

# 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*

8 de agosto de 2016

# Contenido

- 1 Introducción
- 2 Trabajo Relacionado
- 3 Método Propuesto
- 4 Evaluación del método propuesto
- 5 Conclusiones

- Hoy en día los sistemas de recomendación son una componente importante en el comercio electrónico.
- Estos sistemas, buscan recomendar ciertos ítemes a un usuario en base a su perfil. El perfil de usuario se intenta inferir de acuerdo a su historial de compras, y preferencias.
- A pesar del amplio uso de estos sistemas de recomendación, el no considerar la situación actual del usuario podría resultar en malas recomendaciones. (Ej: Comprar un juguete para un sobrino, etc.)
- Para solucionar este problema, se introduce la noción de contexto y los sistemas de recomendación que consideran contexto (*Context-Aware Recommender Systems CARS*)

## 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:

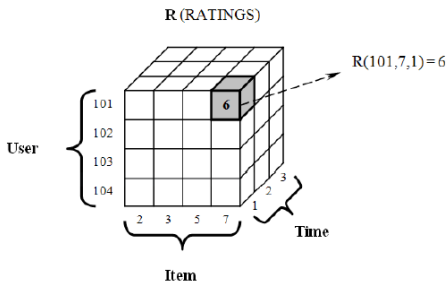
- Representan los textos como un conjunto de características que se utilizan como medidas indirectas de propiedades del texto tales como: Coherencia, cohesión, contenido.
- La nota se calcula como una suma ponderada de estas características (generalmente se utiliza regresión). El primer sistema exitoso (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



- Los sistemas de recomendación descritos previamente sólo utilizan información básica del usuario y del ítem.
- Se agrega el “contexto”, que podría ser una época del año (ejemplo: Recomendaciones de lugares para viajar en invierno y verano).
- El contexto y sus atributos dependerá del dominio en particular (ejemplo: música y estado anímico).

# Trabajo Relacionado

- En general el contexto puede ser explícito, implícito o puede ser inferido.
- En el caso explícito, se obtiene información directa del usuario, por ejemplo mediante cuestionarios, o preferencias. (Problema: Se requiere participación directa del usuario)
- En el caso implícito, el contexto puede obtenerse detectando “cambios” en el usuario, por ejemplo cambios en su ubicación (GPS), o fechas de últimas transacciones. (Problema: No siempre se obtiene información relevante)
- Se puede inferir el contexto aplicando técnicas de minado de datos. Se requiere construir un modelo predictivo (por ejemplo un clasificador).



# Método Propuesto

- El método propuesto se utiliza en el dominio de recomendaciones para viajes.
- Se extrae el contexto del usuario a partir de una consulta. Ejemplo:

I'm planning a romantic trip for my anniversary. I'm looking for an all inclusive resort near a beach. I expect the hotel room to be spacious, have a nice view over the sea and to be nicely decorated.

- Se hace el supuesto de que existen etiquetas explícitas que representan el contexto del usuario. En este dominio, el tipo de viaje.
- El contexto se representa como una función de distribución sobre las etiquetas posibles que puede tener un viaje.

# Método Propuesto

- No en todas las consultas estará explícito el contexto, por lo que se requiere un mecanismo de inferencia.
- La información necesaria para realizar las inferencias se extrae de las reseñas dadas por distintos usuarios.
- Por ejemplo, si en la reseña del usuario  $u$  se indica el tipo de viaje para el hotel  $i$  como “business” y “Solo Travel”:  
 $Context_u^i = \{P(family) = 0, P(couples) = 0, P(solo\ travel) = 0.5, P(business) = 0.5, P(friend's\ get\ away) = 0\}$
- El problema de inferencia descrito es similar al problema de clasificación de textos con varias etiquetas, en el cual, un texto puede ser clasificado en una o más categorías.

- Se escogió Labeled-LDA (L-LDA) como método de categorización de textos.
- L-LDA es una extensión de LDA (Latent Dirichlet Allocation).
- LDA representa los documentos en un CORPUS como una mezcla de K tópicos. En la fase de entrenamiento LDA “aprende” la representación de tópicos de cada documento y las palabras asociadas a cada tópico.

# Método Propuesto

## Ejemplo LDA

- Suponga que se tienen las siguientes oraciones:
  - 1 I ate a banana and spinach smoothie for breakfast.
  - 2 I like to eat broccoli and bananas.
  - 3 Chinchillas and kittens are cute.
  - 4 My sister adopted a kitten yesterday.
  - 5 Look at this cute hamster munching on a piece of broccoli.
- Utilizando LDA y 2 tópicos:
  - Oraciones 1 y 2: 100 % tópico A.
  - Oraciones 3 y 4: 100 % tópico B.
  - Oración 5: 60 % tópico A, 40 % tópico B.
  - Tópico A: 30 % broccoli, 15 % bannanas, 10 % breakfast, 10 % munching... (Se podría interpretar como un tópico que habla sobre comida)
  - Tópico B: 20 % chinchillas, 20 % kittens, 20 % cute, 15 % hamster... (Se podría interpretar el tópico como animales tiernos)

# Método Propuesto

- LDA es un método no supervisado. L-LDA impone restricciones sobre LDA para clasificar documentos según tópicos previamente definidos.
- Luego se puede clasificar una consulta dada, e “inferir” el contexto del usuario para realizar las recomendaciones.
- Por otro lado, utilizando el historial de “reviews” de los usuarios, se puede predecir el contexto de un usuario:

## Similaridad Contextual

$$\text{contextualSimilarity}(i, j) = \frac{\sum_u \text{commonLabels}(i, j)}{\sqrt{\sum_u |\text{labels}(i)| \times |\text{labels}(j)|}}$$

## Predicción del contexto

$$\text{predictedContext}(u, i) = \frac{\sum_{k \in \text{Neighbors}(i)} \text{context}_u^k \cdot \text{contextualSimilarity}(k, i)}{\sum_{k \in \text{Neighbors}(i)} |\text{contextualSimilarity}(k, i)|}$$

Finalmente, utilizando el contexto previsto y el contexto inferido, se puede estimar cuánta utilidad le aporta el ítem al usuario:

### Puntaje Contextual

$$\text{contextScore}(u, i) = \frac{IC_u \cdot PC_u^i}{\|IC_u\| \|PC_u^i\|}$$

### Utilidad

$$\text{utility}(u, i) = \alpha \text{predictedRating}(u, i) + (1 - \alpha) \text{contextScore}(u, i)$$

Donde  $\text{predictedRating}(u, i)$  se obtiene mediante métodos tradicionales (como filtrado colaborativo), y  $\alpha$  es una constante que representa el peso del puntaje previsto para el ítem.

# Evaluación del Método propuesto

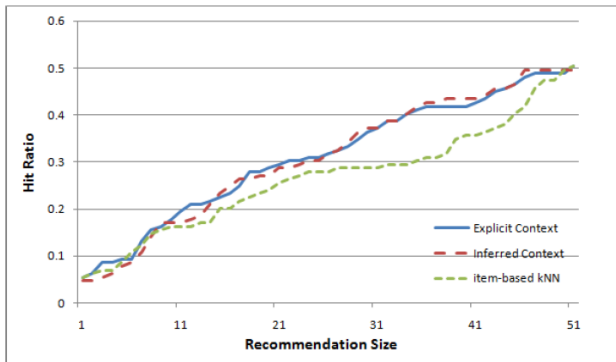
- Como no se está estimando la nota que el usuario le daría a un ítem en particular, no tiene sentido utilizar métricas como MAE u otras que comparan la nota estimada con la real.
- Se propone utilizar la métrica Hit Ratio.
- Se utilizó validación cruzada dejando uno fuera. Se calculó el *Hit Ratio* para dicho ítem.

## Hit Ratio

$$\text{Hit Ratio}_i = \sum_{u \in U_T} \frac{H_{ui}}{|U_T|}$$

Donde  $U_T$  son los usuarios en el conjunto de pruebas,  $H_{ui} = 1$  el ítem  $i$  es recomendado al usuario  $u$  (0 en caso contrario). En palabras simples, el *hit ratio* es la probabilidad de que el ítem  $i$  sea recomendado en  $N$  conjuntos de recomendaciones.

# Evaluación del Método propuesto





- Se propuso un nuevo enfoque para el minado de contexto desde un texto en lenguaje natural, y se utilizó para producir recomendaciones que consideraran contexto.
- Las evaluaciones indican que utilizar información contextual mejora las recomendaciones en términos del *hit ratio*.

# The End