

Escuela de Ingeniería en Computación

Inteligencia Artificial - IC6200 - Grupo 2

**Proyecto de Investigación: Vector Representations of Words**

**Estudiante:**

Alanis Pineda, Stephannie - 201263425

**Profesor:** Luis Carlos Hernández Vega

Cartago, Costa Rica

II Semestre, 2017

**Conclusiones**

* La idea principal del tutorial es poder transformar las palabras en vectores, debido a que asociar una palabra con un identificador no es suficiente información por almacenar y se necesitarían más datos para poder formar modelos estadísticos.

Parte de las ventajas que tiene este método (convertir palabras en vectores) es que las palabras que aparecen en un contexto semántico similar, significan que comparten ciertas propiedades entre ellas (Comparten un significado semántico). Son cercanas semánticamente.

* Hay dos categorías en el principio (compartir un significado semántico):
  + Métodos basados en recuento
  + Métodos predictivos
* Para este tutorial se utiliza el método predictivo porque en éste se intenta predecir directamente una palabra de sus vecinos, simplificando la inserción de una palabra en oraciones.
* El modelo word2vec es computacionalmente eficiente para aprender palabras embebidas.
* Vector Space Models representan palabras “empotradas” en un espacio contínuo en un vector donde palabras semánticamente similares son mapeadas a puntos cercanos.
* Es posible visualizar vectores aprendidos en dos dimensiones usando t-SN3 dimensionality reduction technique. Esta técnica muestra que vectores capturan información general, útil y semántica sobre palabras y sus relaciones a otras.
* Implementación dada muestra que TensorFlow posee gran flexibilidad, entre los elementos que brindan esta flexibilidad está que el cambiar el objetivo de entrenamiento es sencillo, ya que con cambiar la llamada de una función es posible.
* El modelo muestra cómo TensorFlow aporta la flexibilidad necesaria para una experimentación temprana y el control que luego se va a necesitar para la implementación.

**Aplicaciones a un problema real**

Se puede utilizar para recuperación textual. Cuando se pierden datos de un archivo de texto, se pueden calcular o estimar los datos que no pudieron ser recuperados, también al agrupar palabras es mucho más fácil realizar una búsqueda de palabras al estar alineadas por similitudes y en vectores.

**Screenshots**

Entrenamiento de la red neuronal word2vector\_basic.py.







