

Emerging Topics from the Fifth Italian Conference on Computational Linguistics

Roberto Basili*
Università di Roma, Tor Vergata

Simonetta Montemagni**
ILC - CNR

Il primo numero del quinto anno della rivista *Italian Journal of Computational Linguistics (IJCoL)*, la rivista italiana promossa dall'Associazione Italiana di Linguistica Computazionale (AILC - www.ai-lc.it), è un volume miscelaneo i cui articoli documentano lavori di ricerca risultati particolarmente promettenti nell'ambito della Conferenza *CLiC-it 2018* (Torino, 10-12 dicembre 2018) insieme ad un contributo originale sottomesso per la pubblicazione sulla rivista. La selezione di contributi qui rappresentata documenta la ricerca condotta in diversi ambiti, che spaziano da studi sulle applicazioni di modelli neurali all'apprendimento in sistemi di Natural Language Processing, sui modelli vettoriali della semantica distribuzionale sino a metodi complessi di inferenza linguistica, utili in applicazioni moderne, quali la traduzione automatica. I temi affrontati dunque coprono sviluppi recenti e molto fecondi della ricerca in linguistica computazionale, come ad esempio la trasparenza epistemologica dei modelli induttivi basati su reti neurali, l'annotazione sintattica automatica dell'italiano, l'utilizzo di tecniche avanzate di apprendimento neurale, quali gli algoritmi basati sui "trasformer", per arrivare ai modelli distribuzionali del significato.

Gli articoli sui temi emergenti sono stati, come sempre, selezionati attraverso un processo iterativo di *peer-review*: come contributo alla conferenza e come candidato ai premi di "Best Young Paper" e "Distinguished Young Paper". Infine, la versione estesa degli articoli selezionati da *CLiC-it 2018* e gli articoli proposti per la pubblicazione su *IJCoL* sono stati oggetto di revisione paritaria come articolo di rivista scientifica.

Nel contributo di Rossini, Croce e Basili viene discusso il problema della leggibilità dei modelli neurali del linguaggio naturale, che rappresenta una questione centrale quando si considera l'impatto di tali modelli all'interno di scenari applicativi del mondo reale. In particolare, viene proposta una tecnica di *embedding* basata su un metodo matematico per la rappresentazione di spazi caratterizzati da un numero elevato di dimensioni (denominata *metodologia di Nyström*), finalizzata a supportare inferenze analogiche relative alla relazione causale tra i dati di addestramento e le decisioni della rete neurale. I metodi *kernel* utilizzati per costruire gli *embeddings* di *Nyström* sono in grado di catturare informazione semantica, sintattica e lessicale, fornendo spiegazioni fondate e percepite come naturali dall'utente. La valutazione quantitativa riportata si riferisce a un compito di inferenza semantica, il *Semantic Role Labeling*, condotto per l'italiano e l'inglese qui usato come caso di studio. I risultati raggiunti mostrano che le spiegazioni basate su *kernel* semantici sono particolarmente affidabili ed efficaci, consentendo alle

* Dept. of Enterprise Engineering - Via del Politecnico 1, 00133 Roma
E-mail: basili@info.uniroma2.it

** Istituto di Linguistica Computazionale "A. Zampolli", CNR - Via Moruzzi 1, 56124 Pisa
E-mail: simonetta.montemagni@ilc.cnr.it

strutture linguistiche di contribuire in modo significativo alla percezione dell'utente riguardo all'affidabilità della decisione.

Nel secondo articolo, Antonelli e Tamburini riportano i risultati di una vasta sperimentazione condotta con architetture neurali per l'analisi sintattica a dipendenze di testi italiani. Gli esperimenti fanno uso di due *treebanks*, rappresentative di diversi generi testuali, sintatticamente annotate secondo lo schema delle *Universal Dependencies*, ad oggi uno standard "de facto" per l'annotazione sintattica a dipendenze. Viene proposta una nuova architettura di analisi basata su sistemi che combinano il risultato di diversi analizzatori sintattici a dipendenze (*ensemble systems*), che differisce dalle precedenti tecniche di combinazione sia per i parsers utilizzati sia per il metodo di combinazione degli output. Con questo studio è stato possibile valutare l'impatto di diversi algoritmi e tecniche di combinazione degli output rispetto sia a misure standard di valutazione sia all'analisi della qualità dell'albero sintattico generato. I risultati ottimali sono stati ottenuti applicando lo schema di voto a maggioranza, nonostante il fatto che tale metodo non garantisca che l'albero generato sia ben formato.

Seguono due articoli che condividono l'approccio distribuzionale all'analisi semantica. Il contributo di Aina, Bernardi e Fernández riporta i risultati di uno studio di semantica distribuzionale dedicato alla modellazione del fenomeno della negazione di un aggettivo, analizzato in rapporto alla corrispondente coppia di antonimi per la lingua inglese (es. *not cold* vs *cold - hot*). Dopo aver costruito la rappresentazione vettoriale di un insieme di antonimi e delle relative negazioni a partire dalla loro distribuzione nei contesti linguistici in cui ricorrono, l'analisi ha studiato le similarità contestuali osservabili riguardo agli aggettivi oggetto di negazione e gli aggettivi della corrispondente coppia di antonimi. I risultati raggiunti dimostrano che la negazione di un aggettivo (es. *not cold*) è maggiormente simile in termini distribuzionali all'aggettivo stesso che al suo antonimo (ovvero a *cold* piuttosto che a *hot*). L'analisi è stata condotta in relazione a diversi tipi di coppie di antonimi, ovvero *antonimi morfologici* che condividono la radice lessicale, e *antonimi lessicali* contraddistinti da radici lessicali diverse (all'interno dei quali si distinguono antonimi *contrari* vs *contraddittori*). Mentre nel caso di antonimi morfologici la maggiore similarità distribuzionale dell'aggettivo negato rispetto all'aggettivo stesso risulta meno marcata, non sono emerse differenze di comportamento significative tra i diversi tipi di antonimi lessicali e di negazione (semplice vs doppia).

Il contributo di Pannitto e Lenci si pone come obiettivo la valutazione del contributo dell'informazione eventiva nel processo di comprensione della frase. Partendo dalla constatazione che la maggior parte dei modelli proposti nell'ambito della semantica distribuzionale compositiva si basa sull'utilizzo dei soli vettori lessicali, gli autori propongono di arricchire il miglior modello presente in letteratura, costituito dalla somma di vettori (assunta come *baseline*), con informazione distribuzionale sulla conoscenza degli eventi attivata dalle parole nella frase. Recenti studi in ambito psicolinguistico mostrano che il lessico attiva conoscenza sulla struttura degli eventi e sui loro partecipanti prototipici. I risultati degli esperimenti confermano che i vettori distribuzionali non contengono informazione eventiva sufficiente e che, in linea con l'evidenza psicolinguistica, la conoscenza attivata dal lessico svolge un ruolo centrale nella costruzione della rappresentazione semantica della frase. L'approccio proposto è stato valutato su due risorse, RELPRON e un insieme di frasi transitive annotate sintatticamente e semanticamente, con risultati promettenti che mostrano un netto e sistematico miglioramento rispetto ai risultati della baseline, confermando così l'ipotesi di partenza sul ruolo dell'informazione eventiva attivata dal lessico nel processo di comprensione della frase.

Chiude il volume l'articolo di Tebbifakhr e colleghi che propone un nuovo approccio alle fasi di *post-editing* automatico (APE) di un sistema di traduzione automatica neurale. I metodi APE operano sull'output del sistema di traduzione automatica considerando anche la corrispondente frase di origine. Nel contributo gli autori propongono un'architettura neurale basata su *transformer* per APE, i cui vantaggi computazionali e ingegneristici sono significativi, in quanto possono essere facilmente parallelizzati e mantenuti nel tempo. Nell'articolo è anche indagata l'ipotesi relativa al ruolo della quantità dei dati usati per l'addestramento. Relativamente a ciò viene dimostrato che insiemi di dati annotati di grandi dimensioni non implicano sempre prestazioni migliori, poiché l'addestramento del sistema sulla base di piccoli e mirati sottoinsiemi di dati del dominio può portare a una precisione maggiore.

Questa breve sintesi non fa giustizia ai molti stimoli forniti dai lavori del presente volume e alla loro variegata rassegna di metodologie e applicazioni innovative nel trattamento automatico della lingua. Lasciamo, come sempre, al lettore il piacere di approfondirli direttamente nell'interesse di questo volume.

Editorial Note Summary

It is with great pleasure that we introduce the first volume of the fifth year of the *Italian Journal of Computational Linguistics* (IJCoL) promoted by the *Associazione Italiana di Linguistica Computazionale* (AILC - www.ai-lc.it). This is a miscellaneous volume which collects a selection of particularly promising research contributions inspired by young researchers at the CLiC-it 2018 Conference (Turin, 10-12 December 2018). This selection of contributions from CLiC-it 2018 focuses on different paradigms and research aspects, ranging from neural learning for NLP to distributional semantics and applications such as machine translation. As for the other miscellaneous issues, the papers have been selected through an iterative peer-review process. Each selected conference article underwent the evaluation as a contribution to the conference and as a candidate for the "Best Young Paper" and "Distinguished Young Paper" awards of CLiC-it 2018. Finally, both extended and newly submitted papers have been peer-reviewed as journal articles.

The paper by Rossini, Croce and Basili discusses the problem of the readability of neural natural language models, a relevant research issue, as for its impact in real world NLP applications. An embedding technique (based on a mathematical method to represent large dimensional feature spaces, called the *Nyström* methodology) is here proposed as a way to support analogy-driven inferences about the causal relationship between training data and a neural net decisions. Kernel methods used to build the Nyström embedding capture grammatical and lexical semantic information, making explanations meaningful and naturally perceived by the user. The reported quantitative evaluation uses a semantic inference task, i.e. Semantic Role Labeling, both in English and Italian, as a reference test bed. Results suggest that explanations based on such kernels let semantic and syntagmatic structures to contribute to convincing arguments. They are in fact rather effective in the user assessment about the correctness of machine decisions.

In the second paper, Antonelli and Tamburini present an extensive experimentation on neural architectures for parsing Italian texts. The experiments use two Italian treebanks with different text types annotated according to the Universal Dependencies project standards. A new parsing architecture based on ensemble systems is proposed, extending previous ensemble techniques by a voting capability based on individual dependency relations. The aim has been to assess the impact of different algorithms and ensemble techniques with respect to standard parsing quality measures as well as

to the linguistic quality of the resulting parse tree. Optimal results are obtained when the majority-voting ensemble scheme is applied, whose major drawback is the fact that it does not always ensure the well-formedness of the generated dependency tree.

Two articles follow that share the distributional approach to semantic analysis. The contribution by Aina, Bernardi and Fernández reports the results of a distributional semantics study devoted to the modeling of the negation of an adjective analyzed in relation to the corresponding pair of antonyms for the English language (eg *emph not cold* vs *emph cold* - *emph hot*). After having constructed the vector representation of a set of antonyms and their negations starting from their distribution in the linguistic contexts in which they occur, the analysis has concerned the similarities found between the contexts of the adjectives object of negation and the adjectives of the corresponding pair of antonyms. The results achieved show that the negation of an adjective (e.g. *not cold*) is more similar in distributional terms to the adjective itself than to its antonym (i.e. to *cold* rather than to *hot*). The analysis was carried out in relation to different types of pairs of antonyms, namely *morphological antonyms* sharing the lexical root, and *lexical antonyms* characterized by different lexical roots (within which *contrary vs contradictory* antonyms are distinguished). While in the case of morphological antonyms the greater distributional similarity of the negated adjective with respect to the adjective itself is less marked, no significant differences emerged between the different types of lexical antonyms and negation (simple vs double).

The paper by Pannitto and Lenci is aimed at evaluating the contribution of event information in the process of sentence comprehension. Starting from the observation that most of the models proposed in the compositional distributional semantics literature are based on the use of lexical vectors only, the authors propose to enrich the best performing model in the literature, consisting in vector addition (assumed as baseline), with distributional information about events which is activated by the words in the sentence. Recent studies in the psycholinguistic literature show that the lexicon activates the knowledge about the structure of events and their prototypical participants. The results of the experiments presented in the paper confirm that distributional vectors do not contain sufficient event information and that, in line with psycholinguistic evidence, the knowledge activated by the lexicon plays a central role in the construction of the semantic representation of the sentence. The proposed approach has been evaluated against two resources, RELPRON and a set of transitive sentences annotated syntactically and semantically, with promising results showing a clear and systematic improvement with respect to the selected baseline. This confirms the hypothesis about the role played by event information which is activated by the lexicon in the sentence comprehension process.

The volume closes with the paper by Tebbifakhr and colleagues proposing a novel approach to the automatic post-editing stages (APE) of a neural machine translation system. APE methods are competitive as they correct the raw translation output by also considering the corresponding source sentence. In the paper the authors propose a Transformer based neural architecture for APE, whose computational and engineering advantages are significant as they can be easily parallelized and maintained. In the paper assumptions about the role of the size of the training dataset are also validated. It is shown that the widely assumed assumption suggesting that larger data sets imply better performances does not always hold, as fine-tuning the system on small in-domain data can yield higher accuracy.

This synthetic view does not exhaust the suggestions and nuances inspired by the papers here collected. We leave the reader the pleasure to discover them through a thoughtful sailing across the rest of the volume contents.