Applied Data Science and Analysis
Vol.2024, **pp**. 165–172
DOI: https://doi.org/10.58496/ADSA/2024/013; ISSN: 3005-317X
https://mc04.manuscriptcentral.com/adsa

Research Article

Lexicon annotation in sentiment analysis for dialectal Arabic: Consensus Expert Standardized Criteria

Sameh M. Sherif [1], A.H. Alamoodi [1, *]

[1] *Faculty of Computing and Meta-Technology (FKMT), Universiti Pendidikan Sultan Idris (UPSI), Perak, Malaysia.*

ARTICLEINFO

Article History

Received 28 Jun 2024
Revised 21 Aug 2024
Accepted 02 Sep 2024
Published 24 Sep 2024

Keywords

Sentiment analysis
Lexical Criteria
Arabic language
Opinion mining
Dialect
Annotation
Lexicon

ABSTRACT

Sentiment Analysis (SA) in Natural Language Processing (NLP) involves analyzing perceptions, attitudes, and emotions from text. It is crucial for decision-making and consumer insights. Recent studies focus on developing Lexicons for SA research. Understanding the construction and evaluation of existing lexicons is key to advancing development efforts. Evaluation and benchmarking of lexicons are vital for identifying the most suitable ones and establishing best practices. Factors like effectiveness and importance must be considered when building or selecting lexicons. This research outlines three key phases: Determining Lexicons, Identifying Evaluation Criteria, and Engaging Experts. The study aims to enhance understanding of lexicon development processes and improve future guidelines. Efforts in lexicon development can benefit from a structured approach that considers various criteria for evaluation. The research emphasizes the importance of expert input in refining lexicons for optimal performance. Evaluating lexical criteria helps in identifying gaps and areas for improvement in sentiment analysis tools. Benchmarking different lexicons aids in selecting the most appropriate ones for specific applications or domains. Establishing best practices in lexicon development involves thorough evaluation against predefined criteria to ensure quality and reliability. Expert opinions play a crucial role in validating the significance of developed lexicons for sentiment analysis tasks. The research methodology involves systematic identification of lexicons relevant criteria, and experts to inform best practices in the field of sentiment analysis. By focusing on these three key phases, this study aims to contribute valuable insights into enhancing sentiment analysis through improved lexicon development processes.

1. INTRODUCTION

Sentiment Analysis (SA) is a branch of natural language processing (NLP), which also known as the method of analyzing people's perceptions, attitudes, and emotions by extracting and analyzing sentiment and polarity of a text [1]. SA has evolved into an essential component for decision-makers, business leaders, and everyday consumers [2]. Many studies have recently developed and used Lexicons for Sentiment analysis research. To establish lexicons it is essential to understand how the current lexicons are built and evaluated, accordingly more development efforts can be achieved. Evaluation and benchmarking of these Lexicons are important towards understanding the most suitable for fulfilling all essential requirements and to establish future guidelines and best practices in Lexicons development. When building, developing lexicons, or choosing which criteria is more effective or important than others, several factors need to be taken in consideration. Towards that end, this research presented on the basis of the three identification phases Determine Lexicons, Identify Evaluation Criteria and Identify Experts.

2. LITERATURE REVIEW

This study was conducted considering the academic systematic literature review (SLR) procedure in our previous research [3], which proposed a thorough discussion of the preliminary Study Phase, that included the Comprehensive Literature Investigation for ASA, SLR Protocol, and Identified Gaps and Challenges. While this research will go through a deep investigation on the Literature. Towards that end, the methodology proposed in this research will be presented on the basis of the three phases. Firstly, Determine Lexicons were only Arabic lexicons that are open to the general public will be taken into consideration; restricted or inaccessible lexicons won't be considered, followed by the second phase Identify Evaluation Criteria were four main criteria “Labelled Data, Labelling Type, Labelling Techniques, Labelling Targets “including 28 sub

*Corresponding author. Email: alamoodi.abdullah91@gmail.com

criteria have been developed to evaluate the targeted lexicons, lastly the third phase Identify Experts from the literature that we referenced [3].

2.1 Language Lexicons Identification

This study mainly concentrates on Arabic Sentiment Analysis (ASA). Several inclusion and exclusion criteria were established during the process of selecting the most relevant lexicons which are to be used in the case study of this research. Relying on many scholarly works that have been investigated [3] towards identifying available ASA lexicons for the experiment of this research. After a thorough investigation of 74 scholarly papers [3], final lexicons found are shown in Table 1 as follows.

TABLE I. ASA LEXICONS LIST

Lexicon Full Name	Short Name	Alternative	Reference
Arabic Speech and Sentiment corpus of tweets	ArSAS	Lex01	[4-6]
Tweets Emoji Arabic Dataset	TEAD	Lex02	[7, 8]
Arabic Sentiment Tweets Dataset	ASTD	Lex03	[8, 9]
Arabic Tweets Sentiment Analysis Dataset	ATSAD	Lex04	[8]
Opinion Corpus for Arabic	OCA	Lex05	[10]
Arabic movie review dataset	ARMD	Lex06	[10]
Books Reviews in Arabic Dataset	BRAD	Lex07	[1]
Nile University Lexicon	NileULex	Lex08	[11-13]
SLANG SENTIMENTAL WORDS & IDIOMS	SSWIL	Lex09	[1, 14]
Arabic Sentiment Lexicon	ArSeLEX	Lex10	[15, 16]
Multi Arabic Dialect Applications and Resources	MDAR	Lex11	[17]
Tunisian Sentiment Analysis Corpus	TSAC	Lex12	[18, 19]
Integrated Arabic Dialect Dataset	IADD	Lex13	[18, 19]
Shami Dialect Corpus	SDC	Lex14	[20]
Dialectal Arabic Tweets Dataset	DART	Lex15	[21]
Arabic Online Commentary Dataset	AOC	Lex16	[20, 22, 23]
Parallel Arabic Dialect Corpus	PADIC	Lex17	[18, 20, 23]
Saudi corpus for NLP Applications and Resources	SUAR	Lex18	[24]
Qatar Arabic Language Bank	QALB	Lex19	[25]
Sudanese Arabic Dialect, BERT	SudaBERT	Lex20	[24]
Single-label Arabic News Articles Dataset	SANAD	Lex21	[26]
News articles datasets in Arabic with multi-labels	NADiA	Lex22	[26]
Large Scale Arabic Book Reviews Dataset	LABR	Lex23	[9, 27, 28]
Open-Source Arabic Corpora	OSAC	Lex24	[29]
Arabic Jordanian General Tweets	AJGT	Lex25	[24]
Arabic Sentiment Twitter Dataset for the Levantine Dialect	ArSenTD	Lex26	[24]
Hotel Arabic-Reviews Dataset	HARD	Lex27	[24]
Moroccan Dialect Electronic Dictionary	MDED	Lex28	[30]

The chosen lexicons that meet the assessment criteria possess a number of common characteristics and lexical features as illustrated in Table 2 and 3.

TABLE II. COMMON LEXICON ATTRIBUTES

Common Properties	Description
Lexicon Name	The full name of the lexicon.
Year of Creation	The time period for the lexicon development.
Data Size	The final size of the dataset entities.
Source of Data	The source for extracting the data such as: Tweets, Facebook, or other source.
Targeted Dialect	The Arabic dialect used in the lexicon.
No. of Dialect	How many Arabic dialect used in the lexicon.
Dialect Country	Translates the Arabic dialect used in the lexicon into its origin country.
Annotation Type	The used type of annotation, for instance (Manual, Automatic, Hybrid).
Annotation Technique	The used technique of annotation, for instance (Crowdsourcing, Volunteers, Experts).
Annotation Classes	The rating classes of annotation, such as (Positive, Negative, Neutral).
Used Reference	The reference of the lexicon in our work.

TABLE III. LEXICON IDENTIFICATION

Lexicon Name	Domain	Size	Duration	ATY	ATQ	AC	Dialects
Lex01	1	21K	11/2017	M	CR	P, N,Nu,M	NA
Lex02	1	5.6M	6/2017 - 11/2017	M	NS	P, N,Nu	Egy, Gul, Mrc , Lev
Lex03	1	10K	11/2013	M	CR	P, N,Nu	Egy
Lex04	1	58,751	4/2019	A	NA	P,N	Sud, Egy, Syr, MSA
Lex05	2	500	NA	A	NA	P,N	NA
Lex06	3	NA	NA	NA	NA	NA	NA
Lex07	4	54.7K	6/2016 - 3/2018	NA	NA	P, N,Nu	Egy, Lev, Gul, MSA
Lex08	1	6k	12/2014	M	NS	P, N, Nu, M, Sr, A	Egy, Sar, MSA
Lex09	2	1350	NA	A	SY	St , Ds	Egy
Lex10	5	5244	1/2012 - 6/2013	H	NA	NA	Egy, MSA
Lex11	6	1,045	NA	M	NS	NA	NA
Lex12	2	17K	1/2015 - 6/2016	M	NS	P,N	NA
Lex12	1, 2	NA	NA	A	SY	NA	Egy, Lev, Mrc, Irq, Gul
Lex14	1	117,805	NA	A	SY	NA	Lev
Lex15	1	25k	NA	M	CR	NA	Egy, Mrc, Lev, Gul, Irq, Otr
Lex16	2	1.4M	4/2010 - 10/2010	M	CR	NA	Egy, Gul , and Lev
Lex17	6	6.4K	NA	NA	NA	NA	Mrc , Lev, MSA
Lex18	5	104,079	NA	H	MD	NA	Sar, Njd, Hjz, Gul

Domain: 1: Tweets, 2: Comments, 3: Movies Reviews, 4: Books Reviews, 5: Words, 6: Sentences, 7: Texts, 8: Hotels Reviews. Annotation Type [31]: Manual {M}, Automatic {A}, Hybrid {H}. Annotation Technique {ATQ}: Crowdsourcing {CR}, Native Arabic Speakers {NS}, Authors {AU}, Experts [32], Identification Systems {SY}, MADAMIRA Tool {MD}. Annotation Class {AC}: Very Positive {VP}, Positive {P}, Very Negative {VN}, Negative {N}, Neutral {Nu}, Mixed {M}, Sarcastic {Sr}, Ambiguous {A}, Satisfaction {St}, Dissatisfaction [33]. Dialects: Egyptian {Egy}, Sudanese {Sud}, Jordanian {Jor}, Iraqi {Irq}, Gulf {Gul}, Moroccan {Mrc}, Levantine {Lev}, Saudi Arabia {Sar}, Syrian {Syr}, Others {Otr}, Najdi {Njd}, Hijaz {Hjz}.

The Gulf dialect is spoken in the United Arab Emirates, Saudi Arabia, Kuwait, Oman, Qatar, and Bahrain. Due to a number of shared characteristics, this dialect also includes Yemeni and Iraqi dialects [5, 14, 34]. Palestine, Lebanon, Syria, and Jordan all speak a Levantine dialect [5, 34, 35]. The Egyptian dialect spoken in Egypt and Sudan is known as the Nile valley dialect [5, 34]. dialect spoken in Libya, Algeria, Morocco, Tunisia, and Mauritania that is North African (Maghrebi/Moroccan) [5, 34].

2.2 Lexical Criteria Identification

The term "criteria" refers to a variety of measurements or standards that are utilized to evaluate or benchmark principles for lexical data-related attributes or characteristics, such as the data source, Volume, type of labelling, and different strategies employed on the data; this information is crucial for giving further details on the characteristics of the data and how it was created. Nevertheless, not all criteria behave in the same way; some, for instance, are associated with cost-benefit analyses, the cost criteria has an inverse proportion; as the value increases, the performance indicator decreases: such as the value of the amount of time or money spent rises, the performance indicator decreases. Whilst the benefit criteria has a direct proportion, such as the amount of data or resources increases the performance value also increases. In addition, there are some criteria with (Yes/No) values. The criteria used in this study to evaluate lexicons have been classified as two categories: main criteria and sub-criteria. A collection of common characteristics and lexical attributes that were sought out for extraction from the most prevalent lexicon attributes in the literature [3] were the focus of the suggested lexicons criteria that we have chosen for evaluation to be addressed in order to determine whichever is most suited for evaluation and benchmarking of SA lexicons in the Arabic language. Based on the literature analysis [3], there were 32 criteria, all of which have been explored and extensively examined before being submitted to a panel of experts for standardized procedures. There are two hierarchy levels for this 32 criterion: the main criteria level has 4 criteria based on the factors of their appraisal, and its sub-criteria level has 28 criteria as demonstrated in Fig 1.

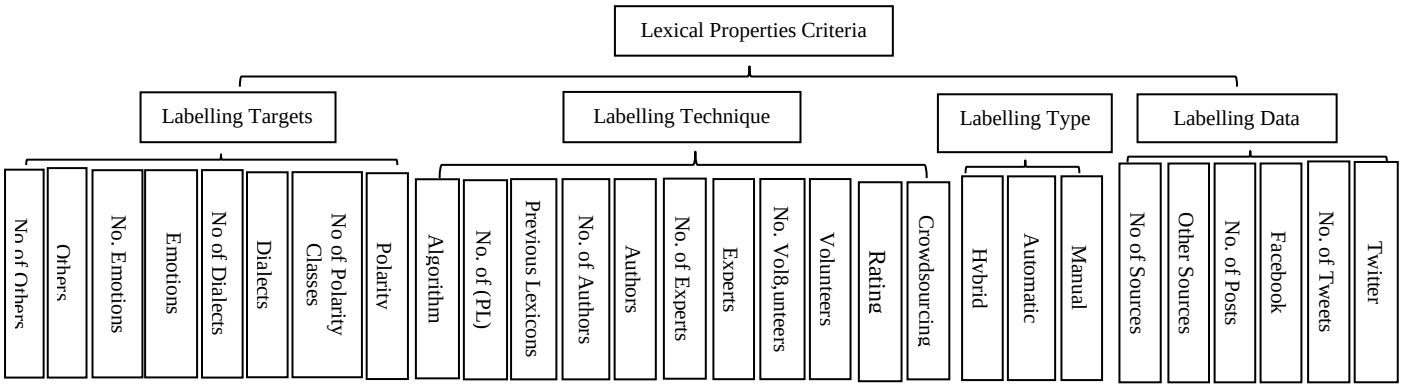

Fig. 1. Definitive Collection Of Used Criteria

2.2.1 Main Criteria

During the (SLR) we have defined some attributes we believed they are essentially common to all the lexicons, those attributes are what we are aiming to assign a weight based on the level of importance for each attribute. We categorized these attributes under four main groups, each of them called main criteria.

- 1st main criterion Labelling Data, refers to the collection of Linguistic data or material which have been acquired either from comments, notes, reviews, posts, corpora, surveys, or other types of texts as audio or video scripts, the data extraction source or platform, such as Facebook, Twitter, or other website or source of data, and the volume of the collected data number of posts, tweets.
- 2nd main criterion Labelling Type, describes the process used in identifying and classifying data by adding one or more meaningful and informative labels to it, such as positive, negative polarity, whether this process was done Manually, Automatically or Hybrid.
- 3rd main criterion Labelling Techniques, discusses the labelling techniques which means the way of carrying out a certain task or a piece of scientific procedure. In this work the labelling techniques are particularly concerned with annotation or labelling techniques applied to the Linguistic data.
- 4th main criterion is Labelling Targets have eight sub-Criteria concerned about the Polarity, number of Polarity Classes, Dialects of the lexicon, how many Dialects in the lexicon, whether the lexicon includes Emotions or not, and what are the Emotions if exists.

2.2.2 Sub-Criteria

This section provides an extensive review of 28 sub-criteria derived from the main criteria. All of these criteria have been extracted during the comprehensive analysis of the literature review [3]. As illustrated below in Table 4.

TABLE IV. SUB-CRITERIA ATTRIBUTES

Main-Criteria	Sub-Criteria	Description
Labelling Data	**Twitter**	Concerns whether the data was extracted from Twitter or not.
	No. of Tweets	Concerns with the volume or number of tweets extracted from Twitter.
	Facebook	Whether the data was extracted from Facebook or not.
	No. of Posts	Concerns with the count or amount of data extracted from Facebook.
	Other Sources	Concerns whether the data was extracted from other sources such as; reviews, comments from service websites, news websites, blogs, or portals.
	No of Sources	concerns with the size or volume of data extracted from other sources.
Labelling Type	**Manual**	Manual Labelling refers to the human utilization of labelling the data associated with individual factor who participated in the annotation process whether they were volunteers, Experts or the Authors themselves.
	Automatic	Automatic Labelling interested in the automatic approach or procedure of simultaneously carrying out the annotation or labelling process, using automatic or non-manual techniques, which can be applied concurrently to all entities, such as machine learning or Algorithm
	Hybrid	Hybrid Labelling is the overlaps between automatic and manual labelling, when more than single labelling technique been applied into the labelling process, such as merging any of the manual and automatic techniques with one another
Labelling	**Crowdsourcing**	The crowdsourcing is comprised of the phrases crowd and outsourcing. Crowdsourcing is the method of enrolling a "crowd" or group of individuals for a shared sense of purpose, such the use of social media and Web portals. It is also referred to as problem-solving, wit

Technique		"the technique of turning to a group of Individuals to gain required expertise or services about a given instance," is used in the first manual annotation strategy; in this study, it refers to the annotation process conducted through online platforms.
	Rating	Rating system, refers to any application, portal, system, or algorithm incorporating methodologies, procedures, data control tools that allow the implementation of classification rules, rating plans, and rating values implemented by humans.
	Volunteers	In the context of manual human approach annotation, volunteers are recognized as ordinary members of the general public who participated in part of the study (the annotating process), but no specific information about their educational background not expertise was supplied.
	No. Volunteers	Refers to the act of counting up determine the total number used to perform a particular task, i.e. the total count of volunteers participated in the labelling process.
	Experts	Experts are identified as Humans who have expertise in linguistics fields linked to sentiment analysis and data annotation, and who have been enlisted to do a particular activity as part of a scientific experiment or study.
	No. of Experts	Refers to the act of counting up determine the total number used to perform a particular task, i.e. the total count of experts participated in the labelling process.
	Authors	Descries the process were authors have participated or fully performed the annotation, together with persons who have been included in the list of authors for a certain paper, piece of study, article, manuscript, or publication because they made significant contribution to its creation.
	No. of Authors	Refers to the act of counting up determine the total number used to perform a particular task, i.e. the total count of authors participated in the labelling process.
	Previous Lexicons	This criteria were assessed depending on the existence of former lexicons been applied through expanding or merging these lexicons.
	No. of (PL)	Refers to the act of counting up determine the total number of entities involved in a particular task, i.e. the total count of Previous Lexicons used in the labelling process.
	Algorithm	These criteria were assessed depending on the existence of mathematically-based coding commands, programming rules or an algorithm that are supplied to a computer in the term of the annotation process.
Labelling Targets	**Polarity**	Polarity refers to the process of identifying and classifying raw data (images, text files, videos, etc.) by adding one or more meaningful and informative labels to it, such as positive, negative polarity.
	No of Polarity Classes	Refers to the act of counting up determine the total number of entities involved in a particular task, i.e. the total count of Polarity Classes used in the labelling process.
	Dialects	Dialects are a regional speech patterns of an informal register, common in spoken conversation but avoided in formal writing, which is peculiar to a specific region or social group who live in one nation or a group of nations with similar cultures.
	No of Dialects	Refers to the act of counting up determine the total number of entities involved in a particular task, i.e. the total count of Dialects determined in the Linguistic data.
	Emotions	Emojis are pictogram, logogram, ideogram or smiley embedded in text and used in electronic messages and web pages. The primary function of emoji is to fill in emotional cues missing from typed conversation, as well as to express important information, which sometimes reflects a special or indicated meaning, Considering these sub- Criteria s will introduce a new measure that reflects emotional meanings and will lead to better results and common sense concepts.
	No. Emotions	Refers to the act of counting up determine the total number of entities involved in a particular task, i.e. the total count of Emotions determined in the Linguistic data.
	Others	Refers to the act of entities involved in a particular task, i.e. other entities determined in the Linguistic data.
	No of Others	Refers to the act of counting up determine the total number of entities involved in a particular task, i.e. the total count of other entities determined in the Linguistic data.

2.3 Expert Identification Phase

The process of selecting experts is crucial, that entails identifying experts and requesting their opinion on the criteria for evaluating. In our attempt to define standards about the experts whom will be involved in the lexicon criteria evaluation process, we reverted to the articles we have used as references, to determine the experts details from the authors, we came up with 211 authors, but only 188 were sharing their contact details (email address, ORICD No., LinkedIn, webpage….etc.). This stage necessitates a thorough selection procedure in which each expert involved must have sufficient and relevant knowledge with the case study in question. We started our communication with the aforementioned 188 experts, with two-step approach. Firstly, by sending 188 introduction emails about the sender, included a clear definition of the author's article that was being used as a reference, seeking the author's permission to participate in a consultation with the lexicon criteria evaluation. Secondly, we only sent an email including the questionnaire's URL and a brief synopsis to those experts who conveyed a willingness to participate in the process. The years of academic experience was one of the questions that were asked from the expert to ensure that they have the experience in the respective field of study. All the experts have prolonged

and intense minimum 2 and up to more than 42 years of experience or extensive knowledge, through occupation practice or based on research publication, in a Sentimental Analysis field and Arabic Language. Academically recognized as a reliable source of techniques or skills, whose faculty for judging or deciding rightly, justly, or wisely is accorded authority and status by peers or the public in the Natural Language Processing domain. Among all the experts we have decided to use the responses of only those have 10 up to more than 42 years of experience which ended up with 28 experts.

2.3.1 Constructing Expert Form

We formed these Criteria measures into a questionnaire form based on the multiple-choice grid with Likert scale of five levels of importance for each measured point (Huge Importance, Big Importance, Important, Slight Importance, No Importance) as shown in Table 5. Which helps to gather a wide range of feedback without making the survey too complex. The 5-points Likert Scale offers five different options for the respondents to choose from which include two extremes, two intermediate, and one neutral opinion. Using this 5-points scale makes it easier to understand responses and draw useful conclusions from the data. We distributed the questionnaire form to the experts who took part in the consultation involving lexicon criteria assessment.

TABLE V. LIKERT SCALE APPLIED TO THE LINGUISTIC

Linguistic Scoring Scale	Numerical Scoring Scale
Huge Importance	5
Big Importance	4
Important	3
Slight Importance	2
No Importance	1

The expert questionnaire can be developed in different ways, such as through (I) literature review and (II) interviews, the development of questionnaire items can be done based on literature review, pilot studies, and experiences [36]. Creating the expert questionnaire was a down to the last detail and time-consuming process. After designing the questionnaire from based on the extracted criteria. The questionnaire was developed using Google Forms and is divided into 7 sections as descried below, it includes 36 questions, first four questions were concerning the expert details which granted to be protected such as name, University, years of experience and emails, the rest of the questions were one question about the level of importance of each one of the four main criteria with total four questions, followed by another 28 questions for each of the 28 sub-criteria we have retrieved from the literature review. Section One: Is an Introduction to the aim behind the questionnaire, to compare Features used in Lexicons criteria evaluation for specifying the importance of these features against each others in order to evaluate which of them has more significance and importance in determining which lexicon is better. Taken into consideration there are four main criteria with twenty eight sub-criteria. Section Two: Is about the participant biography which promised to remain private for the privacy concerns. It requires the personal information such as name, university or institute and years of experience of experts. This part is done using long text and checklist menu where experts be able to write their position and choose the working experiences from given choices. In the following Sections Three to Six: we introduced our four main criteria which need to be assessed; Section Three Is about rating the Labelled Data Main Criteria along with its 6 sub- criteria. Section Four is for rating Labelling Type Main Criteria along with its 3 sub- criteria. Section Five is for rating Labelling Techniques Main Criteria along with its 11 sub- criteria., Section Six is for rating Labelling Targets Main Criteria along with its 8 sub- criteria. The last section, section seven is a thanks to the experts for taking the effort to participate in the questionnaire. The final set of introductions emails were drafted on 24th September and scheduled to be sent to the 188 experts on September 27th 2022 followed by the questionnaires to those whom responded with acceptance of participations. A checklist was created for respondents. Final number of responses was 28 and all the data has been collected for further analysis.

3. DISCUSSION

By analyzing and assessing the experts opinions data, in terms of the first criteria the majority (over 90%) of experts deem labelled data as either of "Huge Importance" or "Big Importance." This reflects the critical role labelled data plays in various tasks. The second criteria Labelling type is seen as essential by a significant portion of experts, with the majority indicating that it is of "Huge Importance." A small percentage consider it less critical, but no one dismisses its relevance entirely. Experts also value the third main criteria labelling techniques, with more than half marking it as having "Huge Importance." The diverse distribution across the other categories indicates some flexibility in opinion, though the "Huge" and "Big" categories are predominant. Whilst most experts consider labelling targets crucial, with over half giving it the highest importance rating. However, some experts find it less critical compared to other factors, indicating a more varied perspective on this criterion. On the other hand, there is a noticeable dissimilarity in the sub-criteria results were the Experts' opinions are evenly distributed here, with a notable percentage considering the number of experts involved to be only of slight or no

importance. However, a substantial portion (about 55%) views it as more important. The number of authors is not seen as crucial by many experts. Only 13.8% consider it of "Huge Importance," while a good proportion finds it less significant or irrelevant. The distribution shows that many experts place considerable weight on the number of polarity classes, with 68.9% considering it of "Huge" or "Big Importance." Nonetheless, some experts think this criterion has only slight relevance. Emotions are generally regarded as highly important, with almost half of the experts assigning "Huge Importance" to this criterion. A majority find it at least important, signaling its relevance in most applications as illustrated below in Fig. 2.

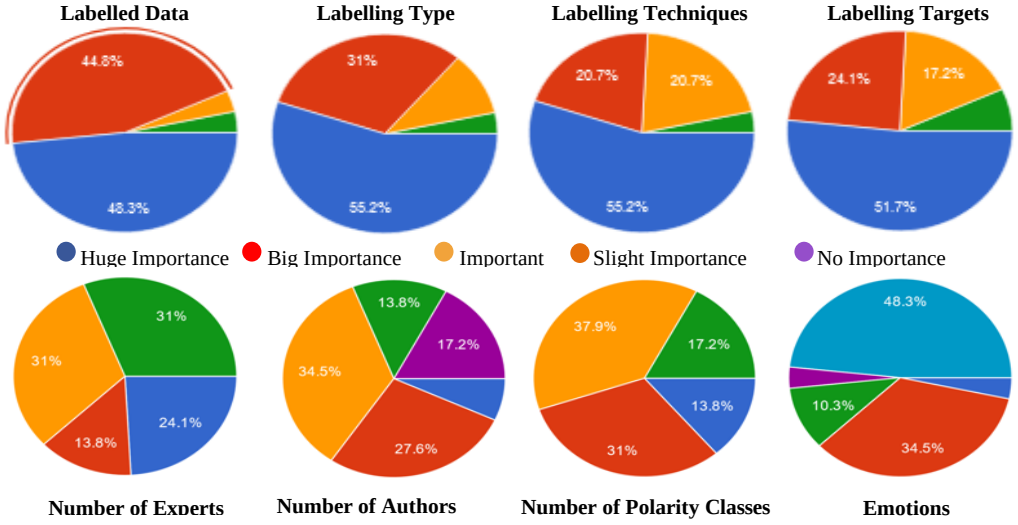

Fig. 2. Sample Of Expert Votes on The Criteria Importance

4. CONCLUSION

While the results showed a high degree of agreement between the experts, indicating that their opinions were aligned on the relative importance of the given criteria. However, there were still some discrepancies in the votes, with some experts placing greater emphasis on certain aspects than others. The criteria that stand out as most important to the experts are Labelled Data, Labelling Type, and Emotions, where the majority consider these of "Huge Importance." On the other hand, Number of Authors and Number of Experts are seen as less critical, with some experts marking them as having slight or no importance. This diverse range of importance levels across different criteria suggests a nuanced approach to each, reflecting the specific context and applications experts have in mind, while there is a general consensus on the importance of certain criteria in a lexicon, the relative importance of these criteria's may vary depending on the individual's perspective and background. Additionally, the study revealed that the significance of lexicons was influenced by several factors, including the data volume, dialects complexity, and annotation. Overall, this study provides valuable insight into the importance of lexical criteria and how they are perceived by experts in different fields. It also highlights the need for further research to better understand the factors that affect lexicon significance and how these factors may differ.

Conflicts Of Interest

The disclosure declaration of the author demonstrates to the absence of any conflicts of interest.

Funding

The author acknowledge that, this research was carried out without funding of any financing resources.

Acknowledgment

The author extends appreciation to the institution for their unwavering support and encouragement during the course of this research.

References

[1] A. Farha and W. Magdy, "Mazajak: An online Arabic sentiment analyser," in Proceedings of the fourth arabic natural language processing workshop, 2019, pp. 192-198.

[2] A. H. Ombabi, W. Ouarda, and A. M. Alimi, "Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks," Social Network Analysis and Mining, vol. 10, pp. 1-13, 2020.

[3] S. M. Sherif, A. Alamoodi, O. Albahri, S. Garfan, A. Albahri, M. Deveci, et al., "Lexicon annotation in sentiment analysis for dialectal Arabic: Systematic review of current trends and future directions," Information Processing & Management, vol. 60, p. 103449, 2023.

[4] A. Abdelli, F. Guerrouf, O. Tibermacine, and B. Abdelli, "Sentiment Analysis of Arabic Algerian Dialect Using a Supervised Method," in 2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS), 2019, pp. 1-6.

[5] S. Al-Azani and E.-S. M. El-Alfy, "Audio-Textual Arabic Dialect Identification for Opinion Mining Videos," in 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019, pp. 2470-2475.

[6] M. Al-Ayyoub, A. A. Khamaiseh, Y. Jararweh, and M. N. Al-Kabi, "A comprehensive survey of arabic sentiment analysis," Information processing & management, vol. 56, pp. 320-342, 2019.

[7] O. Oueslati, E. Cambria, M. B. HajHmida, and H. Ounelli, "A review of sentiment analysis research in Arabic language," Future Generation Computer Systems, vol. 112, pp. 408-430, 2020.

[8] A. Alawami, "Aspect terms extraction of Arabic dialects for opinion mining using conditional random fields," in International Conference on Intelligent Text Processing and Computational Linguistics, 2016, pp. 211-220.

[9] A. Assiri, A. Emam, and H. Al-Dossari, "Real-time sentiment analysis of Saudi dialect tweets using SPARK," in 2016 IEEE International Conference on Big Data (Big Data), 2016, pp. 3947-3950.

[10] S. Albukhitan, A. Alnazer, and T. Helmy, "Framework of Semantic Annotation of Arabic Document using Deep Learning," Procedia Computer Science, vol. 170, pp. 989-994, 2020.

[11] F. Sadat, F. Mallek, M. M. Boudabous, R. Sellami, and A. Farzindar, "Collaboratively constructed linguistic resources for language variants and their exploitation in NLP application–the case of Tunisian Arabic and the social media," in Proceedings of workshop on Lexical and grammatical resources for language processing, 2014, pp. 102-110.

[12] K. Darwish, "Arabizi detection and conversion to Arabic," arXiv preprint arXiv:1306.6755, 2013.

[13] A. Bies, Z. Song, M. Maamouri, S. Grimes, H. Lee, J. Wright, et al., "Transliteration of arabizi into arabic orthography: Developing a parallel annotated arabizi-arabic script sms/chat corpus," in Proceedings of the EMNLP 2014 workshop on Arabic natural language processing (ANLP), 2014, pp. 93-103.

[14] A. Assiri, A. Emam, and H. Al-Dossari, "Towards enhancement of a lexicon-based approach for Saudi dialect sentiment analysis," Journal of information science, vol. 44, pp. 184-202, 2018.

[15] I. Guellil, F. Azouaou, and M. Mendoza, "Arabic sentiment analysis: studies, resources, and tools," Social Network Analysis and Mining, vol. 9, pp. 1-17, 2019.

[16] R. Baly, G. El-Khoury, R. Moukalled, R. Aoun, H. Hajj, K. B. Shaban, et al., "Comparative evaluation of sentiment analysis methods across Arabic dialects," Procedia Computer Science, vol. 117, pp. 266-273, 2017.

[17] S. M. C. Loureiro, J. Romero, and R. G. Bilro, "Stakeholder engagement in co-creation processes for innovation: a systematic literature review and case study," Journal of Business Research, vol. 119, pp. 388-409, 2020.

[18] A. B. Soliman, K. Eissa, and S. R. El-Beltagy, "Aravec: A set of arabic word embedding models for use in arabic nlp," Procedia Computer Science, vol. 117, pp. 256-265, 2017.

[19] S. Almouzini and A. Alageel, "Detecting Arabic depressed users from Twitter data," Procedia Computer Science, vol. 163, pp. 257-265, 2019.

[20] A. Soumeur, M. Mokdadi, A. Guessoum, and A. Daoud, "Sentiment analysis of users on social networks: overcoming the challenge of the loose usages of the Algerian Dialect," Procedia computer science, vol. 142, pp. 26-37, 2018.

[21] I. Alsarsour, E. Mohamed, R. Suwaileh, and T. Elsayed, "Dart: A large dataset of dialectal arabic tweets," in Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018.

[22] J. Younes, E. Souissi, H. Achour, and A. Ferchichi, "Language resources for Maghrebi Arabic dialects' NLP: a survey," Language Resources and Evaluation, vol. 54, pp. 1079-1142, 2020.

[23] H. Rahab, A. Zitouni, and M. Djoudi, "SANA: Sentiment analysis on newspapers comments in Algeria," Journal of King Saud University-Computer and Information Sciences, 2019.

[24] N. Al-Twairesh, R. Al-Matham, N. Madi, N. Almugren, A.-H. Al-Aljmi, S. Alshalan, et al., "Suar: Towards building a corpus for the Saudi dialect," Procedia computer science, vol. 142, pp. 72-82, 2018.

[25] W. Zaghouani, N. Habash, and B. Mohit, "The qatar arabic language bank guidelines," Technical Report CMU-CS-QTR-124, School of Computer Science, Carnegie Mellon …2014.

[26] M. N. Al-Kabi, A. A. Al-Qwaqenah, A. H. Gigieh, K. Alsmearat, M. Al-Ayyoub, and I. M. Alsmadi, "Building a standard dataset for Arabie sentiment analysis: Identifying potential annotation pitfalls," in 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), 2016, pp. 1-6.

[27] T. Almanie, A. Aldayel, G. Alkanhal, L. Alesmail, M. Almutlaq, and R. Althunayan, "Saudi Mood: a real-time informative tool for visualizing emotions in Saudi Arabia Using Twitter," in 2018 21st Saudi Computer Society National Computer Conference (NCC), 2018, pp. 1-6.

[28] N. Boudad, R. Faizi, R. O. H. Thami, and R. Chiheb, "Sentiment analysis in Arabic: A review of the literature," Ain Shams Engineering Journal, vol. 9, pp. 2479-2490, 2018.

[29] M. Heikal, M. Torki, and N. El-Makky, "Sentiment analysis of Arabic Tweets using deep learning," Procedia Computer Science, vol. 142, pp. 114-122, 2018.

[30] R. Tachicart, K. Bouzoubaa, and H. Jaafar, "Building a Moroccan dialect electronic dictionary (MDED)," in 5th International Conference on Arabic Language Processing, 2014, pp. 216-221.

[31] I. Guellil and F. Azouaou, "Arabic dialect identification with an unsupervised learning (based on a lexicon). application case: Algerian dialect," in 2016 IEEE Intl Conference on Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES), 2016, pp. 724-731.

[32] G. Imane, D. Kareem, and A. Faical, "A set of parameters for automatically annotating a Sentiment Arabic Corpus," International Journal of Web Information Systems, 2019.

[33] A. Vallenari, A. G. Brown, T. Prusti, J. H. De Bruijne, F. Arenou, C. Babusiaux, et al., "Gaia data release 3-summary of the content and survey properties," Astronomy & Astrophysics, vol. 674, p. A1, 2023.