MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

Implementing Literature-Based Discovery System using Semantic Relations

Dimitar Hristovski[a], Carol Friedman[b], Thomas C. Rindflesch[c]

[a] *Institute of Biomedical Informatics, Medical Faculty, University of Ljubljana, Slovenia*
[b] *Department of Biomedical Informatics, Columbia University, New York*
[c] *National Library of Medicine, Bethesda, Maryland*

Abstract

Literature-based discovery (LBD) is a method for automatically generating hypotheses from the research literature. Current LBD systems depend on co-occurrence for finding relations between concepts. In previous research we proposed an enhanced method using semantic relations supplied by two natural language processing systems, BioMedLEE and SemRep. The use of these relations reduces the number of hypotheses proposed and simplifies their evaluation. Here, we describe an implementation (based on BITOLA LBD system) of our methodology using semantic relations.

Keywords:

literature-based discovery, natural language processing, text mining, Medline, UMLS, knowledge discovery

Introduction

Current literature-based discovery systems (for example [1]) use concept co-occurrence as their primary mechanism. No semantic information about the nature of the relation between concepts is provided. The use of co-occurrence alone has several drawbacks, since not all co occurrences underlie "interesting" relations: (a) Users must read large numbers of Medline citations when reviewing candidate relations; (b) systems tend to produce large numbers of spurious relations; and, finally, (c) there is no explicit explanation of the discovered relation.

In our previous research [2] we proposed a theoretical methodology to address the deficiencies of using co-occurrence only, which involved augmenting co-occurrences with explicit predications obtained through use of two natural language processing systems: SemRep and BioMedLee. Here we report on the implementation of LBD system based on the methodology described in [2].

Methods and results

We implemented the *Maybe_Treats1* discovery pattern which can be used to find new treatments for a disease or to find new therapeutic applications for an existing drug. *Maybe_Treats1* is satisfied when there is a change in a substance, body function, or body measurement (concept Y) associated with the starting disease X, and there is also an opposite change in concept Y associated with the drug or substance Z; additionally, drug or substance Z should not be already used to treat disease X. The *change* relations were extracted by BioMedLee and the *Treats* relations by SemRep.

The new web interface (Figure 1) that we developed integrates the co-occurrence-based interface of the BITOLA LBD with the semantic relation extracted by BioMedLee and SemRep. We take a top-down approach. We consider co-occurrence of two concepts as an indication of a relation, however then we present in an aggregated fashion the semantic relations extracted from the Medline citations. At the next level the user is shown as evidence the sentences from which the semantic relations were extracted (Figure 2). In other words, we go from concept co-occurrence to semantic relations to sentences supporting these relations.

The user first selects a starting disease X (Parkinson disease in Figure 1). All co-occurring Y concepts are provided in the blue area; information about those that have a semantic relation is also given, so that the user can select concepts that are changed in relation to the starting disease (for example, gamma-Aminobutyric Acid in Figure 1). Further use of the interface displays drugs Z that cause an opposite change to the intermediate concepts Y. The sentences from which the semantic relations were extracted are also accessible (Figure 2). The new interface, enriched with semantic relations, makes it easier for the user to evaluate a smaller number of proposed hypotheses.

References

[1] Hristovski, D., Peterlin, B., Mitchell, J. A. and Humphrey, S. M. (2005), 'Using literature-based discovery to identify disease candidate genes', Int. J. Med. Inform., Vol. 74(2–4), pp. 289–298.

[2] Hristovski D, Friedman C, Rindflesch TC, Peterlin B. Exploiting Semantic Relations for Literature-Based Discovery AMIA Annu Symp Proc. 2006;:349-353.

Address for correspondence

Dimitar Hristovski. Institute for Biomedical Informatics, Faculty of Medicine, University of Ljubljana
Vrazov trg 2/2 1104 Ljubljana, Slovenia
E-mail: Dimitar.Hristovski@mf.uni-lj.si

BITOLA - Biomedical Discovery Support System (Authors: D. Hristovski, B. Peterlin, C. Friedman, T.C. Rindflesch)

Find Starting Concept

Concept: Parkinson Disease **Chr.Loc.:**

Semantic Types: Disease or Syndrome/ **CUI:** C0030567

Find Related Ys

Limit Ys

Contains:

Semantic Group: Any Semantic Type: any

Frequency >= 0 Confidence >= 0 Page size: 50 Change

Order by (Ys)

Frequency, Confidence, Semantic type, Concept name

Descending, Ascending

Related Concepts Y: (first 35 of 35)

Selected	Concept Name	Semantic Type	X-Relation-Y	Changes Y	Freq	Conf(%)
☐	Levodopa	Neuroreactive Substance or Biogenic Amine	SE(inv): TREATS(31), TREATS(INFER)(14),	BL(inv): increase(1),	4640	17.241
☐	Dopamine	Neuroreactive Substance or Biogenic Amine	BL: decrease(2), SE(inv): TREATS(10), TREATS(INFER)(6),		2602	9.669
☐	Alzheimer's Disease	Disease or Syndrome		BL(inv): increase(2),	1664	6.183
☐	Dementia	Mental or Behavioral Dysfunction		BL: decrease(1), BL(inv): increase(1),	1406	5.224
☐	Dyskinesia, Drug-Induced	Disease or Syndrome	BL(inv): increase(2), decrease(2),	BL: decrease(2),	675	2.508
☐	Movement	Organism Function	BL(inv): increase(2), decrease(2),		670	2.490
☐	Huntington Disease	Disease or Syndrome		BL: decrease(1),	663	2.464
☐	Aging	Organism Function		BL: increase(1),	568	2.111
☐	Depressive disorder	Mental or Behavioral Dysfunction		BL: increase(2), decrease(2), BL(inv): increase(1),	385	1.431
☐	Homovanillic Acid	Organic Chemical	BL: increase(1), decrease(1),		299	1.111
☐	Syndrome	Disease or Syndrome		BL(inv): decrease(1),	208	0.773
☐	gamma-Aminobutyric Acid	Amino Acid, Peptide, or Protein	BL: decrease(1),	BL(inv): increase(1),	168	0.624

Figure 1 - Web interface for BITOLA LBD enriched with semantic relations. Parkinson disease is the starting concept. Shown are related (intermediate concepts) Y that are changed (e.g. gamma-Aminobutyric Acid) in association to Parkinson or in TREATS relation (current treatments)

SemBT - Show Relations - Mozilla Firefox

Datoteka Urejanje Pogled Zgodovina Zaznamki Orodja Pomoč del.icio.us

http://ibmi-ts.mf.uni-lj.si/pls/bitola2/showRelations?p_s

Show Relations

CUI1: 'C0023570' Concept1: Levodopa

CUI2: 'C0030567' Concept2: Parkinson Disease

Concept1	Change1	Chg1.Neg.	Relation	Rel. Neg.	Concept2	Change2	Chg2.Neg.
Levodopa			TREATS		Parkinson Disease		
The effect of L-dopa on monoamine metabolites in Parkinson's disease							
Levodopa			TREATS		Parkinson Disease		
Effect of L-dopa on brain monoamines and their metabolites in Parkinson's disease							
Levodopa			TREATS		Parkinson Disease		
Treatment of Parkinson's disease with amantadine and L-Dopa.							
Levodopa			TREATS		Parkinson Disease		
Treatment of Parkinson's disease with L-DOPA and decarboxylase inhibitor.							

Najdi: Najdi naprej Najdi nazaj Označi vse Razlikuj velike/male črke

Figure 2 - Some of the sentences from which the TREATS relations between Levodopa and Parkinson disease are extracted