1 Introduction

Automatic coherence assessment has typically examined the effect of local discourse features on coherence. These features are typically sentence-to-sentence transitions of entities, discourse relations, or other sentential information. Additionally, this task is often framed as selecting the most coherent ordering of a set of sentences in a document, where the original ordering as written by a human is assumed to be the most coherent ordering.

2 Entity Grid Model

The *entity grid* model is a popular local coherence model where a document rendering is represented by a grid of discourse entities and their syntactic role, such as ..., in each sentence. A document rendering ρ is an ordered sequence of sentences $s_1, \ldots, s_n$. The possible set of renderings $\mathcal{P}$ consists of all permutations of these sentences and has $n!$ elements.

> informal

We formally define the entity grid model as a document's entity set $\mathcal{E}$, a finite set of syntactic roles $\mathcal{Q}$, and the set of transition vectors t_i for all $i \in \{2, 3, \ldots, |\rho|\}$.

> more formal. resolve what a "sentence" is.

A sentence s is a vector of length $|\mathcal{E}|$ with elements $q \in \mathcal{Q}$. The jth element of s_i corresponds to the sytnactic role $q_{i,j}$ of entity e_j in sentence i. A transition vector $t_i \in \{\mathcal{Q} \times \mathcal{Q}\}^{|\mathcal{E}|}$ is a vector of entity role transitions from s_{i-1} to s_i where each element corresponds to a transition $r \in \{\mathcal{Q} \times \mathcal{Q}\}$.

A rendering ρ is then represented as a vector $\boldsymbol{\phi}$ of probabilities for each transition r in the document where each element $\phi(r)$ is defined as

$$\phi(r) = \frac{\sum_{i=2}^{n} \sum_{j=1}^{|\mathcal{E}|} I(t_{i,j} = r)}{|\mathcal{E}|(n-1)}.$$
(1)

> clear up vector notation around phi

Given a vector of weights $\mathbf{w} \in \mathbb{R}^{|\mathcal{Q} \times \mathcal{Q}|}$, we define our objective function $f : \mathbb{R}^{|\mathcal{Q} \times \mathcal{Q}|} \to \mathbb{R}$ where $f(\boldsymbol{\phi}; \mathbf{w}) = \mathbf{w} \cdot \boldsymbol{\phi}$. The sentence ordering task can further be defined as finding the rendering that maximizes this function, or

$$\max_{\rho \in \mathcal{P}} f(\boldsymbol{\phi}_\rho; \mathbf{w})$$
(2)

Solving this problem for larger documents is intractable as the time complexity of $f = O(|\mathcal{P}||\mathcal{Q}|^2) = O(n!|\mathcal{Q}|^2)$.

> Clear up why this is slow. Also be careful about f = O

3   TSP Formulation

We formulate this problem as a Traveling Salesperson Problem where each sentence s is a vertex in a fully connected graph. Adding an edge (s_i, s_j) to the solution path represents selecting s_j as the next sentence in the document rendering. Edge weights are determined by the function

$$g(\boldsymbol{\phi}_t^{(-Z)}; \mathbf{w}) = \mathbf{w} \cdot \boldsymbol{\phi}_t^{(-Z)} = \sum_{i=1}^{|\mathcal{Q}|^2} w_i \left(\sum_{j=1}^{|\mathcal{E}|} I(t_j = r_i) \right) \tag{3}$$

where t is the transition vector obtained from the rendering

Rendering is a full thing

$\langle s_i, s_j \rangle$ and $\boldsymbol{\phi}^{(-Z)}$ is the unnormalized version of the probability vector representation from the entity grid model.

Under this formulation, optimal path is equivalent to the solution to the optimal rendering in the entity grid formulation. The total edge weights for the optimal path is the sum of $g(t)$ for all t in the path.

$$\text{Optimal Path Edge Weight} = \sum_t^{\rho} g(\boldsymbol{\phi}_t^{(-Z)}; \mathbf{w})$$

$$= \sum_t^{\rho} \sum_{i=1}^{|\mathcal{Q}|^2} w_i \Big(\sum_{j \in 1, \ldots, |\mathcal{E}|} I(t_j = r_i) \Big)$$

$$= \sum_{i=1}^{|\mathcal{Q}|^2} w_i \Big(\sum_t^{\rho} \sum_{j \in 1, \ldots, |\mathcal{E}|} I(t_j = r_i) \Big)$$

Equivalently the max of the Entity Model objective function is

$$\arg\max_{\rho \in \mathcal{P}} f(\boldsymbol{\phi}_\rho; \mathbf{w}) = \arg\max_{\rho \in \mathcal{P}} \sum_{i=1}^{|\mathcal{Q}|^2} w_i \Big(\frac{\sum_t^{\rho} \sum_{j \in 1, \ldots, |\mathcal{E}|} I(t_j = r)}{|\mathcal{E}|(n-1)} \Big)$$

$$= \arg\max_{\rho \in \mathcal{P}} \sum_{i=1}^{|\mathcal{Q}|^2} w_i \Big(\sum_t^{\rho} \sum_{j \in 1, \ldots, |\mathcal{E}|} I(t_j = r) \Big)$$

give an example.

Ignore this junk

Let $\mathcal{E}$ be the set of entities in a document. Let $\mathcal{Q}$ be the set of possible roles an entity can have.

Let $\mathcal{Q}^2$ be the set of transitions.

$r_i \in \{1, \ldots, |\mathcal{Q}|^2\}^{|\mathcal{E}|}$

$s_i \in \mathcal{Q}^{|\mathcal{E}|} \ \forall i \in \{1, \ldots, n\}$ where n is the length of the document.

$s_{i-1} = [-, -, \times, -]$

$s_i = [s, -, o, \times]$

$r_i = [(-, \times), (-, -), (\times, \times), (-, \times)]$

$\forall r \in |\mathcal{Q}|^2,\ \mathbf{w} \in \mathbb{R}^{|\mathcal{Q}|^2}\ f(r; \mathbf{w}) = \mathbf{w} \cdot \phi(r)$

$\max_{r \in T} f(r; \mathbf{w})$

Y Solving this is $O(M\mathcal{Q}^2) \Rightarrow O(n!\mathcal{Q}^2)$