

Optimización online de los pesos del modelo log-lineal en traducción automática interactiva

Autor: Mara China Rios

**Directores: Francisco Casacuberta Nolla
Germán Sanchis Trilles**

Doctorado en Informática
Pattern Recognition and Human Language Technology
Universidad Politécnica de Valencia, España

machirio@doctor.upv.es



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Resumen

- La intervención de los traductores humanos en un escenario de post-edición para corregir las traducciones obtenidas a partir de los sistemas de traducción automática estadística (statistical machine translation SMT) es aún muy necesaria para lograr la calidad deseada.
- El paradigma de la traducción automática interactiva (Interactive Machine Translation, IMT), es capaz de reducir el esfuerzo y tiempo que el traductor humano tiene que invertir en el proceso de corrección.
- En este trabajo se analiza el algoritmo de aprendizaje online Discriminative Ridge Regression, para el aprendizaje de los pesos del modelo log-lineal dentro del paradigma de traducción automática interactiva.

Introducción

En este trabajo nos centramos en adaptar los pesos del modelo log-lineal α , muy empleados en los sistemas SMT en la actualidad. La ecuación estándar de un sistema SMT [1] es:

$$\hat{y} = \arg \max_y \sum_{m=1}^M \lambda_m h_m(\mathbf{x}, \mathbf{y}) = \arg \max_y \boldsymbol{\lambda} \cdot \mathbf{h}(\mathbf{x}, \mathbf{y}) = \arg \max_y s(\mathbf{x}, \mathbf{y}) \quad (1)$$

$h_m(\mathbf{x}, \mathbf{y})$ es la puntuación de una función que representa una característica importante en la traducción de \mathbf{x} a \mathbf{y} , M es el número de modelos o características y λ_m representa un peso de la combinación log-lineal. $h(\cdot|\cdot)$ se calcula empleando un conjunto de entrenamiento y los valores λ son ajustados mediante un conjunto de desarrollo. $s(\mathbf{x}, \mathbf{y})$ representa la puntuación de la hipótesis y dada una oración de entrada \mathbf{x} .

Traducción automática interactiva

- Ecuación de SMT aplicada a traducción automática interactiva (IMT):

$$\hat{y} = \arg \max_y Pr(\mathbf{y} | \mathbf{x}) \Rightarrow \hat{s}_h = \arg \max_{s_h} Pr(s_h | \mathbf{x}, \mathbf{p}, k) \quad (2)$$

Aprendizaje online de los pesos

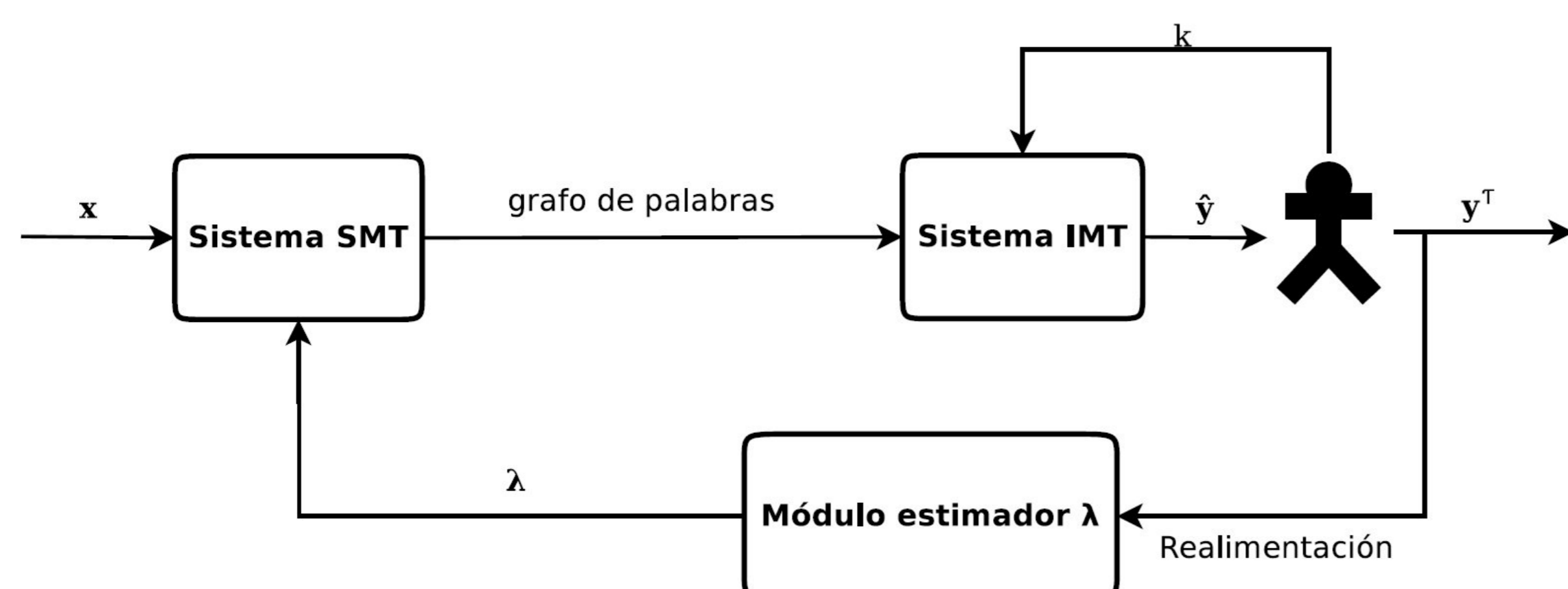


Figura 1: Esquema del paradigma de aprendizaje online dentro del marco de la IMT

Discriminative ridge regression

- Algoritmo de adaptación online discriminativo
- Tiene en cuenta todas las hipótesis dentro de la lista N-best (n-mejores traducciones)

El algoritmo Discriminative ridge regression (DRR) [3] pretende encontrar el vector $\tilde{\lambda}$ que relacione las diferencias en valores como la diferencia en la calidad de las diferentes hipótesis.

$$\tilde{\lambda}_t = (\mathbf{R}'_x \cdot \mathbf{R}_x + \beta \mathbf{I})^{-1} \cdot \mathbf{1}_x \quad (3)$$

donde β es un valor pequeño que representa el termino de regularización para estabilizar el producto $\mathbf{R}'_x \cdot \mathbf{R}_x$ y asegurar que este sea invertible.

- \mathbf{R}_x se define de la siguiente manera: $\mathbf{R}_x = \mathbf{H}_x^* - \mathbf{H}_x$
- $\mathbf{H}_x = [\mathbf{h}(\mathbf{x}, \mathbf{y}_1), \dots, \mathbf{h}(\mathbf{x}, \mathbf{y}_N)]'$, matriz de talla $N \times M$ donde M es el número de características que contienen las funciones de características \mathbf{h} de cada frase y N el número de frases dentro de la lista N-best.
- $\mathbf{H}_x^* = [\mathbf{h}(\mathbf{x}, \mathbf{y}^*), \dots, \mathbf{h}(\mathbf{x}, \mathbf{y}^*)]'$, matriz donde sus filas van a ser iguales al vector de características de la mejor hipótesis \mathbf{y}^* de la lista N-best.

Discriminative ridge regression en escenario interactivo

- En IMT la métrica de calidad está relacionada a un grafo de palabras (no a una hipótesis)
 - ⇒ Es necesario obtener N-best grafos
 - ⇒ No está claro como realizarse, por lo tanto lo aproximamos mediante la generación de conjuntos de pesos
- Necesitamos las características asociadas al grafo de palabras
 - ⇒ No existe un único conjunto de características para un grafo de palabras
 - ⇒ Representamos al grafo de palabras con su mejor camino

Generación de los conjuntos de pesos

- El conjunto de pesos empleados para la generación del grafo de palabras es de gran influencia para su calidad
- Se generó un conjunto finito de nuevos conjuntos de pesos Λ , partiendo del conjunto de pesos baseline $\hat{\lambda}$
- Se emplearon dos aproximaciones para obtener Λ
 - Aproximación gaussiana
 - Aproximación simplex

Experimentos

Todos los experimentos realizados emplearon el mismo conjunto de datos Inglés-Español partición obtenida para el Taller de 2013 en la conferencia ACL. El conjunto de pesos log-lineales iniciales se ajustó por medios la técnica MERT **referencia**, a partir del conjunto de desarrollo del 2008 al 2010. Un verdadero experimento IMT es demasiado costoso para fines de experimentación porque requeriría una evaluación humana, simulamos este procedimiento mediante un conjunto de test del año 2011. Para los experimentos SMT se empleo el toolkit de código abierto Moses [2] y para la simulación del proceso IMT el toolkit de código abierto Thot [4].

Corpus y métricas de evaluación

- Corpus empleados: Europarl y News Commentary
- Traducir: Inglés-Español

La métrica de evaluación empleada fue:

- KSMR (Keystroke Mouse-action Ratio)
KSMR = número total de pulsaciones de teclado + clicks al ratón que realiza el usuario / número total de caracteres de la oración

Resultados

Principales resultados de los experimentos:

Método	α	N	KSMR
baseline			40,6
original	0,01	5000	42,9
gaussina	0,001	201	40,9
simplex	0,0001	70	40,4

Cuadro 1: Resultados de los experimentos

Conclusiones

- Se empleó el algoritmo Discriminative Ridge Regression en un escenario IMT simulado
- Para la generación de los conjuntos de pesos se emplearon dos aproximaciones: la aproximación gaussiana y la aproximación simplex
- Se logran obtener resultados positivos disminuyendo el número de iteraciones con la aproximación simplex.
- Los resultados obtenidos no proporcionan grandes ganancias, pero demuestra que hay margen de mejora en esta dirección.

Referencias

- Peter F Brown, Vincent J Della Pietra, Stephen A Della Pietra, and Robert L Mercer. The mathematics of statistical machine translation: Parameter estimation. *Computational linguistics*, 19(2):263–311, 1993.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. Moses: open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, ACL '07*, pages 177–180, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics.
- Pascual Martínez-Gómez, Germán Sanchis-Trilles, and Francisco Casacuberta. Online adaptation strategies for statistical machine translation in post-editing scenarios. *Pattern Recognition*, 45(9):3193–3203, 2012.
- D. Ortiz-Martínez, I. García-Varea, and F. Casacuberta. Thot: a toolkit to train phrase-based statistical translation models. In *Tenth Machine Translation Summit. AAMT*, Phuket, Thailand, September 2005.