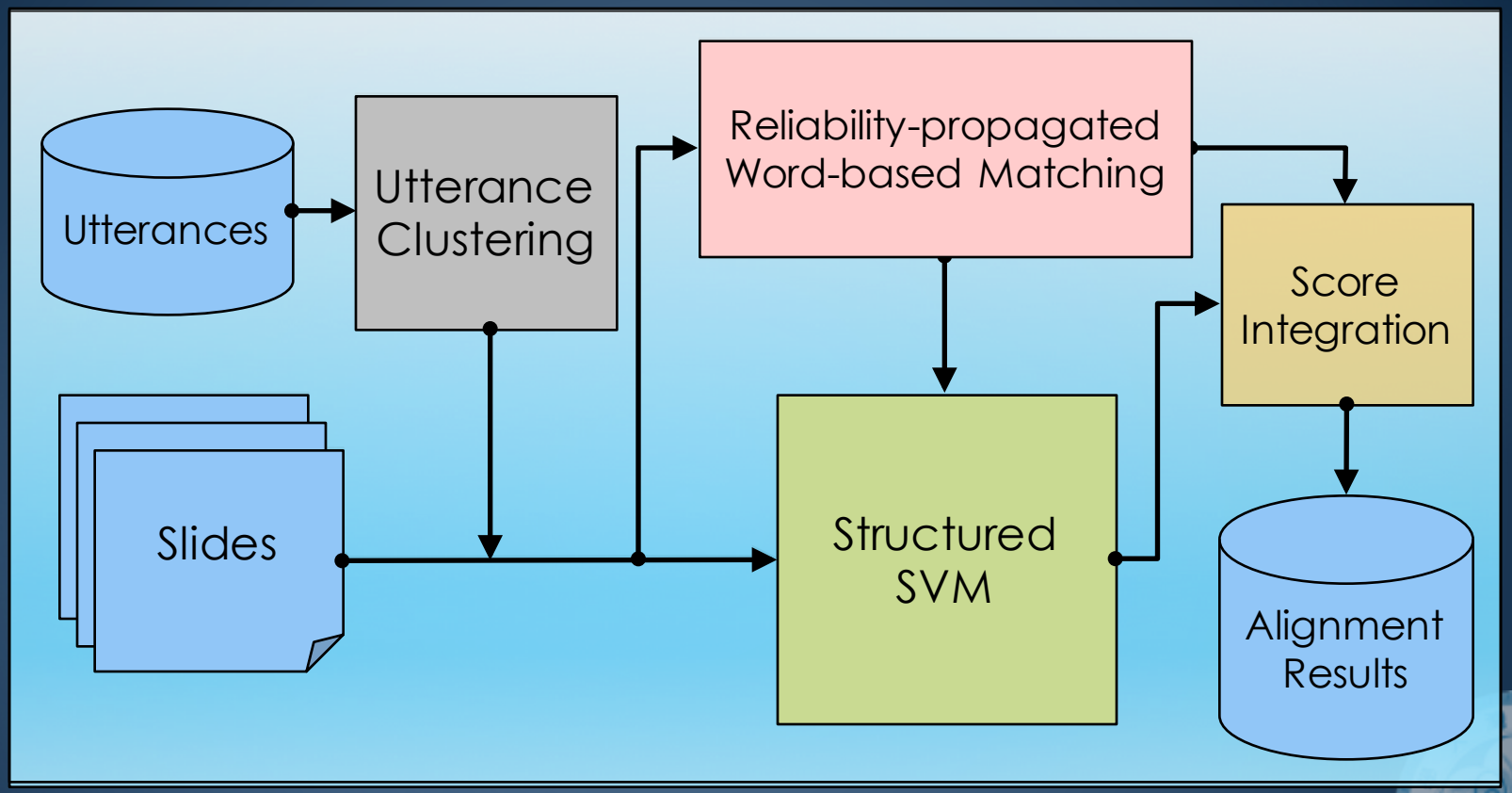
Within A Course

- Target :

  Align utterance cluster → Slide section

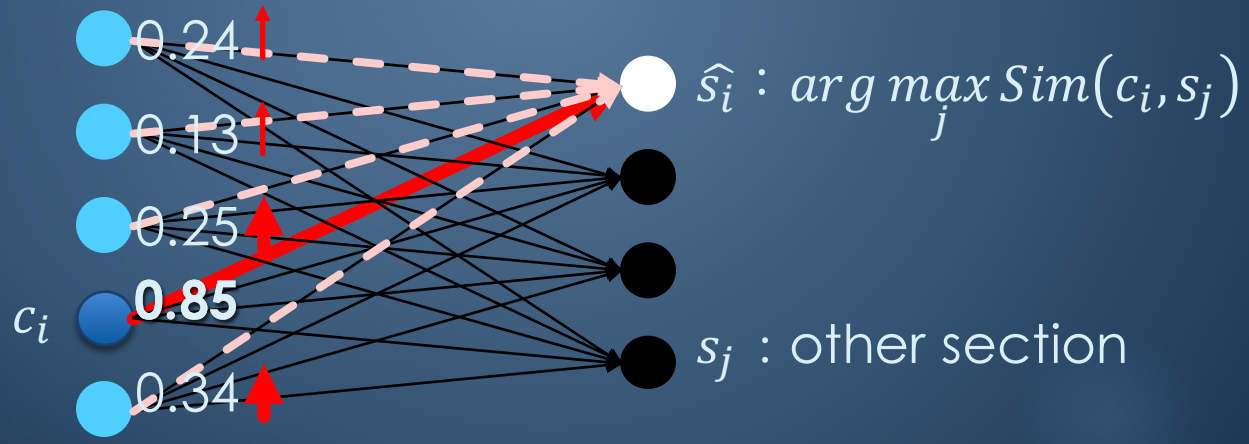


System Overview



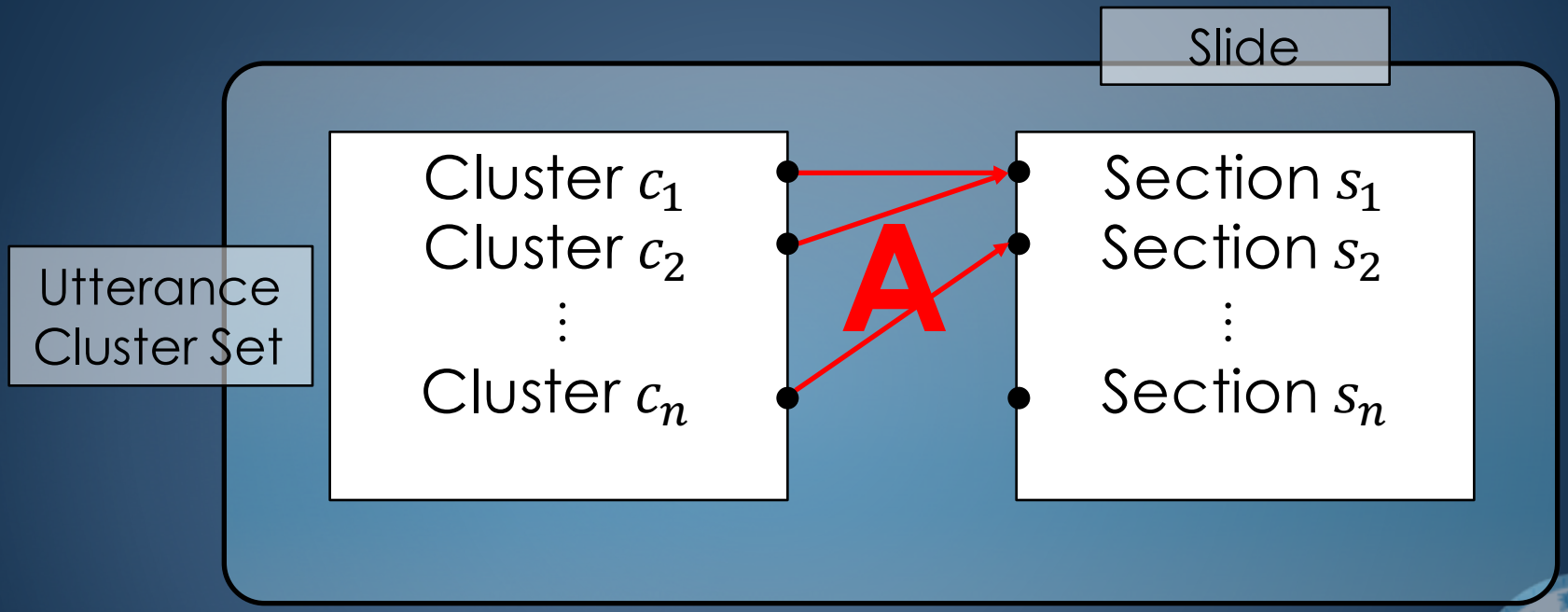
Reliability-Propagated Word-based Matching

- Word-based matching
  - Calculate lexical similarity based on tf-idf vectors.
- Reliable alignment
  - $Sim(c_i, \widehat{s_i})$ much larger than $Sim(c_i, s_j)$ for other s_j
- Reliability-Propagated
  - Neighbors of reliable alignment should increase their scores.

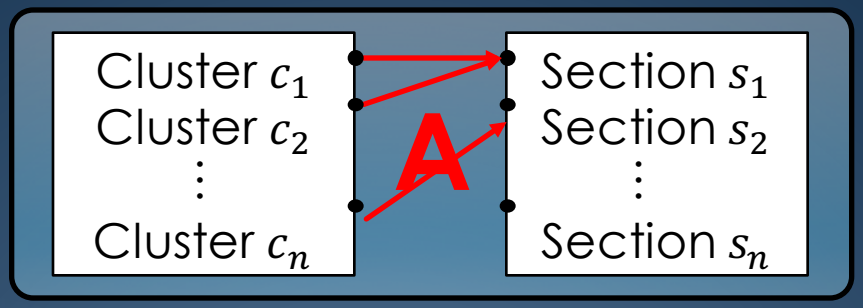


Structured SVM

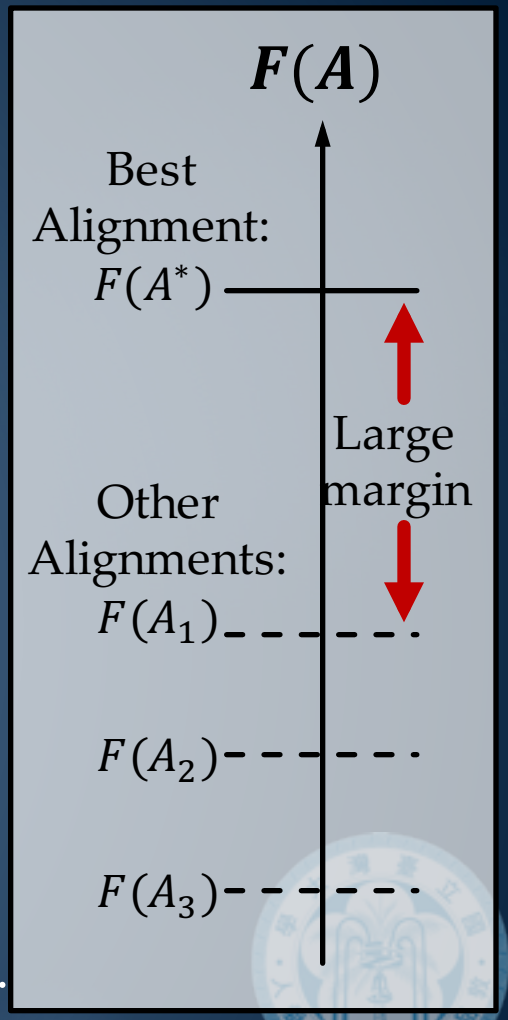
- Consider the global alignment for a slide as a whole.



Structured SVM

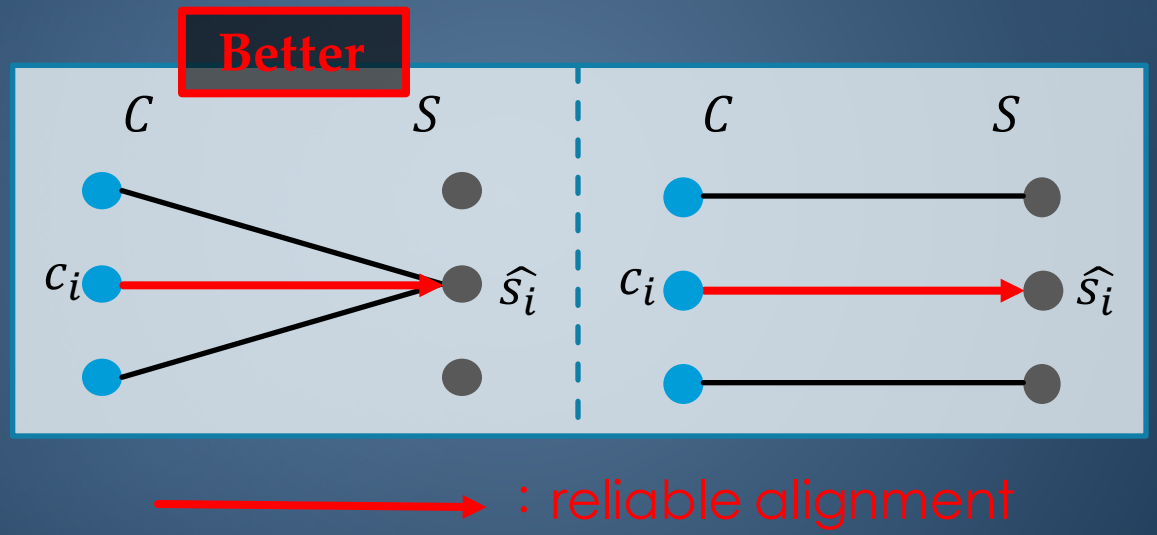


- Target
  - The alignment set A of every slide
- "**A**" is a structure
- Model design
  - $\vec{v}(A)$: Feature vector of the alignment
  - $\vec{w}$: To be learned
  - $F(A) = \vec{w} \cdot \vec{v}(A)$: Evaluation Function
- Training goal
  - Maximize $\vec{w}$ with SVM for maximum margin.



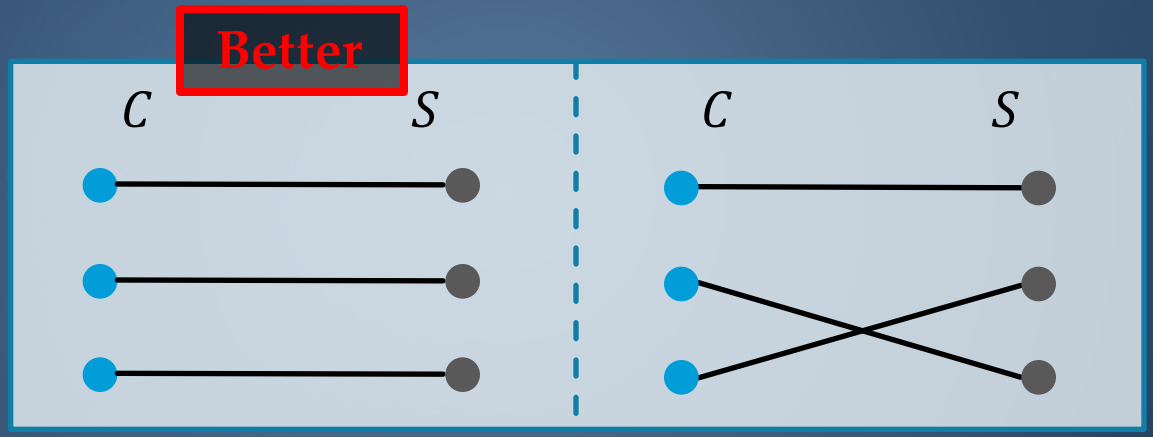
Feature Selection

- Local feature examples
  - Summation of lexical similarity
  - Neighbors of alignment link also align to same section.

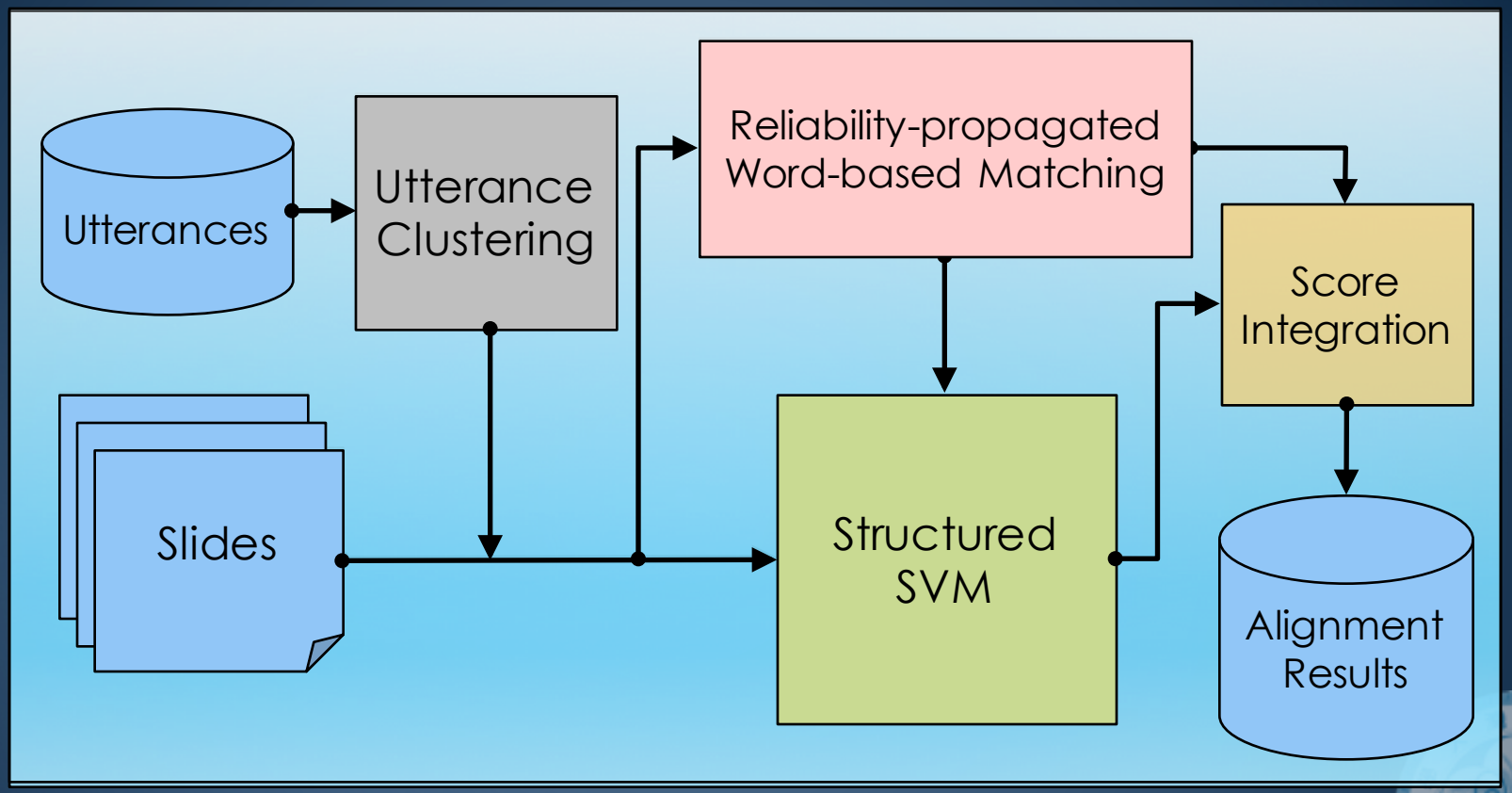


Feature Selection

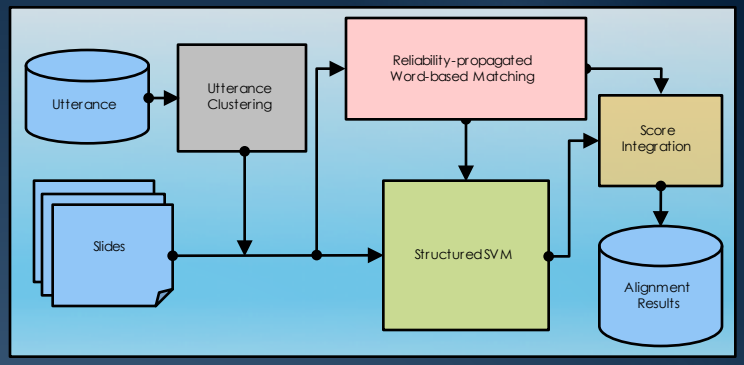
- Global feature examples
  - Number of crossed alignment
  - Longer section should be explained with more utterance clusters.



System Overview



Experimental Results



Approaches		Accuracy
Word-based Matching	Without Propagated	58.43%
	With Propagated	69.50%
Unsupervised	Structured SVM	70.28%
	Score Integration	72.86%
Supervised	Structured SVM	71.26%
	Score Integration	73.15%

- The propagation of reliability makes the performance better.
- Score Integration makes better results.
- Without using the real answers as training target, we could still have high accuracies.

Structuring Lectures

BETWEEN COURSES

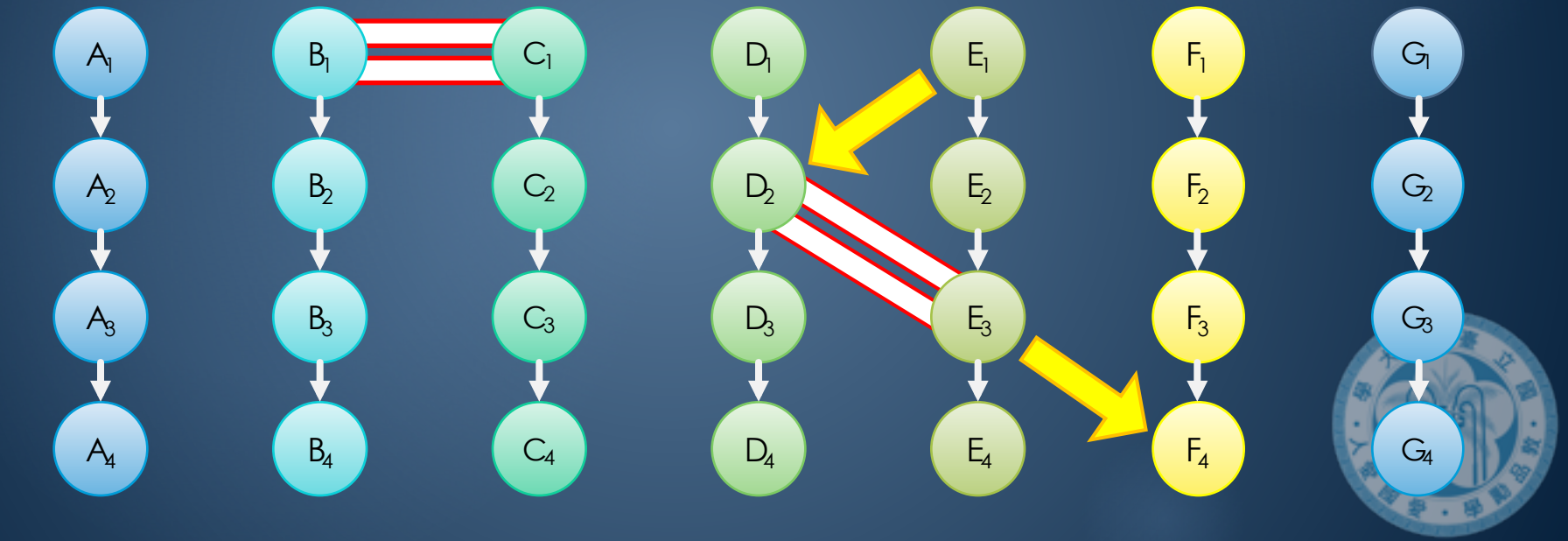
Between Courses

Link lectures with similar content

Predict prerequisite relationship

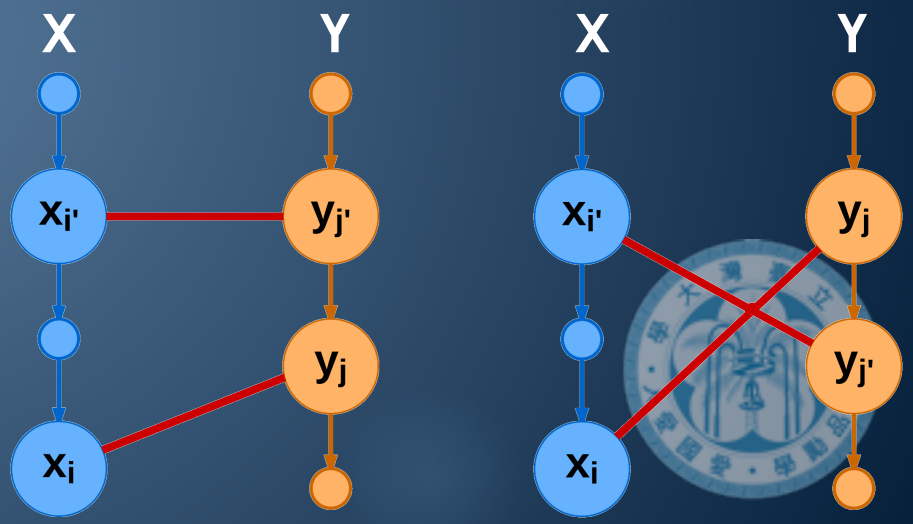
Reconstruct the learning map

Query



Linking Lectures with Similar Content

- Individual pair similarity
  - Calculate cosine similarity for lectures
  - Feature vector examples
    - Tf-idf for all words / key terms only
    - topic vectors by latent topic analysis
- Global structure considerations
  - Crossover may imply something wrong



Linking Lectures with Similar Content

- Maximize the objective function

$$F(L) = \sum_{(x_i, y_i) \in L} S(x_i, y_i) - \lambda_1 C(L) - \lambda_2 |L|$$

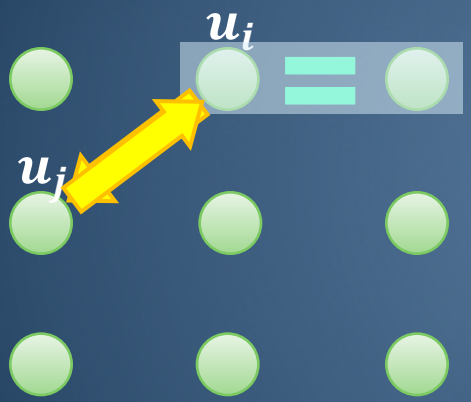
Crossover constraint

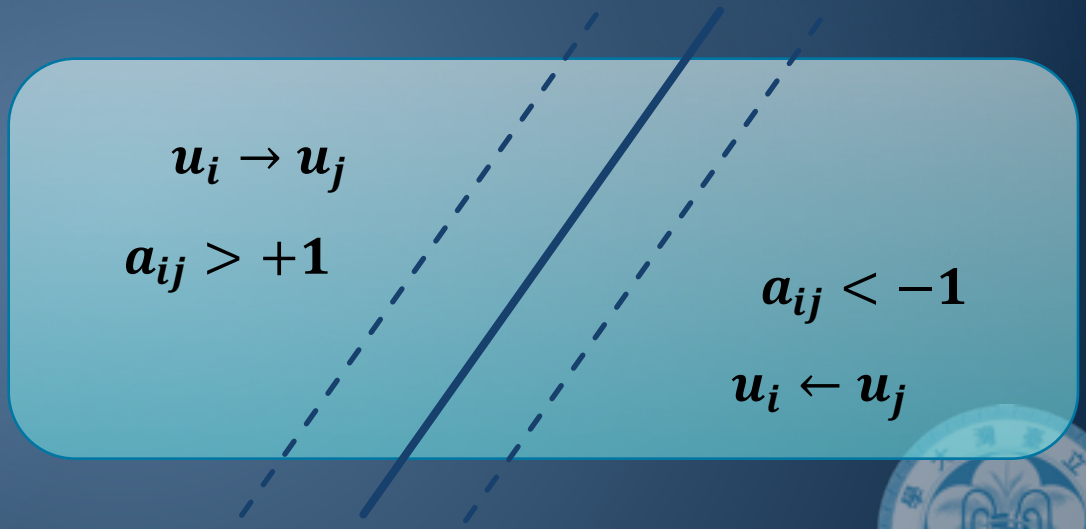
L: set of link relationships

Not including too many links

Prerequisite Prediction

- SVM classification
  - Difference vector : $a_{ij} = M \cdot (u_i - u_j)$
  - Cross-term Matrix : $a_{ij} = u_i^T M u_j$

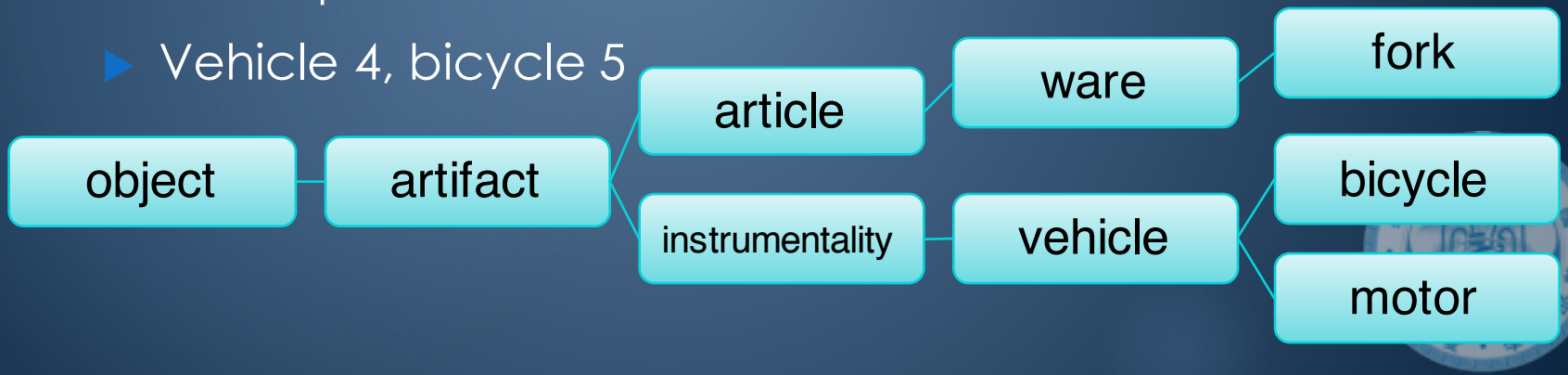




Prerequisite Prediction

- Feature vector representation
  - Weighted Bag-of-word (BOW) : $\{s(w_k)tf(w_k), k = 1,2, \dots, n\}$
  - Weighted Word embedding : $\frac{1}{N}\sum s(w_k)\, tf(w_k)v_k$

    $W = \{w_1, w_2, \cdots, w_n\}$, v_i : Mikolov's $word2vec$
- Semantic weights for keywords
  - WordNet semantic depth : Deeper words in WordNet are more specific
  - Vehicle 4, bicycle 5



Experimental Results

Linking Lectures		Precision	Recall	F-measure
Individual	Audio Transcripts (a) Tfidf - all	13.8	24.6	17.3
	Audio Transcripts (b) Tfidf – key	33.8	26.5	28.8
	Audio Transcripts (c) Topics	48.9	30.2	37.2
	(a)+(b)+(c)+Lecture Title Features	42.9	52.7	47.2
(a)+(b)+(c) + Title Features + Global		53.6	54.6	54.1

- Assembling all the individual features could make better performance than just considering one feature.
- Considering global structure is necessary.

Experimental Results

	Prerequisite Prediction	NLP	Chemistry
Differ	(a) Bag-of-word (BOW)	68.1	61.4
	(b) Weighted BOW	70.0	63.3
	(c) Word Embedding	73.3	65.2
Cross	(d) Word Embedding	76.1	67.0

Difference vector : $a_{ij} = M \cdot (u_i - u_j)$
Cross-term Matrix : $a_{ij} = u_i^T M u_j$

- Semantic weights make better representation.
- Word embedding is a better way for word representation comparing to traditional bag-of-word vectors.
- Cross-term Matrix is a better SVM weight matrix.

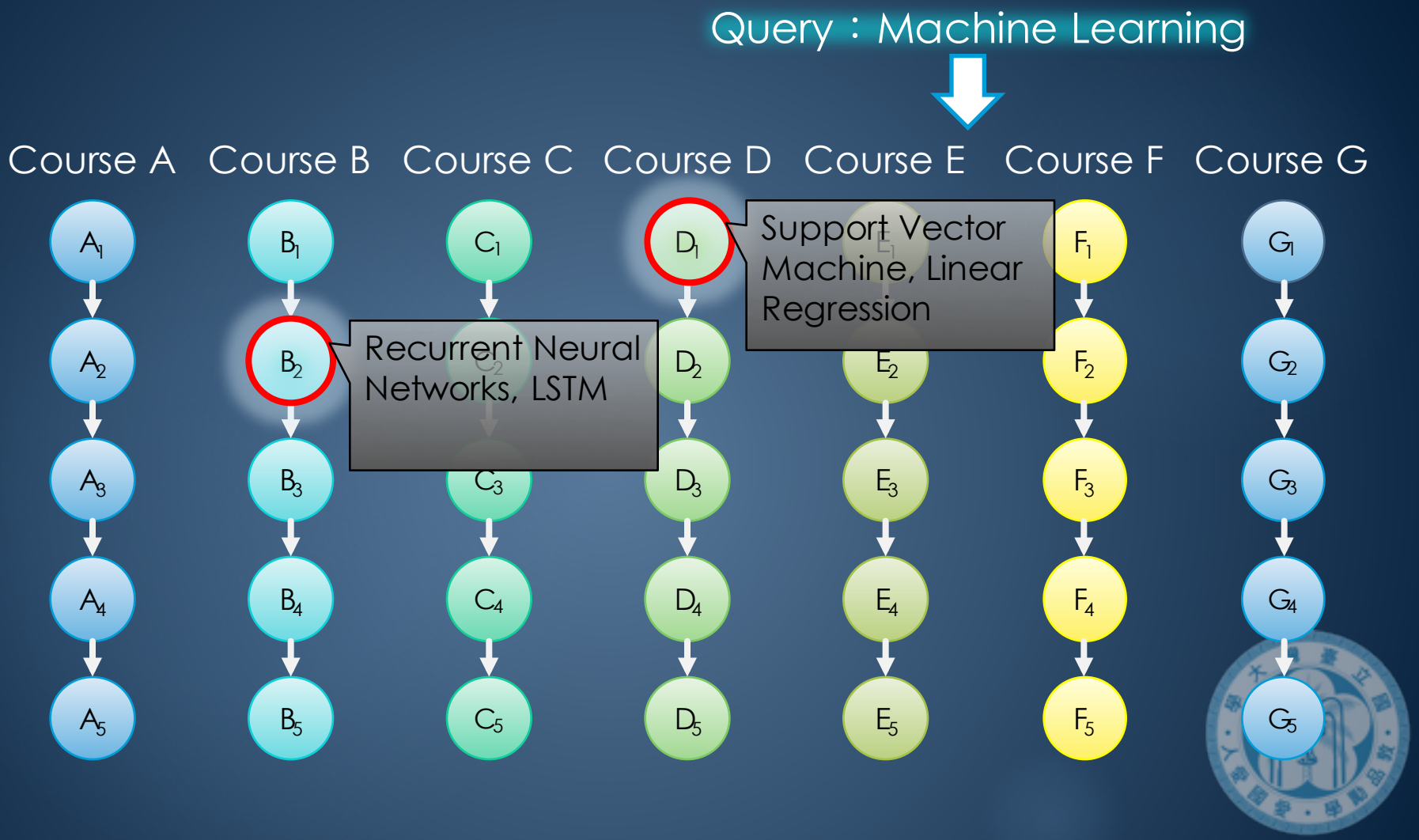
Outline

- Introduction
- Structuring Lectures（課程內容結構化）
  - Within A Course
  - Between Courses
- Classifying Lectures（對課程小段做內容精確分類）
- Understanding Lectures（機器對課程小段內容做了解）
- Conclusion

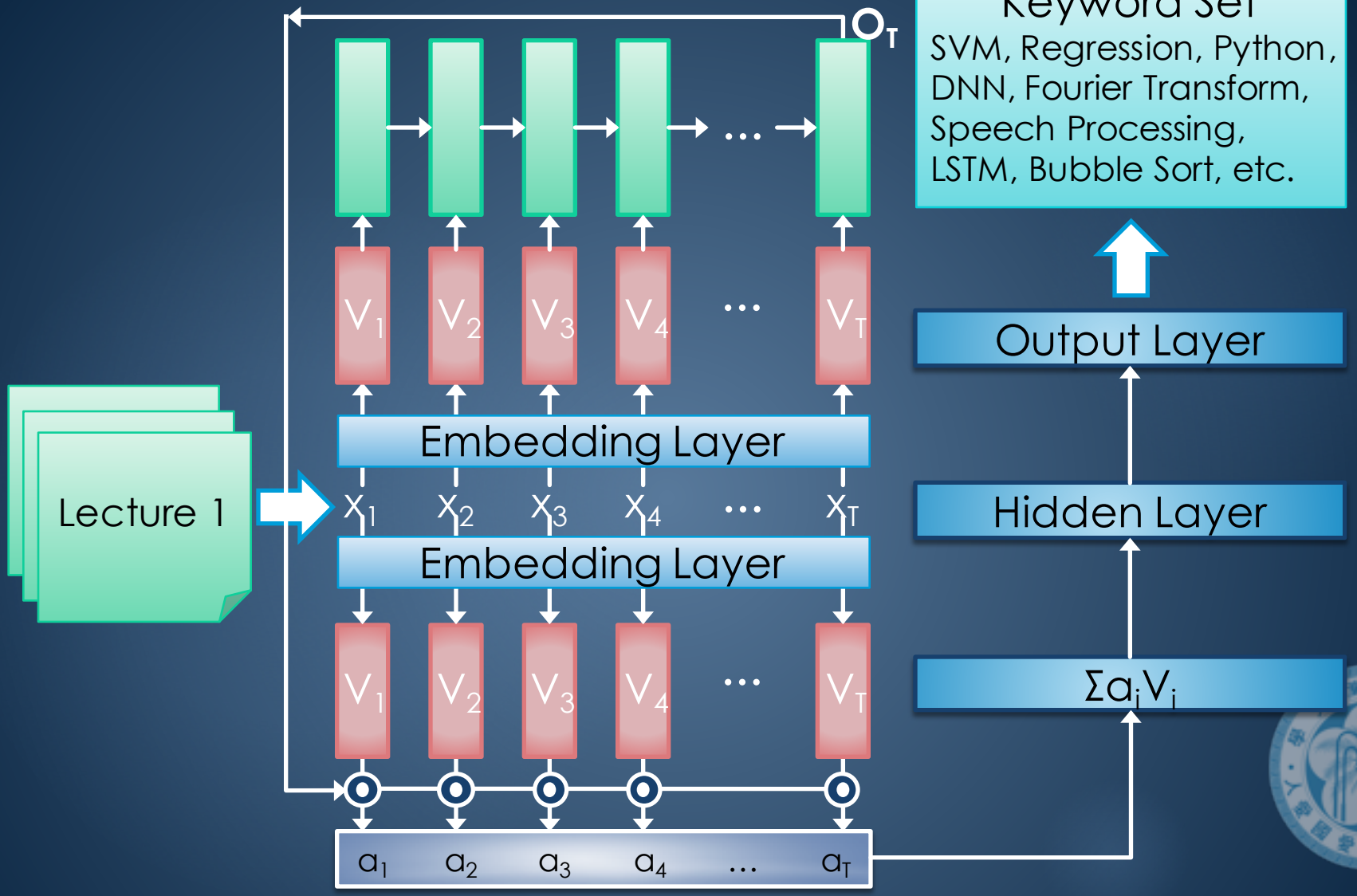
Classifying Lectures

對課程小段做內容精確分類

Classifying Lectures



Proposed Model



Normalization Methods for Attention Mechanism

- Attention mechanism score list $e = (e_1, e_2, \dots, e_T)$

$$e_i = O_T \odot V_i$$

- Sharpening normalization

$$\alpha_i = \frac{\exp(e_i)}{\sum_{i=1}^{T} \exp(e_i)}$$

- Smoothing normalization

$$\alpha_i = \frac{\sigma(e_i)}{\sum_{i=1}^{T} \sigma(e_i)}$$

Experimental Setup

- 290,000 Stack Overflow articles
  - 250,000 for training
  - 40,000 for testing
- 2~6 labeled keywords for each article

Experimental Results

Model		MAP(%)	P@R(%)
(a) Tf-idf Sorting		9.9	8.9
(b) Multiple Layer Perceptron		33.1	29.7
(c) Long Short-term Memory		43.1	40.2
Proposed Model	(d) Sharpening	39.3	36.2
	(e) Smoothing	50.5	46.4

- LSTM > MLP > TF-IDF
- Sharpening normalization eliminates too many information, so it perform worse.

Analysis

Sharpening Normalization

(A) Ground truth : python, numpy, matrix
5-best predict : python, numpy, python-2.7, pandas, python-3.x

I have a huge matrix that I saved with savetxt with numpy library. Now I want to read a single cell from that matrix, e.g.,

```
cell = getCell (i, j); print cell
return the value 10 for example.
```

I tried this:

```
x = np.loadtxt("fname.m", dtype="int", usecols=([i]))
cell=x[j]
```

but it is really slow because I loop over many index. Is there a way to do that without reading useless lines ?

Smoothing Normalization

(B) Ground truth : python, numpy, matrix
5-best predict : python, numpy, arrays, matrix, indexing

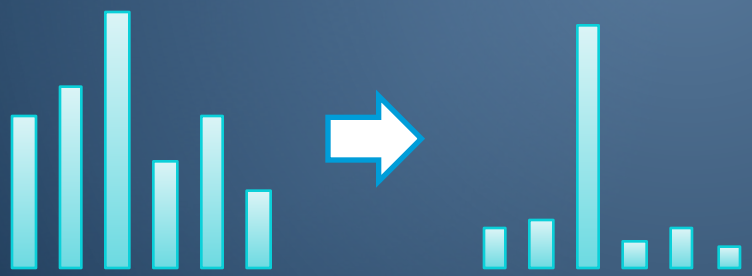
I have a huge matrix that I saved with savetxt with numpy library. Now I want to read a single cell from that matrix, e.g.,

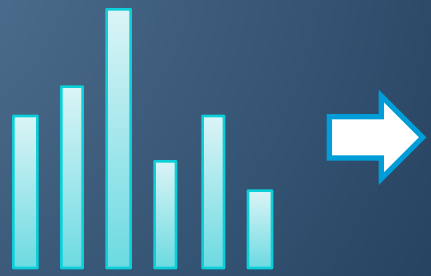
```
cell = getCell (i, j); print cell
return the value 10 for example.
```

I tried this:

```
x = np.loadtxt("fname.m", dtype="int", usecols=([i]))
cell=x[j]
```

but it is really slow because I loop over many index. Is there a way to do that without reading useless lines ?





Outline

- Introduction
- Structuring Lectures（課程內容結構化）
  - Within A Course
  - Between Courses
- Classifying Lectures（對課程小段做內容精確分類）
- Understanding Lectures（機器對課程小段內容做了解）
- Conclusion

Understanding Lectures

機器對課程小段內容做了解

Understanding Lectures

- Our previous works, e.g., structuring lectures and classifying lectures, rely on the understanding of lecture contents.
- Do machines really understand lecture contents ?
- Initial goal
  - Listening comprehension test in TOEFL

Task Definition

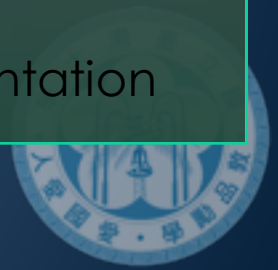
Story: …… I just wanted to take a few minutes to meet with everyone to make sure your class presentations for next week are all in order and coming along well. And as you know, you're supposed to report on some areas of recent research on genetics, something, you know, original. …… (manual transcription)

Question: Why does the professor meet with the student ?

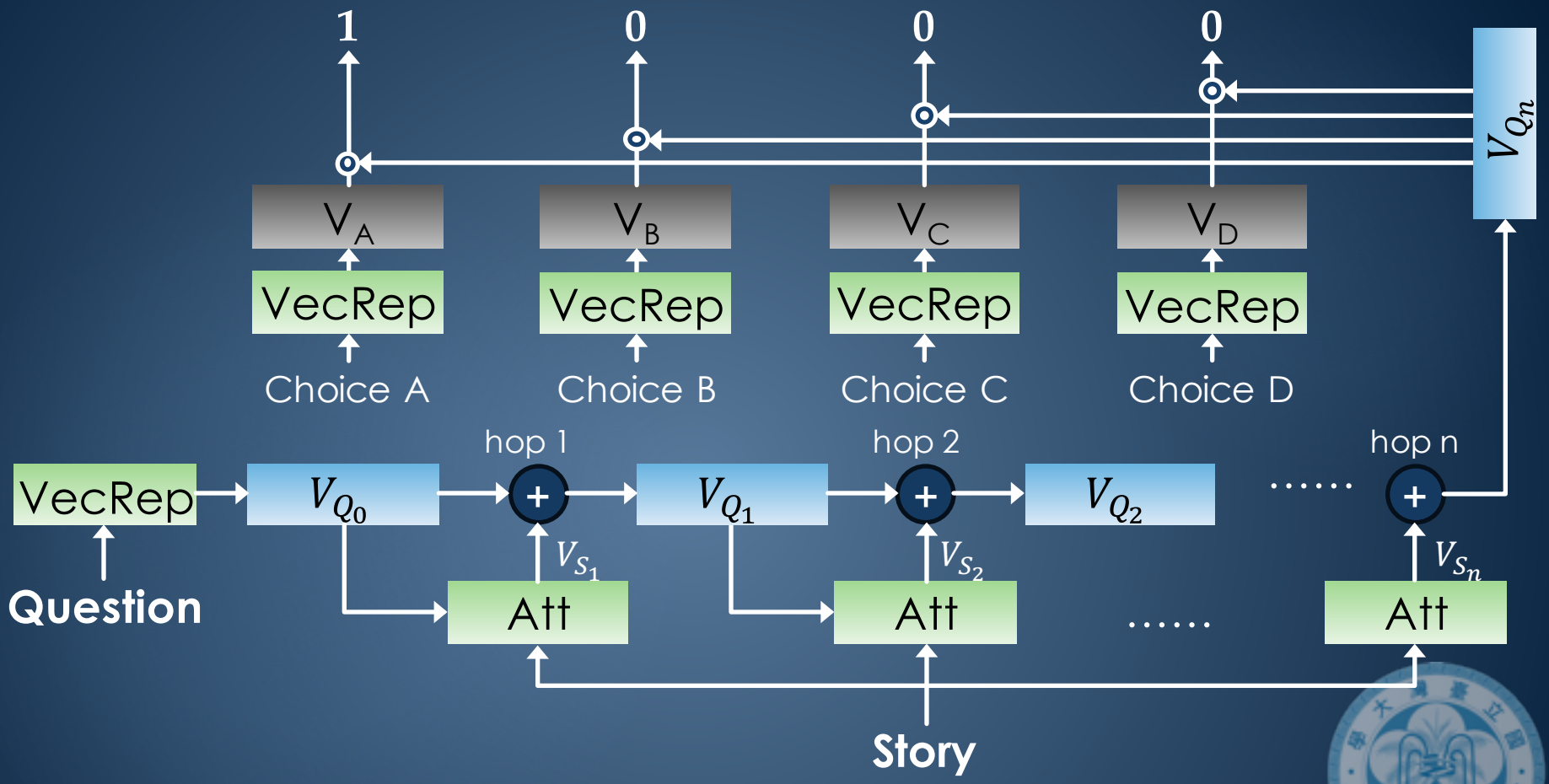
We can't just answering this question by finding matched sentence in the story. In contrast, we need to understand the whole contents.

Choices:

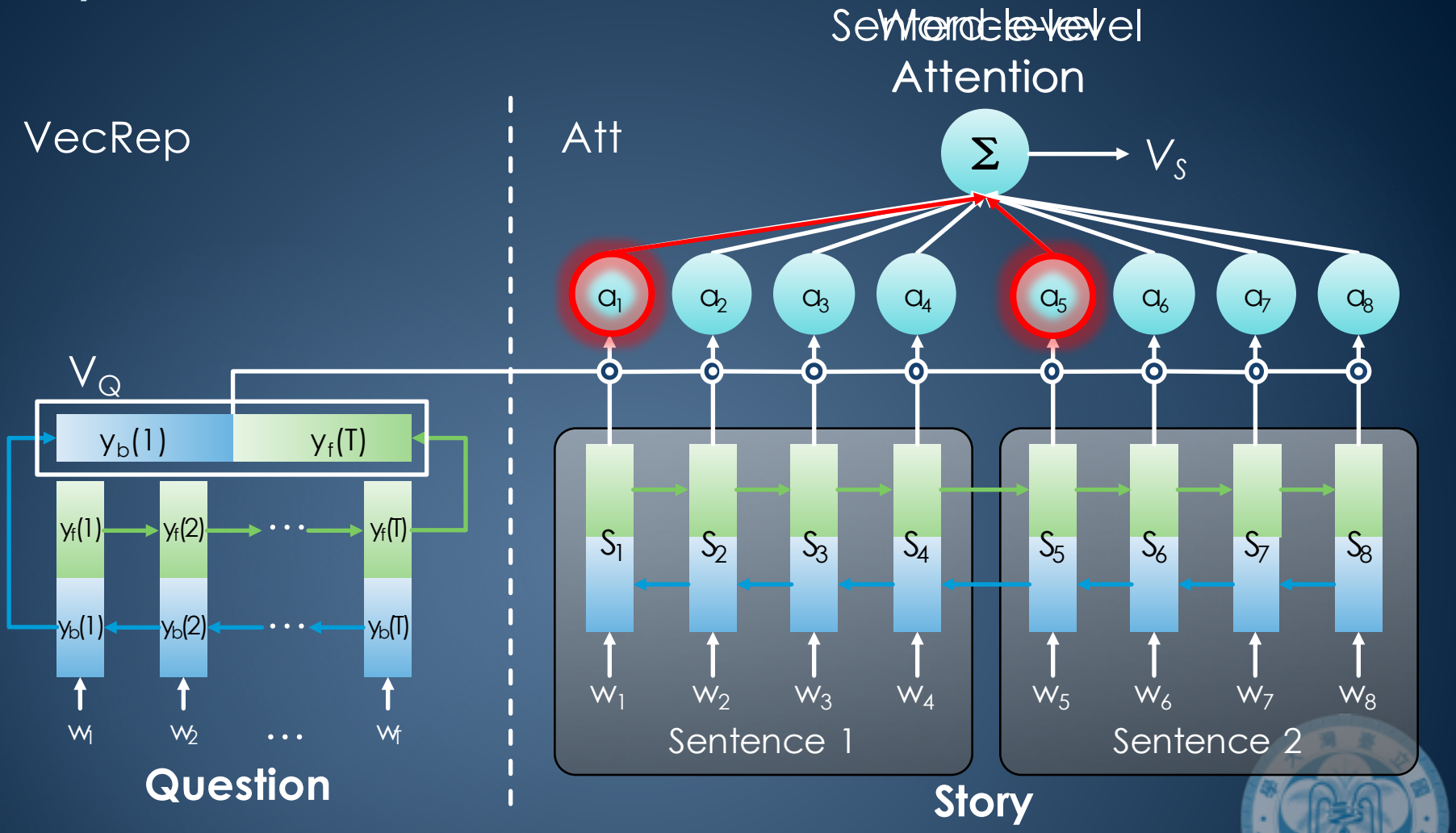
A. **To determine if the student has selected an appropriate topic for his class project**
B. To find out if the student is interested in taking part in a genetics project
C. To discuss the student's experiment on taste perception
D. To explain what the student should focus on for his class presentation



System Overview



System Detail



Experimental Results

Model		Manual	ASR
(a) Random Guess		25%	
(b) Memory Network		39.17%	39.17%
(c) Proposed Model	word	49.16%	48.33%
	sentence	51.67%	46.67%

- The proposed model gain much better performance than the state-of-the-art model.
- Word-level attention mechanism has higher tolerance while errors occur.

Outline

- Introduction
- Structuring Lectures（課程內容結構化）
  - Within A Course
  - Between Courses
- Classifying Lectures（對課程小段做內容精確分類）
- Understanding Lectures（機器對課程小段內容做了解）
- Conclusion

Conclusion

結論

Conclusion

- We propose three kinds of techniques for helping user learn more efficiently on MOOCs.
  - Structuring lectures
  - Classifying lectures
  - Understanding lectures
- Structured SVM is capable of handling structure information in the case of alignment prediction.
- Semantic weights from WordNet provide more information for words.
- Attention-based RNN works better than RNN.

Thanks for your attention.

謝謝各位口試委員的聆聽

Q & A

問答時間