Self-evolving Disease Ontology for Medical Domain Based on Web

Ishara Sandun, Sagara Sumathipala, and Gamage Upeksha Ganegoda

Faculty of Information Technology, University of Moratuwa, Katubedda, Moratuwa, Sri Lanka

Ijfis

Abstract

In last decade information technology has gained a rapid development, and today it plays a crucial role in everyone's life. It makes the life more comfortable for professional to do their work. Every performance and the innovating task will become more comfortable if there is a proper and accurate knowledge base containing up to date information. It will be an added advantage if the so-called knowledge base could shrink and expand dynamically. Especially in the medical domain, there is a higher demand and necessity for such kind of knowledge base which evolves dynamically with time and data because medical field is rapidly evolving and new biomedical entities such as diseases, symptoms, proteins, and so forth are frequently introducing. This study proposes a mechanism to generate dynamically evolving ontology for the biomedical domain which evolves with new relations explores from web data and patient history records. Proposed approach retrieves information from the ontology and generates probabilistic values for each relationship in the disease ontology. This approach used to create a dynamically evolving ontology for the medical domain to manage the relationship between diseases and symptoms more effectively. Furthermore, it retrieves data from the ontology to answer user queries related to the diseases and symptoms.

Keywords: Medical ontology, Biomedical relationship extraction, Named entity recognition, Text mining

1. Introduction

There is a famous proverb saying that "Health is Wealth." Health and well-being are two key factors to live a happy life. They are essential since it affects directly to one's life and to live happily. In last decade information technology has gained a rapid development, and today it plays a crucial role in everyone's life. It makes the life more comfortable for professional to do their work. With the development of technology, people have come up with new methods and inventions to make the life easy for the professionals. Every performance and the innovating task will become more comfortable if there is a proper and accurate knowledge base containing up to date information. It will be an added advantage if the so-called knowledge base could shrink and expand dynamically. Especially in the medical domain, there is a higher demand and necessity for such kind of knowledge base which evolves dynamically with time and data. In medical domain new diseases, viruses, genes, etc. are identified very often, and relationships among them are rapidly changing.

If the new biomedical relationships are not identified accurately and adequately, it will take a longer duration to identify a disease, its causes, methods to cure, and prevention. Therefore,

Received: Dec. 11, 2017
 Revised : Dec. 18, 2017
 Accepted: Dec. 20, 2016

Correspondence to: Sagara Sumathipala
 (sagaras@uom.lk)
 ©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

maintaining a dynamic and evolving knowledge base has become a pressing need. This process can be done manually but it will be difficult, and humans can make mistakes, which can affect the accuracy of the final results. Because of these reasons, there is high demand for a knowledge base which dynamically evolves itself.

This study proposes a mechanism to generate dynamically evolving ontology for the biomedical domain which evolves with new relations explores from web data and patient history records. Proposed approach retrieves information from the ontology and generates probabilistic values for each relationship in the disease ontology. This approach used to create a dynamically evolving ontology for the medical domain to manage the relationship between diseases and symptoms more effectively. Furthermore, it retrieves data from the ontology to answer user queries related to the diseases and symptoms.

Difference between existing approaches and proposed approach is proposed ontology can learn from each entry to the system. It calculates probability values of occurring in every relationship. For example, it calculates the probability value of “symptom S” being a symptom of “disease D” out of “N” number of patient records. When a person gets infected with a disease, he/she will not have all the symptoms of that particular disease. So there is a probability that he/she might not have those symptoms. So by generating a probability value for those relationships, users can get a good idea about the occurrence of that particular relationship.

As a solution to this, the proposed approach collect information and relationships from web articles and patient history records or patient entries. In this approach, ontology evolves with every entry to the system. Numerical probability values will be calculated for each relationship according to the probability of occurring that particular relationship. Since this ontology develops and evolves dynamically, there is no need to update the ontology manually also the ontology will contain the latest information with the probability of occurrences of relationships annotated to the relationships. In this approach, users can retrieve data dynamically by asking questions.

2. Related Work

With the expansion of the medical field, every year scientists and medical researchers find many new diseases. Since a dynamic knowledge base for the medical domain has been a vital need, there have been several types of research done in this area. For relationship extraction from several domain-specific rela-

tionships, extraction approaches can be found. Some of these approaches are “Gene-disease relation extraction”, “Protein name recognition using Gazetteer” [1], “Dictionary-based bio-entity recognition in biomedical literature” [2], etc. However, the problem with those approaches has used a dictionary-based approach to identify named entities in the medical domain which lacks in accuracy after several years because of the arrival of new biomedical entities to the field. There are machine learning based and web-based biomedical entity recognition systems proposed [3–6]. However, most of these approaches only consider the specific type of biomedical entities such as proteins.

Dynamic ontology construction is a complicated and tedious process using existing approaches such as Text-to-Onto [7], and Text2Onto [8]. These systems extract relations using association rule mining and predefined regular expressions to expand the ontology dynamically. Since they identified semantic relationships using only POS tagging information, there is a difficulty in extracting domain specific concepts.

In past few years, several researchers have developed dynamic ontologies for medical domain. Many ontologies that have been done on diseases can be found in BioPortal web page. Some of the researches done on developing an ontology for the medical domain are Infectious Disease Ontology (IDO) [9], Galen ontology [10], Human Phenotype Ontology (HPO) [11], Translational Medicine Ontology (TMO) [12] etc. But the problem with these researchers is that are developed for a specific purpose, and so they cannot be dynamically used for disease diagnosis. Apart from above Disease Ontology Identifiers (DOID) [13] was started in 2003 by Northwestern University and Symptom Ontology (SYMP) [14] was started in 2005 by the Institute for Genome Sciences at the University of Maryland can be used for general disease diagnosis.

3. Proposed Solution

The proposed solution is a dynamically evolving ontology using web data and patient feedback. This solution also contains a process to dynamically retrieve data from the ontology by querying the system in natural language text. Approach for this solution has four main modules. They are as follows:

- 1) Named Entity Tagger for medical domain
- 2) Relationship Extraction from web data and user feedback
- 3) Dynamic evolving ontology construction

4) Retrieval of information from ontology dynamically

This approach collects data from recognized websites from biomedical domain such as PubMed [15], Medicine net [16], CDC [17] etc. For that, it scrapes data from the web articles in reputed websites. Furthermore, it collects data from patients or users of the system. Therefore with some sources (past patient records, web article sources, patient entries) for a particular disease, a probability of occurrence can be generated for every relationship.

Consequently, in this approach, the method calculates probabilities for relationships from every entry to the system. So that anyone can get an idea about the probability of “symptom S” being a symptom of “disease D”. The method gets the patient records or the user symptom entries it automatically generates the probability of a particular symptom being a symptom of a particular disease and it annotates that probability to the relationship. Therefore this lead to giving an idea about most probable relationships and also a threshold value can be used to stabilize the probabilistic relationships which can happen due to wrong information or due to the evolution of diseases with time. Subsequently, a person can get an idea about the diseases he/she can have from the symptoms since this method indicates the probability occurrences as well. Because of these reasons, our ontology approach can self-learn, and no one has to expand the ontology manually. It dynamically evolves itself with every entry to the system. Figure 1 illustrates the basic structure of the proposed solution.

3.1 Named Entity Recognizer for Biomedical Domain

It is essential to identify biomedical named entities in web articles to extract relationships among them. Initially, a dictionary-based approach was used to identify biomedical named entities in the web articles. The dictionary was created with disease names, virus and symptom names. However, it fails to identify entities that are not listed in the dictionary. Therefore a Named Entity Tagger for the biomedical domain was trained to identify the named entities.

In order to train the named entity recognizer for the biomedical domain, a corpus was made using sentences picked from web articles about diseases and patient records. Then the corpus was tokenized into unigrams and tagged with IO encoding. However, then again the results were not accurate as expected. It was observed that since entities have multiple words per name unigram tagger did not produce accurate results as expected. Therefore to achieve results with higher accuracy level, the

Figure 1. Basic structure of the proposed solution.

Table 1. Some of the extracted relationships

Relationship type	Relationship pattern
Synonymic	“is a”, “is equivalent to”, “also known as”, “is also called”
Hyponymic	“such as”, “or example”, “for instance”
Casual	“is caused by”, “causes”

corpus was improved by adding more sentences and instead of the unigram tagger bigram tagger was used with IOB encoding. Figure 2 shows part of the corpus which was tokenized and tagged with IOB encoding.

3.2 Relationship Extraction from Web Data and User Feedback

Named Entity Tagger was trained to extract relationships between entities from web articles. POS tagging and regular expressions were used to identify the relation between two named entities in each sentence. Some of the common relationships extracted from web documents are shown in Table 1.

Coreference resolution is the task of identifying all the expressions that are referred to the same entity in the given context.

1	Dengue	B-DISEASE
2	fever	I-DISEASE
3	is	0
4	a	0
5	disease	0
6	caused	0
7	by	0
8	a	0
9	family	0
10	of	0
11	viruses	B-MICROBES
12	that	0
13	are	0
14	transmitted	0
15	by	0
16	mosquitoes	B-ANIMAL
17	.	0
18	Symptoms	0
19	of	0
20	dengue	B-DISEASE
21	fever	I-DISEASE
22	include	0
23	severe	0

Figure 2. Part of the tokenized and tagged corpus.

```

Main_Class_list= [[NER entity type of entity 1], [NER
entity type of entity 2]]
Properties_list= [[ NER entity type of entity
1,relation,NER entity type of entity 2]]
ClassifiedWords_list= [[entity1, NER entity type of
entity 1], [entity2, NER entity type of entity2]]
Relationship_list= [[entity1, relation, entity2,
parameter1]]

```

Figure 3. Some lists generated for each relationship.

In this study, coreference resolution was conducted to the web pages before extract relationships. Tokenized sentences are then being tagged with the trained Named Entity Tagger to identify the entities in the sentence. Words between two identified entities in a given sentence were used to extract the relationship between two named entities.

POS tagging and regular expressions were used to identify the relationship. Two identified named entities and the extracted relationship is used as the subject, object and predicate respectively. Then the extracted relationship triple is passed to the ontology creation class to update the ontology. Figure 3 indicates the way those four lists are generated.

Figure 4. Flow diagram of the proposed ontology.

3.3 Dynamically Evolving Ontology Construction

The Web Ontology Language (OWL) ready tool for Python was used to create the dynamic evolving ontology. Whenever a new relationship is added, it dynamically generates the primary classes, properties, classified words and relationships to four sets. The primary class name was used as the type of the named entity. Before each relationship addition, the system dynamically generates a list of primary classes in those relations. Then it generates a list of properties in those relations. Figure 3 indicates the way those four lists are generated.

In order to generate the list of properties, the entity types of subject, object entities and relation information was passed. List of classified words that are generated by entity names and entity types of each entity. Finally, the system generates a list of relationship triples where two entities as subject and object and the relation extracted as the predicate. Figure 4 shows the basic flow diagram of how the ontology was created and expanded.

Before passing the relationships, our approach uses SPARQL queries to get all the similar relationships about that particular disease and then they are passed to the ontology construction class with a parameter containing information whether that relationship is present in patient record or source. Therefore, it can calculate probability values for relationships. If a particular relationship not found in the ontology it is a new relationship.

New relationships were annotated as “1.0; 1”, which indicates that it has a probability of 1.0 out of the single source. With the time annotation probability will change. The way a new relationship is annotated is displayed in Figure 5.

For example, if a particular relationship ‘X’ is a symptom of

Figure 5. Flow diagram of the proposed ontology.

disease 'D' was not presented in a patient record of disease 'D' then the probability value of the relationship will be dropped and the number of sources will be incremented by 1. For example, if relationship 'X' is a symptom of disease 'D' had a probability of 0.95 out of 10 sources and then that particular relationship is absent in the next patient record probability will be recalculated as shown in Eqs. (1) to (3) and will re-annotate to the relationship.

If the relationship is positive and is an existing relationship:

$$\begin{aligned} & \text{New Probability} \\ &= \frac{(\text{Existing Probability} \times \text{Previous no of records}) + 1}{\text{Previous no of records} + 1}. \end{aligned} \quad (1)$$

If the relationship is positive and is an existing relationship:

$$\begin{aligned} & \text{New Probability} \\ &= \frac{(\text{Existing Probability} \times \text{Previous no of records})}{\text{Previous no of records} + 1}, \end{aligned} \quad (2)$$

$$\begin{aligned} & \text{New Annotation} \\ &= \text{New Probability} : (\text{Previous no of records} + 1). \end{aligned} \quad (3)$$

For the same relationship if the next source confirms that 'X' is a symptom of disease 'D' the probability value of the relationship is recalculated as shown in eqEq. (1) and the number of sources will be incremented by 1.

Apart from dynamically creation and expansion of the ontology the proposed solution annotate the relationships of the ontology with probability values for each relationship. Probability values change with every patient history record. So the ontology learns with every entry to the system. For each patient history record, new probabilities will be calculated for the symptoms of the disease. With time when the no of records for a particular disease increases probability values will be more accurate. Therefore with time, a person who uses the ontology can get an idea about the probability of a particular symptom

```

setMainclass (main_class_list){
  For each in main_class_list{
    If main_class is there nothing is done if not main_class is
    added
  }
}

setProperty (properties_list){
  For each in properties_list{
    If properties are there nothing is done if not properties are
    added
  }
}

setClassifiedWords(classifiedWords_list){
  for each is classifiedWordsList{
    if classifiedWords are there nothing is done if not they are
    added
  }
}

setRelationships (relationships_list){
  For each in relationship_list{
    If relationship is not there{
      Add relationship and annotate it 1.0;1
    }else{
      Read previous annotation
      Val=probability value in annotation
      No=no of articles/people in annotation

      If relationship present in extractedRelationships or Patient
      record{
        New annotation=((Val*No)/(No+1));(No+1)
      }else{
        New annotation= (((Val*No)+1)/(No+1));(No+1)
      }
    }
  }
}

```

Figure 6. Flow diagram of the proposed ontology.

being a symptom of that disease and the no of records used to calculate the probability. So by using this approach, the ontology will learn and evolve with every patient history record. Pseudocode of the logic is shown in Figure 6.

3.4 Retrieval of Information from the Ontology Dynamically

SPARQL queries are generated dynamically to retrieve information. To retrieve data from SPARQL queries first, it converts the OWL file to RDF format. A question classifier was trained to recognize the type of question asked by the user.

The proposed ontology has six main classes. In this proposed solution the question classifier is trained according to those classes and with different types of ways a user can ask questions. Using the question classifier, it classifies the questions of the user and the computer will recognize the type of question user is asking from the system.

Trained Named Entity Tagger is used to identify the named entities in the question. To map the discovered named entities in the questions with the named entities in the ontology another text classifier was used which contains the entities in the ontology.

SPARQL queries will be generated dynamically according to those type of question. A Naïve Bayes classifier is used to train the questions classifier.

A threshold probability value was used when retrieving relationships from the system. So that relationship with lesser probabilities will not be in results. They will be considered as unstable relationships. This will make the system output only the most probable relationships but, those results can change with time as the probability values of relationships change with new records. So that with time existing relationships can become unstable relationships and unstable relationships can become active again.

For example, if the probability of “symptom A” being a symptom of “disease D” drops below threshold value system will not display “symptom A” as a symptom of “disease D”. It will be considered as a unstable relationship. But the system will keep counting patient records for that relationship and it will be displayed as a relationship if probability value increases above the threshold value.

Furthermore, wrong information by patients will not be count by this approach since probabilities of false relationships will become lower with time. Another advantage of this method is due to the evolution of diseases symptoms or relationships of a particular disease can change from time to time. With this approach, stable or active relationships will contain the most probable relationships during that period. Because of this feature, this ontology and results generated change over the time and ontology learn and gradually evolves with time.

4. Results and Evaluation

One essential part of this research project is the named entity recognition tagger for biomedical domain since it creates the relationship triples based on the trained named entity recognition tagger results. So that has to produce more accurate results for our domain. Initially, we created a corpus of 6000+ words and tagged it with IO encoding. These words have classified by entities into four classes. However, the accuracy of the system was low. We identified that the tagger showed less accuracy since most of the disease names and virus names have more than one word for a name. It showed a 0.3989 accuracy with IO encoding initially. Results are shown in Figure 7.

Therefore further improvements were done to the corpus and created a corpus of 10000+ words. It contained sentences and paragraphs taken from web articles from websites like CDC and PubMed. Instead of IO encoding, IOB encoding was used

Entity	P	R	F1
	0	0.0000	0.0000
ANI	0.6667	0.5000	0.5714
DIS	0.1667	0.3333	0.2222
ORG	1.0000	0.7500	0.8571
VIR	0.0000	0.0000	0.0000
Totals	0.3889	0.3889	0.3889

Figure 7. Results with IO encoding.

Entity	P	R	F1
ANIMAL	1.0000	0.6000	0.7500
DISEASE	0.6667	0.7273	0.6957
MICROBES	1.0000	0.4286	0.6000
NOUN	0.8571	0.9231	0.8889
ORGAN	1.0000	0.2500	0.4000
Totals	0.8182	0.6750	0.7397

Figure 8. Results with IOB encoding.

Question: What are the symptoms of Dengue?

Result:

dengue has_symptom difficulty_breathing
dengue has_symptom abdominal_pain
dengue has_symptom easy Bruising
dengue has_symptom muscle_pain
dengue has_symptom weakness
dengue has_symptom headache
dengue has_symptom pain_behind_the_eyes
dengue has_symptom bone_pain
dengue has_symptom bleeding_gums
dengue has_symptom rash
dengue has_symptom fever
dengue has_symptom nosebleeds
dengue has_symptom skin_hemorrhages
dengue has_symptom swollen_lymph_nodes
dengue has_symptom joint_pain
dengue has_symptom lower_back_pain
dengue has_symptom vomiting

Figure 9. Results with IOB encoding for a sample questions.

and an accuracy level of 0.8182 was achieved. The results are shown in Figure 8.

In order to train the question classifier, Naïve Bayes classifier was trained with around 150 questions. Final results were generated by retrieving values from the ontology using SPARQL queries. However, since it output only the subject-predicate and object results are not grammatically accurate sentences. Results for a sample question is shown in Figure 9.

5. Discussion

In this research, in general, tried to find a way to create a dynamically evolving ontology for the biomedical domain with a method to dynamically retrieve data from that particular ontology. To achieve the result one of the central parts is the Named Entity Tagger for medical domain. Because of the Named Entity Tagger was used to identify the named entities in the web articles and user questions the accuracy of the Named Entity Tagger directly influences the accuracy of the entire system.

Our approach used dictionary-based and IO encoding initially. However, none of the methods were accurate enough to achieve good results. Therefore the tagger was trained with IOB encoding. We identified that it is because named entities in the medical domain can come with multiple words for one entity. For example “Dengue hemorrhagic fever” named entity has three words. With IO encoding it showed a lesser accuracy in finding entities with multiple words. Furthermore, the accuracy of the tagger increased with the quality and the size of the corpus used. In this study, an improvement in the accuracy of the named entity recognizer can be observed with the improvement of the corpus.

Here probability of occurrence values was generated for every relationship of the ontology. Those values were updated with every entry to the system those probability values are generated with no of sources.

A question classifier is used to retrieve data from the ontology dynamically. A Naïve Bayes classifier has been used to identify the type of question trained with real user questions about diseases collected from real users of the system. Here the corpus has been made with real questions collected from real users and the accuracy increased with the size of the corpus and the way questions were collected. By collecting user questions from different types of people a higher accuracy level can be obtained.

Since the final results for user questions are relationship triples from the ontology, it was not displayed in syntactical English sentences. It only displays the subject, object and the relationship between subject and object.

6. Conclusion

This research proposed a novel approach for construction of dynamic ontology with learning retrieval abilities for the medical domain. The method used a semantic relationship extraction approach for medical domain. Compared to existing ontologies,

the proposed ontology can learn and evolve with every entry as it calculates probabilistic values for relationships with every entry. So the user can also get an idea about the probability of occurrence of that particular relationship.

The proposed solution included a way to dynamically generate SPARQL queries and provide answers to user questions about diseases. Therefore this whole solution can be used as a complete system which consists of a dynamically evolving ontology and a way to retrieve information dynamically.

Therefore this whole system can be useful to collect, store information/knowledge in medical domain as a dynamic knowledge base.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

- [1] Y. Tsuruoka and J. I. Tsujii, “Boosting precision and recall of dictionary-based protein name recognition,” in *Proceedings of the ACL 2003 Workshop on Natural Language Processing in Biomedicine*, Sapporo, Japan, 2003, pp. 41-48. <http://doi.org/10.3115/1118958.1118964>
- [2] Z. Yang, H. Lin, and Y. Li, “Exploiting the performance of dictionary-based bio-entity name recognition in biomedical literature,” *Computational Biology and Chemistry*, vol. 32, no. 4, pp. 287-291, 2008. <https://doi.org/10.1016/j.compbiolchem.2008.03.008>
- [3] S. Sumathipala, K. Yamada, M. Unehara, and I. Suzuki, “Protein named entity identification based on probabilistic features derived from GENIA corpus and medical text on the web,” *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 15, no. 2, pp. 111-120, 2015. <https://doi.org/10.5391/IJFIS.2015.15.2.111>
- [4] S. Sumathipala, K. Yamada, M. Unehara, and I. Suzuki, “Protein entity name recognition using orthographic, morphological and proteinhood features,” *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 19, no. 6, pp. 843-851, 2015. <http://doi.org/10.20965/jaciii.2015.p0843>
- [5] S. Sumathipala, K. Yamada, and M. Unehara, “Protein named entity classification with probabilistic fea-

tures derived from GENIA corpus and MEDLINE,” in *Proceedings of 15th International Symposium on Soft Computing and Intelligent Systems (SCIS)*, Kitakyushu, Japan, 2014, pp. 1257-1261. <http://doi.org/10.1109/SCIS-ISIS.2014.7044640>

- [6] S. Sumathipala, K. Yamada, and M. Unehara, “Protein name classification using probabilistic information of orthographic and morphological features,” in *Proceedings of 22nd Symposium of SOFT Hokushin’etsu Chapter*, Nagaoka, Japan, 2013, pp. 51-54.
- [7] A. Maedche and S. Staab, “Mining ontologies from text,” in *Knowledge Engineering and Knowledge Management Methods, Models, and Tools*. Heidelberg: Springer, 2000, pp.169-189. https://doi.org/10.1007/3-540-39967-4_14
- [8] P. Cimiano and J. Vlker, “A framework for ontology learning and data-driven change discovery,” in *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB)*, Alicante, Spain, 2005, pp. 227-238. http://doi.org/10.1007/11428817_21
- [9] Infectious Disease Ontology, Available <http://infectiousdiseaseontology.org/>
- [10] Galen Ontology, Available <https://bioportal.bioontology.org/ontologies/GALEN>
- [11] Human Phenotype Ontology, Available <http://human-phenotype-ontology.github.io/>
- [12] Translational Medicine Ontology, Available <https://bioportal.bioontology.org/ontologies/TMO>
- [13] Disease Ontology, Available http://do-wiki.nubic.northwestern.edu/do-wiki/index.php/Main_Page
- [14] Symptom Ontology, Available <https://bioportal.bioontology.org/ontologies/SYMP>
- [15] PubMed, Available <https://www.ncbi.nlm.nih.gov/pubmed>
- [16] Centers for Disease Control and Prevention (CDC), Available <https://www.cdc.gov/>
- [17] MedicineNet, Available <http://www.medicinenet.com>

Ishara Bandara received B.Sc. degree in Information Technology from University of Moratuwa, Sri Lanka in 2016. His research interests include Bioinformatics and Biomedical Text Mining.
E-mail: isharabnd@gmail.com

Sagara Sumathipala received, Ph.D. in Engineering and M.Eng. in 2015 and 2012 respectively, from Graduate School of Engineering, Nagaoka University of Technology, Japan. He received B.Sc. degree in Computer Science and Technology from the Sabaragamuwa University of Sri Lanka, Sri Lanka in 2009. He is currently working as a senior lecturer in Department of Computational Mathematics, Faculty of Information Technology, University of Moratuwa, Sri Lanka. His research interests are Biomedical Text Mining, Fuzzy Systems, and Mobile Robotics. He is a member of the Sri Lanka Association for Artificial Intelligence (SLAAI).
E-mail: sagaras@uom.lk

Gamage Upeksha Ganegoda received Ph.D. and M.Sc. degrees from Central South University, China, in 2015 and 2011 respectively. Currently, she is working as a senior lecturer in Department of Computational Mathematics, Faculty of Information Technology, University of Moratuwa, Sri Lanka. She able to publish research papers in high end journals such as *BMC Systems Biology*, *IEEE Transaction on Nanobioscience* and *BioMed Research International*. Her current research interests include prediction of disease genes under Bioinformatics, Systems Biology, and Biomedical data analysis.
E-mail: upekshag@uom.lk