Scientometrics (2019) 121:1563–1582
https://doi.org/10.1007/s11192-019-03250-5

Characterizing human summarization strategies for text reuse and transformation in literature review writing

Kokil Jaidka[1] · **Christopher S. G. Khoo**[1] · **Jin-Cheon Na**[1]

Received: 15 April 2019 / Published online: 8 October 2019

Abstract

Citations are useful signals of information salience, but little research has identified the patterns of information selection, transformation, and organization that they espouse. This paper investigated the summarization strategies followed in the writing of literature review sections of information science research papers. We found that the summarization strategies followed are different for the two major styles of literature review writing, descriptive versus integrative literature reviews. Descriptive literature reviews, which focus on individual descriptions of research papers, are more likely to reference the Method and the Result sections of the cited paper and copy-paste text the referenced text. In contrast, integrative literature reviews, which synthesize the main ideas for many papers together, have more critiques and focus mainly on the Conclusion sections. These findings, based on a hand-annotated dataset, have the potential to scale up into a transformation-invariant neural architecture for scientific summarization that can generate different summaries of the input text with integrative or descriptive characteristics.

Keywords Literature review writing · Scientific summarization · Discourse analysis · Citance · Abstracting · Citation analysis

Mathematics Subject Classification 62H20

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11192-019-03250-5) contains supplementary material, which is available to authorized users.

✉ Kokil Jaidka
jaidka@ntu.edu.sg

Christopher S. G. Khoo
assgkhoo@ntu.edu.sg

Jin-Cheon Na
tjcna@ntu.edu.sg

[1] Wee Kim Wee School of Communication and Information, Nanyang Technological University, 31 Nanyang Link, Singapore 637718, Singapore

Introduction

The past few decades have witnessed an exponential growth in the scientific literature, making it a challenge for res earch scholars looking to stay abreast of recent trends and developments in their fields. One of the problems faced by researchers is that they are inundated with thousands of results and find themselves in the position of having to understand large amounts of technical material quickly. A second problem they encounter is that although digital libraries curate scientific papers, there are few tools available to summarize previous work and thus to help researchers write good literature reviews. The computational linguistics and summarization research communities have demonstrated that a variety of lexical, graphical, and knowledge-based approaches can be used to generate scientific summaries automatically. More recently, there is a growing interest in using trained neural networks for generating textual summaries of scientific papers (Rush et al. 2015; Nomoto 2016; Abura'ed et al. 2018). However, an open problem in automatic scientific summarization is that it is hard to emulate the specialized summary structure of a literature review that contextualizes a research problem through several explicit and implicit references to the general knowledge domain. In this paper, we conduct a content analysis of how humans summarize scientific literature, to offer insights and suggestions for the design of automatic summarization systems.

To our knowledge, no study has looked at the citation relationships to identify the linguistic strategies followed to summarize the major findings of previous work or provide an overview of research trends. In previous work, studies have instead viewed citations as clues to track the provenance of research ideas, or as single-sentence summaries of research papers that can be regarded as inputs for automatic summarization systems (Kan et al. 2002; Nanba 2000; Nanba and Okumura 1999, 2005; Jaidka et al. 2010). On the other hand, this study considers citations as an output of human summarization and identifies *where* and *how* humans select information to include into citing sentence. Our study sought to answer the following questions:

1. What are the characteristics of text reuse in integrative as compared to descriptive literature reviews?:
    a. What are the meta-differences in source and citance characteristics?
    b. Which sections of the source papers are reused the most?
    c. What kind of information is likely to be reused the most?
    d. How are the different kinds of information transformed?
    e. What are the types of edits performed for copy-pasting and paraphrasing?
2. Insights:
    a. How can the findings be applied to train contemporary neural network-based scientific summarization systems?
    b. What are the summarization strategies followed in creating an integrative vs. a descriptive literature review?

We analyzed the literature review sections of 30 articles in the domain of information science, sampled from the Journal of the American Society for Information Science and Technology, Journal of Information Science, and Journal of Documentation to answer these questions. The present study follows up on a preliminary analysis of 20 literature reviews in articles from the Journal of the American Society for Information Science and Technology (Jaidka et al. 2013a, b), and extends the analysis to a larger sample size from three different journals. Despite offering a content analysis on a small dataset, we show how our results are also relevant to the ongoing efforts for neural text generation systems.

In the following section, we provide more background about our study and the study of academic writing. This is followed by details on our research method, including data collection and text analysis method. We then present the results of our analysis, describing and comparing the summarization strategies followed in different styles of literature reviews. As a brief interlude, we describe how the findings are applicable to contemporary forays into neural text generation for summarization. We conclude with a discussion of the major implications, takeaways, and recommendations for future work.

Background

This study focuses on the citances in the literature review sections of research papers (hereafter referred to as "literature reviews"). Citances comprises the set of sentences in a citing paper which refer to a "source paper" (Mohammad et al. 2009; Jaidka et al. 2018). A citance can be mapped to a set of sentences in the source paper, which we refer to as 'source sentences.' A literature review uses citances to select and integrate information from various sources to present an assessment and critique of the state of the art. In doing so, citances help literature reviews to serve the following functions:

i. Provide an overview of a research problem common across many papers (Rowley and Slack 2004; Bourner 1996).
ii. Highlight the research contributions of previous papers (Jönsson 2006; Massey 1996; Bruce 1994).
iii. Identify relationships among papers by comparing and contrasting their approaches and contributions (Hart 1998).

This study is a part of a larger project for the multi-level discourse analysis of the literature review sections of research papers. A previous study reported on the characteristics of two distinct styles of literature review writing, namely the *integrative* and *descriptive literature reviews* (Khoo et al. 2011). Integrative literature reviews provide a high-level overview and critique of the recent work by comparing related papers against each other. Descriptive literature reviews provide more in-depth information about each mentioned study. The two styles of literature reviews also have different macro-level discourse structures (Khoo et al. 2011) and employ different rhetorical functions to frame their argument (Jaidka et al. 2010). All literature reviews comprise integrative and descriptive elements in different proportions.

Related work

A major contribution of our work lies in bringing the insights of text reuse and citation analysis approaches to a primarily data-driven automatic summarization community. Studies of the text reuse in citances have used automatic text matching on large scientific corpora, such as the arXiv database (Citron and Ginsparg 2015) and the Microsoft Academic Search dataset

(Singh et al. 2017). They have been supplemented with an empirical investigation of research fields, author influence, countries of origin, and article impact to obtain insights into the prevalent practices in different communities. Their findings suggest that text reuse is more prevalent in theoretical rather than applied fields. In studies of plagiarism, text reuse was found to reduce the impact of the plagiarizing paper in terms of the number of subsequent citations. The paper by Buchanan and McKay (2017) offered a characterization of text reuse behavior at the high level, in terms of the kinds of legitimate text reuse (use of definition, reuse of literature review, reuse of method description etc.) and the levels of text reuse in different sections of a research paper for a small digital libraries corpus. Previous studies of citations have applied rhetorical analysis to identify the different purposes of citing a paper, leading to classifications schemes which annotate their purpose and polarity (Chubin and Moitra 1975; Teufel et al. 1999). These studies consider questions involving *why* scholars make citations. For instance, the study by Chubin and Moitra (1975) explored citations in applied physics papers, and identified that citations fall in one of three broad categories: *affirmative*—to support or agree with a cited study; *negational*—to disprove or disagree with a cited study, and *critical*—to criticize and raise questions about previous methods and results.

In the context of previous work on text reuse, our analysis focuses instead on *how* and *where* citations select, transform, and re-purpose information from the source paper. It is thus similar to the study by Jing and McKeown (1999) who identified the kinds of transformations on source sentences before including them into a news summary. It is different from the study by Teufel (1999, pp. 209) which identified the cue phrases and other textual markers underlying 17 specific relationships between citing sentences and source papers because our focus is not on such cue phrases but on the information which they frame.

Within the summarization community, a body of recent work (Jha et al. 2017; Silva et al. 2016; Tandon and Jain 2012) and a scientific summarization Shared Task (Jaidka et al. 2018) have exemplified the use of citation text for scientific summarization. Citations are a useful signal for summarization because they are more "self-cohesive," and comprising additional information that does not appear in abstracts (Elkiss et al. 2008). Citations are also useful to determine the content and search relevance of scientific articles (Bradshaw 2003).

Citances are essentially short human summaries, and thus have higher per-word salience as compared to abstracts (Mohammad et al. 2009). For this reason, they have been used as input to multi-document summarization systems in some recent studies. Nanba (2000) used automatically categorized citances in a tool for survey generation. The studies by Qazvinian and Radev (2010) used citation relationships in a clustering approach to generate a summary of a research paper. Qazvinian et al. (2010) also demonstrated that the key phrases used in citances could be mined to produce single- or multi-document summaries. Beyond the use of the lexical features and citation networks, the work by Mei and Zhai (2008) has also explored the use of discourse features in citances to summarize the impact of a research paper. In contrast to these studies, we propose that citances are useful to consider as an *output* of human summarization and to reverse engineer the set of strategies which go into synthesizing them.

Research method

Construction of the corpus

The dataset used for analysis comprised one that has earlier been used to study the rhetorical structure and layout of literature review sections. It was constructed by downloading

90 research articles from 8 volumes (2001-2008) of JASIST, Journal of Documentation (JDoc) and Journal of Information Science (JIS) by using a random number generator to choose a volume and issue number. The following exclusion criteria were applied to filter articles during sampling, ostensibly because they are likely to follow a different style and format:

- Papers published in special issues.
- Survey articles or literature review articles.
- Case studies, perspective papers, and conceptual papers which did not have a separate literature review section.

Finally, 30 literature review sections (henceforth referred to as 'literature reviews') comprising 10 literature reviews each from the three information science journals were sampled from this set of 90. All the sentences with citances were extracted. These comprised.

- Sentences which cite the study's authors and describe something about their study (e.g., "The paper by Buchanan and McKay (2017) offered a characterization of text reuse behavior at the high level").
- Sentences which cite a paper in a high-level characterization (e.g., "Studies in text reuse have used automatic text matching on large scientific corpora, such as the arXiv database (Citron and Ginsparg 2015)").
- Sentences adjacent to the sentence with the in-text citation which also refer to the source paper.

We extracted a total of 727 citances. In the next step, we downloaded the 472 source papers corresponding to each of these citances from bibliographic and full-text databases and the Web. Of these, 20 research papers were unavailable, and 57 citances did not correspond to research papers. Finally, we conducted our analysis on the remaining 395 source papers. For each literature review, the number of sources analyzed range from 3 to 24, with a median value of 10. Three-quarters of the literature reviews had more than ten source papers analyzed.

Identifying the Style of the Literature Review style

We explored whether text reuse behavior had any relationship with the overall style of the literature review. In our previous work, we identified that some literature reviews are *descriptive* in style and summarize previous studies without much critical comment (Khoo et al. 2011). Knott (1999) characterized descriptive literature reviews as similar to annotated bibliographies, which briefly summarize the research questions of a research study, its major methods of investigation and its main conclusions. On the other hand, other literature reviews are *integrative* in style: they synthesize ideas and research results from different studies into an integrated whole or to support a theory. Torraco (2005, pp. 1) characterized an integrative literature review as "a form of research that reviews, critiques and synthesizes representative literature on a topic in an integrated way" to generate new frameworks and perspectives on the topic. We anticipate that descriptive literature reviews would be more likely to reuse text because they provide more detailed information about each study.

We hired two graduate students in Information Science and familiar with discourse analysis to categorize each of the literature reviews as "integrative", "descriptive" or "mixed". The students were postgraduates with some knowledge and training in discourse analysis. They were also familiar both with reading and with writing academic discourse. In order to avoid biasing them with our expectations, they were only given broad definitions of integrative and descriptive literature reviews to guide them in their classification (Khoo et al. 2011):

- Descriptive literature reviews summarize individual papers/studies and provide details of the research methods and results of cited studies.
- Integrative literature reviews focus on high-level discussions—they extract results and ideas from the papers and synthesize a summary of the research trends and milestones.

The two coders were unanimous on the categorizations of the 30 literature reviews and annotated 16 of them as descriptive and 14 of them as integrative. The classification task was repeated by two of the authors of this paper who were also unanimous in their categorization. In the pilot study reported in Khoo et al. (2011), we had observed that 3 of the set of 20 literature reviews comprised both integrative and descriptive elements and were categorized as "mixed". No such ambiguity was observed in the present dataset.

Text analysis method

For each citance, we conducted a manual inspection of the cited source paper to identify the sentences in the source paper (source sentences) which were the closest match to it. Then we compared the differences between the source sentences and the citance in a multi-level analysis:

- Types of edits performed, i.e., specific editing steps followed in text transformation: sentence reduction, sentence combination, syntactic transformation, generalization/specification, sentence reordering, and insertion.
- Type of transformation performed: copy-paste, paraphrase, high-level summary, or critical reference.
- Type of information: research objective, method, result, or critique.
- Location of the source sentence: abstract, introduction, related work, method, result, conclusion, other, or unknown.

The following sections detail the text analysis steps with illustrative examples.

Types of Transformation The purpose of this part of the analysis was to identify how source sentences were transformed into citances. We noted four main types of transformations performed on the source sentences:

- *Copy-paste*—where there are only minor differences between the source sentence and the citance. The transformation involves simple changes in tense, part of speech, substituting words with their synonyms or rearranging sentences. Clauses, rhetorical devices or adverbs may have been added to or removed from the source sentences. For example,

Citance: "In a series of experiments on designing various interfaces to the Okapi search engine, it was found that *both implicit and explicit use of a thesaurus dur-*

ing automatic and interactive query expansion was beneficial."
Source sentence: "Our experimental results demonstrate that *both implicit and explicit use of a thesaurus can be beneficial* for query reformulation."

- *Paraphrase*—where the source sentences are transformed through more complex changes, where there is no 1-to-1 correspondence between words but the gist of sentences remains the same. For example,

Citance: "Kelly and Cool [27] suggest that the more *familiar* a person is with a search *topic*, the less *time they spend reading*."
Source sentence: "Our results indicate that as one's *familiarity* with a *topic* increases, one's searching efficacy increases and one's *reading time decreases*."

- *High-level summary*—where the source sentence is semantically transformed (i.e., the semantic content is different) by summarization or generalization to provide a higher-level gist of its information. These transformations involve significant modifications to provide information that may span many sentences from different locations in the source text. For example,

Citance: "Their study concluded that questions with higher complexity had an increased probability of retrieving more relevant and precise results."
Source sentence: "To generalize: higher CE scores bring higher chances or odds for low precision or lower chances for high precision."

- *Critical reference*—when the source information is rhetorically and semantically transformed into a critical argument. Often, it provides information that is not mentioned in the source paper. For example,

Citance: "However, they did not investigate the number of queries made per session."

In some cases, the citance was very similar to more than one source sentence. In this scenario, the sentence requiring the least number of transformations was selected as the source sentence. For example, simple edits such as deletion of a word (copy-pasting transformations) were preferred over paraphrasing transformations.

Types of Information Selected At the first level of analysis, we annotated each source sentence according to the type (or types) of research information they provide:

- Research objective—referencing the purpose of the cited study (e.g., "Van Dijk (1979) conducted a study of how humans summarize information").
- Research method—referencing the procedure followed in the cited study (e.g., "Guo and Li (2007) use domain-specific or general cue phrases and leading phrases to identify important content.").
- Research result—reporting a finding or conclusion of the cited study (e.g., "A study by Liu et al. (2008) contended that short sentences could not carry enough information to be a representative summary.").

- Critique—providing the author's critique of the cited study (e.g., "Their approach raises questions concerning the replicability of their experiments and the generalizability of their findings.").

Sometimes, source sentences provided two kinds of information together—for instance, the research objective may be mentioned in combination with the research result or the research method. In these cases, the source sentence was "divided" into two text segments and analyzed separately.

Location of Source Sentences and Source Sections In the next part of the analysis, we annotated every citance according to the location in the source paper, which it was referencing. For example:

> Citance: "Kelly and Cool [27] suggest that the more familiar a person is with a search topic, the less time they spend reading."
> Source sentence: "Our results indicate that as one's familiarity with a topic increases, one's searching efficacy increases and one's reading time decreases."

This citance was annotated as "Results" because the reference text was found in the Results section. Citances were annotated with "Other" if they referred to non-typical source text such as headings, captions, titles, and tables. They were annotated as "Unknown" if the information could not be found in the source, or if it were so general that it could not be pinpointed to a single source sentence (e.g., "Studies of summarization (Van Dijk 1979)…").

A one-to-many mapping relationship exists between a citance and their sources. Sometimes, the same information occurred in more than one location in the source paper. In this case, the citance was annotated according to the first occurrence of the information after the Abstract. In case a single citance referred to source sentences in more than one source section, it was "divided" into two parts, and different segments were annotated with different sources.

Results

Type of edits

Table 1 characterizes the types of edits performed for copy-paste, paraphrase, and high-level summary in terms of the categories described by Jing and McKeown (1999). While sentence reduction and sentence combination help to compress information into a concise summary, we also observed evidence of syntactic transformation, sentence reordering, generalization and specification, and insertions. We anticipate that changing the tense and voice of source sentences may have been done to conform to the academic writing genre. Rearrangement and paraphrasing may have been done either to highlight certain aspects of the information or to make it more readable; trivial word-substitution would avoid plagiarism. Changing singular nouns into plurals would make a claim or description more general.

Clauses are added to contextualize the cited information into the literature review, or to elaborate with additional details, and numerical values.

Table 1 The types of edits performed for copy-paste, paraphrase and high-level summaries

Sentence reduction
Deletion of modals, e.g., might
Deletion of introductory phrases, e.g., It was observed that
Deletion of affective words or qualitative judgments, e.g., surprising
Deletion of independent second clauses and dependent clauses
Sentence combination
Conjunctions between related sentences
Syntactic transformation
Change of tense to past tense
Change of voice to passive voice in the case of research results
Change of voice to active voice in the case of research objectives and methods
Substitution of nouns with synonyms
Substitution of adjectives with dependent clauses, e.g., *two-part study* changed to *study in two parts*
Generalization/specification
Substitution of singular nouns with plural nouns
Substitution of anaphoric references
Abstract nouns, or nouns referring to processes, are changed to verbs, e.g., *document retrieval* changed to *retrieving documents*
Substitution of personal pronouns with authors' names
Reordering
Rearrangement of clauses
Substitution of personal pronouns with authors' names
Insertion
Insertion of introductory clauses
for example, <authors>
<authors> <verb>
More recently, <authors>
In these schemes,
Based on their surveys,
They found that,
Results were compared using
The study found that
Some authors have demonstrated
Insertion of elaborating clauses, e.g., For this study, conducted in Australia, …

Characteristics of text reuse

Table 2 presents the meta-differences between integrative and descriptive literature reviews in terms of the length of the citances, the number of sources they cite, and the average edit distance between the source information and the citance. We calculate the edit distance as the absolute difference in the number of characters between the citance and the source text over all the citance-source pairs. We observe that integrative literature reviews cite more sources on average and have longer citances on average, with lesser variance as compared to citances in descriptive literature reviews. They

Table 2 Profile of integrative and descriptive literature reviews: meta-differences

	Type of Literature review		
	Integrative	Descriptive	Overall
Number of sources (Mean)	15.6	11.3	13.4
Citance length (Mean)	179.6	171.4	175.5
SD	98.8	104.9	
Source information length (Mean)	252.8	207.7	230.2
SD	158.1	156.5	
Edit distance (Mean)	−80.6	−50.5	−65.5
SD	155.9	143.2	

Profile of integrative and descriptive literature reviews: type of information cited

Type of information		Type of literature review		
		Integrative	Descriptive	Overall
Research objective	Count	32	103	135
	Expected count	34.9	100	
	% of column	18.7%	21.0%	20.4%
Research	Count	52	141	193
Method	Expected count	49	143	
	% of column	30.4%	28.8%	29.2%
Research	Count	73	225	298
Result	Expected count	77.1	214	
	% of column	42.7%	45.9%	45.1%
Critique	Count	14	21	35
	Expected count	9.1	25.9	
	% of column	8.2%	4.3%	5.3%

Pearson Chi Square *p* value 0.22

summarize information from longer source sentences on average, with a greater average edit distance than descriptive literature reviews.

Keeping these basic differences in mind, Tables 3, 4, 5, 6, and 7 delve more deeply into the profile of integrative and descriptive literature reviews through cross-tabulation analyses. In terms of the types of information cited against the style of the literature review, we found that critiques are twice as much in integrative literature reviews than descriptive literature reviews, but this relationship was not statistically significant. In these tables, bold values denote significance at $p < 0.05$.

Authors synthesize summaries differently, depending on the style of literature review they are writing. Table 3 gives the cross-tabulation between the type of transformation to the source sentences and the style of the literature review. Overall, the most common kind of transformation is the high-level summary, accounting for nearly half (48%) of the citances. Comparing the two styles of literature review, we found that descriptive literature reviews have a much higher proportion of copy-paste and paraphrase transformations than for integrative literature reviews. Integrative literature reviews have a higher proportion of high-level summary and critical reference transformations. These findings are statistically significant ($p < 0.001$).

Table 3 Profile of integrative and descriptive literature reviews: type of transformation

Type of transformation		Type of literature review		
		Integrative	Descriptive	Overall
Copy-paste	Count	19	**149**	**168**
	Expected Count	43.5	**124.5**	
	% of column	11.1%	**30.4%**	**25.4%**
Paraphrase	Count	27	**113**	**140**
	Expected Count	36.2	**103.8**	
	% of column	15.8%	**23.1%**	**21.2%**
High-level Summary	Count	**110**	207	317
	Expected Count	**82.0**	235.0	
	% of column	**64.3%**	42.2%	47.9%
Critical reference	Count	**15**	21	36
	Expected Count	**9.3**	26.7	
	% of column	**8.8%**	4.3%	5.5%

Pearson Chi Square *p* value 1×10^{-8}

Table 4 Profile of integrative and descriptive literature reviews: source section referenced

Source section		Type of literature review		
		Integrative	Descriptive	Overall
Abstract section	Count	39	**152**	**191**
	Expected count	49.3	**141.6**	
	% of column	22.8%	**31.0%**	**29.6%**
Introduction section	Count	24	66	90
	Expected count	23.3	66.7	
	% of column	14.0%	13.5%	13.9%
Method section	Count	19	**69**	**88**
	Expected count	22.8	**65.2**	
	% of column	11.1%	**14.1%**	**13.7%**
Result section	Count	18	**79**	**97**
	Expected count	25.1	**71.9**	
	% of column	10.5%	**16.1%**	**15.2%**
Unknown	Count	**51**	78	129
	Expected count	**33.4**	95.6	
	% of column	**29.8%**	15.9%	20.0%
Conclusion section	Count	13	36	49
	Expected count	12.7	36.3	
	% of column	7.6%	7.3%	7.6%

Pearson Chi Square *p* value 0.002

Authors prioritize different kinds of information in descriptive vs. integrative literature reviews. Table 4 presents a cross-tabulation between source section selected and the style of the literature review. Descriptive literature reviews provide more details about individual

Table 5 Cross-tabulation between source section and type of information

Type of source section		Type of information			
		Research objective	Research method	Research result	Overall
Abstract section	Count	**80**	42	68	190
	Expected count	**40.9**	54.2	94.8	
	% of column	**60.6%**	22.4%	23.6%	37.1%
Introduction section	Count	**24**	27	38	89
	Expected count	**19.1**	25.1	43.9	
	% of column	**18.1%**	14.4%	13.1%	17.4%
Method section	Count	3	**71**	14	88
	Expected count	18.9	**25.1**	43.9	
	% of column	3.3%	**37.9%**	4.8%	17.2%
Result section	Count	0	2	**94**	**96**
	Expected count	20.6	27.4	**47.9**	
	% of column	0.0%	1.1%	**32.6%**	**18.7%**
Conclusion section	Count	3	4	**41**	**48**
	Expected count	10.3	13.7	**23.9**	
	% of column	2.2%	2.1%	**14.2%**	**9.4%**

Pearson Chi square p value 8.8×10^{-57}

Table 6 Cross-tabulation between type of information and type of transformation

Type of information		Type of transformation			
		Copy-paste	Paraphrase	High-level summary	Overall
Research objective	Count	30	**41**	63	134
	Expected count	36.1	**29.9**	68/0	
	% within row	22.4%	**30.6%**	47.0%	21.5%
Research method	Count	45	31	**116**	**192**
	Expected count	51.8	42.8	**97.4**	
	% within row	23.4%	16.1%	**60.4%**	**30.8%**
Research result	Count	**93**	67	137	297
	Expected count	**80.1**	66.3	150.6	
	% within row	**31.3%**	22.6%	46.1%	47.7%

Pearson Chi Square p value 0.002

studies' experiments and findings than integrative literature reviews ($p < 0.01$). On the other hand, 30% of the references in an integrative literature review cannot be mapped to a particular source sentence. Such citances often represent a high-level summary of one source paper or multiple source papers to support a particular argument (Hart 1998).

Authors select different kinds of information from different places in the cited paper. Table 5 shows the cross-tabulation between the type of information and the source section. The relation was significant at the 0.001 level. Research objective information was very often taken from the Abstract Sect. (61% of the time), and sometimes from the Introduction

Table 7 Cross-tabulation between type of transformation and style of literature review, for the different source sections

Source section	Type of transformation		Style of literature review		
			Integrative	Descriptive	Overall
Abstract section	Copy-paste	Count	6	**54**	**60**
		Expected count	12.3	**47.7**	
		% of row	15.4%	**35.5%**	**31.4%**
	Paraphrase	Count	17	63	80
		Expected count	16.3	63.7	
		% of row	43.6%	41.4%	41.9%
	Summary	Count	**16**	35	51
		Expected count	**10.4**	40.6	
		% of row	**41.0%**	23.0%	26.7%
Pearson Chi square p value 0.02					
Introduction	Copy-paste	Count	9	18	27
		Expected count	7.2	19.8	
		% of row	37.5%	27.3%	30.3%
	Paraphrase	Count	3	13	16
		Expected count	4.3	11.7	
		% of row	12.5%	19.7%	17.9%
	Summary	Count	12	34	46
		Expected count	12.3	33.7	
		% of row	50.0%	51.5%	51.7%
Pearson Chi square p value 0.57					
Conclusion	Copy-paste	Count	1	**14**	**15**
		Expected count	3.8	**11.2**	
		% of row	8.4%	**38.9%**	**31.2%**
	Paraphrase	Count	0	**12**	**12**
		Expected count	3	**9**	
		% of row	0%	**33.3%**	**25.0%**
	Summary	Count	**11**	10	21
		Expected count	**5.2**	15.8	
		% of row	**91.6%**	27.8%	43.7%
Pearson Chi square p value 0.0005					
Method section	Copy-paste	Count	0	**31**	**31**
		Expected count	6.7	**24.3**	
		% of row	0%	**44.9%**	**35.2%**
	Paraphrase	Count	4	6	10
		Expected count	2.2	7.8	
		% of row	21.1%	8.7%	11.4%
	Summary	Count	**15**	32	47
		Expected count	**10.1**	36.9	
		% of row	**78.9%**	36.4%	53.4%

Table 7 (continued)

Source section	Type of transformation		Style of literature review		
			Integrative	Descriptive	Overall
Pearson Chi square p value 0.001					
Abstract section	Copy-paste	Count	2	**27**	**29**
		Expected count	5.1	**23.8**	
		% of row	11.8%	**34.2%**	**30.2%**
	Paraphrase	Count	1	**16**	17
		Expected count	3.0	**13.9**	
		% of row	5.8%	**20.3%**	17.7%
	Summary	Count	**14**	36	50
		Expected count	**8.8**	41.1	
		% of row	**82.4%**	45.6%	52.1%
Pearson Chi square p value 0.02					

Sect. (18%). As expected, research methods information was often taken from the Method Sect. (38%), and research result information from the Result Sect. (33%) and Conclusion Sect. (14%). The Abstract is a major source of referenced information for research method (22%) and research result (24%).

Authors transform information differently depending on where they select it. Table 6 presents a cross-tabulation between types of information and types of transformation. The relation is significant at the 0.001 level. Overall the research objective and research result are summarized at a high-level about half the time. In 60% of the cases, research method information is provided as a high-level summary. The research result is copy-pasted more often than expected (31%), and the research objective is paraphrased more often than expected (31%).

Authors can summarize information from different places in the cited papers, to suit the style of literature review they are writing. Table 7 compares the types of transformations for different source sections in either style of the literature review. It reinforces the finding that integrative literature reviews apply summary transformations more often than expected, on all source sections except the Introduction. In contrast, descriptive literature reviews have more copy-pasting than expected. Interesting patterns include the following:

- Information from the Abstract section is paraphrased about 42% of the time for both styles of literature review.
- Information from the Introduction section is summarized at a high-level about 50% of the time for both styles of the literature review.
- Information from the Result section is generally summarized at a high-level (82% of the time for integrative literature reviews, and 46% of the time for descriptive literature reviews).
- Information from the Conclusion section is almost always summarized at a high-level in integrative literature reviews (92% of the time).

Applications to Neural Text Generation

We anticipate that the findings of our content analysis can improve the design thinking and hence the quality and variety of synthesized summaries, by proposing the concept that *style* and *content* can interplay in different ways. In this section, we demonstrate an application of our findings to contemporary scientific summarization systems. Neural architectures are designed to generate one *style* of a summary. However, if we consider different styles as different domains, then existing domain (or in our case, style) transfer neural architectures can generate different kinds of summary sentences from the input text, with varying levels of editing such as cut-pasting, paraphrasing, and high-level summarizing (Table 1).

A challenge in fulfilling data requirements is typically the need for parallel corpora (e.g., for the same input text, the need to have three matched outputs which show cut-pasting, paraphrasing, and high-level summarizing). Recent work (Zhang et al. 2017) has attempted to overcome this limitation by allowing a single encoder and multiple decoders to generate summaries without the availability of a parallel corpus in what are known as adversarially regularized auto-encoder (ARAE) architectures (Zhao et al. 2017). Figure 1 illustrates such an architecture with an encoder ($enc\Phi$ in Fig. 1) and a decoder network (decψ in Fig. 1). Building upon this idea, we show that the idea of style transfer, too, can benefit from ARAE architectures. In the following paragraphs, we briefly describe the architecture implementation, which we plan to follow in future work for generating automatic scientific summaries.

Model architecture

We can assume our training data consists of (x, transform) pairs, where x is the sequence of the input text, and transform refers to the type of transformation that should be performed. The encoder accepts a token sequence of features of the input text, $x=(w_1, w_2, \ldots w_n)$ as input and maps it to a lower-dimensional latent space, $z=enc\Phi\ (x)$. The decoder aims to reconstruct the input x from this z. A generator network Pg maps a random variable sampled from $N(0,1)$ distribution to a latent space $\hat{z}$. The domain classifier (transformation classifier) is represented as Pc in Fig. 1 and will remove all the transformation-specific

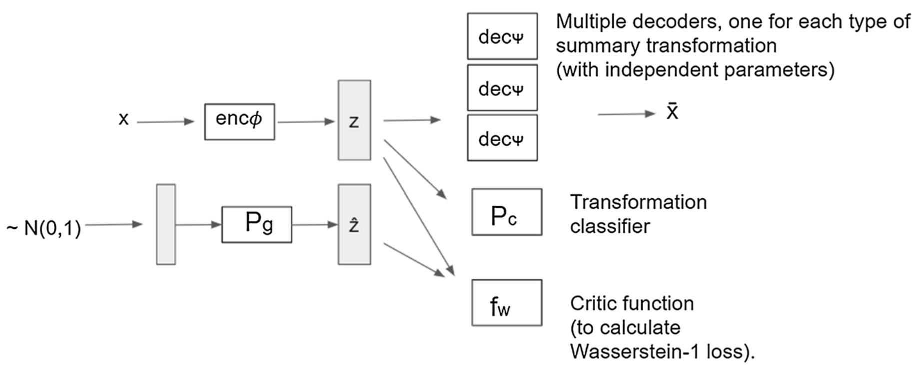

Fig. 1 The proposed neural model architecture for scientific summarization. It consists of the ARAE model for generation, with separate decoders used for each type of transformation industry. A transformation classifier network Pg is additionally trained to predict the type of transformation and ensures domain invariance in z through adversarial regularization

encodings from the latent representation. Finally, in order to minimize the Wasserstein-1 distance between the two latent vectors z and $\hat{z}$, adversarial training is performed on the network using the critic function f_w.

While an adapted ARAE is the proposed architecture we would plan to take, actually applying our findings to generate different styles of summaries is envisioned as future work and recommended as a possible direction for others in the community to explore.

Insights

This study has analyzed a sample of literature reviews in information science journal papers, focusing on the relationships between citances in the literature reviews and source sentences in the source papers. Specifically, the study analyzed the types of information referenced, sections of the source papers they are taken from, text transformations performed on the source sentences for incorporation in the literature review, and how these differ in descriptive versus integrative styles of literature reviews.

The overall findings can be summarized as follows:

1. Authors of literature review sections in information science papers synthesize summaries around research results (45%) more often than around their research objectives (20%) or methods (29%).
2. While authors summarize information from cited papers into high-level summaries more than half the time, they are more likely to copy-paste research results information (31%) and paraphrase research objectives information (31%).
3. Authors select information from the Abstract of the cited paper about one-third of the time, and when they do so, they are likely to paraphrase the text that they choose (42%). On the other hand, if they choose information from the Introduction or the Result section, they are likely to synthesize it into a high-level summary about half the time.

Authors who are writing descriptive reviews present experimental detail about previous studies, such as the approach followed, their results, and evaluation. The focus is on providing important details of previous studies in a concise form. Their writing is also found to have the following characteristics:

- They have a higher proportion of copy-paste and paraphrase transformations.
- They reference the Abstract, Method, and Result sections of the source papers more often than integrative literature reviews.

In contrast, authors writing integrative literature reviews present information from several studies in a condensed form as a critical summary, possibly complemented with a comparison, evaluation, or comment on the research gap. The focus is on highlighting relationships amongst concepts or comparing studies against each other. They are also found to have the following characteristics:

- They have twice the proportion of critiques as in descriptive literature reviews. 8% of citances contain critiques, whereas the proportion is 4% for descriptive reviews.
- They have a higher proportion of high-level summary and critical reference transformations.
- The Conclusion section is summarized at a high-level in integrative literature reviews.

- 30% of the references in an integrative literature review cannot be mapped to particular source sentences.

Our findings require us to revisit the definition of a literature review section, and question how many of its objectives are actually met in the course of our analysis. Did we find that humans writing literature reviews are successfully able to collect, synthesize and integrate the key information from different source papers? In fact, we found that the style of literature review reflects a subset of objectives which the author may have prioritized in their writing. The choice of style may also depend on the type of research studies the author is reviewing. For experimental research, it may be more useful to construct a descriptive literature review which provides relevant details of previous studies. On the other hand, a theory-based paper may require more argumentation and interpretation in an integrative literature review. When the author wants to draw a narrower comparison, then the studies that are similar to the author's work or provide a foundation for the author's work may be described in greater detail. On the other hand, peripheral or distantly related studies may only be synthesized as a high-level summary. We also observed that integrative and descriptive literature reviews are not always poles apart. In fact, we do expect that most literature reviews could have both descriptive and integrative elements. In previous work, we found that three out of a set of twenty literature reviews taken from JASIST were "mixed" with both integrative and descriptive elements (Khoo et al. 2011).

Conclusion and Future Work

This study is a part of a larger project on multi-document scientific summarization. Our findings suggest that an author begins a literature review with an overall strategy in mind, which conforms to either a descriptive or integrative style. They consequently choose the discourse structure and rhetorical arguments to implement the selected style. The contribution of the current study is to show that the author also selects and edits the information content based on the style of the literature review. These ideas, obtained from a fine-grained content analysis of information science research articles, can be synthesized in the design specifications of the next generation of neural text summarization systems. The results complement our analysis of the macro-level document structure and rhetorical functions found in the two styles of literature reviews. Descriptive literature reviews have a significantly greater number of the method, result and interpretation elements, in which writers prefer rhetorical arguments which build a description of cited studies, providing information on their research methods, results and interpretation. Integrative literature reviews have fewer study elements and instead have significantly more meta-summary and meta-critique elements, wherein writers employ rhetorical arguments which help to build a critical summary of topics and provide examples illustrating the author's argument. A challenge we faced was to categorize the processes involved in producing the high-level summaries or critiques of source papers. We recommend that future scholars could explore the semantic, cognitive, and linguistic processes invoked by authors in writing such summaries.

Although the data for this study ended in 2008, nevertheless, we would expect them to be applicable to understand the writing of contemporary literature review sections in information science papers. Furthermore, we were limited in our ability to extend the dataset due to the considerable human effort involved. The challenge, therefore, lies in harvesting the large volumes of scientific literature with appropriate fine-grained annotations; for instance, the ACL Anthology Network (Yasunaga et al. 2019) does automatically identify the list of citing papers for each source paper; however, it does not identify the source and cited text, or the transformation invoked.

Beyond automatic multi-document summarization, we also expect our findings to be useful in the evidence-based teaching of report-writing to undergraduate students. Previous studies have used Toulmin's (2003) Argument Pattern to develop qualitative and quantitative measures of argumentation in classroom discussions. Inspired by this application, we propose that a similar measure of literature review quality can be used to score student assignments and as a teaching rubric to junior research scholars.

Applying our findings to neural text generation in this manner would mean integrating human summarization strategies into automatic text generation functions, which is an exciting prospect. However, a challenge remains to scale up the fine-grained content analysis in automatic ways, so that there is enough data to train our neural architecture. We invite other scholars to devise ways to scale up corpus creation so that with larger datasets, we can innovate in approaches for scientific summarization, and earn a place on the frontier of neural text generation.

References

Abura'ed, A., Bravo, A., Chiruzzo, L., & Saggion, H. (2018). LaSTUS/TALN + INCO@ CL-SciSumm 2018-using regression and convolutions for cross-document semantic linking and summarization of scholarly literature. In *Proceedings of the 3nd joint workshop on bibliometric-enhanced information retrieval and natural language processing for digital libraries* (*BIRNDL2018)*. Ann Arbor, Michigan (July 2018).

Bourner, T. (1996). The research process: Four steps to success. *Research methods: guidance for postgraduates, Arnold, London*, pp. 7–11.

Bradshaw, S. (2003). Reference directed indexing: Redeeming relevance for subject search in citation indexes. In *International conference on theory and practice of digital libraries* (pp. 499–510). Springer, Berlin, Heidelberg.

Bruce, C. S. (1994). Research students' early experiences of the dissertation literature review. *Studies in Higher Education, 19*(2), 217–229.

Buchanan, G., & McKay, D. (2017). The lowest form of flattery: characterising text re-use and plagiarism patterns in a digital library corpus. In *Proceedings of the ACM/IEEE joint conference on digital libraries* (pp. 1–10). IEEE.

Chubin, D. E., & Moitra, S. D. (1975). Content analysis of references: Adjunct or alternative to citation counting? *Social Studies of Science, 5*(4), 423–441.

Citron, D. T., & Ginsparg, P. (2015). Patterns of text reuse in a scientific corpus. *Proceedings of the National Academy of Sciences, 112*(1), 25–30.

Dijk, T. A. (1979). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. New York: L. Erlbaum Associates.

Elkiss, A., Shen, S., Fader, A., Erkan, G., States, D., & Radev, D. (2008). Blind men and elephants: What do citation summaries tell us about a research article? *Journal of the American Society for Information Science and Technology, 59*(1), 51–62.

Guo, Q., & Li, C. (2007). The research on the application of text clustering and natural language understanding in automatic abstracting. In *Fourth international conference on fuzzy systems and knowledge discovery, 2007. FSKD 2007*. (vol. 4, pp. 92–96). IEEE.

Hart, C. (1998). *Doing a literature review*. London: Sage.

Jaidka, K., Chandrasekaran, M. K., Rustagi, S., & Kan, M. Y. (2018). Insights from CL-SciSumm 2016: The faceted scientific document summarization shared task. *International Journal on Digital Libraries, 19*(2–3), 163–171.

Jaidka, K., Khoo, C., & Na, J. C. (2010). Imitating human literature review writing: an approach to multi-document summarization. In *Proceedings of the international conference on asian digital libraries* (pp. 116–119). Springer, Berlin, Heidelberg.

Jaidka, K., Khoo, C., & Na, J. C. (2013a). Deconstructing human literature reviews–a framework for multi-document summarization. In *proceedings of the 14th European workshop on natural language generation* (pp. 125–135).

Jaidka, K., Khoo, C. S. G., & Na, J. C. (2013b). Literature review writing: How information is selected and transformed. *Aslib Proceedings, 65*(3), 303–325.

Jha, R., Jbara, A. A., Qazvinian, V., & Radev, D. R. (2017). NLP-driven citation analysis for scientometrics. *Natural Language Engineering, 23*(1), 93–130.

Jing, H., & McKeown, K. R. (1999). The decomposition of human-written summary sentences. In *Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval* (pp. 129–136). ACM.

Jönsson, S. (2006). On academic writing. *European Business Review, 18*(6), 479–490.

Kan, M. Y., Klavans, J. L., & McKeown, K. R. (2002). Using the annotated bibliography as a resource for indicative summarization. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*.

Khoo, C. S., Na, J. C., & Jaidka, K. (2011). Analysis of the macro-level discourse structure of literature reviews. *Online Information Review, 35*(2), 255–271.

Knott, D. (1999). *Writing an annotated bibliography*. Retrieved January 2009. http://www.writing.utoronto.ca/advice/specific-types-of-writing/annotated-bibliography.

Liu, Y., Wang, X., Zhang, J., & Xu, H. (2008). Personalized PageRank based multi-document summarization. In *IEEE international workshop on semantic computing and systems, 2008. WSCS'08.* (pp. 169–173). IEEE.

Massey, A. (1996). *Using the literature: 3×4 analogies*. *The Qualitative Report*, *2*(4). Retrieved from January 2009. http://www.nova.edu/ssss/QR/QR2-4/massey.html.

Mei, Q., & Zhai, C. (2008). Generating impact-based summaries for scientific literature. In *Proceedings of the ACL conference on human language technologies* (pp. 816–824). Association for Computational Linguistics.

Mohammad, S., Dorr, B., Egan, M., Hassan, A., Muthukrishan, P., Qazvinian, V., & Zajic, D. (2009). Using citations to generate surveys of scientific paradigms. In *Proceedings of human language technologies: The 2009 annual conference of the north american chapter of the association for computational linguistics* (pp. 584–592). Association for Computational Linguistics.

Nanba, H. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. In *Proceedings of the American Society for Information Science (ASIS)/the 11th SIG classification research workshop, classification for user support and learning, Chicago, USA, 2000* (pp. 117–134). Morgan Kaufmann Publishers.

Nanba, H., & Okumura, M. (1999). Towards multi-paper summarization reference information. In *Proceedings of the 16th international joint conference on Artificial intelligence-Volume 2* (pp. 926–931). Morgan Kaufmann Publishers Inc.

Nanba, H., & Okumura, M. (2005). Automatic detection of survey articles. In International Conference on Theory and Practice of Digital Libraries (pp. 391–401). Springer, Berlin, Heidelberg.

Nomoto, T. (2016). NEAL: A neurally enhanced approach to linking citation and reference. In *Proceedings of the joint workshop on bibliometric-enhanced information retrieval and natural language processing for digital libraries* (BIRNDL) (pp. 168–174).

Qazvinian, V., & Radev, D. R. (2010). Identifying non-explicit citing sentences for citation-based summarization. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 555–564). Association for Computational Linguistics.

Qazvinian, V., Radev, D. R., & Özgür, A. (2010). Citation summarization through keyphrase extraction. In *Proceedings of the 23rd international conference on computational linguistics* (pp. 895–903). Association for Computational Linguistics.

Rowley, J., & Slack, F. (2004). Conducting a literature review. *Management research news, 27*(6), 31–39.

Rush, A. M., Chopra, S., & Weston, J. (2015). A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 379–389).

Silva, F. N., Amancio, D. R., Bardosova, M., Costa, L. D. F., & Oliveira, O. N., Jr. (2016). Using network science and text analytics to produce surveys in a scientific topic. *Journal of Informetrics, 10*(2), 487–502.

Singh, M., Niranjan, A., Gupta, D., Bakshi, N. A., Mukherjee, A., & Goyal, P. (2017). Citation sentence reuse behavior of scientists: A case study on massive bibliographic text dataset of computer science. In *Proceedings of the ACM/IEEE joint conference on digital libraries (JCDL)* (pp. 1–4). IEEE.

Tandon, N., & Jain, A. (2012). Citation context sentiment analysis for structured summarization of research papers. In *35th German conference on artificial intelligence* (p. 98).

Teufel, S. (1999). Argumentative Zoning: Information Extraction from scientific text. *Ph.D. Thesis, University of Edinburgh*.

Teufel, S., Carletta, J., & Moens, M. (1999). An annotation scheme for discourse-level argumentation in research articles. In *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics* (pp. 110–117). Association for Computational Linguistics.

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review, 4*(3), 356–367.

Toulmin, S. E. (2003). *The uses of argument*. Cambridge: Cambridge University Press.

Yasunaga, M., Kasai, J., Zhang, R., Dan, A. R. F. I. L., & Radev, F. D. R. (2019). ScisummNet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks. In *Proceedings of the AAAI annual meeting*.

Zhang, Y., Barzilay, R., & Jaakkola, T. (2017). *Aspect-augmented adversarial networks for domain adaptation*. arXiv preprint arXiv:1701.00188.

Zhao, J. J., Kim, Y., Zhang, K., Rush, A. M., & LeCun, Y. (2017). *Adversarially regularized autoencoders for generating discrete structures*. CoRR, abs/1706.04223.