Expert Systems With Applications 46 (2016) 262–273

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

An entropy-based evaluation method for knowledge bases of medical information systems

Christian F. Hempelmann [a,*], Unal Sakoglu [b], Varadraj P. Gurupur [c], Seetaramaraju Jampana [c]

[a] Department of Literature and Languages, Texas A&M University-Commerce, TX, USA
[b] Department of Computer Science and Information Systems, Texas A&M University-Commerce, TX, USA
[c] Department of Health Management and Informatics, University of Central Florida, FL, USA

ARTICLE INFO

Keywords:
Knowledge base evaluation
Knowledge representation
Information entropy

ABSTRACT

In this paper we introduce a method to develop knowledge bases for medical decision support systems, with a focus on evaluating such knowledge bases. Departing from earlier efforts with concept maps, we developed an ontological-semantic knowledge base and evaluated its information content using the metrics we have developed, and then compared the results to the UMLS backbone knowledge base. The evaluation method developed uses information entropy of concepts, but in contrast to previous approaches normalizes it against the number of relations to evaluate the information density of knowledge bases of varying sizes. A detailed description of the knowledge base development and evaluation is discussed using the underlying algorithms, and the results of experimentation of the methods are explained. The main evaluation results show that the normalized metric provides a balanced method for assessment and that our knowledge base is strong, despite having fewer relationships, is more information-dense, and hence more useful. The key contributions in the area of developing expert systems detailed in this paper include: (a) introduction of a normalized entropy-based evaluation technique to evaluate knowledge bases using graph theory, (b) results of the experimentation of the use of this technique on existing knowledge bases.

1. Introduction

This article describes a method that can be used to evaluate knowledge bases by comparing their information density against other knowledge bases. Information density is operationalized based on a normalized entropy measure between all concept pairs in the knowledge base. Our materials are an in-house knowledge base that is used to develop an information system for caregivers and families of individuals suffering from dementia and related diseases and the backbone of one of the most widely used medical information systems, the Unified Medical Language System (UMLS) (Bodenreider, 2004). Our own knowledge base is part of a framework named Ontological Personal Health Information System (OPHIS), a newly devised ontological-semantic approach towards developing a Personal Health Information Systems (PHIS) (Gurupur, Sakoglu, Jain, & Tanik, 2014).

Our particular aim is the development of the knowledge base for the proposed system which will be used in educating and empowering the caregivers of dementia patients by providing required recommendations. The purpose is not directly to further research on the disease itself, but new insights are expected to be indirectly gained from observations of the interaction between the system and its users. In the past, we have developed prototypes for developing a PHIS using knowledge bases generated from concept maps. However, the knowledge bases thus developed had limitations mainly in terms of the lack of richness of their properties and connections among concepts. To overcome these issues, OPHIS can accommodate larger amounts of required information in richer and more rigidly formalized structures because it is based on established knowledge-engineering principles (Hempelmann, Raskin & Taylor, 2010; Wei, Sung, Doon, & Ng, 2006), thereby reducing the amount of acquisition effort and computation required to develop and use the system.

As part of the development of OPHIS, we felt it desirable to be able to compare its knowledge base to those of other health-related information systems in order to assess its potential usefulness a priori, i.e., before we could evaluate it through its application in an actual system. This seemed desirable in order to guide our ontological engineering effort while it was ongoing. Our present effort in knowledge base evaluation is based on a previous simpler scheme described in (Gurupur et al., 2014). The present paper focuses on this advancement of knowledge base evaluation from the previous systems and in relation to existing methods. In this paper, we build on, improve and

* Corresponding author. Tel.: +1 3477216977.
E-mail addresses: c.hempelmann@tamuc.edu (C.F. Hempelmann), unal.sakoglu@tamuc.edu (U. Sakoglu), varadraj.gurupur@ucf.edu (V.P. Gurupur), sjampana@leomail.tamuc.edu (S. Jampana).

http://dx.doi.org/10.1016/j.eswa.2015.10.023
0957-4174/

apply one of the two methods, i.e., the “simple connectivity-matrix method” which was introduced in our previous work (Gurupur et al., 2014), to two actual knowledge-bases, UMLS and OPHIS, as described in detail in the next section.

2. Background

2.1. Using information entropy to evaluate knowledge bases

Information density is operationalized based on the entropy measure between all concept pairs in the knowledge base graph as normalized with the harmonic mean of the number of relations available and the number of their actual uses. Knowledge bases are the main means to provide the most crucial resource to information technology applications: knowledge. Evaluation of knowledge bases is as important as its development in serving expert systems (Nirenburg & Raskin, 2004). By now, knowledge base alignment as the main approximation of knowledge base evaluation is a mature subfield of knowledge base engineering and has been discussed in detail at least since the late 1990s. In contrast to mere knowledge base verification, i.e., confirming a knowledge base’s adherence to certain formalisms, knowledge base evaluation should be based on the information in the knowledge base itself. While highly desirable to guide their engineering and acquisition from early stages on, i.e., before the knowledge base can serve an application, an abstract level of their evaluations is not trivial to find. In addition to the advantages of evaluating knowledge bases under construction, evaluating general knowledge bases that are intended to serve a number of different applications, not just one specific application, could provide a means to reflect the knowledge base’s potential value.

Existing comparative methods focus more on the similarity of knowledge bases than their respective power in providing information. The purpose of these methods is mapping onto, or aligning with each, other different knowledge bases in order to facilitate the exchange of information. A popular approach of this type (Maedche & Staab, 2002) uses edit distance between the labels on the concepts combined with a structural comparison based on two aligned pivot concepts, one in each knowledge base. Similarly, Hovy, 2001 aims to compare knowledge bases for alignment, not evaluation, and does so by repeated cycles of suggesting aligned concepts in terms of name match, definition match, and taxonomy match, the latter in terms of overlap in direct ancestor concepts. In contrast to this, our method aims not to align, but to evaluate information density of any pair of knowledge bases that are selected as covering the same domains in comparable density and at comparable grain size.

In sum, the ultimate measure of a knowledge base’s quality should be how well it is able to fulfill the function of providing information to applications, not how much they conform to a given type of formalization, such as RDF, graph types (lattice vs. directed), or indeed even if the information captured by them is non-contradictory by adding meta-properties as in OntoClean (Welty & Guarino, 2001; Guarino & Welty, 2002) or is not reflecting reality (Taylor & Raskin, 2011). Because the purpose of knowledge bases is to provide information, their evaluation should in some way measure the information itself.

Calmet and Daemi (2004) proposed the use of entropy and mutual information for estimating the “distances on ontologies”, i.e., the amount of reduction of uncertainty in a given concept due to knowledge of another target concept, even though they did not apply it to any ontologies or knowledge bases. They proposed one of the graph theoretical concepts of centrality measures, the *node degree*, for estimating the probability mass function of connections in ontologies. There, the *node degree*, i.e., the ratio of the number of edges a node has divided by the total number of edges in the graph, was used to measure the ambiguity or probability for that node. However, all the edges were assumed to have same weight.

Doran, Tamma, Palmisano, Payne, and Iannone (2008) used a reformulated version of the entropy-based method proposed by Calmet and Daemi (2004) by accommodating different weights of edges, thereby accounting for different types of relationships between concepts, in order to evaluate ontology modules and their reuse in ontologies. They showed that their method could differentiate between structurally different modules of the same size, and that their metric provided a finer grain differentiation than the original metric by Calmet and Daemi (2004).

In our previous work (Gurupur et al., 2014), we also introduced an entropy-based measure for an ontology that we developed. In this paper, we have aim to improve this with a normalized version of the overall entropy, normalized entropy value per assertion, and applied it to evaluate two knowledge bases. In contrast to alignment approaches evaluating similarity, our evaluation based on the approaches described in this section derives an independent entropy metric for a given knowledge base, which can be compared to that of another knowledge base. Therefore we named it Normalized Entropy Value per Assertion (NEVA).

2.2. Existing knowledge bases for medical information systems

The demand for doctor, caregiver, and patient access to web-based information systems makes it increasingly difficult to discover, organize, and integrate available legacy health information resources. In order to overcome this situation many standards, methods, and technologies have been developed. Some of these include the Semantic Web, Health Level 7, Personal Health Information Systems (Gurupur et al., 2012), and technologies based on controlled vocabularies, often called knowledge bases despite the fact that they are merely two-dimensional cladistic hierarchies. These controlled vocabularies are usually very “upper-level” (i.e., shallow) networks, in the case of the UMLS over a “metathesaurus” based centrally on the SNOMED CT (Rector & Iannone, 2012). Typical for the technologies based on the UMLS are MetaMap (Bodenreider, 2004.) and cTakes (Savova et al., 2010). MetaMap maps words in natural language text to “concepts” in the UMLS. It even does probabilistic disambiguation and is a project that continues to see improvement. Apache cTakes (Savova et al., 2010) is a modular NLP system to annotate text with UMLS entities. Another popular shallow controlled vocabulary is MeSH (Avansino, Goldman, Sawin, & Flum, 2005). None of these resources have the rich texture required to non-probabilistically process natural language, which is characterized by ambiguity and lack of specification, in a knowledge-based fashion. Their applications have to be very limited for this reason and are usually restricted to terminology unification and integration, annotation, but don’t attempt parsing of natural language for its meaning and representation of that meaning.

The technologies based on these resources commonly work in three dimensions

- Standard knowledge representation supporting structured annotation or organization of information;
- Standard knowledge protocols supporting semantic interoperation such as dynamic clustering and integration of decentralized information/knowledge resources (Cannataro & Talia, 2004);
- On-demand intelligent services satisfying professional needs and expectations (Guinard, Trifa, Karnouskos, Spiess, & Savio, 2010).

Existing applications developed on these and other technologies include products as diverse as

- PredictAD developed by VTT Technical Research Centre of Finland which promises to enable earlier diagnosis of the disease on the basis of patient measurements and large databases (www.vtt.fi);
- Physician decision-support tools such as those provided for early detection, clinician practice tools for impairment assessment, and diagnosis through electronic medical record analysis by Minnesota’s ACT on Alzheimer’s volunteer caregivers group (http://www.actonalz.org);