Sentiment Analysis Model Based-CNN and extended transition-based Medical Concepts Recognition

HANANE GRISSETTE [1], NFAOUI ELHABIB [2], .

[1,3]Sidi Mohamed ben Abdellah University, Department of computer science , Morocco
[2]Laboratory of informatics, imaging and numerical analysis (LIIAN) .
e-mail(s): [1]Hanane.grissette@usmba.ac.ma, [2]elhabib.nfaoui@usmba.ac.ma,

OVERVIEW

About twenty billion are typed in Twitter every day, a new Drugs, Diseases, Adverse Drug Reactions (ADRs) keep appearing in new Unicode versions. Generalized Model for both Sentiment analysis (SA) and Medical Concepts Recognition (MCR) Requires human labeled data or make use of resources for weak supervision. In this work, we propose a Sentiment Analysis Model based-CNN and extended transition-based Medical concepts recognition. In order to define a related-sentiment discrimination ability regarding Medical concepts from daily patients self-reports.

INTRODUCTION

Patients and health consumers shared their experiences and their related-treatment opinions on Social media, whereby they describe all the incredibly complex processes happening in the real-time treatment.
the tweet in figure .1 3. could contain various related-medical concepts, such as drug names, Doctors, ADRs.

Figura 1 – a tweet Posted on 2019-05-26 at 13:56:37.864000
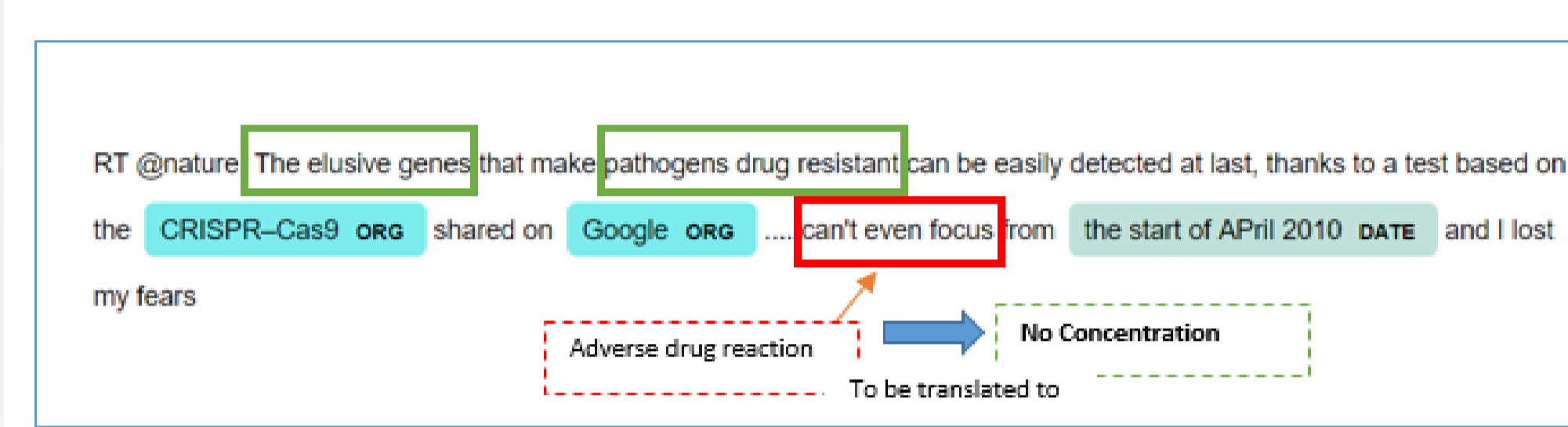
Aim :
- *Define a dynamic model vectorization regarding medical corpora from online messages For Sentiment prediction Model.*

Hypothesis :
- *Deep Neural network learns more the semantic and conceptualized information under distributed representations by incorporating external knowledge.*
- *transition-based dependency an efficient strategy with unkown subword fearures that is used to support huge medical vocabularies from social media.*

MAIN CONTRIBUTIONS

as dispected in figue 2. The proposed Neural Network layers is shared between medical concept recognition model and Sentiment prediction.

- **sentiment prediction :** Stacked-LSTM is used to understand and leverage related sentiment information context behind vectors of each subword features.

RESULTS AND DISCUSSION

- Datasets are collected from real world scenarios where limited resources in this case, such as Twitter and AskPatient.
- we evaluate the proposed method using the accuracy measure.

Our Experiments results shwo that improved hybrid vectorization regarding medical components such as ADRs are effective for Sentiment Analysis. several algorithm have been applied on this vectors representation for training a sentiment prediction Model, as shown in table 1

Tabela 1 – Corpus u

Algorithm /Vectorization scheme	MedicalKnowledge / Corpora	dataset used	Accuracy %
SVM+word2vec	ADRMine+PubMed	Twitter	0,77
RNN	PubMed+ADR corpora	Twitter/Facebook	0,71
RNN+ proposed vectorization	PubMed+ADR corpora	Twitter/Facebook	**0,78**
LSTM	ADRMine+PubMed	Twitter	0,68
stachek LSTM + proposed vectorization	PubMed+ADR corpora	Twitter	**0,81**

Fonte: Best Achieved Accuracy and corresponding Algorithms and Medical Knowledge.

the sentiment detection is assessed through experiments on Three non annotated datasets figure 3. Both LSTM and stacked LSTM and markedly and consistently outperform all of the other datasets for all the three datasets.
the baselines requires (a lot of) training data and are domain-dependent, which do not adopt new drug/ADRs corpora for efficient Matching.

CONCLUSION

- Turning patients' opinions into actionable information is having a profound impact on healthcare and pharmaceutical development.
- CNN and Stacked-LSTM can effectively capture temporal semantics of social media texts regarding medical concepts.

- *detecting potential drugs issues or ADRs from social media through sentiment analysis is of critical importance.*

Figura 2 – sentiment Analysis Model: model vectorization and sentiment prediction
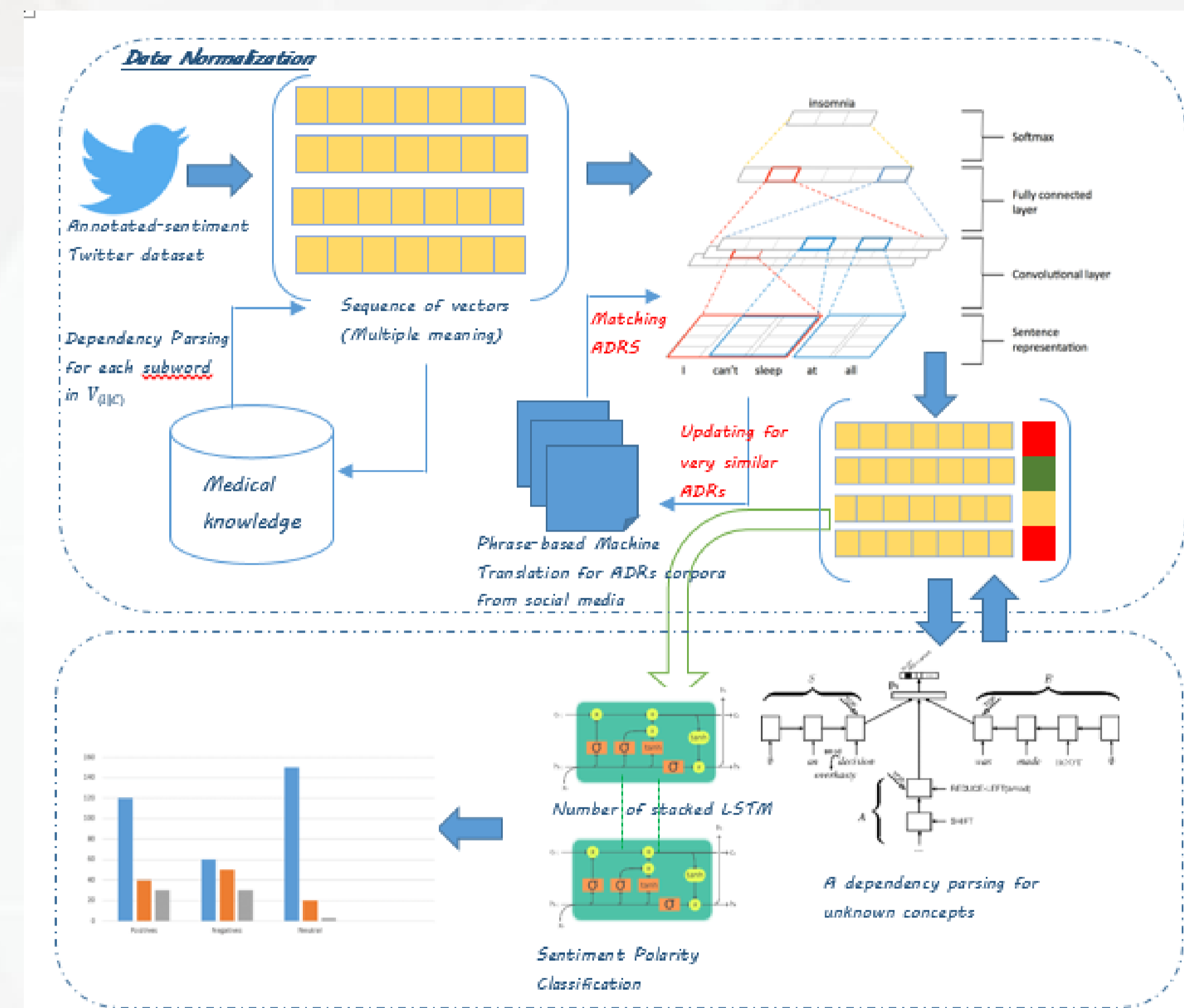
- **model vectorization :** that can recognize efficiently medical concepts Expressions from everyday patient self-reports on social media.
  - a based-matching Medical concepts recognition incorporates primarily the PubMed by MedLine database of references and abstract on life sciences and biomedical topics. also a Corpora of ADRs constructed by (LIMSOPATHAM; COLLIER, 2016).
  - a based-CNN method is used to discriminate text items into defined concepts and entities.
  - a based-transition dependency method (KIPERWASSER; GOLDBERG, 2016) parsing the unknown ADRs and diseases, which subword is associated with very similar ones.
  - the model makes a temporal vector by concatenating the vector representations appear in the same context.

RESULTS AND DISCUSSION

Baselines :
- (GIATSOGLOU et al., 2017) hybrid vectorization using word embedding for sentiment analysis.
- SenticNet (CAMBRIA et al., 2018): Conceptual primitives from text and link them to commonsense concepts and named entities for sentiment analysis, is the the most effective baseline.

Figura 3 – Tweet on Monday .
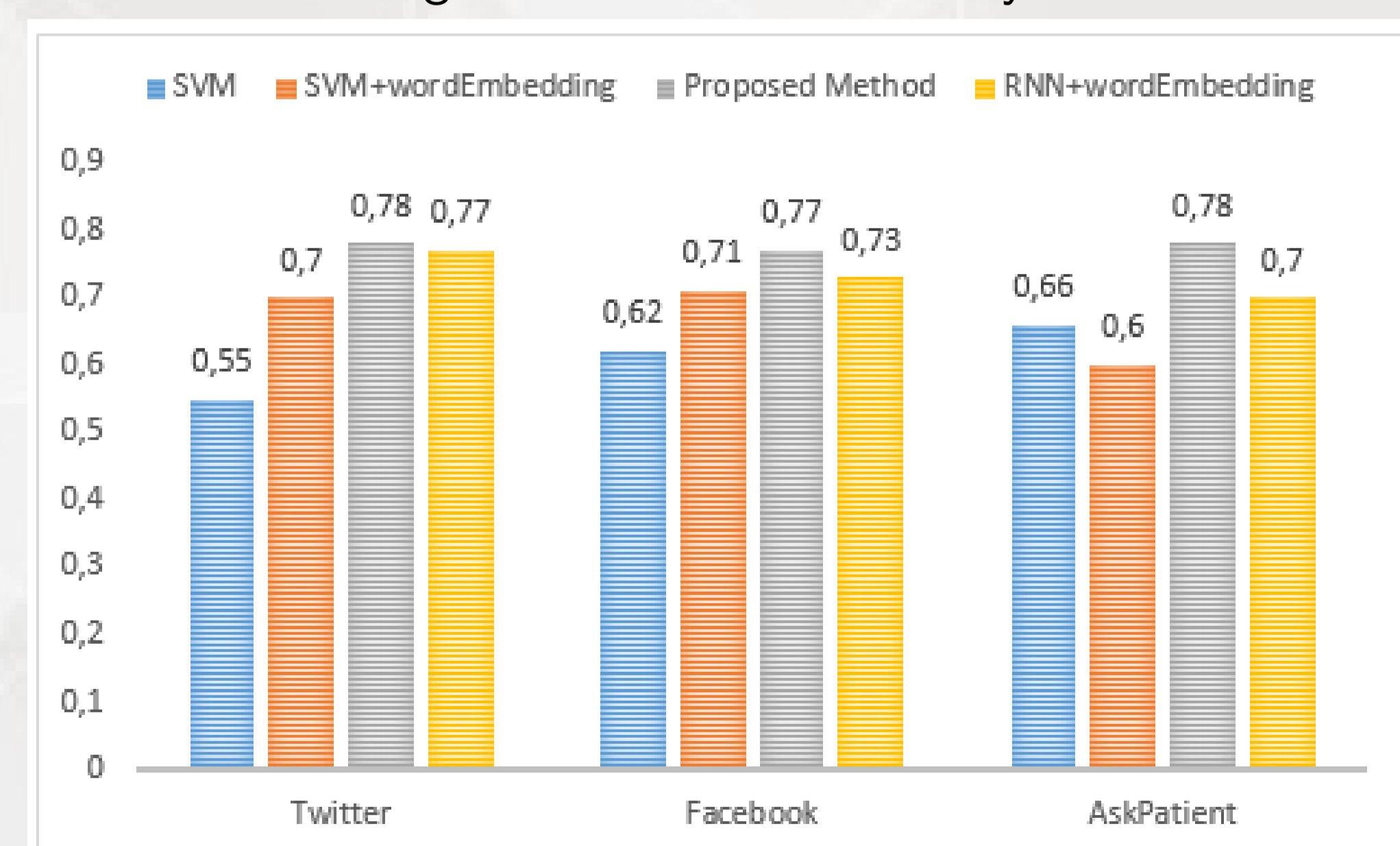
Stacked-LSTM + proposed vectorization significantly outperforms existing methods, which can explore medical knwoledge regarding expressed Sentiment.

REFERENCES

CAMBRIA, E. et al. **SenticNet 5: Discovering Conceptual Primitives for Sentiment Analysis by Means of Context Embeddings**. [S.l.: s.n.], 2018. Disponível em: <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16839>.

GIATSOGLOU, M. et al. Sentiment analysis leveraging emotions and word embeddings. **Expert Systems with Applications**, v. 69, p. 214–224, 2017. ISSN 0957-4174. DOI: