* **Problem definition**

Opinion mining task “detecting the people opinion about something” is still an open research area especially in Arabic language. Recently, neural network models have been used to learn semantic representations for NLP tasks (Le and Mikolov, 2014; Tang et al., 2015), achieving highly competitive results and we trying to get the advantages from those techniques in the opinion mining task. For example distributed word representations “Word Embedding” (Mikolov et al., 2013) have been used as the basic building block by most models for NLP. Numerous methods have been proposed to learn representations of **phrases and larger text segments** from distributed **word representations.**

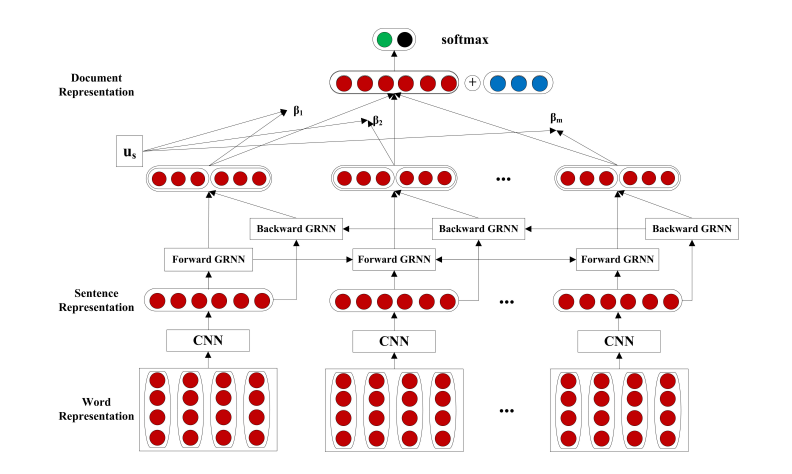
Convolutional neural networks have been widely used for semantic composition automatically capturing n-gram information. Sequential models such as recurrent neural network or long short-term memory (**LSTM**) (Li et al., 2015a; Tang et al., 2015) have also been used for recurrent semantic composition. The attention mechanism was first proposed in machine translation (Bahdanau et al., 2014). We explore CNN and recurrent neural networks with attention mechanism to learn document representation for detecting opinion mining in Arabic language.

* **Solution**

The main target of this work is to explore the opinion mining problem in Arabic specifically for the **Khalij dialects**. So we going to build a framework that can automate this task.

As shown in figure 1 the solution will be:

1. Crawl the khalij tweets from twitter and store them into intermediate data storage.
2. Apply some data cleansing rules like removing numbers, removing elongation, lemmatization, POS and NER “Named Entity Recognition”.
3. Create some features that will contribute in the model building step.
4. Run the model as depict in figure 2 which is a hybrid of neural network feature and discrete feature as attention where red nodes represent neural features, and blue nodes represent discrete features the first layer for the model will detect the word embeddings, the second layer we use the CNN to learn the sentence embedding trying to detect the n-gram and the final layer (LSTM) to classify the tweets actually this step will be confirmed during the experiments + the discrete features in the attention mechanism.



**Figure 1: Neural network model structure for opinion mining detection**

**Crawl Reviews**

**Review Database**

**Lemmatization**

**Remove numbers**

**Remove elongation**

**Data cleaning**

**Discrete Features**

**Run The Model**

**Output**

**Figure 2: Opinion spamming framework architecture**

* **References**

[1] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems.

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[3] Abu Bakr Soliman, Kareem Eisa, and Samhaa R. El-Beltagy, “AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP”, in proceedings of the 3rd International Conference on Arabic Computational Linguistics (ACLing 2017), Dubai, UAE, 2017.