A Check on Annotation in Sentiment Research

Fitrah Rumaisa, Halizah Basiron, Zurina Saaya

Abstract**: *The research literature on sentiment analysis methodologies has exponentially grown in recent years. In any research area, where new concepts and techniques are constantly introduced, it is, therefore, of interest to analyze the latest trends in this literature. In particular, we have chosen to primarily focus on the literature of the last five years, on annotation methodologies, including frequently used datasets and from which they were obtained. Based on the survey, it appears that researchers do more manual annotation in the formation of sentiment corpus. As for the dataset, there are still many uses of English language taken from social media such as Twitter. In this area of research, there are still many that need to be explored, such as the use of semi-automatic annotation method that is still very rarely used by researchers. Also, less popular languages, such as Malay, Korean, Japanese, and so on, still require corpus for sentiment analysis research.***

Index Terms**: *Survey, Sentiment-Annotated, Methodology, Dataset.***

I. INTRODUCTION

Research in the field of sentiment analysis is now a very interesting topic to be discussed. The researchers conducted in-depth research on various datasets, languages, and methodologies. The study of this sentiment, which uses various languages as its dataset, requires a dictionary or commonly known as a corpus, as a reference to the sentence classification process. Annotation is known as one of the most important things needed in a corpus. Annotations are considered capable of enriching a corpus as a reference in the field of linguistic research going forward. Annotation itself is a practice of adding interpretive and linguistic information into an electronic corpus of oral language and written language of data[1]. Clear and easy-to-understand information is necessary to get a good annotation. It also applies to a fundamental explanation such as comment or review, in which one can judge something as positive, negative or neutral[2]. As an annotation word, researchers always set limits on a case like "Does this word a positive, negative or neutral sentiment?"[3], “Does this word have anything to do with positive, negative, or neutral sentiments?”[2], or “which word is more positive?”/”which word has a more prominent relationship with positive conclusion”[4,5]. The purpose of this study is to survey what methodologies have been used by previous researchers in sentiment-annotated, including frequently used datasets and from which they were obtained. This paper will discuss 23 kinds of literature on sentiment-annotated last five years. In this paper will be explained the theory of granulation of annotation along with several studies that have discussed each level, and then proceed with the discussion of methodologies that have been used by researchers and also the dataset. The discussion section explains methods, tools, and datasets obtained from the discussion of the previous section.

II. ANNOTATION METHODOLOGY

In this section, we will explain in detail methodologies that have been used by previous researchers. The methods used are conceptual metaphor theory, manual and automatic annotation, a lexicon-based system for both singular and multilingual corpus from various languages in the world.

Before discussing the methodologies that have been used by previous researchers, we will discuss first what an annotation is, and what are manually annotation and automatic annotation.

Annotation is a methodology for adding information to words, phrases, or entire documents. The purpose of annotations is to speed up data retrieval in search of documents or applications, or to add captions to documents, or to connect certain texts within a document to a broader concept or background. Figure 1 depicts 3 (three) connection of annotation.

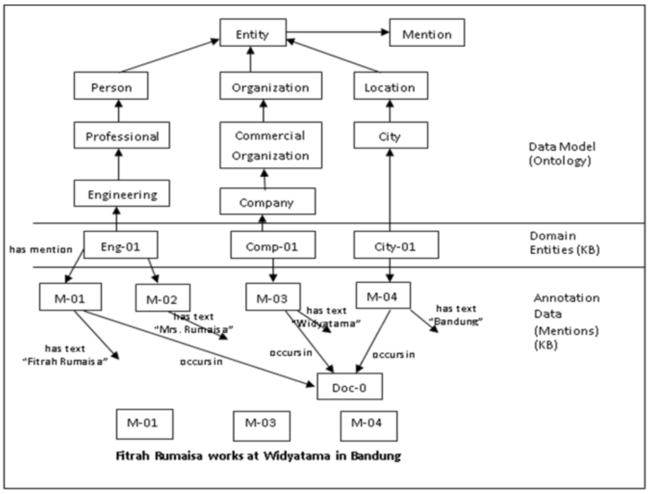

Figure 1 Three connection of annotation

Revised Manuscript Received on August 19, 2019.

Fitrah Rumaisa, Information Technology Department, Widyatama University, Indonesia.

Halizah Basiron, Faculty of Information And Communications Technology Universiti Teknikal Malaysia Melaka (UTeM) Melaka, Malaysia

Zurina Saaya, Faculty of Information And Communications Technology Universiti Teknikal Malaysia Melaka (UTeM) Melaka, Malaysia.

There are two processes for adding data that is manual and automatic. Automated annotations are deemed to be less precise but can be operated on many documents than people do. While the manual annotation process is considered more appropriate but requires a solid process and this process is often used to train the engine of automatic annotation.

Here are the steps to do in the annotation process[6]:

1. Read the entire document. Read the documents as well as possible to understand the contents of the document.
2. Put a mark on the entity. Read the document a second time, annotate this basic entity.
3. Reviewing. Review what has been done before to make sure nothing is missing, especially for the types and features of the annotations should be correct.
4. Write down any additional information found. Each annotates a document; note any additional information that is considered important.

Researchers do not always use annotation methods in corpus development. This is explained by Wallis & Nelson, that there are three kinds of methodologies for building corpus called 3A (Annotation, Abstraction, and Analysis). The annotation consists of applying the schema to the text. Annotations may include structural markup, part-of-speech tagging, parsing, and many other representations. Abstraction consists of translation (mapping) terms in the scheme to mention motivation models or data sets that are theoretically motivated. Abstraction usually includes searches devoted to linguistic languages but may include, for example, rule-learning for parsers. The analysis consists of probing statistics, manipulations, and generalizations of the dataset. The analysis may include statistical evaluation, optimization of basic rules methods or knowledge discovery methods[7]. Below are discussed several methodologies that have been used by previous researchers.

III. Survey on Methodology

The first researcher to be discussed is the study by Shutova, 2013. In his paper, he described the conceptual metaphor theory (CMT) which produces a very significant echo of the fields of philosophy, linguistics, cognitive science and artificial intelligence and still underlies most modern research on metaphors. The research questions of his research as follows:

1. How to describe conceptual metaphor intuitively on metaphor linguistic human annotators and whether to consistently annotate inter-conceptual mapping;
2. What difficulties will be faced by an annotator;
3. Whether a conceptual metaphor is sufficient to explain linguistic metaphors or require some conceptual metaphor sequences.

The experimental results show that the main difficulty faced by an annotator is to determine the appropriate level of abstraction of the domain (very difficult to consistently define labels on domains and targets), although inter-conceptual associations exist in some ways and are intuitive for humans[8].

Furthermore, the second researcher to be discussed is the study by Schulz (2010). In this research, the methodology of manually annotation development for multilingual corpus in very fine level or often called aspect-level on opinion mining and target annotation is discussed. The first step is to use the English corpus as the basis of the multilingual corpus used. To ensure that the comparability between corpuses made with the English language corpus, the same annotation scheme is used for documents with a little refinement. Then if the product feature does not appear in the sentence, where the opinion or pronoun is used, then the experiment should be able to capture the intention implicitly. For example in the phrase "The camera is designed very well." It is known that the opinion explains a review of the design, although the noun "design" is not contained in the sentence. Then, to facilitate the annotation process, a Java-based graphical user interface (GUI) tool is used, which is divided into three sections. The first section to present product review meta-information, the second part is to show the whole review text and annotate annotations to see the context of sentences, and the third part is the core of this tool that serves to show where product features along with their polarity and power of opinion can be recorded. The last step he did was an inter-annotator agreement. The assignment of annotations is a very subjective step. At least two annotators are required to get a reliable and objective perspective on an opinion. There are two agreements are used to deal with the accuracy of the string level and the accuracy of the content in the sense that both annotators use different phrases or use different writings for product features[9].

Another research performed by Hovy (2010) is also focused on annotation corpus. According to him, the annotation corpus can add preventive information into a collection of texts. Thus, he proposes a methodology that can break down the constraints faced while building the corpus annotation using human code. In fact, Hovy (2010) uses 8 (eight) basic steps of building a corpus based on Natural Language Processing (NLP) as follows:

1. Identify and prepare representative text options as starting materials for 'training corpus' (sometimes called 'training suite')
2. Instantiation of linguistic theories or linguistic concepts given, to determine the set of tags to be used, the conditions of their application, etc.
3. Annotation of some fragments from the training corpus, to determine the feasibility of both instantiation and the manual annotator.
4. Measure the results (compare annotator decisions) and determine which actions are appropriate, and how they apply.
5. Determining what level of agreement should be considered satisfactory (too little agreement means too little consistency in the annotation to allow the machine learning algorithm to be successfully trained). If the agreement is not satisfactory, the process will recur from step 2, with changes consistent with the theory, its instantiation, its Manual, and its annotator. Otherwise, the process goes to step 6.

6. Annotation of the most corpuses, perhaps for several months or years, with many checks, improvements, etc.

7. When enough material has been described, practice the NLP machine learning technology automatically on some training corps and measure its performance on the remainder (compare with the result when applied to the remaining text, often called 'data held heavily,' for annotator decisions).

8. If the agreement is satisfactory, the technology can be applied to additional materials, without labels, of the same type, thus helping future analysis. If the deal is unsatisfactory, the process recurs, perhaps from step 2 or perhaps from step 6 if more training data is needed.

Furthermore, he also poses 4 (four) basic steps which can be regarded as additional science in annotations. Here are 4 (four) proposed steps[10]:

1. Develop a more specific description of what is being studied, including the required pre-theory explanation of the type and range of classes and values observed.

2. Provide a clear record for the corpus to be annotated, as well as information on possible bias effects.

3. Establish clear and detailed procedures, including training for annotators, independent annotations that occur in at least two person annotators, discussions between supervised annotators and environmental concerns.

4. Developing an appropriate evaluation system, with a clear understanding of values and issues with the agreement.

Annotation activities on the formation of the corpus are urgently needed. This is because the corpus is formed not only in English. In the current era of microblogging, people can express their opinions using a variety of languages and informal languages. For example, posting an opinion on Twitter. All people from different countries can post their opinions on Twitter using their language.

The next study to be discussed is by Perez-Rosas et al. (2012). This research creates a framework for obtaining lexicon of sentiments in the Spanish by using annotation data either manually or automatically in languages with large data, such as English. The first step is to try to use manual annotations that are already in the Finder Opinion available at word-level. Then transfer it into English WordNet by applying SentiWordNet based on some constraints. The next step is to take advantage of the fact that multilingual WordNet uses syns (a collection of entities that have different meanings, and their members can be used interchangeably in the same context) appropriate for building blocks capable of explaining the map level of the language. Then, because the resource of the manual annotation process remains limited, the automatic annotation process is used in the English language to produce higher coverage and lower decision-making costs[11].

Next is the research by Lobur, 2012, his study discusses the making of sentiment-annotated in Ukraine. The process proposed by him is as follows[12]:

1. Collect text data for the corpus to be created
2. Specifies the software to be used for annotation manuals
3. Build annotation schemes
4. Annotate text data sets

Steinberger (2012), also discusses the creation of sentiment-annotated. But he uses semi-automatic annotations for dictionaries of several languages at once. First, the standard corpus sentiment is generated for two languages and then automatically translates into a third language. Then the annotation result is evaluated research to verify the result of the triangulation hypothesis[13].

Almost similar to Steinberger, Morgan, et al. (2013), also uses three languages in the formation of his annotated sentiment, namely English, Mandarin, and Russian. But Morgan presents a more subtle level. The dataset is taken from a conversation on Wikipedia and small group chats IRC. The datasets are annotated on two social actions: alignment and authority claim. The results of the annotations are then evaluated to the most subtle level using the inter-annotator agreement[14].

Next, Szabo et al. (2016) create a corpus sentiment from Hungarian using manually annotated aspect-level. The dataset used derives from Hungarian opinions on various product reviews. The purpose of this research is to establish a suitable database for software development in the future. There are two stages performed have been performed. In the first stage several annotations are carried out, namely, the overall construction that expresses positive and negative opinions, sentiments expressing positive and negative opinions at the lexeme level, sentiment targets, elements that modify the previous polarity (semantic orientation) of sentiment. In the second stage, the entire database has been created according to the new annotation design. The difference in this method is that entities and their aspects are noted with different tags and provide the same explanation for consistent corpus document targets[15].

Another researcher who also uses manual annotations is Mohammad (2016). He proposes two annotation schemes using questionnaires for simple sentiment annotations with more precise annotation directions and can provide additional labels. Both using semantic-based questionnaires with additional questions on accounts of user opinions or expressions and activity descriptions[16].

Along with the development of research, since 2013, Pustejovsky introduced a methodology model for annotation. This model is known as MATTER.

Details of MATTER are as follows[17]:

a. Model – The first step is the Phenomena Model. The required steps vary depending on the nature of the assigned task. The parameters used may also vary. In his research, the parameters used consisted of the vocabulary of the term, T, the relationship between terms, R, and interpretation, I. So, the model, M formed triple M = <T, R, I>.

b. Annotation – The next step after determining the phenomenon model of the specific document, it is necessary to train the human annotators to dot the dataset according to the important record.

c. Train – This step is used to train the algorithm to be used. If an error is found, the algorithm will be repaired and re-done. If complete, the algorithm is executed to test the task.

d. Test – This step is used to analyze errors. Once the algorithm performs the training, it will also be tested, and the error list can be generated to find out where the error lies.

e. Evaluation – In general, the most suitable method for evaluating the accuracy of the performance of the algorithm used is to calculate how accurate the data label is used. This can be done by measuring the fraction of results from properly labeled datasets using standard "relevance assessment" techniques called Precision and Recall metrics.

f. Revision – If there is an error in the evaluation stage, then the next step is to make revisions to correct the errors found.

In the next section will be described on the type of database and from where it gets.

IV. DATASETS

This section describes what sort of the datasets always used by previous researchers, including where the dataset came from. Twitter and other microblogging media are often used to get datasets. Some researchers use Twitter posts and other microblogging media as corpus datasets, such as [5,18,19,20,21,21,22,23,24,25,26,27]. Also, the dataset is also obtained from several articles, such as financial reports, police reports, news, textbooks, translations of scriptures and others, as discussed by [28,29,30,31]. The studies that have been presented in this section are very useful as the basis for classifying sentiments in the process of analytical sentiments. More on analytical sentiments will be discussed in subsequent chapters.

V. DISCUSSIONS & RESULTS

Based on the above description, it appears that previous researchers discussed various methods, tools, and data resource platform of each annotation method. For more details appear in the table below:

Table 1 Method, Tools And Datasets Of Annotation

Annotation	Method	Tools	Data Resource Platform
Manual	Human Annotator	WordFreak	– Microblogging (Twiter, Facebook, Foursquare, etc.) – Financial Reports – Police Reports – News – Textbooks – Translation of scriptures
	Inter-Annotation Agreement	GATE	
	Conceptual Metaphor Theory	BRAT	
	Simple Questionaire		
	Semantic-Role Based Sentiment Questionaire		
Automatic	SentiWordNet	WordFreak	
	WordNet	GATE	
	Lexicon method	Domeo	
	Machine Learning		
	SVM		
	ANN		
Semi-Automatic	Triangulation hypothesis	Domeo	
	Inter-annotator Agreement		

Table 1 shows that inter-annotation agreement can be used for two types of annotations, namely manual and semi-automatic annotation. For the tools used, GATE can be used for two types of annotations: manual and automatic, while Domeo can be used by automatic and semi-automatic annotations. As for the dataset, Twitter remains a prima donna to get the dataset.

The use of semi-automatic annotations in research is still very rarely studied. This can be a good field for developing research topics. In addition, the languages in used for this research is still around popular languages in the world such as English, China, and Arabic. While for the less popular languages are still very less used in research.

VI. CONCLUSION

The purpose of this study is to know the methodologies, datasets, and tools that have been used by some previous researchers. To find out all that was surveyed 23 kinds of literature from several years earlier. The literature surveyed is written in English and drawn from several papers published in journals and proceedings.

From the discussion, section mentioned there are still some research challenges that can be discussed more deeply in subsequent studies that are expected to enrich science. Based on the survey, it appears that researchers do more manual annotation in the formation of sentiment corpus. As for the dataset, there are still many uses of English language taken from social media such as Twitter. In this area of research, there are still many that need to be explored, such as the use of semi-automatic annotation method that is still very rarely used by researchers. Also, less popular languages, such as Malay, Korean, Japanese, and so on, still require corpus for sentiment analysis research.

The next research that will be done by the author based on the above review is the semi-automatic annotation method of Bahasa Melayu and Bahasa Indonesia corpus which has the same vocabulary but different meanings and polarities, where the data is taken from social media twitter and facebook.

REFERENCES

1. T. Garside, Roger; Leech, Geoffrey; McEnery, "Introducing corpus annotation," Corpus Annotation: Linguistic Information from Computer Text Corpora. pp. 1–292, 1997.
2. S. M. Mohammad and P. D. Turney, "Crowdsourcing a Word – Emotion Association Lexicon," Comput. Intell., vol. 59, no. 0, pp. 1–24, 2012.
3. M. Hu and B. Liu, "Mining and summarizing customer reviews," Proc. 2004 ACM SIGKDD Int. Conf. Knowl. Discov. data Min. KDD 04, vol. 4, p. 168, 2004.
4. S. Kiritchenko and S. M. Mohammad, "Capturing Reliable Fine-Grained Sentiment Associations by Crowdsourcing and Best--Worst Scaling," Proc. 2016 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol., pp. 811–817, 2016.
5. S. Kiritchenko and S. M. Mohammad, "Sentiment Composition of Words with Opposing Polarities," Naacl, pp. 1102–1108, 2016.
6. M. Petrillo and J. Baycroft, "Introduction to Manual Annotation," no. April, 2010.
7. S. Wallis and G. NELSON, "Knowledge Discovery in Grammatically Analysed Corpora," Data Min. Knowl. Discov., vol. 5, pp. 305–335, 2001.

8. E. Shutova, B. J. Devereux, and A. Korhonen, “Conceptual metaphor theory meets the data: A corpus-based human annotation study,” Lang. Resour. Eval., vol. 47, no. 4, pp. 1261–1284, 2013.
9. J. M. Schulz, C. Womser-hacker, and T. Mandl, “Multilingual Corpus Development for Opinion Mining,” Proc. Seventh Conf. Int. Lang. Resour. Eval., no. 2003, pp. 3409–3412, 2010
10. E. Hovy and J. Lavid, “Towards a ‘ Science ’ of Corpus Annotation : A New Methodological Challenge for Corpus Linguistics,” Int. J. Transl., vol. 22, no. 1, p. 25, 2010.
11. V. Perez-Rosas, C. Banea, and R. Mihalcea, “Learning Sentiment Lexicons in Spanish,” Proc. Eighth Int. Conf. Lang. Resour. Eval., pp. 3077–3081, 2012.
12. M. Lobur, M. Romanyshyn, and A. Romaniuk, “SENTIMENT-ANNOTATED CORPUS OF REVIEWS IN UKRAINIAN,” pp. 131–139, 2012.
13. J. Steinberger et al., “Creating sentiment dictionaries via triangulation,” Decis. Support Syst., vol. 53, no. 4, pp. 689–694, 2012.
14. J. Morgan, M. Oxley, E. Bender, L. Zhu, V. Gracheva, and M. Zachry, “Are We There Yet?: The Development of a Corpus Annotated for Social Acts in Multilingual Online Discourse,” Dialogue & Discourse, vol. 3, no. 2, pp. 1–33, 2013.
15. M. K. Szabó, V. Vincze, K. Simkó, V. Varga, and V. Hangya, “A Hungarian Sentiment Corpus Manually Annotated at Aspect Level,” Lang. Resour. Eval. Conf., pp. 2873–2878, 2016.
16. S. M. Mohammad, “A Practical Guide to Sentiment Annotation : Challenges and Solutions,” pp. 174–179, 2016.
17. J. Pustejovsky and A. C. Stubbs, “Natural Language Annotation for Machine Learning,” pp. 1–343, 2013.
18. G. Xiang, B. Fan, L. Wang, J. Hong, and C. Rose, “Detecting offensive tweets via topical feature discovery over a large scale twitter corpus,” Proc. 21st ACM Int. Conf. Inf. Knowl. Manag. - CIKM ’12, p. 1980, 2012.
19. A. F. Wicaksono, C. Vania, T. B. Distiawan, and M. Adriani, “Automatically Building a Corpus for Sentiment Analysis on Indonesian Tweets,” 28th Pacific Asia Conf. Lang. Inf. Comput., pp. 185–194, 2014.
20. M. A. Saloot, N. Idris, A. T. Aw, and D. Thorleuchter, “Twitter corpus creation: The case of a Malay Chat-style-text corpus (MCC),” Digit. Scholarsh. Humanit., vol. 31, no. 2, pp. 227–243, 2016
21. A. Pak and P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining,” Proc. Seventh Conf. Int. Lang. Resour. Eval., pp. 1320–1326, 2010.
22. M. Al-Kabi, M. Al-Ayyoub, I. Alsmadi, and H. Wahsheh, “A prototype for a standard Arabic sentiment analysis corpus,” Int. Arab J. Inf. Technol., vol. 13, no. 1A, pp. 163–170, 2016.
23. Y. Yu, H. Lin, J. Meng, and Z. Zhao, “Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks,” Algorithms, vol. 9, no. 2, p. 41, 2016.
24. P. Aliandu, “Sentiment Analysis on Indonesian Tweet,” Proc. Int. Conf. Information, Commun. Technol. Syst., pp. 203–208, 2013.
25. I. Mozetič, M. Grčar, and J. Smailović, “Multilingual Twitter Sentiment Classification: The Role of Human Annotators,” PLoS One, vol. 11, no. 5, p. e0155036, 2016.
26. M. A. Siddiqui, “Building A Sentiment Analysis Corpus With Multifaceted Hierarchical Annotation,” no. August, 2015.
27. W. He, L. Guo, J. Shen, and V. Akula, “Social Media-Based Forecasting:,” J. Organ. End User Comput., vol. 28, no. 2, pp. 74–91, 2016.
28. A. Moreno-Ortiz and J. Fernández-Cruz, “Identifying Polarity in Financial Texts for Sentiment Analysis: A Corpus-based Approach,” Procedia - Soc. Behav. Sci., vol. 198, no. Cilc, pp. 330–338, 2015.
29. L. Wah Lee and H. Min Low, “Developing an online Malay Language Word Corpus for primary schools,” Int. J. Educ. Dev. Using Inf. Commun. Technol., vol. 7, no. 3, pp. 96–101, 2011.
30. M. P. Hamzah and A. Mathematics, “Part of Speech Tagger for Malay Language Based,” vol. 2014, no. October, pp. 1499–1502, 2014.
31. H. T. Sukmana, R. H. Gusminta, Y. Durachman, and A. F. Firmansyah, “Semantically Annotated Corpus Model of Indonesian Question Answering System Performance,” Proc. 2016 4th Int. Conf. Cyber IT Serv. Manag. CITSM 2016, 2016.

Retrieval Number: B10650882S819/2019©BEIESP
DOI:10.35940/ijrte.B1065.0882S819

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication