

Enhancing Domain Specific Sentiment Lexicon for Issue Identification

Madhusudanan N^(✉), B. Gurumoorthy, and Amaresh Chakrabarti

Virtual Reality Lab, Centre for Product Design and Manufacturing,
Indian Institute of Science, Bangalore, India
{madhu, bgm, acl23}@cpdm.iisc.ernet.in

Abstract. The research work reported here is part of a larger project aimed at acquiring knowledge about issues in assembly, from documents. In order to do so, the first step is to identify the presence of issues. For this, sentiment analysis is proposed as a means. The presence of issues is proposed to be found by detecting negative sentiment. However, general English sentiment lexicons are not enough to detect negative sentiment in specialized domains, in this case, aircraft assembly. This paper studies an existing sentiment analysis tool, and creates an enhanced sentiment lexicon on the basis of the performance of the tool. The enhanced domain lexicon is proposed as the solution to identifying domain specific issues.

Keywords: Natural language text · Issues in assembly · Sentiment analysis

1 Introduction

Knowledge is an important asset in an organization and is generated through many complex processes. Hence it is useful to capture and reuse this knowledge, especially where it can improve time or costs. The aim of the research, which the present work is a part of, is to acquire knowledge about manufacturing - in particular, knowledge of assembly issues from documents. Such acquired knowledge is intended to be used to detect potential issues in current or future assembly plans. The domain under study is assembly of aircraft structures. The knowledge to be acquired is that of issues that arise in assembly stage of manufacturing. This knowledge is then expected to be fed back to assembly planners to foresee issues in their current assembly plans. In the context of a product's lifecycle, this amounts to reuse of knowledge from a later stage of the lifecycle in an earlier stage. Such a reuse of knowledge is an important factor for PLM systems [1].

The overall process of acquiring knowledge is shown in Fig. 1, and the focus of this paper is shown in the dotted rectangle. From previous assembly processes, documents about problems in assembly may have been generated. These input documents are processed in the first step to segregate portions of documents that are related to aircraft assembly. Among these relevant portions, issues or problems that are present in the text are to be identified. This implies finding parts of text that talk about these issues in the domain. Once such issues are identified, the causes of these issues and

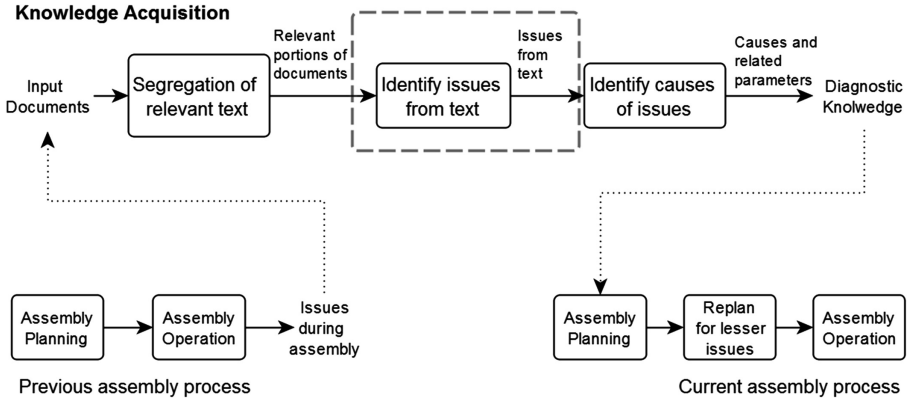


Fig. 1. Overview of the process to acquire knowledge of issues

parameters related to these causes are to be found. This knowledge about issues, their causes, and the parameters leading to the causes, should be structured as diagnostic knowledge. This structured knowledge would become the source for predicting assembly issues in the current assembly process.

To illustrate the knowledge reuse, consider an example where a document has been written about problems faced during an earlier assembly operation. An issue that a *particular riveting gun does not provide enough force for clean riveting* (although specifications say otherwise) is described in this document. Hence a riveting gun with a higher force specification is prescribed. If there is another assembly that is currently in the planning stage, and it also involves riveting, knowledge of this issue is relevant. Hence the planners can use a riveting gun of a higher force in practice and thus avoid a revision in assembly planning, due to this knowledge being available beforehand.

In order to identify issues from text, the first step is to identify sections that talk about issues. In an earlier paper [2], the authors have evaluated various methods and finally identified two possible means for such an identification. As a more practical solution, the method of sentiment analysis, which is a set of natural language processing techniques, was chosen.

1.1 Sentiment Analysis

Sentiment analysis (SA) is a set of natural language processing techniques whose aim is to determine whether a given piece of text is carrying a positive, negative, or neutral sentiment [3] (i.e. the sentiment polarity of the text), and a numeric value to indicate the strength of the sentiment. For example, ‘happy’ is positive, ‘very happy’ is more positive, and ‘not happy’ is negative. Sentiment analysis has found use in domains like movie reviews, and consumer electronics [4]. Sentiment can be calculated at various levels in text - at document, sentence, phrase or word-levels.

1.2 Research Problem

Although sentiment analysis is useful, adapting it to a specific domain is not a straightforward process. This process of adapting this technique to the domain of aircraft assembly is described in this paper.

The contribution of this paper is a means of enhancing domain lexicon in the field of aircraft assembly, in order to detect negative sentiment in documents. Such a negative sentiment, it is hoped, will be linked to a potential problem description in the document of interest.

2 Tools for Sentiment Analysis

2.1 Different Types of Sentiment Analysis Tools

There are two major groups of sentiment analysis techniques, depending on the practical application [5]. The first group is that of supervised classification methods based on training with large amounts of positive and negative text. The second group is to construct lexicons of predefined words for a given sentiment task in a domain. Both these groups of methods have their distinct advantages and disadvantages. For example, the former is sensitive to training data supplied (and hence can adapt well to a domain, given enough data), but it also demands the availability of large amounts of data. The latter does not necessitate such labelling of data, but requires detailed study to build suitable sets of positive and negative words.

The need for large amount of training data is also prevalent in the domain of aircraft assembly. The currently available limited set of sources of documents are from the World Wide Web, and there is no single coherent source of such data.

2.2 Choice of Tool

Due to the difficulties in finding sufficiently large training data in the domain of interest, we chose a lexicon-based approach to our task of identifying sections mentioning issues in text. From the available choices in this approach, two tools, namely SentiWordNet [6] and SO-CAL [3] were considered. SentiWordNet is a lexicon based on WordNet with sentiment values assigned to words. SO-CAL, on the other hand, is based on a detailed theory of how sentiment is not just dependent on single words, but also gets modified, using valence shifters [7]. Since this could judge overall sentiment for sentences, SO-CAL has been used as the tool for Sentiment Analysis in this research. For further details on this tool, readers are referred to Taboada et al. [3].

3 Shortcomings of General Sentiment Lexicons

As mentioned in the previous section, we now have a choice of a Sentiment Analysis (SA) tool to identify locations in text where issues are being described. The next step in the research was to verify if the lexicon, which was developed for a general English

language texts (or for a different domain) would be still applicable to the domain of aircraft assembly. As mentioned in Kanayama and Nasukawa [8], it is more difficult to prepare domain dependent lexicons than domain independent ones.

Fahmi and Klenner [9] have developed a combination of target nouns and adjectives that bear sentiment. This requires large, organized resources such as Wikipedia, and documents which bear the correct sentiment polarities (i.e. whether the text is positive, negative or neutral) for the objects. We are currently not having such resources for the domain of aircraft assembly. Denecke [10] compared lexicon based methods with machine learning based methods, concluding that the latter were better with SentiWordNet scores, doing well in multi-domain classification. However, sentence level sentiment was not studied, and machine learning based methods require labelled training data. Yue et al. [11] proposed an optimization based method to learn target-specific sentiment words by combining multiple knowledge sources. It, however assumes the availability of *aspects* (set of words describing a topic), either from experts or from an automatic method, which requires it to be combined for operation. This method is also capable of handling clause level sentiment. Muhammed et al. [12] attempted to extend a general lexicon to a social media lexicon and then combining the general lexicon with the domain-specific lexicon. This, once again, was dependant on a distant supervision dataset to be labelled and available. Ohana et al. [13] suggest the use of many different lexicons with a score adjustment based on term frequencies, in order to improve domain-independent sentiment classification. However, there are no directions to generate a lexicon for a given domain, in the absence of one.

As seen in the various methods discussed above there are several practical issues in adapting them, such as the availability of training data, or availability of other information that complements the lexicon building process. From a practical perspective, the simplest, yet useful method seemed to be to extend the lexicon manually for the chosen SA tool.

3.1 Study of Existing Lexicon on Domain Documents

In order to understand the current performance of the chosen SA tool on domain specific documents, a set of documents was initially chosen. These were documents available over the World Wide Web, and were about issues in manufacturing. SO-CAL was then used to analyze these documents, and the results were studied.

For every sentence, the researchers compared the polarity of sentiment assigned, with what was perceived to be the actual polarity. It is important to state here that the strength of sentiment (how strongly positive/negative) could not be considered, as that requires considerably more subjects and efforts to arrive at commonly agreed numbers.

A total of 357 sentences from 5 different documents were studied. Out of these, true positives and true negatives, as well as false positives and false negatives were identified. When we say “True Positive” here, it means that the sentence was positive in sentiment, and was also marked positive by the SA tool. The numbers are presented in Table 1. Since the original focus is to identify only negative sections of text, even sentences with SO-CAL score 0 were considered positive for this study (24 sentences were indecisive, hence they were not counted here).

Table 1. Initial performance of the tool, without adding any specific domain lexicon.

		Classification by SA tool	
		Positive	Negative
Classification by researchers	Positive	144 (TP)	20 (FN)
	Negative	68 (FP)	101 (TN)

Some examples of these four categories were:

True Positive (TP):

“A good rule to be used is that the number of blind rivets needs to be increased roughly in the proportion of 5 blind rivets for 3 solid rivets.”

True Negative (TN):

“In my opinion, this defeats the purpose of these rivets in the first place.”

False Positive (FP):

“But the mechanics say managers keep pressuring them to fix the planes faster.”

False Negative (FN):

“From race cars to airplanes, the blind rivet is the fastener of choice for joining sheet metal.” (It may be noted that the word ‘blind’ triggers the classification as negative sentiment)

3.2 Inadequacies of General Sentiment Lexicons

As seen in Table 1, it was observed that there are large number cases where the negative (or positive) sentiment in a sentence is correctly identified by the SA tool. However, there were still other cases where it was not identified correctly (68 and 20 sentences for the positive and negative cases by tool). Each of these cases where the assigned sentiment did not match the opinion of the researchers was studied. The observations were classified into the following categories:

- Ambiguity of word meaning: The sense of a word differs, based on the context in which it is used. For example, the word ‘issue’ was marked negative even when it was used in the context of a magazine’s date. Also, domain-specific use is a huge contributor to ambiguity, since a term that is used in general English would have a different meaning in manufacturing. For example, a ‘blind’ rivet is not negative in meaning; ‘upset’ a rivet is not negative either; however, ‘crossed’ wires is negative. At times, this can be seen more as domain-specific meaning, rather than ambiguity.
- Missing entries in the lexicon: There were many words in the SA tool’s in-built lexicon which did not have a corresponding entry for sentiment score. Fortunately, SO-CAL provides a list of such missing entries which cannot be scored, for a corresponding Part-Of-Speech. Some examples were ‘non-conformance’, ‘openness’, and ‘carcinogenic’. Also missing are certain entries that are phrases indicative of sentiment, such as ‘got to our head’ and ‘build up’.

- Clause level sentiment change: In the current work, the unit chosen is sentences, since the SA tool being used can handle sentences. However, even within sentences, there may be opposing sentiments in different clauses, which are finally summed up. For example, consider the following sentence:

‘The program had been the gold standard of industrial design tools in the 1980s but was only capable of producing two-dimensional blueprints’.

In this sentence, the first part of the sentence appears to be largely positive whereas the second part is negative, and they are connected by what is called as the ‘but’ connective in literature [14].

4 Enhancing Lexicon for Domain Specificity

The previous section described various reasons as to why a general sentiment lexicon did not perform as expected on text that is specific to a domain. In this section, we describe means of resolving some of these concerns.

The list of missing entries was the first means for improving the lexicon to result in better sentiment analysis. SO-CAL outputs a list of words that could be resolved by its tagger, but marked as missing in its dictionaries. The list is classified into four categories on the basis of part-of-speech, namely nouns, verbs, adjectives and adverbs. For the initial set of 5 chosen documents, this list consisted of 2160 nouns, 1080 verbs, 484 adjectives, and 152 adverbs. This list was then collected and manually analyzed. The objective was to assign a single number (between -5 and $+5$) to each word that indicates its sentiment specific to the current domain. The sentiment values were only *prior* values, which meant they were values for context-independent use of the words.

4.1 Assignment of Domain-Specific Sentiment Values

Based on the above list, assignment of sentiment value for the list of words was to be chosen. Since there was no specific guideline that was available, we chose the following scheme to assign these values. Since the maximum sentiment value was $+5$ and minimum was -5 , we decided to limit our values between $+4$ and -4 for the extreme cases, with one point for a buffer. Some generic guidelines were,

- If the word indicates high efficiency, or solution to a problem, it was given a score of 4 (“*much-lauded*”).
- If it reflects cause for improvements or progress, it was given a score of 2 (“*completion*”).
- An object description of name was given a neutral score (“*hydraulic*”).
- If something is not hazardous but still problematic, it got a score of -2 (“*inaccessible*”).
- If there is a hazard involved the word got a score of -4 (“*burst*”).
- Any word which was felt to be in between these categories was given an appropriate middle value, although such a value is subjective.

4.2 Evaluating the Effects of Adding User Lexicon

The user specific dictionary was then tested. This first iteration of testing involved testing the effects only for the additional dictionary. The same set of documents were run through SO-CAL, after configuring it to use the additional dictionary. The results of this first iteration of testing are shown in Table 2 (Please note that the ‘Zero’th iteration in the table refers to testing without any domain lexicon being added, from Table 1).

Table 2. Effects of enhanced dictionary (first, second iteration) and changed settings (second iteration)

Iteration	TP	TN	FP	FN	Others
Zero	144	101	68	20	24
First	140	112	57	24	24
Second	145	138	31	19	24

There was improvement in terms of reduced False Positives and False Negatives. There was also an increase in True negatives. However, there was also a minor reduction in number of True Positives.

The FN and FP cases were then analysed using the same approach described in Sect. 3.1. Multiple reasons were identified for the current performance of the tool. Some of the sentiment values needed a modification for the sentence level sentiment to be reflected correctly. In other cases, we realized that not only the specialized dictionary, but also the settings of the tool itself had played a role in deciding sentiment. These settings were related to the ignoring of sentiment words if they were in quotes or in irrealis mode (e.g. “*may forget*”). However, for our purposes both these modes were required to be present. Also, to a minor extent, some words from the existing SO-CAL dictionary itself had to be assigned a modified score, so that their prior sentiment is appropriate for the domain.

In the second iteration, there were three changes made in the analysis: the modified extra dictionary added by the user (19 instances), ignoring the quotes and irrealis factors (15 instances), and a small number of modifications to the original sentiment dictionary (2 instances). SO-CAL was again re-run on the same test documents. The results can be found in the third row of Table 2.

It is observed that the best improvements are present in the True Negatives and False Positives. Between the initial state and the second iteration the number of true negatives has increased by 37 instances, which is a 10.3% improvement over the total number of sentences. Similarly the number of false positives has fallen by the same number of instances.

5 Conclusions

This paper has discussed a method to improve the performance of a sentiment classification tool for the domain of aircraft assembly. Since there is a specific purpose for which sentiment analysis was used (to detect the presence of issues), the study focused more on the negative sentiment identification.

The study led to two means of improving the performance of the tool for the domain of aircraft assembly. The first is the construction of a dictionary of sentiment terms with prior sentiments assigned to them. The second was, to a lesser extent, the ignoring of *irrealis* and *modal* factors in text. By testing the tool’s performance on sample documents from the domain of interest repeatedly over two iterations, the dictionary has also been improved. Though we expected a larger number of modifications in the original English dictionary, there were not many that were eventually made.

The results clearly establish the feasibility of the tool to perform well to detect negative sentiment in domain specific documents. This would enable us to detect presence of issues by using sentiment analysis as the method of choice.

6 Future Work

The research reported in this paper can be subjected to many improvements. As discussed in Sect. 4.2, the major improvement that is immediately possible is to assign finely tuned values of sentiment priors to words in the dictionaries. The number of entries in the dictionary will have to increase once more documents are studied, and might reach a steady state once a large number of documents have been studied.

From a larger perspective, an important part would be to have target specific sentiment lexicon, and a means to use it. The sentiment value of a word may be of two types - prior, or context (target) dependant. We have currently addressed only the prior values in the aircraft assembly domain. As seen in the example of “cold pizza” vs “cold coke” by Fahrni and Klenner [9], it is necessary to associate specific words which are context (target) sensitive. Although there were no concrete cases in our test examples which suffered because of this, we can foresee that this may well be a problem (e.g. “*blind rivets*” is not negative, but “*blind spot*” is).

The other issue, seen in some cases, is that of ambiguity of word sense (By “late” autumn...). This is a commonly occurring issue during processing of natural language text, and methods like Word Sense Disambiguation (WSD) are suggested as means to resolve it.

From the perspective of creating the domain specific sentiment dictionary, creating a lexicon in a manual way is usually subjective and might be prone to errors. Automatic methods, many of which have also been described in literature here, may be used to improve the size and quality of the dictionary, since it remains to be seen how the size of the dictionary would grow over larger numbers of documents.

Acknowledgments. The authors are very grateful to Prof. Maite Taboada of the Simon Fraser University for making available the SO-CAL tool and the in-built dictionaries of SO-CAL for our research purposes, since the tool is the foundation for building the sentiment analysis work. This project was carried out with funding from The Boeing Company USA, under SID Project PC 36030.

References

1. Ameri, F., Dutta, D.: Product lifecycle management: closing the knowledge loops. *Comput.-Aided Des. Appl.* **2**(5), 577–590 (2005)
2. Madhusudan, N., Gurumoorthy, B., Chakrabarti, A.: Evaluation of methods to identify assembly issues in text. In: Bouras, A., Eynard, B., Fofou, S., Thoben, K.-D. (eds.) *PLM 2015. IAICT*, vol. 467, pp. 495–504. Springer, Heidelberg (2016). doi:[10.1007/978-3-319-33111-9_45](https://doi.org/10.1007/978-3-319-33111-9_45)
3. Taboada, M., et al.: Lexicon-based methods for sentiment analysis. *Comput. linguist.* **37**(2), 267–307 (2011)
4. Blitzer, J., Dredze, M., Pereira, F.: Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification. In: *ACL*, vol. 7 (2007)
5. Gonçalves, P., et al.: Comparing and combining sentiment analysis methods. In: *Proceedings of the first ACM Conference on Online Social Networks*. ACM (2013)
6. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: *LREC*, vol. 10 (2010)
7. Polanyi, L., Zaenen, A.: Contextual valence shifters. In: Shanahan, J.G., Qu, Y., Wiebe, J. (eds.) *Computing Attitude and Affect in Text: Theory and Applications*, vol. 20, pp. 1–10. Springer, Heidelberg (2006)
8. Kanayama, H., Nasukawa, T.: Fully automatic lexicon expansion for domain-oriented sentiment analysis. In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics (2006)
9. Fahrni, A., Klenner, M.: Old wine or warm beer: target-specific sentiment analysis of adjectives. In: *Proceedings of the Symposium on Affective Language in Human and Machine*. AISB (2008)
10. Denecke, K.: Are SentiWordNet scores suited for multi-domain sentiment classification? In: *Fourth International Conference on Digital Information Management, ICDIM 2009*. IEEE (2009)
11. Lu, Y., et al.: Automatic construction of a context-aware sentiment lexicon: an optimization approach. In: *Proceedings of the 20th International Conference on World Wide Web*. ACM (2011)
12. Muhammad, A., et al.: Domain-based lexicon enhancement for sentiment analysis. In: *SMA@ BCS-SGAI* (2013)
13. Ohana, B., Tierney, B., Delany, S.-J.: Domain independent sentiment classification with many lexicons. In: *2011 IEEE Workshops of International Conference on Advanced Information Networking and Applications (WAINA)*. IEEE (2011)
14. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics (1997)

Product Lifecycle Management for Digital
Transformation of Industries

13th IFIP WG 5.1 International Conference, PLM 2016,
Columbia, SC, USA, July 11-13, 2016, Revised Selected
Papers

Harik, R.; Rivest, L.; Bernard, A.; Eynard, B.; Bouras, A.
(Eds.)

2016, XIV, 655 p. 241 illus., Hardcover

ISBN: 978-3-319-54659-9