A Fuzzy Graph-Based Retrieval System for Health Research

UNB Computer Science Fredericton

Mahsa Kiani, Virendrakumar C. Bhavsar, and Harold Boley

{Mahsa.Kiani, Bhavsar, Harold.Boley}@unb.ca

Faculty of Computer Science, University of New Brunswick, Fredericton, NB, Canada

Introduction

Previously called secondary use, the utilization of clinical electronic data for system evaluation, planning, policy development and management has now been recognized as a critical legitimate use of that data. Administrative data, electronic health records, and electronic medical records are stored in databases.

To support health researchers, the data and the related database software require to be structured with the ability to query and extract information at the record level. A retrieval system compares the query with the stored records in the database and retrieves the most similar records to the query

Challenges of retrieval in health domain

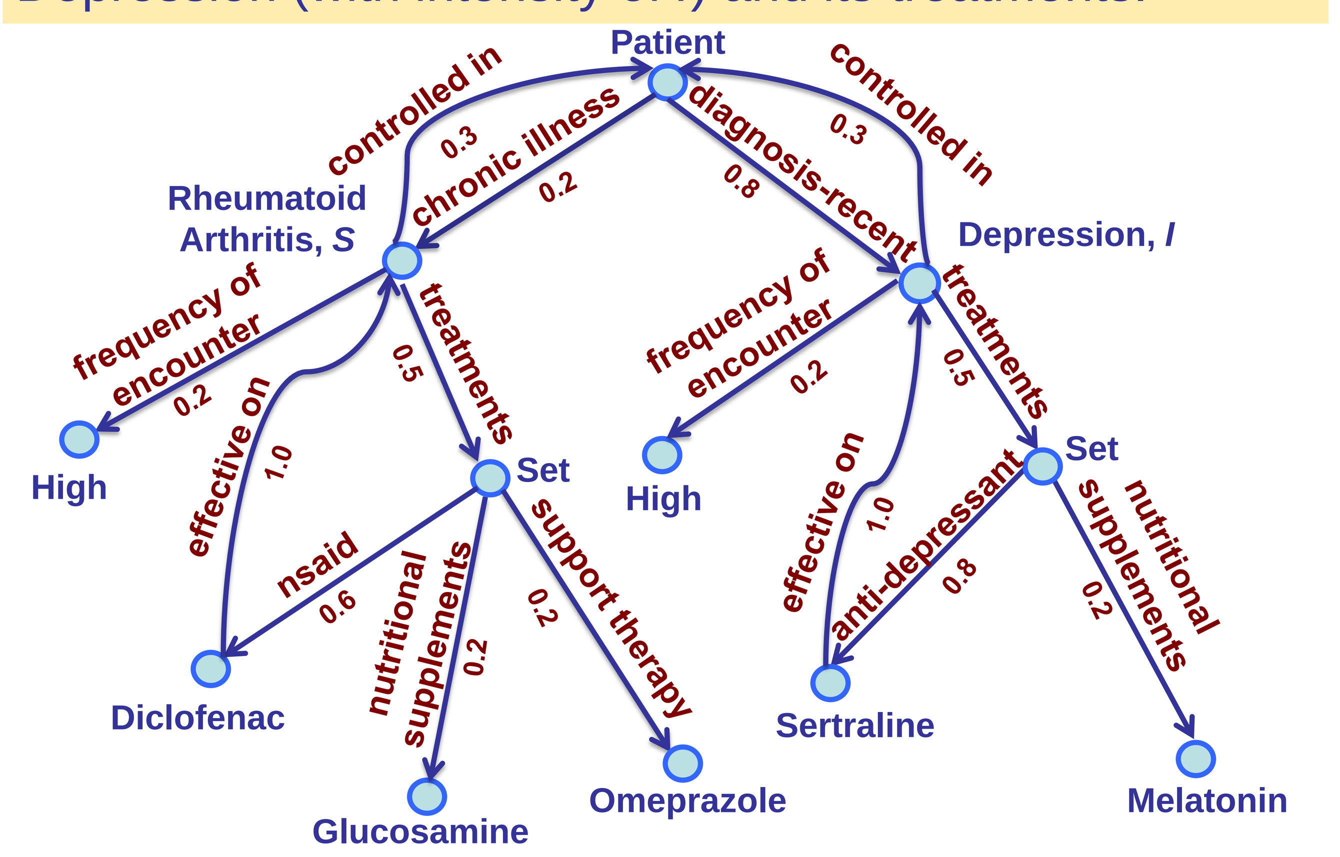
- Complex semantic relationships among concepts,
- Different importance of various features of clinical cases,
- Ambiguous medical concepts and records.

We have introduced Attributed Generalized Tree (AGT), which is a hierarchical graph having vertex labels, edge labels and edge weights. AGT is able to represent complex relationships among concepts as well as different importance of various features.

Our vertex-attributed edge-attributed generalized tree structure proposed earlier is augmented using fuzzy attributes. Labels of vertices represent objects, while edge labels express fuzzy attributes. Edge weights represent the (percentage-) relative importance of fuzzy attributes, a kind of pragmatic information. Our similarity approach integrates the similarity of corresponding fuzzy attributes with the structure similarity.

Representation of Data

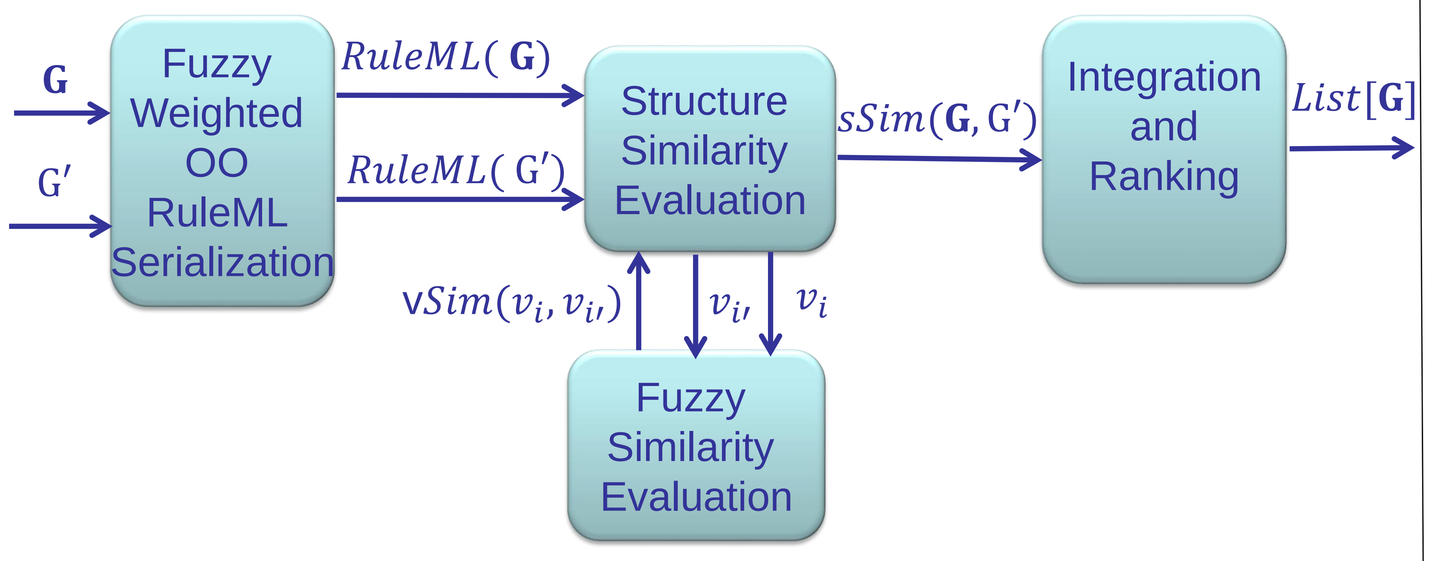
The health record illustrates the existence of chronic Rheumatoid Arthritis (with severity of *S*), which is treated by Diclofenac, Glucosamine, and Omeprazole. Diclofenac has been reported to be effective on Rheumatoid Arthritis and the condition is controlled. The record also represents the recent diagnosis of Depression (with intensity of *I*) and its treatments.



Metadata of Health Record

Architecture

- $\mathbf{G}$ and G': A set of stored metadata extracted from health records, and a given query being matched
- $sSim(\mathbf{G}, \mathrm{G}')$: Structure similarity values
- $v_i, v_{i'}$: Vectors of membership degrees
- $vSim(v_i, v_{i'})$: Fuzzy similarity value
- $List[\mathbf{G}]$: Ranked metadata of health records



System Architecture

Fuzzy Weighted OORuleML Serialization Module

Given query and each stored metadata are uniformly represented using the integration of a weight extension of OO RuleML [1] and a fuzzy extension of OO RuleML. This approach preserves all structural information of fuzzy generalized trees.

Structure Similarity Module

Structure similarity of the query to each metadata is computed using an extension of our AGT similarity algorithm, which visits the vertices of two structures simultaneously, starting from their roots. In the process of matching two structures, a set of membership degrees related to the linguistic terms of fuzzy sets is assigned to each vertex using fuzzification of the numeric data of vertex labels. The fuzzy similarity values computed by the Fuzzy Similarity Module are integrated with the structure similarity value.

Fuzzy Similarity Module

Membership degrees related to each pair of corresponding vertex labels are inputs to the Fuzzy Similarity Module. The similarity between two vectors of membership degrees is computed using weighted fuzzy semantic distance measure.

Integration and Ranking Module

The Integration and Ranking Module ranks metadata in $\mathbf{G}$ based on the structure similarity.

- $sSim(\mathrm{G}_i, \mathrm{G}')$: Structure similarity of G_i to G'
- $\mathrm{G}_1 << \mathrm{G}_2$: G_1 appears before G_2 in the ranked list

$\mathrm{G}_1 << \mathrm{G}_2$ if and only if [$sSim(\mathrm{G}_1, \mathrm{G}') > sSim(\mathrm{G}_2, \mathrm{G}')$]

$\mathrm{G}_1 \ll \mathrm{G}_2$ or $\mathrm{G}_2 \ll \mathrm{G}_1$ if [$sSim(\mathrm{G}_1, \mathrm{G}') = sSim(\mathrm{G}_2, \mathrm{G}')$]

Computational Experiments

- Generalized tree dataset: $\mathbf{G} = \{\mathrm{G}_1, \mathrm{G}_2, \ldots, \mathrm{G}_6\}$
- Given query: G'
  - Structures given in the Representation of Data section

Rheumatoid Arthritis Severity (*S*) and Depression Intensity (*I*)

Generalized Tree	S	I	Generalized Tree	S	I
G'	7	46	G_4	3	42
G_1	7	46	G_5	2	22
G_2	6	45	G_6	1	18
G_3	5	43			

Membership Functions

$$Severity = \begin{cases} Light = \begin{cases} 1.0, & if\ x \le 1 \\ -0.3333\,.x + 1.3333, & if\ 1 < x \le 4 \\ 0, & if\ x > 4 \end{cases} \\ Mild = \begin{cases} 0.3333.x - 0.3333, & if\ 1 < x < 4 \\ 1.0, & if\ x = 4 \\ -0.25\,.\ x + 2, & if\ 4 < x < 8 \end{cases} \\ Severe = \begin{cases} 0.25\,.x - 1, & if\ 4 < x < 8 \\ 1.0, & if\ x \ge 8 \end{cases} \end{cases}$$

$$Intensity = \begin{cases} Abstent = \begin{cases} 1.0, & if\ 0 \le x < 16 \\ -0.25\,.x + 4, & if\ 16 \le x < 20 \end{cases} \\ Mild = \begin{cases} 0.25.x - 4, & if\ 16 \le x < 20 \\ 1.0, & if\ 20 \le x < 25.9 \\ -0.2439\,.\ x + 7.31704, & if\ 25.9 \le x < 30 \end{cases} \\ Moderate = \begin{cases} 0.2439\,.x - 6.31701, & if\ 25.9 \le x < 30 \\ 1.0, & if\ 30 \le x < 40 \\ -0.1449\,.x + 6.796, & if\ 40 \le x < 46.9 \end{cases} \\ Severe = \begin{cases} 0.1449\,.x - 5.796, & if\ 40 \le x < 46.9 \\ 1.0, & if\ x \ge 46.9 \end{cases} \end{cases}$$

Similarity Values and Ranked List

Generalized Tree	Structure Similarity	Rank	Fuzzy Structure Similarity	Rank
G_1	1.0	1	1.0	1
G_2	1.0	1	0.7368	2
G_3	1.0	1	0.4337	3
G_4	1.0	1	0.3122	4
G_5	1.0	1	0.1196	5
G_6	1.0	1	0.1	6

Conclusion

We have improved the expressivity of our vertex-attributed edge-attributed generalized tree structure by using fuzzy attributes. The proposed similarity algorithm integrates the fuzzy similarity of vertex labels and the structure similarity. This approach outperforms our earlier generalized tree similarity approach that considers exact string matching for computing the similarity of vertex labels.

The author would like to acknowledge the valuable contribution of Mehrdad Kiani, MD. in this research as the expert in health domain.

Publications:

Kiani, M. and Bhavsar, V.C. and Boley, H. 2014. A fuzzy-structure similarity algorithm for attributed generalized trees, in: 13th IEEE International Conference on Cognitive Informatics and Cognitive Computing (ICCI*CC), London South Bank University, London, pp. 203-210.

Kiani, M. and Bhavsar, V.C. and Boley, H. 2014. Structure similarity of attributed generalized trees, in: 8th IEEE International Conference on Semantic Computing (ICSC2014), Newport Beach, California, USA, pp. 100-107.

References:

[1] H. Boley, B. Grosof, M. Kifer, M. Sintek, S. Tabet, and G. Wagner, "Object-oriented ruleml", http://www.ruleml.org/indoo/indoo.html, 2004.