

Ensamble de Modelos Sintácticos y Semánticos para la Evaluación Automática de Ensayos

Diego Palma S.

Profesor Supervisor: John Atkinson

Universidad de Concepción

dipalma@udec.cl

11 de agosto de 2016

Contenido

- 1 Introducción
- 2 Trabajo Relacionado
- 3 Propuesta
- 4 Metodología
- 5 Plan de Trabajo

Introducción

- En la actualidad, un tema bastante debatido es la capacidad de redacción y comprensión que debiesen tener las personas que egresan del sistema escolar.
- Una mala redacción tiene consecuencias relevantes en la vida de una persona.
- Un buen texto, debe cumplir con dos propiedades fundamentales: *coherencia* y *cohesión*.
- Una forma de ayudar a los estudiantes a mejorar sus capacidades de escritura y expresión de ideas, es produciendo textos que sean corregidos por un experto que le entregue retroalimentación al estudiante.
- Tener personal especializado disponible todo el tiempo es costoso.

Hipótesis

Un método de evaluación de textos que considere características sintácticas y semánticas para evaluar coherencia textual será más efectivo para la tarea de evaluación automática de ensayos en comparación métodos que utilicen medidas superficiales estadísticas.

Objetivos

- Objetivo General
 - Desarrollar un método computacional que permita evaluar automáticamente textos en forma de ensayos considerando aspectos de coherencia textual.
- Objetivos Específicos
 - Analizar estrategias de evaluación de ensayos en forma de texto, basados tanto en modelos de estadísticos, como en teoría de discurso.
 - Desarrollar una estrategia que considere coherencia a nivel de contenido y sintaxis.
 - Crear un prototipo para realizar las pruebas.
 - Evaluar el modelo propuesto.

Trabajo Relacionado

Métodos basados en características superficiales:

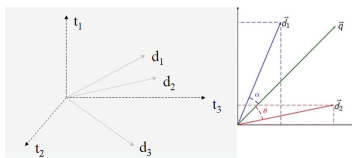
- Representan los textos como un conjunto de características que se utilizan como medidas indirectas de propiedades del texto tales como: Coherencia, cohesión, contenido.
- La nota se calcula como una suma ponderada de estas características (generalmente se utiliza regresión). El primer sistema exitoso (PEG) utilizaba regresión lineal.
- Para obtener la ponderación de cada característica, se requiere tener un conjunto de ensayos previamente evaluados.

Similaridad Contextual

$$Nota = \beta_0 + \sum_{i=1}^n \beta_i P_i \quad (1)$$

Trabajo Relacionado

Modelo de Espacio Vectorial



- Se representan los documentos en un espacio vectorial con muchas dimensiones.
- Se puede definir una medida de similitud entre documentos (por ejemplo coseno).
- Método presenta algunos problemas.

Similitud Coseno

$$\cos(d_i, d_j) = \frac{d_i \cdot d_j}{\|d_i\| \|d_j\|} \quad (2)$$

Reducción Dimensional

Existen diversos problemas en utilizar un método de espacio vectorial, tales como:

- Polisemia ($\text{sim} < \text{cos}$)
- Sinónimos ($\text{sim} > \text{cos}$)
- Relaciones (ej: doctor/paciente/enfermera/tratamiento)
- Matriz dispersa

Análisis Semántico Latente (LSA)

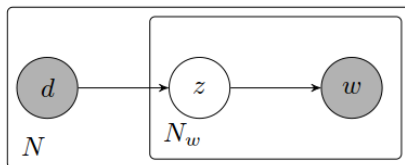
- Utiliza una técnica de reducción dimensional conocida como descomposición de valores singulares.
- El espacio dimensional reducido se le denomina espacio semántico.
- Se identifican conceptos ocultos (latentes).
- Se comparan los documentos en el espacio semántico obtenido.

GLSA (Generalized Latent Semantic Analysis)

- Variación de LSA, que utiliza n-gramas en lugar de documentos.
- Permite, por ejemplo, hacer distinción entre “Literatura Fantástica” con “Fantástica Literatura”.
- Considera orden de las palabras.
- Sin embargo tiene algunas falencias: Alta dimensionalidad del espacio vectorial, alto grado de dispersión. La descomposición de valores singulares es costosa computacionalmente y crece con el orden de los n-gramas (explosión combinatoria).

Probabilistic Latent Semantic Analysis

- Ni LSA ni GLSA logran solucionar el problema de la polisemia.
- Otro problema relevante es que el espacio semántico obtenido es difícil de interpretar.
- PLSA se fundamenta en un modelo probabilístico denominado *Aspect Model*, el cual expresa que las variables ocultas (tópicos) están asociadas a las variables observadas.



Algunas ventajas de PLSA:

- Modelo con base estadística sólida, con parámetros bien definidos que tienen una interpretación probabilística clara (variables latentes como tópicos).
- Se puede utilizar teoría estadística para la selección del modelo (es decir, cantidad de dimensiones).
- Logra solventar el problema de la polisemia.

Trabajo Relacionado

Métodos basados en teoría de discurso:

- Los métodos descritos hasta ahora, consideran los documentos como un conjunto de palabras independientes (modelo de bolsa de palabras).
- El orden importa, pues afecta a la coherencia textual. La *teoría del centrado* establece que cada segmento de un texto (enunciados u oraciones) tiene asociado un centro (los cuales pueden ser pronombres, nombres, objetos, etc.).
- La teoría define cuatro tipo de transiciones, las cuales sirven para evaluar qué tan coherente es un texto.

	$C_b(U_i) = C_b(U_{i-1})$	$C_b(U_i) \neq C_b(U_{i-1})$
$C_b(U_i) = C_p(U_i)$	<i>Continue</i>	<i>Smooth Shift</i>
$C_b(U_i) \neq C_p(U_i)$	<i>Retain</i>	<i>Rough Shift</i>

Trabajo Relacionado

Ejemplo:

- a) Juan tenía un terrible dolor de cabeza.
($C_b = ?$ $C_f = Juan > DolordeCabeza$, $transicion = ninguna$)
- b) Cuando la reunión terminó
($C_b = ninguno$ $C_f = Reunion$, $transicion = Rough Shift$)
- c) él corrió a la farmacia
($C_b = ninguno$ $C_f = Juan$, $transicion = Rough Shift$)

- a) Juan tenía un terrible dolor de cabeza.
($C_b = ?$ $C_f = Juan > DolordeCabeza$, $transicion = ninguna$)
- b) él corrió a la farmacia
($C_b = Juan$ $C_f = Juan > Farmacia$, $transicion = Continue$)
- c) cuando la reunión terminó
($C_b = ninguno$ $C_f = Reunion$, $transicion = Rough Shift$)

Método de evaluación de coherencia textual basado en teoría del centrado

- Se asume que existen ciertas transiciones que ocurren en textos coherentes.
- La idea es, representar un texto como un conjunto de oraciones, donde cada entidad del texto tiene cambios de rol (transiciones) a lo largo de las oraciones. Estos cambios de rol, representan cómo van presentando las ideas del texto.

Trabajo Relacionado

Ejemplo de documento representado como una matriz de entidades.

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	s	o	s	x	o	-	-	-	-	-	-	-	-	-	-	1
2	-	-	o	-	-	x	s	o	-	-	-	-	-	-	-	2
3	-	-	s	o	-	-	-	-	s	o	o	-	-	-	-	3
4	-	-	s	-	-	-	-	-	-	-	-	s	-	-	-	4
5	-	-	-	-	-	-	-	-	-	-	-	-	s	o	-	5
6	-	x	s	-	-	-	-	-	-	-	-	-	-	-	o	6

	s	s	o	s	x	s	-	o	s	o	o	x	-	x	s	o	x	x	x	x	-	s	-	o	x	-
d_1	0	0	0	.03	0	0	0	0	.02	.07	0	0	.12	.02	.02	.05	.25									
d_2	0	0	0	.02	0	.07	0	.02	0	0	.06	.04	0	0	0	.36										
d_3	.02	0	0	.03	0	0	0	.06	0	0	0	0	.05	.03	.07	.07	.29									

Propuesta

Se propone combinar los distintos métodos para obtener un método que considere distintos aspectos de la coherencia textual.

- Se propone utilizar medidas estadísticas que reflejan parcialmente características intrínsecas de los ensayos.
- Utilizando un modelo de espacio vectorial se pueden evaluar algunos aspectos en relación a la semántica del texto.
- Posteriormente, para considerar el orden de las ideas, se propone adaptar los métodos de evaluación de coherencia textual basados en teoría del discurso, para ser utilizados en el contexto de la evaluación automática de ensayos.
- Finalmente, se propone combinar estos métodos.

- ➊ Revisión bibliográfica de métodos para evaluación de coherencia a nivel de discurso.
- ➋ Establecer una representación de textos con la que se pueda modelar la semántica del contenido textual.
- ➌ Recopilación de datos de ensayos evaluados por humanos, los cuales se limpiarán y prepararán para utilizarlos en el método a desarrollar.
- ➍ Desarrollo de un método de evaluación que considere sintaxis y semántica del ensayo a evaluar.
- ➎ Implementación de un prototipo computacional de los distintos métodos propuestos en la literatura, con la finalidad de realizar experimentos que validen la hipótesis.

- 6 Evaluación del método propuesto y comparación con otros métodos existentes en la literatura, para ello se utilizarán métricas tales como:
- *Exact Agreement*: Mide la proporción de ensayos que fueron calificados igualmente por el evaluador humano y el método computacional.
 - *Adjacent Agreement*: Mide la proporción ensayos que fueron evaluados igual por el evaluador humano y el método computacional o que difieren en a lo más 1 punto (de calificación).
 - *Quadratic Weighted Kappa*: Mide el grado de acuerdo entre los puntajes asignados por dos evaluadores. Su valor máximo es 1, si los evaluadores están completamente de acuerdo, si hay desacuerdo, la métrica puede tomar valores negativos.
 - *Correlación de Spearman*: Mide qué tan buena es la relación entre la evaluación humana y la automática.

Plan de Trabajo

- Revisión bibliográfica de métodos de evaluación de coherencia textual, basados en teoría de discurso - Mes 1
- Diseño de un método automático para evaluar ensayos - Meses 2 - 3
- Implementación de métodos en el estado del arte, para posteriormente comparar con lo propuesto.
- Ensamble de modelos que consideren sintaxis y semántica - Mes 2
- Crear prototipo para realizar pruebas - Mes 3
- Evaluar rendimiento del modelo y comparar con modelos del estado del arte - Mes 4

FIN