Medical Informatics Europe '96
J. Brender et al. (Eds.)
IOS Press, 1996

Modelisation of a Criterion of Proximity : Application to Medical Thesauri

Dominique Petiot , Anita Burgun and Pierre Le Beux
Laboratoire d'Informatique Médicale. Faculté de Médecine. Av Pr Léon Bernard
35043 Rennes cedex. France

Proximity is a concept related to distance or common features between things. In the medical domain, proximity is based on semantic characteristics of the terms.
Proximity between medical terms is in relation with the position of the terms within a terminology and with the common features between terms. We suggest a model of calculation of proximity between medical concepts which takes into account these two approaches.

1. Introduction

Searching information becomes an important task of the usual medical practice. The volume of information increases daily and we have more and more sources of information. In order to help physicians to access information, several computer tools have been developed to facilitate the access to the data bases [1]. The conception of a criterion of proximity between medical terms is linked with the semantic features of medical concepts.

It seems natural for any person having some medical knowledge to consider for example that the term angina pectoris is nearer to cardiac insufficiency than to gastric ulcer, that vascular cerebral accident is nearer to carotid thrombosis than to astrocytoma. However, it doesn't exist a way to determine exactly the existing proximities between these terms. A criterion of proximity would allow to locate the semantic settings of the medical terms and then to enlarge the field of research with semantic bases.

2. The concept of proximity

2.1 Definitions

Proximity between things may be defined in two ways :

♦ *in relation with the notion of distance.* This definition is taken from geometric models [2]. This kind of proximity is generally associated with methods in which a criterion of measure allows to associate a value to each thing, the proximity between two things corresponds to the inverse of the difference between the two associated values.

♦ *in relation with common features between things*, the proximity is then characterized by the set of shared properties of the things. Tverski [3] provided this approach and suggested the following definition :

$$S(a,b) = \theta\, f(A \cap B) - \alpha\, f(A-B) - \beta\, f(B-A) \qquad \text{with } \theta, \alpha, \beta \geq 0$$

where S(a,b) represents similarity between a and b, $A \cap B$ corresponding to the features shared by a and b, A-B to the features belonging to a which does not belong to b and B-A to the features belonging to b which does not belong to a.This approach is required when it is not possible to find a physical criterion with associated measurement.

2.2 Semantic proximity

Proximity is connected here to the notion of synonymy but is more shaded : synonymy corresponds to an identical meaning between two concepts whereas proximity allows a significance "little different from" which enlarges the number of possibilities [4]. We are going to detail some problems encountered through the determination of a semantic proximity.

A majority of authors agree with the fact that estimation of proximity between things vary in function of their context. Therefore, telling that two things are similar doesn't have any meaning if we don't specify in which context this proximity has been observed [5]. It ensues that very often it is possible to determine a favorable context for a significant proximity between two things in relation to the searched goal [3,6].

Proximity is bound to the perception and this perception depends on multiple factors related either to the observer (age, intellectual faculties, culture, experience, level of attention...), or to the performances of the "tools" used (instruments of measure, microscopes, scales.) [5]. It depends on the level of knowledge and information that the observer has about things he compares [3.6].

Geometric models introduced the notion of symmetry in the concept of proximity. Nevertheless, at the time of a comparison of two things without an objective criterion of measurement, there is always one thing which is used as a reference, a prototype and the other thing is considered as an instance or a variation. This induces a dissymetry in the relation of proximity between the two things in relation with the thing taken as the reference [3].

The number of properties characterizing one thing is going to modify the proximities : the more the number of common features between two things is important, the more their proximity is going to be admitted [3,5]. However, it is necessary to take into account the importance of these features in the definition of the things : if properties concern only details, it is likely that proximities will be very little modified.

3. Conception of a criterion of proximity in the medical domain

A criterion of proximity is evidently tied with the setting of a semantic proximity. It can be considered according to two different but complementary points of view :

- *the position of a concept within a structure.* The determination of proximity can be ground on the different links between the terms; it corresponds to a quantification of the relations surrounding a given concept.
- *the own meaning of each medical concept* independently of any classification. It amounts to formulate computationaly in an easy way the inferences and heuristics which define proximities between all concepts in the medical world.

3.1 Hierarchical approach

Here, we take into account hierarchical aspects within a terminology. It is really obvious that calculated proximities according to this approach depend closely on the structure of the terminology and could vary with the source in use.

These relations have sometimes a particular meaning making them more explicit. We take into account different meanings of the links in assigning a weight depending on the kind of relation.

The main characteristic of these relations is their orientation and therefore their absence of symmetry. In terms of proximity, we find this asymmetry in function of the direction of the comparison : with a "bottom-up" relation which joins a concept to one of its ancestors, the reference in the comparison being an ascendant of the concept, the proximity

is more important than in the case of a "top-down" relation where the reference is one of the descendants of the considered concept [7].

On the other hand, the more we go down into the hierarchy, the more the differentiation between elements is based on some points of details and therefore the more the proximity is important.

In taking and in modifying the criterion suggested by Botti [8], we can summarize all the properties aforementioned using the formula :

$$p(X,Y)=\frac{U-2\times F\times T\times K\times P\times N(X,Y)}{U}\times 100$$

with : **U** : unitary value. May U be fixed or dependent of the classification ? If it is fixed, we can find some negative values when ***2FTKPN(X,Y)*** is greater than ***U***, which can arise when the number of links between the two concepts is important. We choose to give U a fixed value and negative values are set to 0.

F : coefficient varying with the kind of linkage between X and Y

T : coefficient varying with the direction of the relation. When several links are crossed and if these links have different directions, T was maintained as if it was about a bottom-up relation, F takes at this time a different value.

K : weight dependent on the kind of relation ; it varies with the number of the links between the two concepts : if this number is greater than 1, K is the mean of the values of all the links found on a path

P : inverse of the level of depth of the starting concept with regards to the root of the hierarchy

N(X,Y) : Number of links between X and Y.

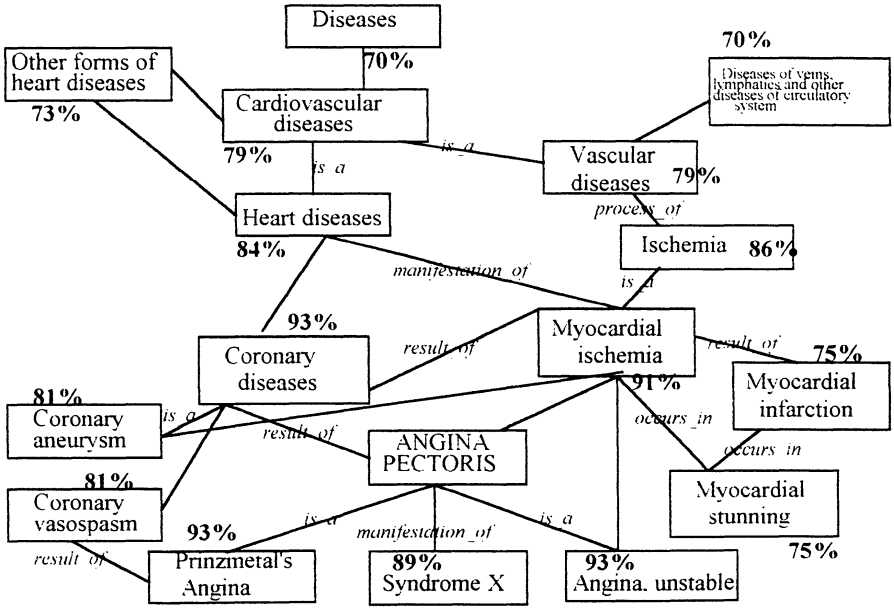

figure 1 example in UMLS : proximities of angina pectoris

With the UMLS structure [9], several proximity values are found between two concepts ; the most important value will be kept. Parameters values are fixed as follows :

U = 10,

F = 1 if X and Y are on a same lineage; 1.5 otherwise

T = 1 if the relation is ascending; 1.5 otherwise

K = 1 for a *is-a* link, 1.2 for a *part-of* link, 1.5 for other links with specified meaning and 1.8 otherwise

P = inverse of depth level

We have to determine the root of the network in order to calculate the depth of the starting concept. On figure 1, three concepts are likely to be the root : *Diseases, Other forms of heart diseases* and *Diseases of...*, we stop with the first encountered root, so, with the smaller depth level.

3.2 Conceptual approach

This approach consists in the characterization of each term with a conceptual representation of its properties and in making an intersection of the graphs in order to determine common features between terms.

This approach is now in development in some other domains [10,11]. It was used in medical domain to structure medical information of a data base [12] and within MAOUSSC architecture in order to characterize precisely diagnostic and therapeutic procedures [13].

Several characteristics allow us to identify determinant factors of proximities between medical terms :

- *anatomical localization* which is an essential factor due to its stability and its precise and well known definition,
- *nature of the concept* which allows to specify if it is about a disease, a symptom, a sign of examination, a biological anomaly, a diagnostic procedure or a therapeutic procedure
- *medical content* allows to specify concepts in function of their nature: if it is about a disease: physiopathology or mechanisms (infection, necrosis, tumor...),if it is about a therapeutic procedure (resection, prosthetic replacement, dilatation....)..etc...

Medical concepts are exploded into elementary concepts and can thus be described according to these three axes. Three axes could be insufficient for completely characterize some concepts. Of course, some other axes can be defined but in a first time we restrict our work to these.

In order to execute the calculation we choose to weight the axes of description, which allow to make proximities vary and therefore to modulate the results : for example, we can weight more the anatomical axis for some researches centered on a topographic region or the axis nature to obtain solely pathologies or therapeutic procedures.

The calculation of this approach with UMLS is not possible due to the absence of some relations within UMLS (particularly, relations between several concepts and their anatomical component), it requires an other source of data having this kind of relations. We use SNOMED III [14] for concepts related to diseases and MAOUSSC for concepts related to diagnostic and therapeutic procedures.

3.3. Global approach

The combination of the two aforementioned approaches allows to get a precise characterization of a medical concept in taking into account at a time the intrinsic properties of each concept (conceptual approach) and its relative position within a terminology (hierarchical approach).

If we consider two concepts, we can search for elementary concepts on each axis of the conceptual structure to determine conceptual proximity. On the other hand, it is possible to calculate the hierarchical proximity between the two starting terms. Therefore, we obtain two different values of proximity between two terms. The problem is then to choose the one that will be kept as proximity between the two terms.

4. Discussion

Proximities between medical terms are based on meaning of the concepts and so are not easy to estimate. Does the defined criterion verify the main characteristics of the concept of proximity given above ?

The context is clearly defined, it corresponds to information in a medical thesaurus. Founded proximities are also reliable and reproducible.

Perception phenomenons are moderate because medical concepts have most of the time a precise definition admitted by the whole medical profession. However, some of the concepts have a different interpretation depending on the specialization of the physicians, for example the concept *Arterial hypertension* is differently estimated by a cardiologist and a nephrologist.

The asymmetry is found when the concepts are situated in a hierarchical structure, it is obvious when we compare *epigastric pain* and *abdominal pain* ; these concepts being linked with a *is-a* relation but between *abdominal pain* and *thoracic pain*, it is difficult to say if the relation is symmetric or not.

The precision of the concepts description can modify conceptual proximities either in enlarging the number of axes, or in modifying elementary concepts founded on each axis. Taking into account the maximal value instead of meaning values allow to limit these variations.

The relevance of the obtained results is actually difficult to appreciate because there is a lack of references about proximity in the medical domain in the literature. The valorizations of all the parameters and weights have to be precised and the description and cutting up of the concepts in elementary concepts must be detailed.

A similar approach could be made by using the GALEN model [15].

References

[1] **Lindberg D.A., Humphreys B.L., McCray A.T.** The Unified Medical Language System *Methods Inf Med, 1993, 32(4) : 281-91*

[2] **Rips L.J.,Shoben E.H., Smith E.E.** Semantic distance and the verification of semantic relations *Journal of verbal learning and verbal behavior, 1973, 12 : 1-20*

[3] **Tversky A., Gati I.** Studies of similarities, *Cognition and categorization, 1978*

[4] **Nelson J.S., Tuttle M.S., Cole W.G., Sherertz D.D., Sperzel W.D., Erlbaum M.S., Fuller L.L., Olson N.E.** From meaning to term . semantic locality in the UMLS Metathesaurus *Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care, 1992 : 209-213*

[5] **Goldstone R.** The role of similarity in categorization : providing a groundwork. *Cognition, 1994, 52, 125-157*

[6] **Delugach H.S.** An exploration into semantic distance.*Conceptual Structures : Theory and implementation, 7th annual workshop, 1992, 124-137*

[7] **Cimino J.J., Octo Barnett G.** Automated translation between medical terminologies using semantic definitions, *MD Comput,1990, 7(2) : 104-107*

[8] **Botti G., Proudhon H., Fieschi M., Joubert M.** Towards aided translation between medical nomenclatures, *Proceedings of MEDINFO'92*

[9] **National Library of Medicine** UMLS knowledge sources 5th experimental edition, *Bethesda, 1994*

[10] **Poole J.** Similarity in legal case based reasoning as degree of matching between conceptual graphs *Preprints of the first european conference on case based reasoning, 1995, 54-58*

[11] **Puder A., Markwitz S., Gudermann F.** Service trading using conceptual structure *3rd International Conference on Conceptual Structure (ICCS'95)*

[12] **Do Amaral M.B., Satomura Y.** Structuring medical information into a language-independant database *Med Inform, 1994,19(3) : 269-282*

[13] **Burgun A., Botti G., Lukacs B., Mayeux D., Seka L.P., Delamarre D., Bremond M., Kohler F., Fieschi M., Le Beux P.** A system that facilitates the orientation within procedure nomenclatures throught a semantic approach, *Med Inform, 1994, 19(4) : 297-310.*

[14] **Systematized Nomenclature of Medicine** *College of American Pathologists, 3rd edition*

[15] **Rector AL, Salomon WD, Nowlan WA, Rush TW, Zanstra PE, Claassen WMA** A medical terminology server for medical language and medical information systems. *Meth Infor Med 1995: 34(1 2):147-57*