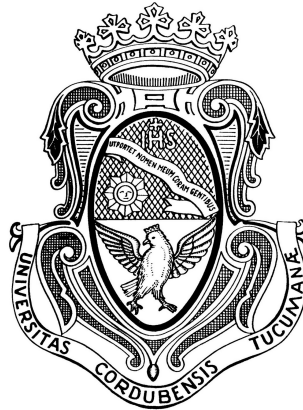


UNIVERSIDAD NACIONAL DE CÓRDOBA

FACULTAD DE MATEMÁTICA, ASTRONOMÍA Y FÍSICA



Trabajo final del Doctorado en Ciencias de la Computación

**Generación de expresiones referenciales: el punto
de vista lógico, el más humano**

Alumna: Ivana Romina Altamirano

Supervisora: Luciana benotti

CÓRDOBA ARGENTINA
Diciembre, 2013

Gracias Lu, Carlos . . .

■ CLASIFICACIÓN DE BIBLIOTECA:

I.2.7

■ PALABRAS CLAVE:

Palabra Clave 1, *Palabra Clave 2*, Palabra Clave 3, Palabra Clave 4, *Palabra Clave 5*, *Palabra Clave 6*.

RESUMEN

Se propone un algoritmo para la generación de expresiones referenciales (REs) que adapta la aproximación de [?, ?] para incluir sobre-especificación y probabilidades aprendidas a partir de corpus. Luego de introducir el algoritmo discutimos como las probabilidades requeridas pueden ser computadas para cualquier dominio para el cual existe un corpus de REs y como las probabilidades pueden ser ajustadas para nuevas escenas en el dominio usando aprendizaje automático. Nosotros ejemplificamos como computar las probabilidades sobre el corpus GRE3D7 de [?] y el TUNA corpus de [?]. El algoritmo resultante es capaz de generar diferentes expresiones referenciales para el mismo target con una frecuencia similar a aquella observada en el corpus. Nosotros empíricamente evaluamos el nuevo algoritmo sobre el corpus GR3D7, y mostramos que la distribución de probabilidad de las expresiones referenciales generadas por el algoritmo machean las encontradas en el corpus con alta precisión. Nosotros también comparamos nuestros resultados con los resultados the [?], los ganadores de la competencia NLG Challenge sobre selección de atributos para generación de expresiones referenciales (ASGRE). También proveemos un análisis de error y conclusiones para ambos corpus.

ABSTRACT

We propose an algorithm for the generation of referring expressions (REs) that adapts the approach of arec2:2008:Areces,arec:usin11 to include overspecification and probabilities learned from corpora. After introducing the algorithm, we discuss how probabilities required as input can be computed for any given domain for which a suitable corpus of REs is available, and how the probabilities can be adjusted for new scenes in the domain using a machine learning approach. We exemplify how to compute probabilities over the GRE3D7 corpus of viet:gene11 and the TUNA corpus of gatt-balz-kow:2008:ENLG. The resulting algorithm is able to generate different referring expressions for the same target with a frequency similar to that observed in corpora. We empirically evaluate the new algorithm over the GRE3D7 corpus, and show that the probability distribution of the generated referring expressions matches the one found in the corpus with high accuracy. We also compare our results with the results of graph08 the winners of the NLG Challenge on Attribute Selection for Generating Referring Expressions (ASGRE). We provide error analysis and conclusion for both corpus.

Índice general

1. Introducción	3
2. Selección de contenido de expresiones referenciales	5
3. Decisiones metodológicas de corpus y evaluación	7
4. Recolección y análisis del corpus ACREM	9
5. Nuestra propuesta	11
6. Evaluación de nuestra propuesta	13
7. Conclusiones	15

En ling $\tilde{A}_{\frac{1}{4}}$ Astica una expresi \tilde{A}^3 *nreferencial* (RE) es una expresi \tilde{A}^3 *nque identifica un vocamente a un objet de par*

Capítulo 1

Introducción

En ling $\tilde{A}^{\frac{1}{4}}$ stica una expresi \tilde{A}^3 nreferencial(RE)es una expresi \tilde{A}^3 nque identifica unvocamente a un objeto de para un inte

Capítulo 2

Selección de contenido de expresiones referenciales

Capítulo 3

Decisiones metodológicas de corpus y evaluación

Capítulo 4

Recolección y análisis del corpus ACREM

Capítulo 5

Nuestra propuesta

Capítulo 6

Evaluación de nuestra propuesta

Capítulo 7

Conclusiones