

Natural Language Processing in Comprehension SEEDING Project

Comprehension Through Self-Explanation, Enhanced Discussion and INquiry Generation

Clustering Student Responses



Clustering provides teachers with a capability to lead discussions based on the students' responses in groups of belief.

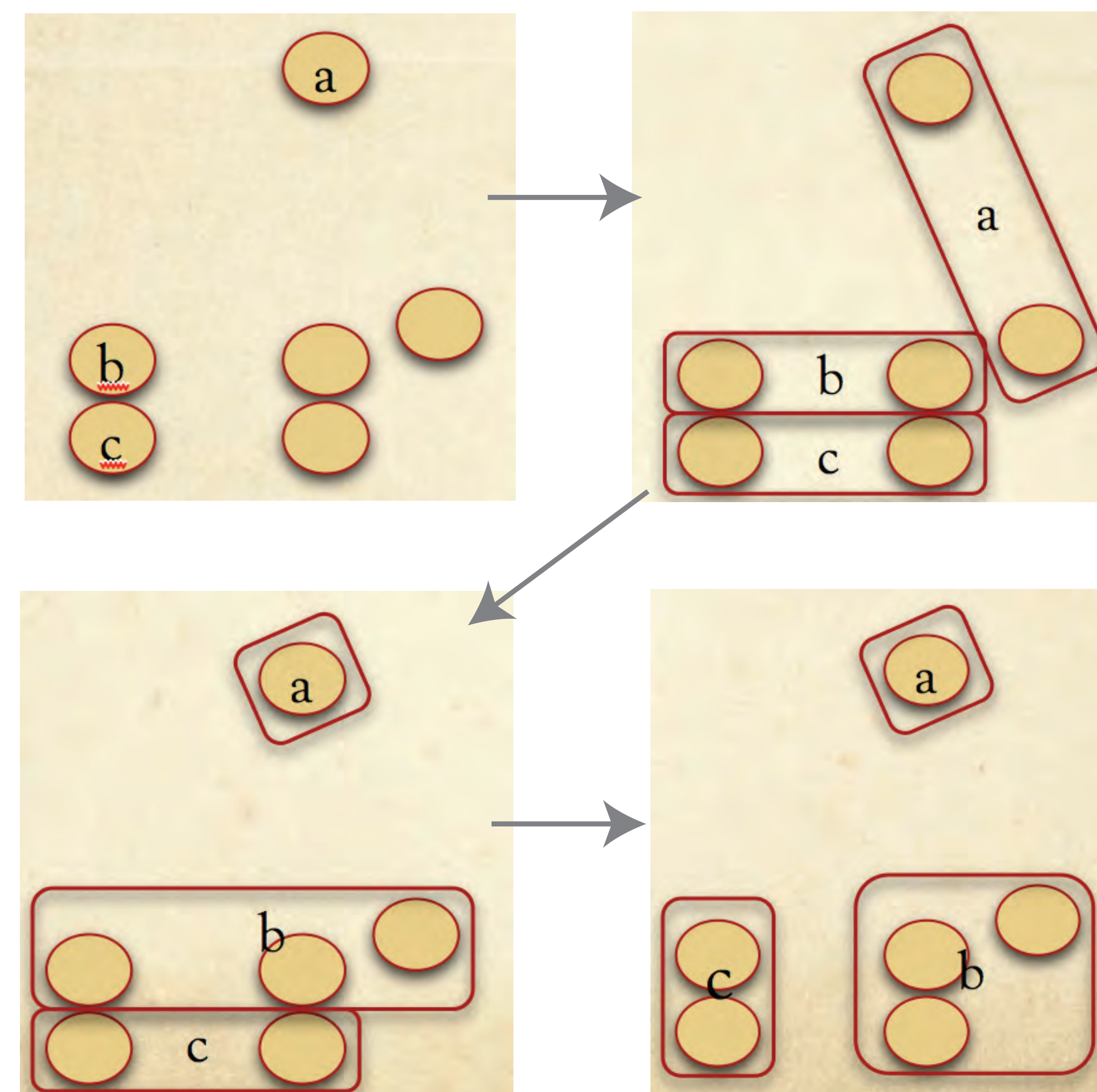
k-means clustering

Iterate 10 times unless convergence, then iterate through each response feature vector and assign each feature vector to the nearest centroid.

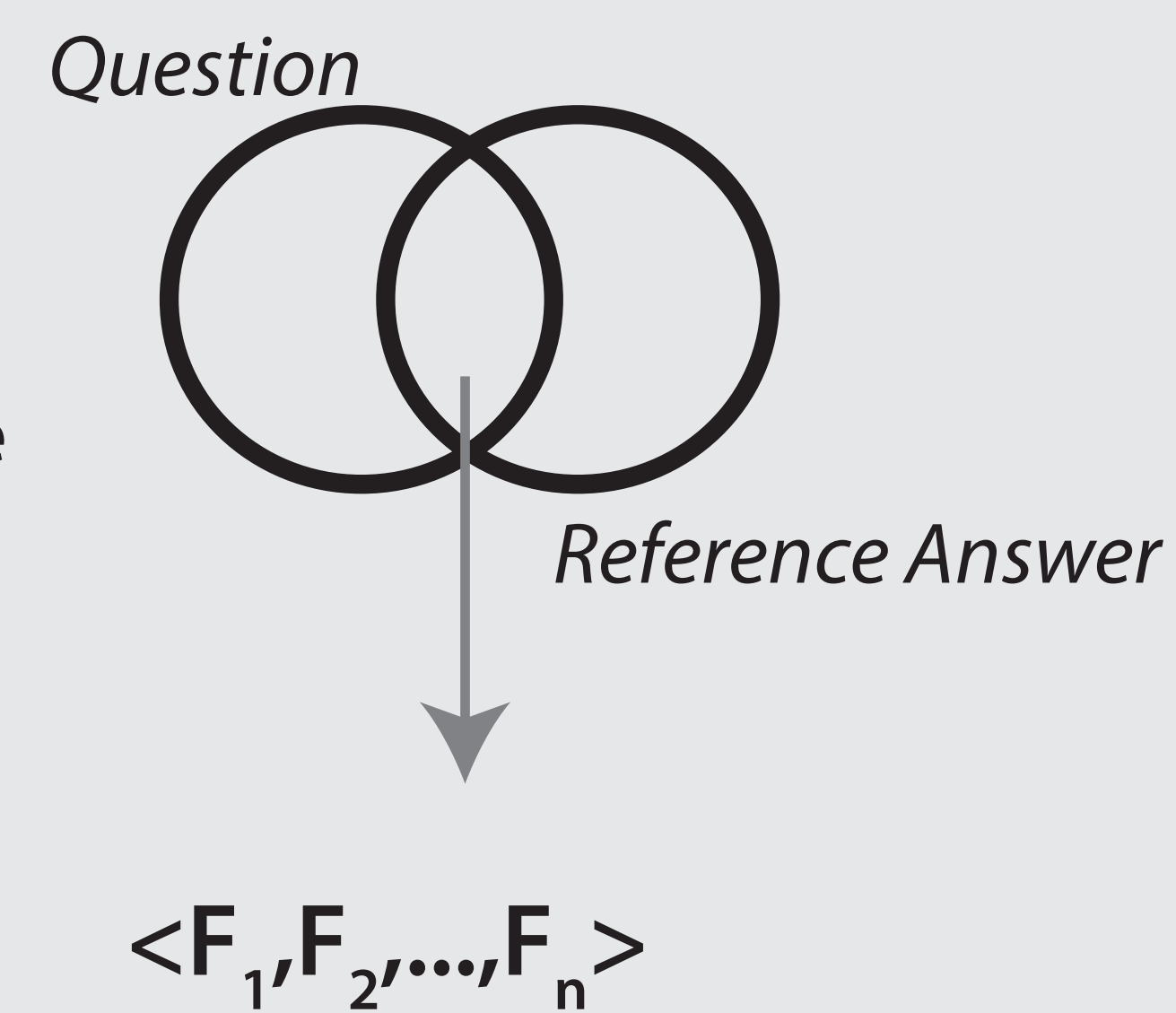
At the conclusion, adjust the centroid of each cluster.

Pseudocode:

```
while i < 10 or Converge:
  for J ∈ Response Feature Vector:
    Assign Closest Centroid(j)
  for K ∈ Centroid
    Adjust Centroid
```



Extract features from Question and Reference Answer to create a feature vector



Weigh Each Student Response against Feature Vector

$$\begin{matrix} \text{student 1} \\ \text{student 2} \\ \vdots \\ \text{student n} \end{matrix} \begin{bmatrix} f_1 & f_2 & \dots & f_n \\ f_1 & f_2 & \dots & f_n \\ \vdots & \ddots & \ddots & \vdots \\ f_1 & f_2 & \dots & f_n \end{bmatrix}$$

Vocabulary for Students



Extract all the content words from all RA_1, \dots, RA_n
List $W = \{RA_1, RA_2, \dots, RA_n\}$

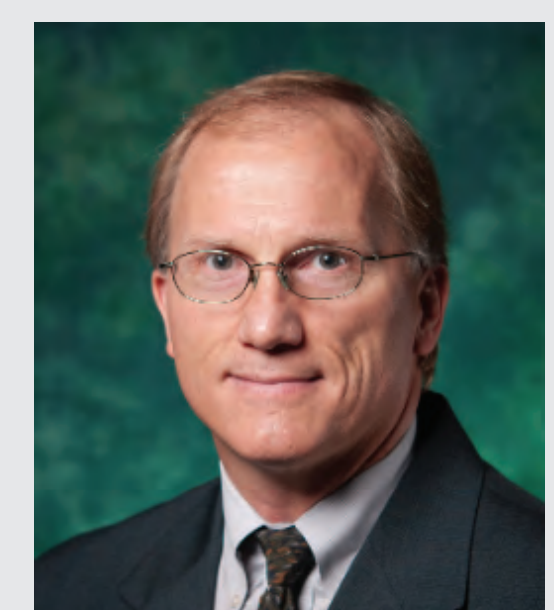
for $i \in \text{listW}$
Synonym(i)
Antonym(i)

As students are responding to a specific question, append to List W the most frequently used words

List W has a max of 5 x Number of Content Words in the current reference answer.

SEEDING Development Team

Dr. Rodney Nielsen,
Associate Professor, CSE
Director, HiLT Lab



UNT CSE Student Developers

Mingyu Lin
James Glenn

Frank Paiva

HiLT Lab



Human Language Technologies
Health Informatics Language Technologies
Dr. Rodney Nielsen

UNT[®]

UNIVERSITY
OF NORTH TEXAS[®]