Combining Lexical and Semantic Methods of Inter-terminology Mapping Using the UMLS

Kin Wah Fung, Olivier Bodenreider, Alan R. Aronson, William T. Hole, Suresh Srinivasan

National Library of Medicine, Bethesda, Maryland, USA

Abstract

The need for inter-terminology mapping is constantly increasing with the growth in the volume of electronically captured biomedical data and the demand to re-use the same data for secondary purposes. Using the UMLS as a knowledge base, semantically-based and lexically-based mappings were generated from SNOMED CT to ICD9CM terms and compared to a gold standard. Semantic mapping performed better than lexical mapping in terms of coverage, recall and precision. As the two mapping methods are orthogonal, the two sets of mappings can be used to validate and enhance each other. A method of combining the mappings based on the precision level of sub-categories in each method was derived. The combined method outperformed both methods, achieving coverage of 91%, recall of 43% and precision of 27%. It is also possible to customize the method of combination to optimize performance according to the task at hand.

Keywords:

Unified Medical Language System, controlled terminology, inter-terminology mapping.

Introduction

The need for mapping between biomedical terminologies commonly arises when data encoded in one terminology is reused for a secondary purpose that requires a different system of encoding. Imagine an electronic patient record system that captures clinical information using SNOMED CT codes. It will be a big efficiency gain if ICD9CM and CPT codes can be generated automatically for billing purposes. For this to happen, mappings from SNOMED CT to ICD9CM and CPT are required.

A lot of research has been done on the creation of inter-terminology mappings by algorithmic methods. Generally speaking, the algorithms can be divided into lexically-based or semantically-based methods [1-7]. Lexical methods rely on the lexical properties of terms in a terminology. Terms are first normalized or broken down into components before they are compared and matched. On the other hand, semantic methods find matches by utilizing semantic links between terms from the terminologies being mapped.

The UMLS, with over 130 terminologies represented in one common format, is a useful resource for creating inter-terminology mappings. We have previously reported on the use of the UMLS to create semantic mappings between two clinical terminologies [7]. In this report, we describe the use of a UMLS-based tool (MetaMap) to generate lexical mappings between the same terminologies. The performance of the two methods is compared and one way to combine the two methods is described.

Semantic mapping (IntraMap)

The IntraMap algorithm (a modification from the Restrict to MeSH algorithm) makes use of semantic relationships between UMLS concepts to find mappings [7, 8]. Starting from the source concept (the UMLS concept containing the term in a source terminology from which mapping is sought), the algorithm looks for target concepts (UMLS concepts containing terms in the terminology being mapped to) which are related to the source concept either through synonymy or explicit mapping relationships provided by some source terminologies. Failing to find a target concept, the search will widen by using ancestors of the source concept as starting points to look for target concepts. If that fails again, ancestors of the children of the source concept and finally, ancestors of the siblings of the source concept will be used for target concept searching.

Lexical mapping (MetaMap)

MetaMap is a program developed by the NLM to map biomedical text to concepts in the UMLS [9, 10]. The algorithm of MetaMap is as follows: the input text is first parsed into noun phrases. For each phrase, lexical variants are generated. A candidate set of all UMLS concept names containing at least one of the variants is retrieved. Each candidate is then evaluated using a linguistically principled evaluation function. Finally, complete mappings are constructed by combining candidates involved in disjoint parts of the phrase. If one uses the terms in one terminology as the input text and limits the mapping to UMLS concepts containing terms from a specific target terminology, one can use MetaMap to find inter-terminology mappings. MetaMap operates on English terms from the UMLS Metathesaurus.

Selected for best paper award.

Methods

To evaluate the two mapping methods, we used them to find mappings from SNOMED CT terms to ICD9CM terms. As the gold standard for comparison, we used the January 2004 version of the SNOMED CT to ICD9CM mappings provided by the College of American Pathologists. All the mappings from one SNOMED CT term to a single ICD9CM term in the gold standard were used (84% of all mappings). The version of UMLS used was 2004AA which contained the same version of SNOMED CT as used in the gold standard. For semantic mapping by IntraMap, we set the target terminologies to ICD9CM or MTHICD9 (a source that provides additional entry terms to ICD9CM codes). The explicit mapping relationships from the source 'SNOMEDCT' were ignored; otherwise all mappings in the gold standard would be found because the gold standard mappings were also included in 2004AA. For lexical mapping with MetaMap, the fully-specified English names of the SNOMED CT concepts were used as input strings. The option of 'term processing' was turned on to bypass parsing of the terms into component phrases. Other settings of MetaMap were the same as the default mode on the MetaMap website. The target terminologies were limited to ICD9CM or MTHICD9. While no semantic type restriction was used in this experiment, we show in the discussion that restricting the output of MetaMap to semantic types of interest can improve precision.

The mappings found by the two mapping algorithms were compared to the gold standard in terms of their coverage (percentage of SNOMED CT terms for which mappings were found), recall (the percentage of mappings in the gold standard that were found) and precision (the percentage of found mappings that agreed with the gold standard). The overlap between the two sets of mappings was analyzed. A method of combining the two sets of mappings based on the precision ranking of the sub-categories was derived and the improvement in performance was analyzed.

Results

Semantic mapping alone

As reported in [7], among the 66,382 SNOMED CT terms that were used, IntraMap managed to find ICD9CM mappings for 57,293 terms (86.3% coverage). Overall recall was 43.3% and precision was 22.1%. On average, there were 2.3 mappings found per SNOMED CT term. The precision of the sub-categories of mappings were: mapping by synonymy 78.4%, mapping by explicit mapping relationships 50.1% and mapping by ancestor expansion 9.2%. The mappings found by children and sibling expansion were too small in number to warrant further consideration. The results are summarized in Table 1.

Table 1 - Overall and sub-category performance of semantic mapping by IntraMap

	Sub-category of mapping			Overall
	Synonymy	Explicit mapping	Ancestor expansion	
Coverage	19.5%	17.0%	47.2%	86.3%
Recall	16.6%	13.0%	13.0%	43.3%
Precision	78.3%	50.1%	9.2%	22.1%
Mapping per term	1.1	1.5	3.0	2.3

Lexical mapping alone

MetaMap was able to find mappings for 44,452 out of 66,382 SNOMED CT terms (70.0% coverage). The overall recall and precision was 28.4% and 14.7% respectively. There were on average 2.9 mappings found per SNOMED CT term. Among those mappings that were considered to be perfect matches (MetaMap score of 1000), the precision was 85.8%. For those SNOMED CT terms with no perfect matches found, if we only used the top ranking mappings (the mappings with the highest MetaMap score), the precision was 22.6%. The results are summarized in Table 2.

Table 2 - Overall and sub-category performance of lexical mapping by MetaMap

	Sub-category of mapping		Overall
	Perfect mapping	Top mapping	
Coverage	9.7%	57.2%	70.0%
Recall	8.6%	18.3%	28.4%
Precision	85.8%	22.6%	14.7%
Mapping per term	1.0	1.4	2.9

Overlap between the two sets of mappings

A total of 29,468 mappings (distinct pairs of SNOMED CT and ICD9CM codes) were common to both sets of mappings. This represented 22.6% and 22.9% of all Intra-Map and MetaMap mappings, respectively. This set of common mappings covered 35.7% of the SNOMED CT terms, with recall and precision of 22.5% and 50.8% respectively. The mappings that were only found in one algorithm but not the other were higher in coverage but lower in precision (Table 3).

Table 3 - Mapping performance according to the method of mapping

	Both IntraMap and MetaMap	**Only from IntraMap**	**Only from MetaMap**
Coverage	35.7%	57.4%	51.9%
Recall	22.5%	20.8%	5.9%
Precision	50.8%	13.7%	3.9%
Mapping per term	1.2	2.6	2.9

Altogether 13,797 correct mappings were found by semantic mapping alone and missed by lexical mapping. Among these, mappings found by synonymy, explicit mapping relationship and ancestor expansion constituted 9%, 48% and 43% respectively. One example was the mapping from SNOMED CT term '3072001: Hormone-induced hypopituitarism' to ICD9CM term '253.7: Iatrogenic pituitary disorders'. The failure of MetaMap to find this mapping was expected as it was unlikely that the similarity in meaning between 'hormone-induced' and 'iatrogenic' could be detected by lexical matching alone.

On the other hand, there were 3,906 correct mappings that were found by lexical mapping but missed by semantic mapping. Among these, one was deemed to be a perfect match in MetaMap. This was unexpected as almost all perfect lexical matches were genuine synonyms and should normally belong to the same UMLS concept. If so, the mapping should be identified by IntraMap as well. Indeed, this was an anomaly caused by an editing error in the UMLS. In 2004AA, the ICD9CM term '241.9: Unspecified nontoxic nodular goiter' was assigned to a UMLS concept (C1313958) which was different from the UMLS concept (C1318500) containing the SNOMED CT term '190236006: Non-toxic nodular goiter'. This error has since been corrected and the two terms now belong to the same UMLS concept (C1318500). All the other cases in which the correct mapping was found only by MetaMap had less than perfect MetaMap scores. On inspection of a small sample, many of these were mappings from a narrower to a broader concept. One example was the mapping from the SNOMED CT term '67600007: Vascular-biliary fistula' to the ICD9CM term '576.4: Fistula of bile duct' by way of the synonym 'biliary fistula' in the same UMLS concept. This mapping was not found by IntraMap because the two UMLS concepts containing the two terms were not linked by any hierarchical or mapping relationships in the UMLS.

Combining the semantic and lexical mapping sets

Since semantic and lexical mapping are fundamentally different approaches, they are orthogonal and thus can be used to validate and complement each other. To make use of both sets of mappings simultaneously, we derived a method based on the precision level of each sub-category of mapping. First we created a 'precision ladder' according to the precision of each sub-category of mapping. (Table 4)

Table 4 - Precision ladder according to precision of each sub-category of mapping

Rank	Sub-category	Precision
1	M-PM (MetaMap perfect match)	85.8%
2	I-S (IntraMap synonymy)	78.3%
3	C-O (Combined overlapping)	50.8%
4	I-EM (IntraMap explicit mapping)	50.1%
5	M-TM (MetaMap top score)	22.6%
6	C-IO (Combined IntraMap only)	13.7%
7	I-AE (IntraMap ancestor expansion)	9.2%
8	C-MO (Combined MetaMap only)	3.9%

Next we pooled all the mappings together and arranged them in descending order of precision. If the same mapping appeared in more than one sub-category, only the one in the highest ranking sub-category was kept. If there were multiple mappings for the same SNOMED CT term, only the one with the highest ranking was kept and the alternative lower ranking mappings were discarded. The combined set contained 107,172 mappings for 60,454 SNOMED CT terms (coverage 91.1%). The overall recall and precision of the combined set was 42.9% and 26.6% respectively.

The fact that the mappings were arranged in the order of precision allowed us to further fine-tune the way in which the mappings could be used. By setting different cut-off points on the precision ladder (i.e. ignoring mappings below a certain sub-category) we could obtain different combinations of coverage, recall and precision. As expected, the further down we go on the precision ladder, the higher the coverage and recall but the lower the precision. (Table 5)

The mapping performance did not change further with inclusion of mappings ranking lower than rank 6. This was expected because the mappings in rank 7 (IntraMap mappings found by ancestor expansion) were already included in higher sub-categories (ranks 3 and 6). Mappings from rank 8 did not contribute to the overall performance because rank 1 and rank 5 already covered every SNOMED CT term for which a mapping was found by MetaMap.

The F-score (the harmonic mean of precision and recall) is frequently used as an overall indicator of performance. We calculated the F-scores for each cut-off point with equal weight to precision and recall by the following formula:

F-score = $(0.5/\text{precision} + 0.5/\text{recall})^{-1}$

Looking at the F-scores, rank 4 is the optimal cut-off point if one wishes to optimize equally on recall and precision.

Table 5 - Mapping performance according to the cut-off point on the precision ladder

	Cut-off point on the precision ladder							
	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**	**Rank 6**	**Rank 7**	**Rank 8**
Coverage	9.7%	19.5%	38.1%	50.8%	71.0%	91.1%	91.1%	91.1%
Recall	8.6%	16.6%	24.4%	34.0%	38.2%	42.9%	42.9%	42.9%
Precision	85.8%	78.4%	52.0%	51.6%	39.3%	26.6%	26.6%	26.6%
Mapping per term	1.0	1.1	1.2	1.3	1.4	1.8	1.8	1.8
F-score	0.16	0.27	0.33	0.41	0.39	0.33	0.33	0.33

Discussion

The two mapping algorithms

Semantic and lexical mapping algorithms have their own strengths and weaknesses. In general, semantic mapping is considered to be more precise. This is confirmed by our results (precision of IntraMap 22.1%, precision of MetaMap 14.7%). However, for semantic mapping to work there needs to be a pre-existing knowledge base to provide the semantic linkages. In our study, the UMLS is used as the knowledge base, containing over a million concepts and tens of millions of semantic relationships. The performance of semantic mapping depends heavily on the density and quality of these relationships. If there is no semantic relationship linking two concepts a mapping cannot be found.

On the other hand, lexical mapping does not depend on a pre-existing knowledge base. One can perform lexical matching based solely on the lexical properties of the terms from the terminologies being mapped. However, the mapping algorithm of MetaMap does utilize resources from the UMLS which include the rich collection of synonyms and the SPECIALIST lexicon, which have undoubtedly bolstered its performance.

Further refinement

There are possible ways to fine-tune the performance of each of the mapping algorithms. For IntraMap, it is possible to selectively exclude the relationships contributed by certain source terminologies, if such relationships are less likely to result in correct mappings. Restricting the extent of ancestor expansion (e.g., limiting just to level 1 and 2 ancestors) has been found to improve precision [7].

In MetaMap, there is a built-in option to restrict the target concepts by semantic type (STY). In the original run, no restriction on STY was used. In a subsequent run, we restricted target concepts to 18 STYs related to diseases or findings (e.g. Finding, Disease or Syndrome, Acquired abnormality and Congenital abnormality). This restriction was appropriate because the SNOMED CT terms being mapped were all disorders or findings. As ICD9CM also contained terms for procedures, by excluding STYs like

Therapeutic or preventive procedure and Diagnostic procedure some incorrect mappings could be avoided. One example of such error was the mapping of the SNOMED CT term '168943003: Abdominal aortogram abnormal' (a finding) to the ICD9CM term '88.42: Aortography' (a procedure). In the re-run, 2,588 mappings that were present before were dropped because of the STY restriction. Among these dropped mappings, most of them (97%) turned out to be incorrect. This resulted in a small increase of precision from 14.7% to 14.9%.

Apart from STY, one can further refine the mappings by term type (TTY) information in the UMLS. In ICD9CM, only the lowest level terms (the leaf nodes) are valid for coding. Therefore, in the gold standard, all mappings are to the lowest level terms. In the UMLS, these lowest level terms are given TTYs of PT (preferred terms) while higher level terms have TTYs of HT (hierarchical term). However, the two mapping algorithms did not distinguish between ICD9CM PT and HT terms. If we discard all mappings to HTs, the precision would increase from 22.1% to 32.1% for IntraMap and from 14.7% to 25.0% for MetaMap, without impact on recall.

Enhancing performance through combination

The prospect of combining semantic and lexical mapping is particularly exciting. We showed that by pooling all the mappings together according to their anticipated precision ranking, while removing duplicates and lower-ranking alternative mappings, the overall mapping performance was significantly better than either mapping method used alone. Another advantage of combining the mappings by the precision ladder approach is the possibility of adjusting the recall-precision profile of the combined mappings to suit the task at hand. For instance, if the task is automatic code translation, one would prefer a mapping algorithm with high precision. One way to achieve this is by taking only the first two rungs of the precision ladder which will give highly precise mappings (precision close to 80%) to about 20% of SNOMED CT terms. On the other hand, a more likely use case of the mapping algorithms is to suggest candidate mappings to human editors to assist them in their task of creating mappings. In that situation, one will need a mapping algorithm with high coverage. If one takes

every rung from the precision ladder, one will find candidate mappings for over 90% of SNOMED CT terms, with a precision of 27%. However, in this particular use case, even the incorrect candidate mappings may serve some useful purpose. If we compare only the first three digits of the ICD9CM codes in the candidate mappings and the gold standard mappings, the precision jumps to 49.5%. This means that one out of two of the candidate mappings will either be exactly correct or will bring the editors closer to the correct mapping.

Generalization

Except for validating the method, there was little need for creating a mapping between SNOMED CT and ICD9CM, since one already exists as part of the SNOMED CT distribution. However, the strategy presented here can be applied to virtually every pair of terminologies in the UMLS. In future work, we plan to test the algorithms on other terminologies that do not already have known mappings e.g. between SNOMED CT and MeSH.

Conclusion

Semantic and lexical mapping between two terminologies can be done using resources available in the UMLS. The performance of semantic mapping is generally better than lexical mapping. When the two sets of mappings are combined according to the anticipated level of precision of individual mappings, the overall performance is better than either algorithm used alone. One further advantage of this approach of combining semantic and lexical mapping is the possibility of customization of the trade-off between coverage, recall and precision according to the task at hand.

Acknowledgments

References

[1] Rocha RA, Rocha BH, Huff SM. Automated translation between medical vocabularies using a frame-based interlingua. Proc Annu Symp Comput Appl Med Care 1993:690-4.

[2] Cimino JJ, Barnett GO. Automated translation between medical terminologies using semantic definitions. MD Comput 1990;7:104-9.

[3] Barrows RC, Jr., Cimino JJ, Clayton PD. Mapping clinically useful terminology to a controlled medical vocabulary. Proceedings - the Annual Symposium on Computer Applications in Medical Care 1994:211-5.

[4] Dolin RH, Huff SM, Rocha RA, Spackman KA, Campbell KE. Evaluation of a "lexically assign, logically refine" strategy for semi-automated integration of overlapping terminologies. Journal of the American Medical Informatics Association 1998;5:203-13.

[5] Masarie FE, Jr., Miller RA, Bouhaddou O, Giuse NB, Warner HR. An interlingua for electronic interchange of medical information: using frames to map between clinical vocabularies. Comput Biomed Res 1991;24:379-400.

[6] Rocha RA, Huff SM. Using digrams to map controlled medical vocabularies. Proc Annu Symp Comput Appl Med Care 1994:172-6.

[7] Fung KW, Bodenreider O. Utilizing the UMLS for semantic mapping between terminologies. AMIA Annu Symp Proc 2005:266-70.

[8] Bodenreider O, Nelson SJ, Hole WT, Chang HF. Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. Proceedings / AMIA Annual Symposium 1998:815-9.

[9] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proceedings / AMIA Annual Symposium 2001:17-21.

[10] MetaMap website http://skr.nlm.nih.gov/.

Address for correspondence

8600 Rockville Pike, MSC-3826
Bethesda, MD 20894
USA
E-mail: kwfung@nlm.nih.gov