Characteristics of Citation Scopes: 
A Preliminary Study to Detect Citing Sentences

In-Su Kang\textsuperscript{1} and Byung-Kyu Kim\textsuperscript{2,*}

\textsuperscript{1}Kyungsung University, Pusan, South Korea
\textsuperscript{2}Korea Institute of Science and Technology Information (KISTI), Daejeon, South Korea
dbaisk@ks.ac.kr, yourovin@kisti.re.kr

Abstract. Citing sentences are gaining much attention in citation-based summarization and article review generation which depend on precisely identifying the scope of citing sentences. This article presents characteristics of citing sentences and citation scopes, which were obtained from the manual analyses of numerous citing sentences.

Keywords: Citing Sentence, Detection of Citing Sentences, Citation Scope Unit.

1 Introduction

Recently, citing sentences receive a growing attention in understanding academic articles cited. The following shows an example of citing sentences which provide a peer’s review on the article cited as ‘Li et al. (2012)’.

\textit{Li et al. (2012) proposed a SVM-based ensemble method for imbalanced data. They used VQ algorithm to segment the majority class to create less-skewed … The method showed the best performance for the well-known UCI datasets.}

A collection of such citing sentences from several different papers citing the same article could be used to generate an article review \cite{4}, or to produce an article summarization \cite{1,3}. In addition, Athar and Teufel \cite{2} employed citation context to discern positive and negative article reviews.

However, detecting the scope of citing sentences within a full-text article is not trivial, since citing sentences may not have explicit citation markers. For example, the first sentence in the above is explicitly indicated to cite something by a citation marker ‘Li et al. (2012)’, while the next sentences implicitly cite the same article without apparent marks. Earlier approaches to identifying citing sentences relied on either citation cue phrases \cite{4} or coreference-chain \cite{3}, reporting performances less than 80\% in F1.

\footnote{Corresponding author.}
As a preliminary study to identify citing sentences, we have attempted to manually analyze numerous citing sentences to determine different types of citation areas and some characteristics of citing sentences. This article presents the results of such analyses.

2 Data

We collected a set of 56 recent articles in PDF format published in 50 different academic journals which belong to broadly science and engineering fields. Next, a total of 1048 citing sentences with one or more explicit citation markers\(^1\) such as ‘(Croft, 2011)’ or ‘[23]’ was identified. For each citation marker found, the sentence containing the marker and its adjacent sentences were analyzed to determine the boundaries that the citation marker spans. This process corresponds to locating a group of citing sentences referring to a particular previous work. We call such group of citing sentences citation area, citation scope or citation context.

3 Characteristics of Citation Areas

Our analysis has identified five types of citation-scope units: phrase, clause, sentence, multi-sentence, and the others. The following (a) through (e) show real examples for such units sequentially, with citation markers bold-faced and citation-scope underlined.

(a) Using BioEdit\textit{(Hall 1999)}, we collated the total sequences and then aligned them visually.

(b) \textit{Lee et al. (2008)} presented symmetrical BVR light curves in the observing season in 2005, while Zhu & Qian (2009) obtained an asymmetrical V light curve with a strong O’Connell effect.

(c) The first GA approach for the structure learning is introduced by \textit{Larranaga et al.} at 1996 [12].

(d) Hoffmann’s light curves were reanalyzed by \textit{Kaluzny (1986)} with the Wilson Devinney binary model (WD: Wilson & Devinney 1971). His solutions showed that WZ Cep is a contact binary with the components of unequal surface temperatures.

(e) \textit{Fig. 13. Import of copper ore [14]}.

In the above, ‘(Hall 1999)’ of (a) was cited only to refer to a term ‘BioEdit’. ‘Lee et al. (2008)’ of (b) affects only the main clause, regardless of the subordinate clause ‘while’ leads. In (c), the whole sentence is all about ‘Larranaga et al. [12]’. In (d), the second sentence continues to describe the work of ‘Kaluzny (1986)’ which was cited in the previous sentence. In (e), a figure (not shown here) from ‘[14]’ and its caption are about ‘[14]’.

\(^1\) Citation markers mean in-text citations.