5-2022

Metaphor Detection in Poems in Misurata Arabic Sub-Dialect : An LSTM Model

Azza Abugharsa

Abstract

Natural Language Processing (NLP) in Arabic is witnessing an increasing interest in investigating different topics in the field. One of the topics that have drawn attention is the automatic processing of Arabic figurative language. The focus in previous projects is on detecting and interpreting metaphors in comments from social media as well as phrases and/or headlines from news articles. The current project focuses on metaphor detection in poems written in the Misurata Arabic sub-dialect spoken in Misurata, located in the North African region. The dataset is initially annotated by a group of linguists, and their annotation is treated as the seed data used in the project. Moreover, the verses in the dataset are annotated by layman native speakers of the sub-dialect who are not acquainted with the rhetorical principles of this kind of poetry. The model applied in the project is built on the Long Short-Term Memory (LSTM) architecture. The aim is to compare the performance of the model to the performance of human annotators who are not experts in the Arabic figurative language used in poetry. The results show that the model outperforms the output provided by the human annotators and scores a higher score of 79%. In addition, the model scores an 80.7 % accuracy score in predicting metaphors from unseen blind data. Since Arabic sub-dialects are acquired as a native language, it becomes important to develop NLP models that can be trained on these informal varieties of Arabic in order to fulfill many tasks such as auto-correction, machine translation, dialogue systems, and sentiment analysis among others.

METAPHOR DETECTION IN POEMS IN MISURATAARABIC SUB-DIALECT

AN LSTM MODEL

A THESIS

Submitted in partial fulfillment of the requirements

For the degree of Master of SCIENCE

by

AZZAABUGHARSA

Montclair State University

Montclair, NJ

2022

Acknowledgments

I would like to express my sincere gratitude and appreciation to the six poets who provided the poems to use in this project:

Mohamed Emmiama

Salah Abaid

Alhusian Elasawadi

Khalid Iqsheerah

Nedal Sowayeb

Omar Abualayem

I also want to thank the faculty in the Arabic Department at Misurata University for annotating the seed data of the dataset. I especially want to thank Mrs. Amna Omar Ben Hamaida, an expert in Arabic rhetoric, literature, and morphology for her great assistance in detecting metaphors in these poems.

In addition, my gratitude is expressed to the three graduate students in the Arabic department at Misurata University for participating in the process of annotating the dataset used in this project.

I would like to thank the three annotators in the layman group for allocating valuable time to do the tedious task of data annotation.

Finally, I would like to Thank Mr. Ramadan Maiteeg, a native speaker of the Misurata sub-dialect who possesses a special interest in Arabic literature, for providing definitions and explanations of the vocabulary used in Misurata traditional poems.

Table of Contents

List of Tables

List of Figures

Metaphor Detection in Poems in Misurata Arabic Sub-dialect

An LSTM Model

Introduction

Over the last years, deep learning architectures have emerged as powerful machine learning models which are capable of learning multiple layers of data features and producing automatic feature extraction, resulting in state-of-the-art predictions. Recently, deep learning methods have been applied extensively in different areas of study, especially in NLP tasks such as the automatic analysis of figurative language. One possible reason for this interest in analyzing figurative language is that it is understudied in the community of computational linguistics even though the figurative language is commonly used in daily conversations and social media discussions.

Another reason is the fact that metaphors are productive constructions that are complex in nature, and this means that every word/phrase can virtually be used metaphorically. Therefore, these words may not be processed accurately in certain machine learning tasks such as cross-language translation. This makes these structures hard to predefine and impossible to list. Moreover, due to their ubiquitous use, common metaphors become more solidified into fixed expressions that add one more sense to the expression's meaning; making it unidentified as metaphorical.

Another reason is related to the scarcity of datasets that focus on metaphorical expressions to be used in different artificial intelligence tasks. This scarcity of resources is noticeable in languages that are understudied especially when compared to English;

more specifically, languages that are highly dialectical such as Arabic. The present project focuses on metaphor identification and detection from poems written in an Arabic sub-dialect spoken in Misurata. Based on my reading and studying of the projects conducted on this topic, no previous work is done to extract figurative text from Arabic poetry written in an Arabic sub-dialect. It is my desire to provide a state-of-the-art model that can be applied to detect metaphors from Arabic sub-dialect poems.

The following sections provide an explanation of the nature of metaphors in Arabic and a discussion of related work of metaphor identification in NLP. Before discussing the project in detail, a brief description of the Arabic language is presented.

Arabic Language

The Arabic language is among the top six official languages around the world which are recognized by the United Nations. Arabic is the official language of twenty-two countries in the Middle East. It is either the only official language or among the official languages of all countries in which the Muslims represent the majority of the population. The reason is that Arabic is the language of the Quran; the Islamic holy book. Therefore, it is the spiritual language that is used to perform the rituals and practice the religion of Islam. The Arabic language is spoken by more than 400 million people, most of whom are populated in these countries (Boudad et al., 2018).

Arabic is generally classified into three categories: The first category is referred to as Standard Arabic, which is also known as Classical Arabic. It is the language of the Quran and the religious practices in Islam. The second category is the Modern Standard Arabic (MSA) which is the formal variety of the language used in formal government documents, the news on TV, the newspapers, and academia. The third category of the Arabic language is dialectical Arabic. This is the informal variety of MSA which is spoken in everyday communications by the populations of the twenty-two countries. Each country speaks a major dialect unique to it and similar to the neighboring countries, but a bit distinct from other dialects spoken in the rest of the region.

Within each Arabic-speaking country, there are sub-dialects that branch out from the major dialect spoken in that country. These sub-dialects are spoken in different regions within the country and are categorized on different grounds. For example, they are usually classified in terms of the four directions (e.g the sub-dialect of the North, the subdialect of the South, etc.), or in terms of the geographical features (e.g the sub-dialect of the mountain area). In some cases, the classification is based on ethnic groups, in that the sub-dialects are classified as the variety spoken by city people vs. the variety spoken by the Bedouins (people who live in the desert areas). These dialects and sub-dialects are adopted by newborns prior to their exposure to MSA which is taken from TV cartoons and school. This, by virtue, makes the Arabic sub-dialect the first acquired variety of the language, and MSA is learned at a later stage of linguistic competence development. Since all these dialects stem out of the standard Arabic, they can be genuinely understood by everyone whose native language is Arabic and lived most of their life in the Middle Eastern region (Abugharsa, 2021, p. 106).

Similar to any other natural language, all the varieties of Arabic experience the process of constant change due to factors such as globalization. Moreover, many new borrowed words are introduced to Arabic and become standardized over time. It is only the Classical variety of the Quran that does not undergo these changes as it is fixed on a permanent style ever since it was created fourteen hundred years ago.

Metaphors in Arabic

Arabic rhetoric primarily take into consideration the semantic stylistics and gives the pragmatic accounts required for effective communicative skills. Gholami et al. (2016) state that “rhetoric is the flesh and blood of the Arabic language” (p.58) as it provides the stylistic mechanisms needed by the language users to fulfill the communicative needs for eloquently effective discourse. Arabic rhetoric is rich with metaphors that contribute to sharpening and upgrading the competence of the language.

A metaphor is one form of linguistic allegory of Arabic figures of speech. It can have a significant impact on general perceptions of reality and can manipulate people’s value systems and ideologies. The word for ‘metaphor’ in Arabic is ‘istiara’ which is derived from the verb ‘yasta’eer - to borrow, and ‘yo-eer’ - to lend’. As the Arabic name implies, the metaphor formation process involves borrowing an attribute from one entity and applying it to another entity. In the figurative speech, this procedure means turning the cognitive or abstract sense of a meaning that belongs to a given entity into concrete by assigning it to another unrelated entity in order to establish a ground on which the two significations (the cognitive and the concrete) meet. More discussion about the nature of a metaphor is provided next.

What is a Metaphor?

Figurative language is the process in which one, or more, senses of a word/phrase is used in a way that changes the standard unmarked literal meaning and adds a new exaggerated sense to that word. Examples of figurative language include the metaphors used in poetic rhetoric as well as regular conversations in order to emphasize a certain idea and make a strong impact on it.

One of the earliest definitions of a metaphor is presented by Aristotle (1997) who defines a metaphor as the "transference of a word of another significance either from genus to species, or from species to genus, or from species to species by analogy or proportion" (p.150). Another comprehensive definition of a metaphor is the one explained by Lakoff and Johnson (1980) who define a metaphor as the structured mapping between two frames of speech to describe an abstract concept in terms of a more concrete one.

A metaphor is one type of figurative language that exists in all human languages and thus is inherently valuable in these languages as it has both cultural and cognitive significance (Alkhatib & Shaalan, 2017). This makes the use of metaphors become more pervasive every day, in language as well as in thought and action, leading to the constant creation of new metaphors to fulfill the different linguistic needs that are called for different cultural communications. Over time, commonly used metaphors become part of a given natural language's lexicon; i.e., their connection to the metaphorical domain fades away due to the high conventionality of these expressions in familiar contexts (Shutova, 2010; Waite, 2018). One example is the word 'honey' used by husbands to address their wives. This word is conceived as a synonym to the word 'wife.'

A metaphorical expression generates from the association of at least two conceptual spaces; namely, the tenor space (or target concept), and the vehicle space (or source concept) within a specific context. This association results from violating the semantic constraints of the literal meanings of a given context. These constraints are referred to by Wilks (1978) as selectional restrictions. Metaphor spaces are represented in at least one word for each domain. The purpose is to associate a certain property of the source domain to the target domain in order to create a new concept in the target domain similar to that of the source domain. This similarity implies the characteristics in common between these two domains (Shutova, 2010). However, according to Yasen (2013), the metaphorical statement cannot be fully equivalent in meaning to the similarity statement because each statement has a truth condition different from the truth condition in the other statement. In this regard, the notion of similarity is crucial for context-dependent metaphor analysis.

The aspect of similarity in metaphors brings the discussion to the difference between a metaphor and a simile. A metaphor is sometimes defined as a simile without using the words ‘like’ and ‘as’. For example, ‘life is like a battle’ is a simile whereas ‘life is a battle’ is a metaphor. The two sentences include examples of figurative speech in which the comparison is used with the element of exaggeration; only in the simile, the comparison is explicitly stated by using the word ‘like’ but in the metaphor, it is implied. In fact, the implied comparison in metaphors can be hidden more when one of the metaphor components is dropped. More discussion on this point is provided in the following two sections.

Components of a Metaphor

Linguistically speaking, there are three components of a metaphor. These components are the topic (tenor or target), vehicle (source), and ground. The topic is an entity described by the metaphor, and the vehicle is the notion to which this topic element is being compared through the mapping process in which relevant information are activated and irrelevant features are eliminated. The similarity base on which this comparison is established is referred to as the ground (Alkhatib & Shaalan, 2017). Zhang and Hu (2009) state that factors such as cultural characteristics, cognitive models, and background knowledge, among others, can all contribute to the structure of the ground in the metaphor. Consider the following example:

- The spine of the book.

The tenor is the hard paper part in the book's binding to which all the book papers are attached from one of their vertical sides. The vehicle is the human spine, and the ground is the similarity between the hard part of the book and the human spine, in that both of them are used as supporters for other parts of the entity to which they belong.

The topic or tenor is also known as the target domain and the vehicle is referred to as the source domain. As seen in the example above, the source domain is about the bodily 'physical' experience that includes the concrete concept (the human spine in the example) which is used to create and describe the abstract idea or concept (the book spine) in the target domain.

Types of Metaphors

Based on the discussion presented by Alkhatib and Shaalan (2017), metaphors can be classified into two major types: cognitive metaphors and declarative metaphors. In declarative metaphors, the vehicle (source) is stated explicitly while the tenor (target) is hidden. For example, one of the common metaphors in Arabic is to compare a beautiful woman's appearance and elegant figure to the deer because it has a beautiful and graceful body. As a declarative metaphor, we can say something like 'I saw a deer at the party' to mean 'I saw a beautiful lady.' In this metaphor, there is no mention of the words 'woman' or 'lady' that carry the target concept of the metaphor. The reason is that the conventional concept of the metaphor is well-known due to the strong connection between the words 'deer' and 'beautiful woman' that this culture makes in terms of beauty and elegance.

On the other hand, in a cognitive metaphor, the process is reversed. The tenor is provided and the vehicle, though not straightforwardly presented, is implied by the use of a noun or a verb that always collocates with it. For example, in the metaphor 'the soldier roars on the battlefield', the verb 'roar', which is associated with the lion, is used to describe the heroic actions of the soldier in order to assimilate his bravery to that of the lion's. This mapping is explained by Alkhatib and Shaalan (2017) in that the focus in cognitive metaphors is "on the denotation rather than the connotation of the metaphor that addresses the receptor in order to highlight its cognitive function." (p. 172).

The dataset used in the current project includes many examples of a declarative metaphor, but not a single example of a cognitive metaphor. This can be used to infer significant information about the rhetorical, linguistic, and cultural characteristics that this variety of Arabic features.

Metaphors in Natural Language Processing

Modeling figurative language is one of the challenging tasks in natural language processing. It is described by Shutova (2010) as a “serious bottleneck in automatic text understanding” (p. 688). One possible reason could be related to the external form of metaphors, the lack of clear-cut semantic distinction between the two domains, and the variation of metaphor use. Algorithms are developed for automatic metaphor identification of text as either metaphorical or literal. This differentiation is required in many applications as a highly important process in natural language understanding. Applications need to have a module that can be trained on the metaphorical knowledge in order to understand the difference between ‘give me the book’ vs. ‘give me a break’, ‘cut the paper’ vs. ‘cut the act’, and ‘a broken glass’ vs. ‘a broken heart’, etc.

One of the first works in automatic processing of figurative language dates back to the early nineties when Fass (1991) developed an approach to distinguish between literary expressions vs. metaphors, metonymy, and anomaly. The distinction is based on the violation of selectional preference presented by Wilks (1978). However, this approach has limitations due to the high conventionality of common metaphors, and other pervasively used metaphors that regularly get conventionalized. This makes it hard for the approach to extract the correct sectional distributions of different expressions.

Another limitation is that text interpretations are context-dependent and are naturally culture-specific. Therefore, hand-coded information is affected by subjectivity which can be a factor in determining the classification of a text as figurative or literal; i.e., humans may provide text classifications that are not similar from one human to another. Moreover, this classification can be cultural and/or context-dependent, and this increases the subjective component. Data annotation in this project is performed by a group of expert linguists who work together as a team to annotate and revise the metaphoricity in the dataset. Since those linguists live in Misurata and speak the Misurata sub-dialect, they are familiar with the contextual and cultural factors associated with the creation and use of metaphors in these poems.

The recognition and interpretation of metaphors are indispensable in any language's semantic-oriented natural language processing. The automatic identification and analysis of metaphors play a crucial role in fulfilling tasks in different fields such as Computational Linguistics, Cognitive Linguistics, and Machine Translation. Tasks in these fields require access to useful information about understanding a given text as literal or metaphorical. Shutova (2010) manually annotated metaphorical text in The British National Corpus which contains different genres and found that more than 90% of the expressions are metaphorical. In this regard, Shutova (2010) states that Semantic-oriented NLP applications can be applied to recognize and interpret figurative language in datasets of any genre.

Since metaphors are culture-specific and highly complex in nature, they can have a semantic effect on the meanings of machine-translated texts in which two languages are involved. Tsvetkov et al. (2014) applied a metaphor identifier model that is originally designed to detect English metaphors and used it to detect metaphors in texts written in

Spanish, Farsi, and Russian. The model was applied to these three languages without making any adoption to it. The results show a high level of accuracy and a state-of-the-art performance of this model in all of these languages including English. However, this experiment is limited to a certain category of metaphors that have two kinds of syntactic relations; namely, word order (Subject-Verb-Object) and Adjective-Noun compositions. Accordingly, different results can be obtained if the model is applied to different categories of metaphors that may not have similar patterns in all the languages used in the project.

It is taken that cross-lingual metaphor detection can be a challenging task since syntactic and semantic patterns differ across languages. Each language is created and developed within a culture that contains norms and life aspects which are specific to that culture, like the ‘dear’ example discussed above.

Related Work

Previous works of research have focused on metaphor detection in English, using both corpus-based approaches (Birke and Sarkar, 2007; Boudad, 2018; Gerow and Keane (2011); Hovy et al., 2013; Shutova, 2010; Neidlein, et al., 2020; Neuman et al., 2013) as well as manually-created linguistic resources (Abugharsa, 2021; Broadwell et al., 2013; Gedigian et al., 2006; Krishnakumaran and Zhu, 2007; Mason, 2004). These studies, in addition to others, were conducted on many kinds of linguistic genres in different languages.

One kind of these studies focuses on identifying metaphors in different kinds of articles. These metaphors play an important role in the impact that the news language plays to emphasize the importance of the event. Waite (2018) investigated the possibility of using modeling algorithms to classify newspaper headlines that contain metaphors. Neuman et al. (2013) also investigated metaphors in newspaper articles by evaluating three algorithms to classify metaphorical vs. literal text and compare the algorithms' performance to that of human annotators. In their research, Gerow and Keane (2011) used a corpus of financial news reports to analyze the distribution of up and down verbs that are used in metaphorical expressions to describe the movements of stocks and shares in hierarchical structures in subordinate and superordinate groups.

The general principle of automatic metaphor identification as discussed in many related projects is to analyze the semantic patterns between the metaphorical expression and its semantic/syntactic dependencies. In other words, the phrase's abstractness and distributional properties are considered in order to capture the contextual nature of the figurative text (Turney et al., 2011). As discussed earlier, the pervasiveness of most metaphors makes it difficult to obtain resources that spot the clear-cut distinction between the metaphorical and literal uses of each word. Metaphor detection at the word level is also investigated by Choi et al. (2021) who focused on detecting the words that make a given text metaphorical by adapting a pre-trained contextualized BERT model. A similar study is presented by Mao et al. (2018) who proposed a method that uses unsupervised data to identify and interpret words at the word level.

Other strategies are mainly based on the association that holds between the source domain and the target domain (Shutova, 2010; Swarnkar & Singh, 2018; Xiao et al., 2016). In their investigation of metaphor automatic extraction, Heintz et al. (2013)

applied the Latent Dirichlet Allocation approach in light of the assumption that metaphorical expressions contain vocabulary related to a source domain and a target domain, and achieved good results. These approaches implement defined mappings of metaphors from the source domain to the target domain. For example, X is like Y in: ‘attitudes are contagious.’ However, such a strategy may not be applicable to general novel metaphors as it may not be feasible to list and categorize all possible mappings of a given word. Accordingly, data processing in the current project does not involve the mapping of the source domain and the target domain of the metaphors in order to avoid the obstacle of limiting the mappings of these words to the dataset under investigation.

In their work, Tsvetkov et al. (2013) based their model on the literal interpretations of the words by hypothesizing that metaphors are manifested as unusual semantic compositions with syntactic relations. These relations are structure-specific in that any violation of the structure generates new syntactic relations that are considered an indication of a potential metaphor. For example, the phrase ‘eat my lunch’ does not have a violation of the syntactic structure that connects the verb ‘eat’ with the noun ‘lunch’. On the other hand, ‘eat my brain’ includes a violation of the rule that does not connect this verb to this noun; accordingly, metaphoricity likelihood arises.

This approach is similar to the one presented by Shutova et al. (2010) who introduced a bottom-up method in which a set of seed metaphors are used to detect possible metaphorical texts that have a verb-noun relation identical to that in the cluster of similar verb-noun seed examples. Additionally, Neidlein et al. (2020) used word synonyms to detect metaphors from unseen data by relating metaphorical words to synonymous words in the trained data. These statistics are also used in other projects to determine metaphoricity candidacy which requires word-sense borrowing in light of the supersense relationship that corresponds to the literal usage of parts of speech (Neuman et al., 2013). A previous study by Turney et al. (2011) also used word-sense borrowing in automatic metaphor extraction with the hypothesis that metaphorical word senses are related to a degree of abstraction in context words. This abstraction is analyzed to classify word senses as either metaphorical or literal.

Metaphor recognition does not only help identify the purpose of the given text based on its figurative vs. literal meanings, but it also helps understand other aspects beyond the linguistic scope. Certain patterns of word/phrase co-occurrence can be detected by analyzing large corpora of data. This, in turn, can reveal significant insights into certain linguistic behaviors in some cultures (Neuman et al., 2013). Such findings show how metaphors can organize our conceptual system which regulates how we think and act (Shutova, 2010).

The Use of LSTM in Automatic Processing of Figurative Language

LSTM, a variant of Recurrent Neural Network (RNN), is an architecture that is pervasively used in sequence predictions. The main reason is that LSTM is capable of handling long-term dependencies in sequential data as well as vanishing potential gradient problems. LSTM computes the context vector at each word in the global word sequence which is encoded by a recursive computation of context vectors.

The LSTM model has been used in other studies to detect metaphors (Pramanick et al., 2018; Swarnkar & Singh, 2018; Wu, 2018; Liu, 2020). It is used in combination with other layers such as Convolutional Neural Network (CNN), Conditional Random Field (CRF), parts-of-speech (POS), and WordNet-based features. In addition, the bi-directional LSTM model (or Bi-LSTM), where both directions are trained, has performed efficiently in many tasks that work on the automatic detection of figurative language. One example is the model proposed by Kuo and Carpuat (2020) whose study on metaphor detection in TOEFL essays shows that Bi-LSTM outperforms feature-rich linear models.

Another example is the study conducted by Gao et al. (2018) in which they used the standard Bi-LSTM model to detect verb metaphor benchmarks at the sentence level and showed a remarkable performance in the task of metaphoricity prediction. In their approach, Sun and Xie (2017) developed an LSTM model to detect metaphors based on the sentences' subject-verb-object relation. Their model constructs text feature representations for metaphoricity identification. Their results show state-of-the-art performance on metaphor detection of LSTM sub-sequence models.

As discussed above, some previous studies on automatic metaphor detection have focused on literal vs. metaphorical senses of words (Goatly, 1997; Karov & Edelman, 1998; Lonneker-Rodman. 2008; Reining & Lonneker-Rodman, 2007). Other studies propose identifying metaphors in light of the POS associations; i.e. the relationship between certain parts of speech used to detect metaphors (Birke & Sarkar, 2006; Gedigan et al., 2006; Krishnakumaran & Zhu, 2007). A third method includes identifying the source domain and target domain of words (Barnden & Lee, 2002; Agerri et al., 2007; Feldman & Narayanan, 2004; Narayanan, 1999).

It can be observed that most automatic metaphor detection methods are based on word-level analysis by selecting certain target words to investigate their metaphoricity. The current study is based on the co-occurrence of words in that the model detects the words that are used in metaphorical texts and learn to detect similar relationships in unseen data.

The Significance of the Current Project (The Motivation)

The dataset used in this project is extended from a shorter dataset used in my previous project (Abugharsa, 2021) which focused on detecting sentiment polarity from poems written in the Misurata Arabic sub-dialect. The aim of the current project is to design a model that can detect metaphors from these poems. As in my previous project, this detection is based on poetic verses which are linguistically and rhetorically related to each other in the poems. In other words, the verses are semantically, linguistically, and contextually connected to the main idea of the poem. As a result, these verses are different from the isolated reviews taken from social media platforms which are produced by many people rather than a single poet as in the case of poetry.

In order for the model in this project to be used to detect metaphors in isolated comments on different social media platforms, the nature of these comments needs to be investigated. Most of the Arabic reviews on social media are written in informal Arabic; i.e. dialects and sub-dialects rather than the standard Arabic. Therefore, it becomes crucial to train computer models to detect different figurative linguistic features in these reviews as an initial step to efficiently perform other tasks such as machine translation, topic modeling, and sentiment analysis.

Other projects about detecting figurative language in Arabic have focused on poems written in Standard Arabic, and aimed at classifying these poems into two categories: MSA poetry (Ahmed et al., 2019) and old Arabic poetry (Alsharif et al., 2013). To the best of my knowledge, no previous NLP research on processing figurative language in Arabic has tackled a sub-dialect variety of the language, not to mention the genre of poetry.

Hopefully, this project can lead to more studies that handle other NLP tasks and highlight more aspects of this variety of Arabic. The following sections discuss the project methodology and provide a description of the LSTM algorithm. The results obtained from using the model and other findings are also discussed.

Methodology

In this section, the dataset is introduced, followed by a discussion about data annotation, then a discussion about data cleaning and preprocessing. After that, the process of building the model is explained. Finally, an evaluation of the model is discussed.

Dataset

The dataset contains verses from poems written in the Misurata Arabic sub-dialect. These poems are written by six poets who speak Arabic as a native language and are all born and raised in Misurata. The poems include terms and expressions that are specially used for this type of literature. This can make data annotation a challenging process for individuals who are not familiar with the linguistic structure of these poems, regardless of the fact that those individuals are native speakers of the language. Therefore, the seed data annotation is performed by professional linguists in the Arabic Department at Misurata University.

Data Annotation

The dataset was initially annotated and revised by faculty and graduate students in the Arabic department at Misurata University; i.e. experts in the field. During the process of annotation, the poets were frequently contacted and asked to provide explanations of some words and expressions used in their poems. This was crucial to determine the metaphoricity of the poem verses and obtain the most accurate annotation possible.

After the initial annotation of the dataset, another annotation is performed by three layman individuals who are also native speakers of the Misurata sub-dialect.

However, those individuals are not professionally acquainted with the morphological and semantic principles that this literature is based on. The three annotators were each provided with a subset of the data that represents 25% of the whole dataset; it is the same percentage of the part of the data used for the test set in the model.

The annotation performance of the layman annotators is revised by the faculty team at Misurata University and an evaluation of the performance is provided. This performance is compared to the output of the model built in the project in order to evaluate the model's level of accuracy in detecting metaphors.

Data Cleaning and Preprocessing

As already discussed, the seed data of the dataset are provided by a group of linguists who are faculty and graduate students in the Arabic Department at Misurata University. The annotation was revised multiple times before it was approved to use in this study. Therefore, the dataset contains no outliers, noise data, special characters, or empty lines.

The data is cleaned by removing diacritics, punctuations, conjunctions, and hyphens in order to normalize all the data text into a unified form. Figure 1 shows the raw data before the preprocessing step. Data normalization includes replacing {أ, إ ,آ} with {ا}, also {ئ ,ي} with {ى}, and {ة} with {ه}. Figure 2 shows the data after being cleaned.

Sentence	Metaphor
في غربة لكن ما نستها	non-metaphorical
و فالقلب تعيش مدينتها	metaphorical
و قالت أمي و ما ننساها	metaphorical
بكحل الروح رسمت ضناها	metaphorical
اللي ضحو بالروح فداها	non-metaphorical
تاريخ موتق بصمتها	metaphorical
فعلاً قلم شموخ و عزة	metaphorical
أصيلة مصراتة يا عزة	metaphorical
إنتي جزء و ما يتجزأ	metaphorical
م الثورة و رسمة حالتها	metaphorical
ف الغربة لكن ما انستها	non-metaphorical
و القلب تعيش مدينتها	metaphorical

Figure 1. Raw Data Before Preprocessing

Sentence	Metaphor
غربه نستها	non-metaphorical
فالقلب تعيش مدينتها	metaphorical
قالت امي ننساها	metaphorical
بكحل الروح رسمت ضناها	metaphorical
اللي ضحو بالروح فداها	non-metaphorical
تاريخ موتق بصمتها	metaphorical
فعلا قلم شموخ عزه	metaphorical
اصيله مصراته عزه	metaphorical
انتي جزء يتجزا	metaphorical
الثوره رسمه حالتها	metaphorical
الغربه انستها	non-metaphorical
القلب تعيش مدينتها	metaphorical

Figure 2. The Preprocessed Data

After cleaning the data, the TF-IDF frequency algorithm is applied to reduce the size of the overused words and convert the text into sequences of vector representations in order for the network to deal with it as input.

Building the Model

The dataset includes more than twenty thousand Arabic poetic verses that are labeled as either metaphorical or non-metaphorical. First, the zero-padding parameter is implemented to convert all the vectorized verses into the same length. The deep learning methods employed in this project include the sequential model which is used as a baseline and the LSTM sequential labeling of the transformed vector sequence into a single vector. This helps detect contextual information that constitutes metaphors in terms of word order and vectorized sequence of words.

Text metaphoricity is determined based on word embeddings that occur together in metaphorical phrases. The layers added to the model include a vocabulary size of 10000 and an embedding dimension which equals 512. Since the LSTM model predicts the classification of the verses based on the co-occurrence of words, the categorical cross-entropy loss function is used, and the model is compiled with adam optimizer. After adding two LSTM layers, a dense layer (softmax) is created to detect metaphoricity. Figure 3 shows the general architecture of the model.

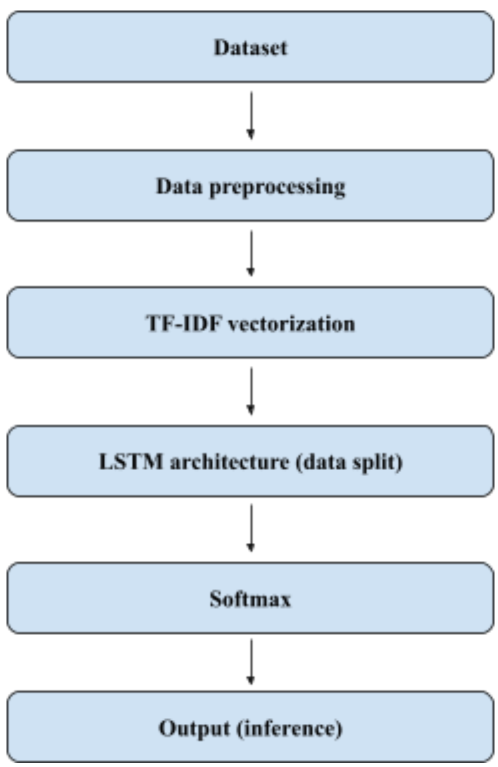


Figure 3. The Architecture of the LSTM Model

The Python libraries used in this project are primarily Keras and Sklearn. At all stages of the research, various default values for the parameters (e.g. number of epochs, vector size, activation functions, and so on) are tuned in order to understand how these modifications affect the overall performance of the model.

Model Evaluation

To train the network, the size of the data is designed to equal 10. The LSTM model is implemented to extract features that can be used to detect metaphors. These features are then fed into two LSTM layers that treat context and word ordering at a more sophisticated level than the usual Bag of Words approach. To fit the model into the data, the network is trained by using 3 epochs. Cross-validation is used during the training phase as the evaluation method according to best practice. The results are measured by accuracy.

Findings

The findings include the level of accuracy of metaphor detection performed by the layman individuals compared to the score of the model performance. Table 1 presents these findings.

<table>
<tr><th rowspan="3">Accuracy</th><th rowspan="3">Layman Annotators</th><th colspan="4">LSTM Model</th></tr>
<tr><th rowspan="2">Accuracy</th><th rowspan="2">Loss</th><th colspan="2">Validation</th></tr>
<tr><th>Accuracy</th><th>Loss</th></tr>
</table>

Total Accuracy Score	**68.2%**	**79%**	**0.5%**	**5%**	**5%**
Metaphorical Accuracy	**51%**	**59%**	_____	_____	_____
Non-metaphorical Accuracy	**56%**	**73%**	_____	_____	_____

Table 1. Performance of Layman Group Annotators and the LSTM Model

According to the data presented in Table 1, the overall accuracy score of the layman group is 68.2% while the model gets an accuracy score of 79% with an overall loss of 5%. It should be stated that the training dataset is unbalanced; it includes 12869 metaphorical phrases vs. 7902 non-metaphorical phrases. This can have an effect on the overall accuracy of the output.

The first epoch gives a 6% accuracy score and the last epoch outputs an accuracy score of 8%; these are good results of the model. Figure 4 illustrates the training and validation accuracy values, and Figure 5 presents the training and validation loss values.

Figure 4. LSTM Model Accuracy

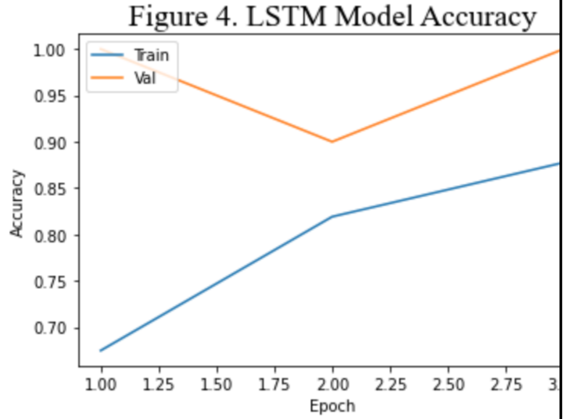
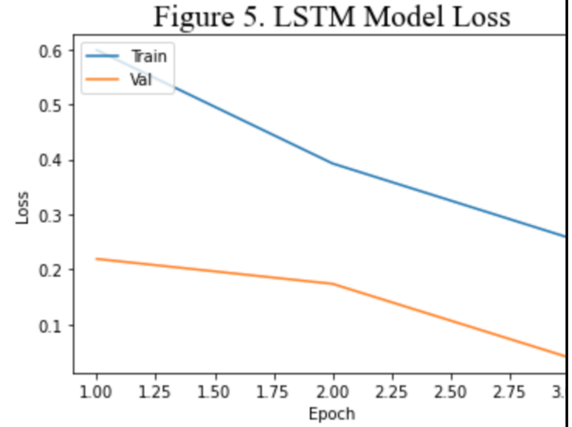


Figure 5. LSTM Model Loss



The model accuracy score was lower when the dataset was smaller in size. When more poems were added to the dataset, the model accuracy score became higher. This means that it is possible for the accuracy score to get even higher if the size of the dataset is larger.

In addition, the layman group obtained a metaphorical accuracy of 51% and a non-metaphorical accuracy of 56%. On the other hand, the accuracy score of detecting metaphorical utterances obtained by the LSTM model is 59% while the non-metaphorical utterances reached a level of accuracy of 73%. As noticed, both the model and the layman group performed better at detecting non-metaphorical text than the metaphorical text.

Nonetheless, the results show that the model performs better than the layman group. One possible interpretation is that the model is trained on this kind of data while the layman individuals are not. As discussed earlier, the text analyzed is highly poetic and is most familiar among those who study Arabic linguistics and literature. In addition, this kind of poetry is mostly known by those who have a refined background in this genre such as the poets themselves and individuals who show special interest in this kind of literature. It is a rich area for more investigation in the future.

Blind Set Evaluation

In addition to the accuracy scores obtained by the model, the model managed to work on a blind dataset to predict metaphoricity in totally new poems. The blind dataset is composed of 1000 verses that represent phrases in poems that the model never worked on. The model predicted metaphoricity with an 80.7% accuracy score. LSTM has proven to be an effective approach for capturing internal representations in Arabic sequential figurative textual data. This is confirmed by Strobelt et al. (2017) whose empirical work on LSTM as a tool to analyze visual hidden states in RNNs indicates that LSTM learns "to capture complex relationships between the words within a sentence or document" (Strobelt, 2017, p. 668).

Additional Findings

Data analysis has resulted in more findings than the accuracy of metaphoricity prediction. Two features are observed: One of them is the syntactic distribution of the metaphorical phrases in light of the source domain and the target domain. The other observation is about the relationship between text sentiment polarity and metaphoricity.

By observing the grammatical description of the words in the source domain and the target domain of the metaphorical phrases, it is noticed that the words included in the source domain are verbs and adjectives. Another observation is that while there are no hidden source domain words, some target domain words are hidden. Based on the discussion of the types of Arabic metaphors in section 3.2 above, the metaphors in the dataset are declarative metaphors. This is one characteristic to learn about Misurata literature.

Besides labeling metaphors, sentiment orientation is also labeled for each verse in the dataset. The sentiment labels are positive, negative, and neutral. The results show that most of the phrases contain positive metaphors. Table 2 shows the breakdown of the relationship between sentiment labels and metaphoricity in percentage.

Metaphoricity	Sentiment Labels		
	Positive	Negative	Neutral
Metaphorical	54%	45%	1%
Non-metaphorical	51%	46%	3%

Table 2. Sentiment Labels and Metaphoricity

Apparently, positive and negative metaphors display significant importance in Misurata sub-dialect poetry with the positive metaphors being relatively more frequent. Figure 6 illustrates the findings in Table 2.

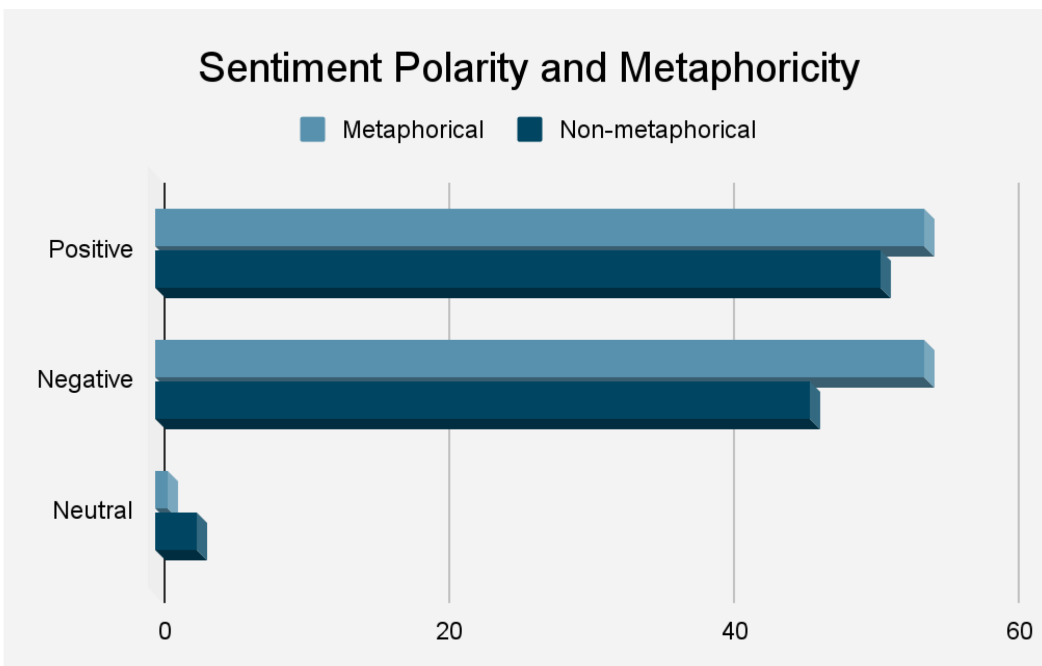


Figure 6. Sentiment Polarity and Metaphoricity

The data provided cast light on the nature of poetry written in one Arabic sub-dialect. It follows that more investigation in this area can reveal more information about this kind of Arabic literature.

Conclusion

This project presents a study about detecting metaphors in poems written in Misurata Arabic sub-dialect; a variety of Arabic that has not been sufficiently investigated in NLP before. The LSTM model achieved an accuracy level of a maximum of 79% with an overall loss of 0.5%. Furthermore, the model has shown a significant performance with a high accuracy score in predicting figurative text in blind data. In addition to these results, dataset analysis shows that the poetic utterances are genuinely positive declarative metaphors.

The model's performance efficiency can be tested by applying it to other varieties of Arabic sub-dialects and comparing the results with the findings of the current project. Furthermore, the text the model works on is poetic which is syntactically and morphologically different from other non-poetic texts. As a result, investigating the performance of the model with different language varieties of Arabic lays out potential avenues for future NLP research in different tasks such as machine translation, topic modeling, and sentiment analysis. This project can stand as a contribution to research in the field of Computational Linguistics in Arabic as a model that serves the needs of Arabic speakers who use their native sub-dialect in their daily conversations.

References

Abugharsa, A. (2021). Sentiment Analysis in Poems in Misurata Sub-dialect, A Sentiment Detection in an Arabic Sub-dialect. *International Journal of Computers and Technology*, *21*(21), 103-114.

Agerri, R., Barnden, J. A., Lee, M.G. and Wallington, A.M. (2007). Metaphor, inference and domain independent mappings. In *Proceedings of RANLP2007*, 17-23, Borovets, Bulgaria.

Alkhatib, M. and Shaalan, K. (2017). Natural language processing for Arabic metaphors: A conceptual approach. *Advances in Intelligent Systems and Computing*, *533*, 170-181.

Aristotle. (1997). *Poetics*. (G. Walley, Trans.). McGill-Queen's University Press.

Birke, J. and Sarkar, A. (2006). A clustering approach for the nearly unsupervised recognition of nonliteral language. In *Proceedings of EACL-06*, 329-336.

Boudad, N., Faizi, R., Thami, R., & Chiheb, R. (2018). Sentiment analysis in Arabic: A review of the literature. *Ain Shams Engineering Journal, 9*(4), 2479-2490.

Barnden, J.A., and Lee, M.G. (2002). An artificial intelligence approach to metaphor understanding. *Theoria et Historia Scientiarum*, 6(1):399-412.

Broadwell, G., Boz, U., Cases, I., Strzalkowski, T., Feldman, L., Taylor, S., Shaikh, S., Liu, T., Cho, K., & Webb, N. (2013). Using imageability and topic chaining to locate metaphors in linguistic corpora. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, 102-110. Springer.

Choi, M. , Lee, S., Choi, E., Park, H., Lee, J., Lee, D., & Lee, J. (2021). MelBERT: Metaphor detection via contextualized late interaction using metaphorical identification theories. *arXiv preprint arXiv:2104.13615*.

Fass, D. (1991). met*: A method for discriminating metonymy and metaphor by computer. Computational Linguistics, 17(1): 49-90.

Feldman, J. and Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language*, 89(2):385-392.

Gao, G., Choi, E., Choi, Y. & Zettlemoyer, L. (2018). Neural metaphor detection in context. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing.*

Gedigian, M., Bryant, J., Narayanan, S., & Ciric, B. (2006). Catching metaphors. In *Proceedings of the 3rd Workshop on Scalable Natural Language Understanding*, 41-48.

Gerow, A., & Keane, M. (2011). Identifying Metaphor Hierarchies in a Corpus Analysis of Finance Articles. 2049- 2054. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*: Boston, MA.

Gholami, S., Montashery, I., & Khorrami, M. (2016). Metaphors in Arabic and English texts: A case study of translation of metaphor in the English version of Al-Sahifah Al-Sajjadiyyah. *Journal of Literature, Languages and Linguistics*, *21*, 56-66.

Goatly, A. (1997). *The Language of Metaphors*. Routledge, London.

Heintz, I., Gabbard, R., Srinivasan, M., Barner, D., Black, D., Freedman, M., and Weischedel, R. (2013). Automatic extraction of linguistic metaphor with LDA topic modeling. In *Proceedings of the First Workshop on Metaphor in NLP, 2013*, 58-66.

Hovy, D., Srivastava, S., Jauhar, S., Sachan, M., Goyal, K., Li, H., Sanders, W., & Hovy, E. (2013). Identifying metaphorical word use with tree kernels. *Association for Computational Linguistics*, 52-57.

Karov & Edelman, Y. (1998). Similarity-based word sense disambiguation. *Computational Linguistics*, 24(1):41-59.

Krishnakumaran, S., & Zhu, X. (2007). Hunting elusive metaphors using lexical resources. In *Proceedings of the Workshop on Computational approaches to Figurative Language*, 13-20.

Kuo, K., & Carpuat, M. (2020). Evaluating a Bi-LSTM model for metaphor detection in TOEFL essays. In *Proceedings of the Second Workshop on Figurative Language Processing*. 192-196.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.

Liu, J., O'Hara, N., Rubin, A., Draelos, R., and Rudin, C. (2020). Metaphor detection using contextual word embeddings from transformers. In *Proceedings of the Second Workshop on Figurative Language Processing*, 250-255.

Lonneker-Rodman, B. (2008). The hamburg metaphor database project: issues in resource creation. *Language Resources and Evaluation*, 42(3):293-318.

Mao, R., Lin, C., & Guerin, F. (2018). Word embedding and wordnet based metaphor identification and interpretation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1222-1231.

Mason. Z. (2004). CorMet: a computational, corpus-based conventional metaphor extraction system. *Computational Linguistics*, *30*(1), 23-44.

Narayanan, S. (1999). Moving right along: A computational model of metaphoric reasoning about events. In *Proceedings of AAAI 99*, 121-128, Orlando, Florida.

Neidlein, A., Wiesenbach, P., & Markert, K. (2020). An analysis of language models for metaphor recognition. In *Proceedings of the 28th International Conference on Computational Linguistics.* 3722-3736.

Neuman, Y., Assaf, D., Cohen, Y., Last, M., Argamon, S., Howard, N., & Frieder, O.(2013). Metaphor identification in large texts corpora. *PloS one*, 8(4):e62343.

Pramanick, M., Gupta, A., & Mitra, P. (2018). An LSTM-CRF based approach to token-level metaphor detection. In *Proceedings of the Workshop on Figurative Language Processing*. 67-75.

Reining, A. and Lonneker-Rodman, B. (2007).Corpus-driven metaphor harvesting. In *Proceedings of the HLT/NAACL-07 Workshop on Computational Approaches to Figurative Language*, 5-12, Rochester, New York.

Shutova, E. (2010). Models of metaphor in NLP. *Association for Computational Linguistics*. 688-697.

Strobelt, H., Gehrmann, S., Pfister, H., and Rush, A. (2017). LSTMVis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. *IEEE transactions on visualization and computer graphics*, *24*(1), 667-676.

Sun, S. and Xie, Z. (2017). Bilstm-based models for metaphor detection. In *The National CCF Conference on Natural Language Processing and Chinese Computing*, 431-442. Springer, Cham.

Swarnkar, K., & Singh, A. K. (2018). Di-LSTM contrast: A deep neural network for metaphor detection. In *Proceedings of the Workshop on Figurative Language Processing*. 115-120.

Tsvetkov, Y., Boytsov, L., Gershman, A., Nyberg, E., & Dyer, C. (2014). Metaphor detection with cross-lingual model transfer. *Association for Computational Linguistics*, 248-258.

Turney, P., Neuman, Y., Assaf, D., & Cohen, Y. (2011). Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the 2011 Conference on the Empirical Methods in Natural Language Processing*. 680-690.

Waite, R. (2018). Using NLP to spot metaphors-The New York Times challenge. Winton. Retrieved September 22, 2021, from https://tech.winton.com/2018/01/metaphor-competition/

Wilks, Y. (1978). A referential pattern-seeking semantics for natural language inference. *Artificial Intelligence*. 6, 53-74.

Wu, C., Wu, F., Chen, Y., Wu, S., Yuan, Z., & Huang, Y. (2018). Neural metaphor detecting with CNN-LSTM model. In *Proceedings of the Workshop on Figurative Language Processing*. 110-114.

Xiao, P., Alnajjar, K., Granroth-Wilding, M., Agres, K., & Toivonen, H. (2016). Meta4meaning: Automatic metaphor interpretation using corpus-derived word associations. *Proceedings of the Seventh International Conference on Computational Creativity*.

Yasen, A. (2013). The commonest types of metaphor in English. *Journal of Al-Frahedis Arts*, *5*(15), 52-69.

Zhang, F., & Hu, J. (2009). A study of metaphor and its application in language learning and teaching. *International Education Studies*, *2*(2), 77-81.