Minimal Video Lecture Sequencing

Problem Formulation

We have n number of lecture videos[1] ("documents") in the entire course, arranged in some pedagogical ordering $1 < 2 < \cdots < n$. Let this ordered set of documents be $\mathbb{N}_n$. Let us introduce the following:

Definition1: Concept Set C refers to the set of all concept phrases extracted from n lectures. (Extracted using POS tagging, TextRank weighing, usage analysis, or something more substantial like from hyperlinked Wikipedia.) C_i refers to concept set of lecture i.

Definition2: Prerequisite concepts Set P_i refers to the set of concepts which are required to understand lecture i. Clearly, $\bigcup_i P_i \subseteq C$. (Annotation done using word-usage, first-occurrence and co-occurrence analysis.)

Definition3: Outcome concepts Set O_i refers to the set of concepts which are outcomes of having read lecture i. Clearly, $\bigcup_i O_i = C$.

Rule1: Outcome Exclusivity says that every concept is an outcome concept of exactly one document.

$$\forall c \in C \; \exists! \, i \in \mathbb{N}_n \; such \; that \; c \in O_i$$

Rule2: Prerequisite Establishment says that a prerequisite concept must be defined before invocation, that is, it must be an outcome concept of a lecture which features before the lecture which cites it as a prerequisite, in the pedagogical ordering.

$$\forall i \; \forall c \in P_i \; \exists j < i \; such \; that \; c \in O_j$$

Definition4: Concept coverage of document i by document j refers to the amount of prerequisites of i which are covered by reading j.

$$\gamma(i,j) = \frac{|P_i \cup O_j|}{|P_i|}$$

Definition5: Concept relevance of prerequisite concept c in document i refers to the importance of the concept in understanding the entire document[2].

$$\alpha(c,i) = \frac{freq(c,i)}{\sum_{c \in P_i} freq(c,i)}$$

Definition6: Prerequisite lectures Set ρ_i refers to the set of prerequisite documents which must be read before reading document i.

Definition7: Reading Burden β refers to the willingness of the student to read more number of lectures so as to understand the learning goal better. $\beta \in [0,1]$

The objective function: We have to maximise the coverage of concepts of high relevance, while minimising the number of lectures to be studied to understand an outcome concept of lecture i. We therefore propose the following formulation:

$$\delta_i = \underset{s \,\in\, \wp(\mathbb{N}_{i-1})}{argmax} \left\{ \beta \sum_{j \in s} \gamma(i,j) \left(\sum\nolimits_{c \in P_i \cap O_j} \alpha(c,i) \right) - (1-\beta) \left| \bigcup_{j \in s} j \cup \rho_j \right| \right\}$$

$$\rho_i = \delta_i \cup \bigcup_{j \in \delta_i} \rho_j$$

---

[1] Videos should be considered at topical granularity for best results.

[2] Scope for improvement in estimating concept relevance.

Algorithm Suggestions

Greedy Selection[3]:

1. $j = 2, \rho_1 = \phi$
2. While $j \leq i$
   a. $\rho_j = \phi$
   b. While (true or $\rho_j \neq \mathbb{N}_{j-1}$)
      i. $k = \underset{k \in \mathbb{N}_{j-1} \backslash \rho_j}{argmax} \left\{ \beta\gamma(j,k) \left(\sum_{c \in P_j \cap O_k} \alpha(c,j) \right) - (1-\beta)|k \cup \rho_k| \right\}$
      ii. $\epsilon_j = \beta\gamma(j,k) \left(\sum_{c \in P_j \cap O_k} \alpha(c,j) \right) - (1-\beta)|k \cup \rho_k|$
      iii. If $\epsilon_j \leq 0$ then break else $\rho_j = \rho_j \cup k \cup \rho_k$
   c. $j = j + 1$
3. Output ρ_i

Greedy Elimination[4]:

1. $j = 2, \rho_1 = \phi$
2. While $j \leq i$
   a. If $\gamma(i,j) \left(\sum_{c \in P_i \cap O_j} \alpha(c,i) \right) > 0$ then $\rho_i = \rho_i \cup j \cup \rho_j$
   b. $j = j + 1$
3. $\sigma_i = \rho_i$
4. While (true or σ_i not empty)
   a. $j = \underset{j \in \sigma_i}{argmin} \left\{ \beta\gamma(i,j) \left(\sum_{c \in P_i \cap O_j} \alpha(c,i) \right) - (1-\beta)\left|j \cup \rho_j\right| \right\}$
   b. $\epsilon_j = \beta\gamma(i,j) \left(\sum_{c \in P_i \cap O_j} \alpha(c,i) \right) - (1-\beta)\left|j \cup \rho_j\right|$
   c. $exclusive_j = (j \cup \rho_j) \backslash \bigcup_{k \in \rho_i \backslash \{j\}} \rho_k$
   d. If $\epsilon_j \leq 0$ then
      i. $\rho_j = \rho_j \backslash exclusive_j$
      ii. $\sigma_i = \sigma_i \backslash \{j\}$

   Else break
5. Output ρ_i

Integer Linear Program[5]:

$$maximise \sum_j x_j \left\{ \beta\gamma(i,j) \left(\sum\nolimits_{c \in P_i \cap O_j} \alpha(c,i) \right) - (1-\beta) \right\}$$

$$subject\ to \boxed{\forall\, j < i, \sum_{x_k \in \rho_j} x_k \geq \left|\rho_j\right| x_j} \ and\ \forall j\ x_j = \{0,1\}$$

$$output\ \rho_i = \{x_j | x_j = 1\}$$

---

[3] Computes $\rho_j \forall j < i$ within the algorithm

[4] Assumes $\rho_j \forall j < i$ have already been computed

[5] Assumes ILP has been solved $\forall j < i$, and takes all prerequisite chains as first-order relationships.

Softening ILP Constraints

There are two issues with the previous formulation:

1. It is an iterative ILP, that is, to know the constraints to solve for lecture n, we must solve the ILP for all lectures in $\mathbb{N}_{n-1}$. If an approximate algorithm is used to solve the ILP in polynomial time, the approximation factor will get exponentially propagated across every iteration.
2. We have hard constraints, that is, if we are to consider a lecture i as a prerequisite for lecture n, then we are considering all ρ_i to be prerequisites as well. This may not be an optimal choice, given that we would rather like to learn order-one prerequisite lectures first, and then worry about learning order-two prerequisite lectures. Also, not all order-two prerequisites may be important.

To soften the constraints, and ensure that only a single ILP needs to be solved, we propose an alternative formulation of the problem.

Definition8 Lecture weight w_{ji} is the weight that we attach to a lecture $j < i$ as contributing to learning concepts of i

$$w_{ji} = \gamma(i,j)\left(\sum\nolimits_{c \in P_i \cap O_j} \alpha(c,i)\right)$$

Now, essentially we have $i-1$ constraints with the following ILP formulation:

$$maximise \sum_j x_j\{\beta w_{ji} - (1-\beta)\}$$

$$subject\ to \boxed{\forall\, j < i, \frac{\sum_{k=1}^{j-1} w_{kj} x_k}{\sum_{k=1}^{j-1} w_{kj}} \geq \beta x_i}\ and\ \forall j\ x_j = \{0,1\}$$