International Journal of IT-based Public Health Management
Vol. 1, No. 1 (2014), pp. 11-16
http://dx.doi.org/10.21742/ijiphm.2014.1.1.02

Classification of Disease-Treatment HML

Reddi Prasadu[1] and E. Adinarayana[2]
[1]*Department of Computer Science and Engineering, KL University, Vaddeswaram, AP, 522502, India*
[2]*Department of ECE, SMCE, Guntur*
[1]*reddiprasad0112@gmail.com*, [2]*adinarayanamtech@gmail.com*

Abstract

Machine Learning (ML) executions can be found in numerous spaces for computerization and simply as of late has turned into a solid apparatus in the medicinal area as well. ML is imagined as an instrument by which machine based frameworks can be coordinated in the health awareness part so as to show signs of improvement, quicker and more productive restorative consideration. This experimental space of programmed learning drives the production of insightful and mechanized applications that helps human services personals to embrace undertakings, for example, therapeutic choice help, restorative imaging, protein-protein connection, extraction of medicinal information, and a general patient administration framework. This paper depicts a Hybrid ML-based technique that is melded with a SVM classifier in blend with Bag-of-Words Representation and NLP assignments for building an application that is equipped for recognizing and spreading health awareness data. In its primary structure it concentrates sentences from restorative data sources, for example, distributed therapeutic papers, patient case sheets that say maladies and medications, and recognizes semantic relations that exist between the ailments and medicines. This key methodology acquires solid conclusions that could be coordinated in an application to be utilized as a part of the therapeutic consideration space. A usage of the proposed methodology accepts the claim.

Keywords: Healthcare, machine learning, natural language processing, SVM classifier.

1. Introduction

Individuals think profoundly about their wellbeing and need to be, presently like never before, accountable for their wellbeing and human services. Life is more frenzied than has ever been, the prescription that is honed today is an Evidence-Based Medicine (hereafter, EMB) in which therapeutic mastery is focused around years of practice as well as on the most recent disclosures too. Instruments that can help us oversee and better stay informed regarding our wellbeing, for example, Google Health1 and Microsoft Health Vault2 are reasons and realities that make individuals all the more effective regarding the matter of health awareness learning and administration. The conventional medicinal services framework is likewise turning into one that grasps the Internet and the electronic world. Electronic Health Records (hereafter, ehr) are turning into the standard in the social insurance domain. Researches and studies demonstrate that the potential advantages of having an EHR framework are3:

Article history:
Received (March, 2014), Review Result (May 07, 2014), Accepted (June 28, 2014)

ISSN: 2205-8508 (Print), eISSN: 2207-3965 (Online) IJIPHM

Health information recording and clinical information storehouses quick get to patient findings, allergies, and lab test outcomes that empower better and time-productive medicinal decisions;

Medicine administration quick get to data as to antagonistic medication responses, immunizations, supplies, and so on;

Choice backing the capacity to catch and use quality restorative information for choices in the work process of human services; and obtain medicines that are customized to particular wellbeing needs-quick get to data that is centered on certain topics.

2. Related works

The most pertinent related work is the work done by Rosario and Hearst. The creators of this paper are the ones who made and conveyed the information set utilized as a part of our research. The information set comprises of sentences from Medline5 modified works clarified with infection and treatment substances and with eight semantic relations in the middle of sicknesses and medicines. The fundamental center of their work is on entity recognition for ailments and medicines. The creators use Hidden Markov Models and most extreme entropy models to perform both the undertaking of substance recognition and the relation discrimination.

Their representation strategies are focused around words in setting, grammatical form data, expressions, and medicinal lexical ontology-Mesh6 terms. Contrasted with this work, our exploration is centered on distinctive representation techniques, different grouping models, and in particular creates enhanced results with less commented data. The undertakings tended to in our examination are information extraction and connection extraction. From the abundance of exploration in these areas, we are going to say some illustrative works.

The undertaking of connection extraction or connection distinguishing proof is long ago handled in the medicinal writing, however with a concentrate on bio medical assignments: sub cell area (Craven, [4]), quality issue affiliation (Ray and Craven), and illnesses and medications (Srinivasan and Rindflesch). For the most part, the information sets utilized as a part of bio medical particular assignments utilize short messages, frequently sentences. This is the situation of the initial two related works said above. The errands regularly involve Identification of relations between substances that co-happen in the same sentence.

3. Existing system

Pervasive computing is the concept that incorporates computation in our working and living environment in such a way so that the interaction between human and computational devices such as mobile devices or computers becomes extremely natural and the user can get multiple types of data in a totally transparent manner. The potential for pervasive computing is evident in almost every aspect of our lives including the hospital, emergency and critical situations, industry, education, or the hostile battlefield. The use of this technology in the field of health and wellness is known as pervasive health care.

• eHealth provides a host of direct benefits such as easier patient record management, reduced paperwork, faster information flow between various departments, greater organizational flexibility, reliable and timely information, minimal inventory levels, reduced wastage, reduced waiting time at the counters for patients and reduced registration time for patients.
• By enabling an automated and intelligent flow of patient information hospitals and doctors can serve their patients better and eHealth is excellent in this aspect.

• To make it better we propose to extend eHealth with automated disease treatment relations classification for a faster understanding of a patient profile from his digital case sheets.

Some Successful famous Remote Monitoring Health Care Projects that can be extended with our proposed approach for a better service

- The Cell-Life Project -South Africa
- Chinese Aged Diabetic Assistant (CADA) -China
- Colecta-PALM -Peru
- Mashavu: Networked Health Solutions for the Developing World -Tanzania
- MediNet Healthcare Management System -Trinidad and Tobago
- Mobile Care, Support and Treatment Manager (MCST) -India
- Mobile Phones for Health Monitoring –India &the United Kingdom
- Phoned Pill Reminders for TB Treatment -Thailand
- SIMpill Solution for TB -South Africa
- Virtual Health Pet –Brazil

The traditional healthcare system is also becoming one that hugs the Internet and the electronic world. Electronic Health Records (EHR) is becoming the standard in the healthcare domain. Researches and studies show that the potential benefits of having an EHR system are:

- Health information recording and clinical data repositories immediate access to patient diagnoses, allergies, and lab test results that enable better and time-efficient medical decisions;
- Medication management rapid access to information regarding potential adverse drug reactions, immunizations, supplies, etc;
- Decision support the ability to capture and use quality medical data for decisions in the workflow of healthcare; and Obtain treatments that are tailored to specific health needs-rapid access to information that is focused on certain topics.

Disadvantage:

- In order to embrace the views that the EHR system has, we need an intelligent better, faster, and more reliable access to information.
- All research discoveries come and enter the repository at high rate, making the process of identifying and disseminating reliable information a very difficult task.
- Absent of an automated system that can understand disease treatment classifications.

4. Proposed system

This paper describes a Machine Learning (ML)-based methodology for building an application that is capable of automated identification and dissemination of healthcare information. It extracts sentences from published medical papers that mention diseases and treatments, and identifies semantic relations that exist between diseases and treatments. Our evaluation results for these tasks show that the proposed methodology obtains reliable outcomes that could be integrated in an application to be used in the medical care domain. We propose and discuss pros and cons of certain ML settings used these days such as NLP, classification algorithms, Bag-of-Words Representation, NLP and Biomedical Concepts Representation, Medical Concepts (UMLS) Representation. The potential value of this paperstands in using any one of the ML settings and developing an automated application that can understand and classify disease treatment classifications in shorter texts.

Earlier approaches use the following classification algorithms. Decision trees are decision-based models similar to the rule-based models that are used in handcrafted systems, and are suitable for short texts. Probabilistic models, especially the ones based on the Naıve Bayes theory, are the state of the art in text classification and in almost any automatic text classification task. Adaptive learning algorithms are the ones that focus on hard-to-learn concepts, usually underrepresented in the data, a characteristic that appears in our short texts and imbalanced data sets. SVM-based models are acknowledged state-of-the-art classification techniques on text. It is evident that SVM-based models are most recent and advanced classification techniques that are optimized to yield better performance irrespective to the voluminous of data required for training sets. SVM has following advantages.

- Prediction accuracy is generally high
- Robust, works when training examples contain errors
- Fast evaluation of the learned target function

The bag-of-words (BOW) representation is commonly used for text classification tasks. It is a representation in which features are chosen among the words that are present in the training data. Selection techniques are used in order to identify the most suitable words as features. NLP tasks that has sentence selection and relation identification. We propose to implement a hybrid model that uses an SVM classification in combination with Bag-of-Words Representation and NLP tasks. This combinatorial hybrid approach implements the best of all the three methods for effective disease treatment classifications in shorter texts at better time frames and performances of individual approaches.

Table 1.

SVM ALGORITHM:

Candidate SV = { closest pair from opposite classes }

while there are violating points
do
Find a violator

Candidate SV = candidate SV U violator
if any $\alpha p < 0$ due to addition of c to S
then candidate SV = candidate SV / p
repeat till all such points are pruned
end if

end while

5. Experimental results

The test set on which the models are assessed contain the genuine classes and the assessment tries to distinguish what number of the genuine classes were anticipated by the model classifier. In the ML settings, exceptional consideration needs to be coordinated to the assessment measures that are utilized. For information sets that are very imbalanced (one

class is [6][2] overrepresented in correlation with an alternate), standard assessment measures like exactness are not suitable. Since our information sets are imbalanced, we decided to report notwithstanding exactness, the macro arrived at the midpoint of F-measure. Test data is 2700 retrieved from original data sets. Specify Training record Count: 50Specify data record count: 5

Table 2. Training and testing datasets comparison results

Abstract Number	SVM classifier	CNB classifier
1	95	92
2	90	94
3	97	90
4	90	90
5	96	97

Figure 1 presents the best comes about acquired as such. An increment of very nearly 5 rate focuses, for both precision and F-measure is acquired when utilizing as representation gimmicks biomedical substances removed by the Genia tagger and [6] [7] CNB as classifier.

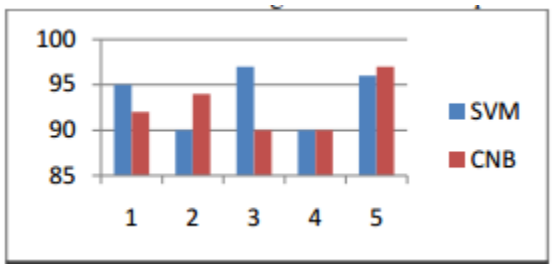

Figure 1. Accuracy and f-measure results when using training and testing medical datasets

The change over alternate settings can be because of the way that the blend of classifier and peculiarities has a decent anticipating worth for a model prepared on the three relations. The results demonstrate that probabilistic models focused around Naive Bayes equation, acquire great results. [1][4] the way that the SVM classifier performs well demonstrates that the current revelations are in accordance with the literature.

6. Conclusion

Machine Learning (ML) field is picking up notoriety in practically any space of exploration and simply as of late has turned into a solid device in the medicinal area. We incorporated the sentences that did not contain any of the three relations being referred to and the results were lower than the one when we utilized models prepared just on sentences containing the three relations of investment. These disclosures accept the way that it is critical

to have the first venture to get rid of uninformative sentences, before looking deeper into characterizing them. This paper depicts a Hybrid ML-based approach that is melded with a SVM classifier in mix with Bag-of-Words Representation and NLP errands for building an application that is equipped for recognizing and dispersing health awareness data. In its unrefined structure it concentrates sentences from medicinal data sources, for example, distributed therapeutic papers, patient case sheets that say illnesses and medicines, and distinguishes semantic relations that exist between the ailments and medications.

References

[1] Oana Frunze, Diana Inkpen, and Thomas Tran, "A Machine Learning Approach For Identifying Disease-Treatment Relations in Short Texts", IEEE Transactions On Knowledge and Data Engineering, Vol. 23, No. 6, **(2011)**.

[2] L. Hunter, Z. Lu, J. Firby, W.A. Baumgartner Jr., H.L. Johnson, P.V. Ogren, and K.B. Cohen,"OpenDMAP: An Open Source, Ontology-Driven Concept Analysis Engine, with Applications to Capturing Knowledge Regarding Protein Transport, Protein Interactions and Cell-Type-Specific Gene Expression," BMC Bioinformatics, Vol. 9, No. 78, **(2008)**.

[3] M. Yusuke, S. Kenji, S. Rune, M. Takuya, and T. Jun'ichi, "Evaluating Contributions of NaturalLanguage Parsers to Protein-Protein Interaction Extraction," Bioinformatics, Vol. 25, pp. 394-400, **(2009)**.

[4] S. Novichkova, S. Egorov, and N. Daraselia, "Med Scan, A Natural Language Processing Engine for MEDLINE Abstracts," Bioinformatics, Vol. 19, No. 13, pp. 1699-1706, **(2003)**.

[5] T. Mitsumori, M. Murata, Y. Fukuda, K. Doi, and H. Doi, "Extracting Protein-Protein Interaction Information from Biomedical Text with SVM," IEICE Transactions on Information and Systems, Vol. E89D, No. 8, pp. 2464-2466, **(2006)**.

[6] C. Giuliano, L. Alberto, and R. Lorenza, "Exploiting Shallow Linguistic Information for RelationExtraction from Biomedical Literature", Proceedings. 11th Conference European Chapter of the Association for Computational Linguistics, **(2006)**.

[7] J. Ginsberg, H. Mohebbi Matthew, S.P. Rajan, B. Lynnette, S.S. Mark, and L. Brilliant, "Detecting Influenza Epidemics Using Search Engine Query Data," Nature, Vol. 457, pp. 1012-1014, **(2009)**.

[8] S.V.N. Vishwanathan, M. NarasimhaMurty, "SSVM: A Simple SVM Algorithm", In Advances in Neural Information Processing Systems, (NIPS*2000), NIPS, Cambridge MA: MIT Press, Vol.13, **(2001)**.

[9] S. V. N. Vishwanathan and M. Narasimha Murty. "Geometric SVM: A Fast and Intuitive SVM Algorithm". Technical Report IISC-CSA-2001-14, Department of CSA, Indian Institute of Science, Bangalore, India, November 2001. Submitted to ICPR **(2002)**.

[10] L. Hunter and K.B. Cohen, "Biomedical Language Processing: What's beyond PubMed?"Molecular Cell, Vol. 21- 5, pp. 589-594, **(2006)**.

[11] S. Ray and M. Craven, "Representing Sentence Structure in Hidden Markov Models for Information Extraction," Proceedings, Internation Joint Conference of Artificial Intelligence (IJCAI '01), **(2001)**.