Automated ICD9CM Coding employing Bayesian Machine Learning: a preliminary exploration

Alan David March

Eitel Lauria
Marist College

Automated ICD9-CM Coding employing Bayesian Machine Learning: a preliminary exploration.

Alan D. March, MD[a-b]; Eitel J. M. Lauría, PhD[c]; Jorge Lantos, MD[d].

[a] *Conceptum SA, Buenos Aires, Argentina.*
[b] *Schools of Medicine and of Business Administration, Universidad del Salvador, Buenos Aires, Argentina.*
[c] *School of Computer Science & Mathematics, Marist College, Poughkeepsie, NY, USA.*
[d] *Swiss Medical Group, Buenos Aires, Argentina.*

Statistical text analysis techniques based on Bayesian machine learning were applied to a set of 2180 free-text discharge diagnoses (including both active and inactive problems) occurring in 767 hospitalizations. All of these diagnoses were previously coded according to ICD9-CM using a computerized version of the 1999 Spanish Edition of the named classification scheme, thus producing a set of codes univocally related to one or more of the section codes occurring for each hospitalization. A bag-of-words was parsed out of each phrase set in order to build a dictionary which would serve as the set of attributes to be measured for each set of discharge diagnoses. The probability weights relating each free-text set of diagnoses to one or more known ICD9 corresponding section codes were computed by means of a Bayesian classifier using naïve Bayes and shrinkage-assisted naïve Bayes. A subset of randomly selected hospitalizations were subsequently classified using the previously derived probability set. 86% of the hospitalizations were correctly classified using this approach. 4% erroneous assignations were due to spelling mistakes and wrongly coded diagnoses. The remaining incorrect assignations were due to low prevalence diagnoses. Given a sufficient number of cases for a training set, Bayesian techniques may be useful for the automated ICD9-CM classification of discharge diagnoses. Determination of optimal probability thresholds, the use of more sophisticated parsing methods, and the availability of more numerous training sets may further increase the success rates of this technique and allow for coding at the 3, 4 or 5-digit level.

Key words: ICD9-CM, Bayesian machine learning,

Introduction

Given the free-text nature of medical diagnoses usually encountered in clinical settings, abstraction, classification and coding are necessary requirements for producing machine readable data which may be subsequently made available for a variety of applications, ranging from statistical analysis to the generation of computer-assisted decision support. Proper coding usually requires trained coders and imposes an important workload on healthcare organizations. In order to lower coding costs, many authors have suggested and experimented with automated coding techniques of very different nature and varying results. Most of these techniques rely heavily on linguistic techniques based on grammar rules and are often confounded by the complex grammatical construction of medical utterances, the sometimes grossly non grammatical nature of spontaneously produced text and other minor details such as local jargon, spelling mistakes and transcription errors. In order to overcome these problems, other authors have suggested the use of statistical techniques for drawing meaning out of natural language texts. Generally speaking, statistical text analysis ignores grammar rules by assuming that words do not appear randomly in written or spoken language, and that meaning can be derived from analyzing the fact that certain words occur together in parts of discourse. Whereas these techniques may not be able to interpret texts in a precise manner, it is none less true that for the purpose of classification they may be effective enough.

In the medical domain, a variety of "coding systems" are available to transform free-text (the most prevalent form of information) into structured data sets. These coding schemes are many and vary greatly with regard to their origin, structure and intent. In some few cases, such as is the Systematized Nomenclature of Medicine (SNOMED), they come under the form of sophisticated thesaurus with an underlying logical structure which allows for lexical operations to be performed. In most cases, however, they represent classification schemes of varying complexity. Such is the case of the "family" of classifications known as the International Classification of Diseases. Originally devised during the late 19th century as a list of causes of death, ICD gradually evolved into a catalogue of diseases and health related conditions, maintained and published by the World Health Organization and adapted by various nations in order to fulfill local requirements (the development of various local adaptations has gained the denomination of "family of classifications" for the ICD).

ICD-based coding is a laborious and cost consuming process requiring specially trained human resources. Although sophisticated tools are available for assistance to coders, the sheer volume of information produced by healthcare organizations often precludes the possibility of coding all of the information available in a timely and low-cost fashion. For this reasons, several authors have suggested and experimented with a number of different techniques for automated coding. Although many of these techniques have met with adequate and sometimes surprisingly high levels of success, the fact that most of them rely on grammar-based rules has hindered there application in settings exceeding the original sites where these techniques were developed. In that sense, language barriers pose the greatest obstacles to reproducibility, as they often require extensive rewriting of basic algorithms. In order to overcome these limitations, the authors have resorted to the above mentioned statistical techniques.

Applying Bayesian Learning to Text Classification.

The automated classification of texts is one of the central problems of the discipline known as *Information Retrieval* (IR). In the most general case, a "learning algorithm" is trained with a set of free text documents, sometimes previously identified or "labeled" as belonging into one or more particular topics or "classes". According to the "learning method" employed, a series of parameters and associated statistics are generated and saved. If the original documents in the learning set were not labeled, a set of natural clusters is generated. In the second step, a test document or collection of documents is processed according to the parameters and statistics generated by the learner in the first step. If the original documents were labeled, in this step the labels corresponding to the test documents are "guessed". If no previous labels exist, the classification of the new documents into one or more of the previously generated clusters is again "guessed".

Several different techniques have been proposed for text classification. Among the best known are Bayesian learners, rule induction and support vector machines.

Bayesian methods constitute a probabilistic approach to machine learning. Bayes' theorem provides the means of calculating the posterior probability of a hypothesis H (i.e $\mathrm{P}(H \mid D)$) based on its prior probability $\mathrm{P}(H)$, the probability $\mathrm{P}(D)$ of the set of observational data D and the likelihood $\mathrm{P}(D \mid H)$ that the observational data D fits the hypothesis H.

$$\mathrm{P}(H \mid D)=\frac{\mathrm{P}(D \mid H)\cdot \mathrm{P}(H)}{\mathrm{P}(D)} \propto \mathrm{P}(D \mid H)\cdot \mathrm{P}(H) \quad (1)$$

Suppose we have a classification problem where the class variable is denoted by C and can take values $C=\{c_1, c_2, ..., c_{|C|}\}$. Consider a data sample $D=\{d_1, d_2, ..d_N\}$ where each instance $d_j \in D$ is represented by n attributes $A_1, A_2, .., A_n$, and corresponds to a class value $c_1, c_2, ..., c_{|C|}$. The Bayesian approach to classifying a new instance would then be to assign the most probable class value by calculating the posterior probability for each class value given the training data set, and from them choosing the one that holds the maximum a posteriori probability.

$$c_{\mathrm{MAP}} = \arg\max_{c_i \in C} \left[\mathrm{P}(D \mid c_i)\cdot \mathrm{P}(c_i)\right] \quad (2)$$

Although we can successfully estimate $\mathrm{P}(c_i)$ from the training data, calculating the joint probability $\mathrm{P}(D \mid c_i)$ is usually not feasible: unless we have a very large training data set, we would end up with estimates which are representative of a small fraction of the instance space and are therefore unreliable. The *naive Bayes* classifier attempts to solve this problem by assuming conditional independence among attributes of the data sample, and therefore computing the joint probability $\mathrm{P}(D \mid c_i)$ as $\prod_{j=1}^{n}\theta_{ij}$, the product of the conditional probabilities θ_{ij} of each attribute value $A_j = a_j$ given the class c_i (i.e. $\theta_{ij} = \mathrm{P}(a_j \mid c_i)$). Replacing $\prod_{j=1}^{n}\theta_{ij}$ in (2) we get:

$$c_{\mathrm{MAP}} = \arg\max_{c_i \in C} \left[\mathrm{P}(c_i)\cdot \prod_{j=1}^{n}\theta_{ij}\right] \quad (3)$$

The maximum likelihood estimates of the conditional probabilities of each individual attribute can be calculated from the frequency distributions of the sample data set D as N_{ij}/N_i, where N_{ij} is the number of training examples for which attribute A_j takes value a_j, and class value is c_i; and N_i is the number of training examples for which the class value is c_i. The prior probabilities $\mathrm{P}(c_i)$ can also be estimated by computing frequency counts on the sample data set.

To solve the cases in which there are very few or no instances in the data set for which $A_j = a_j$ given a certain class value c_i, which would in turn render poor estimates of θ_{ij} or make it equal to zero, a common approach is to estimate θ_{ij} as $\hat{\theta}_{ij} = \frac{N_{ij} + \alpha_{ij}}{N_i + \alpha_i}$, where α_{ij} and α_i can be seen as fictitious counts coming out of our prior estimate of the probability we wish to determine. In rigor, this implies considering a conjugate prior probability given by a Dirichlet distribution (for more details see Ramoni & Sebastiani, 1999). A typical method for choosing α_{ij} and α_i in the absence of other information is to assume uniform distribution of the counts, which means that if an attribute has *r* possible values, $\alpha_{ij} = 1$ and $\alpha_i = r$. This results in

$$\hat{\theta}_{ij} = \frac{N_{ij} + 1}{N_i + r} \tag{4}$$

These assumptions have the effect of substantially reducing the number of distinct conditional probability terms that must be estimated from the training data. Although naive Bayes classification makes a major simplification (conditional independence is a very strong assumption), in practice it often competes well with more sophisticated algorithms. See Rish (2001) for an empirical analysis of the naive Bayes classifier.

Following Mitchell (1997), in order to apply Naive Bayes for text classification (i.e. assigning each document to a set of predefined categories $c_1, c_2, ..., c_{|C|}$), the instances used in the learning stage are given by a group of sample documents, each of which is labeled with a given category. We add to the previous assumptions of Naïve Bayes the following simplifications: (i) each sample document is an instance, and each word in the text of the document is an attribute of the instance. This means that a document with 200 words will have 200 attributes A_1, ..A_{200}; (ii) $\mathrm{P}(c_i)$ is computed as usual from the training set by calculating the proportion of documents that belong to each class. Each $\mathrm{P}(a_j \mid c_i)$ is estimated by accepting that the probability of occurrence of a word w_k in a given document is independent of the position of w_k in the document (w_k identifies the k_{th} word in the vocabulary, the set of all distinct words occurring in all training documents, and denoted as V). Therefore $\mathrm{P}(A_1 = w_k \mid c_i) = ... = \mathrm{P}(A_j = w_k \mid c_i) = \mathrm{P}(w_k \mid c_i)$, which greatly reduces the number of calculations by assuming a position independent probability $\theta_{ik} = \mathrm{P}(w_k \mid c_i)$. We adapt expression (4) to estimate $\theta_{ik} = \mathrm{P}(w_k \mid c_i)$. For this purpose we concatenate all example documents that belong to class value c_i into one document $\mathrm{DOC_i}$. This creates a set of concatenated documents $\{\mathrm{DOC_i}\}$, i=1, |C|. With this we calculate θ_{ik} as:

$$\hat{\theta}_{ik} = \frac{N_{ik} + 1}{N_i + |\mathrm{V}|} \tag{5}$$

where |V| is the size of the vocabulary*;* N_i is the total number of distinct word positions in concatenated document $\mathrm{DOC_i}$; and N_{ik} is the number of times a word w_k appears in concatenated document $\mathrm{DOC_i}$. Finally, given a new document formed by a set of words { a_j }, where a_j denotes the word found at position *j*, we classify it by applying:

$$c_{\mathrm{MAP}} = \arg\max_{c_i \in C} \left[\mathrm{P}(c_i) \cdot \prod_{j=1}^{|\#\text{ of positions in document}|} \mathrm{P}(a_j \mid c_i) \right] \tag{6}$$

with each $\mathrm{P}(a_j \mid c_i)$ estimated as $\hat{\theta}_{ik}$.

Dealing with hierarchies using Shrinkage.

When the number of categories is large compared with the sample training data sets the variance of the probability estimates $\hat{\theta}_{ik}$ increases significantly, which in turn deteriorates the accuracy of the naïve Bayes classifier. But if the set of categories are organized as a hierarchy with leaf nodes $c_1, c_2, ..., c_{|C|}$, this hierarchical structure can be used within the Bayesian learning framework to obtain better probability estimates. Several methods have been proposed to implement hierarchical Bayesian classification, among them Koller & Sahami (1997) and McCallum et al's shrinkage technique (1999). We apply the shrinkage algorithm in our study to improve the estimate of the probability $\theta_{ik} = \mathrm{P}(w_k \mid c_i)$, given the fact that it seems to be more appropriate for documents with both small and large subsets of the total vocabulary V.
For each node in the hierarchy, a maximum likelihood estimate is constructed based on the data associated with that node. The estimate of each leaf node $\hat{\theta}_{ik}$ is then computed by interpolating ("shrinking") its estimate based on the maximum likelihood estimates of its t ancestors $\{\hat{\theta}_{ik}^{(1)}, \hat{\theta}_{ik}^{(2)}, ..., \hat{\theta}_{ik}^{(k)}\}$ in the tree path of length (k-1). $\hat{\theta}_{ik}^{(1)}$ is the estimate at the leaf node, and $\hat{\theta}_{ik}^{(k)} = 1/|V|$ is the uniform estimate for all words w_k, an additional parameter beyond the root, to deal with the fact that even the root may give way to unreliable estimates is the amount of data for some rare words is not enough. This is similar to adding pseudocounts, and dismisses the need to smooth the maximum likelihood estimates, as done in expression (5). See James & Stein (1960) and Carlin & Louis (1996) for more details about shrinkage estimators.

$$\hat{\theta}_{ik} = \lambda_i^{(1)} \cdot \hat{\theta}_{ik}^{(1)} + \lambda_i^{(2)} \cdot \hat{\theta}_{ik}^{(2)} + ... + \lambda_i^{(t)} \cdot \hat{\theta}_{ik}^{(t)} \qquad (7)$$

The interpolation weights $\lambda_i^{(1)}, \lambda_i^{(2)}, ..., \lambda_i^{(k)}$ among the ancestors of class c_i add to 1. The estimates along the path are calculated by substracting each child's data from its parent's before computing the parent's estimate, in order to ensure that the maximum likelihood estimates remain independent

The optimal weights $\lambda_i^{(s)}, s = 1...k$ are calculated by using a simple variation of the EM algorithm (Dempster et al, 1977) that maximizes the likelihood of some unobserved hold out of data. For details of the algorithm see McCallum et (1999)

Experimental Results

Discharge diagnoses, involving both active and inactive problems where obtained for 767 hospitalizations occurring throughout 5 months. The lists of discharge diagnoses were prepared by 8 different experienced physicians as free text phrases. ICD9-CM codes from the 1999 Spanish Edition were assigned to each hospitalization by one trained and experienced coder. Due to the small number of cases, each resulting code was assigned to one ICD9-CM "section", defined as the groupings of 3 digit codes as resulting from chapter subtitles. A total of 112 "section codes" was thus obtained and each 3, 4 or 5-digit code was assigned to one of these groups. The complete list of "section codes" is shown in appendix "A".

Ten runs were generated for each of the two Bayesian learning methods. In each run 100 records were selected at random as a test hold out, and the rest of the records were used for training purposes.

	Accuracy (%)											
Runs	1	2	3	4	5	6	7	8	9	10	Average	SD
Naïve Bayes	67	73	75	79	75	73	78	74	72	74	74.00	3.30
Shrinked Naïve Bayes	83	89	85	89	84	89	85	88	87	85	86.40	2.27

Table 1. Accuracy of naïve Bayes vs. shrinked naïve Bayes

For shrinked naive Bayes, the correct section code appeared in the first candidate code (that with the greatest probability out of the 112 possible groups) in 864 cases (87.5%), in the second candidate code in 46 cases (4.6%), in the third candidate or fourth candidate code for 6 cases respectively (0.6%), in the fifth candidate code for 3 cases (0.5%), and in either of the remaining 107 candidate codes for the remaining 64 cases (6.6%). The average probability for the first, second and third candidate section code when the correct diagnoses is encountered among the first, second or third candidate code respectively is shown in Table XX.

	p for 1st candidate	p for 2nd candidate	p for 3rd candidate
Correct class in 1st place	0.81 ± 0.20	0.04 ± 0.06	0.02 ± 0.02
Correct class in 2nd place	0.53 ± 0.24	0.23 ± 0.11	0.04 ± 0.04
Correct class in 3rd place	0.52 ± 0.28	0.17 ± 0.07	0.10 ± 0.07

Table 2. Correct assignations for candidate order and probability.

In order to explore how absolute probability levels define correct codes in the first candidate, the result database was queried for first candidates with a probability equal or greater than .61 (1 standard deviation below the mean). Through this simple heuristic method, 701 correct codes (71%) were detected. Further research using greater data sets and considering other parameters (ie. the probability gap between the first and second candidate) may serve to further refine an optimal correct code detection method.

Discussion

The above shown experimental results suggest that Bayesian text analysis may be a useful tool in automated coding.

Several methods have been proposed for free text analysis, Bayesian text classification among them. Reviews of these methods have been published by Sebastiani [2000] and by Wilcox and Hripcsak [1999]. To the best of our knowledge, naïve Bayes analysis with shrinkage has not been applied to clinical domains and offers better perspectives than common naïve Bayes, a fact which was predictable given the hierarchical nature of ICD.

One criticism which might be raised against our experimental setting is that of oversimplification. ICD requires coding to be done at the leaf level of the hierarchy, which most often implies the 4- or 5- digit level, and in many cases at least a 3 digit code. Out intent in this preliminary exploration was not to provide a full-fledged automated classifier but rather to explore the possible limitations set by naturally occurring free text annotations (as is the case of discharge diagnoses lists) on Bayesian classifiers, as well as the form of Bayesian Machine Learning best suited to hierarchical classifications. In this sense, the better performance of shrinkage assisted Bayes as compared with simple Bayes has been demonstrated, as have some problems of confounding factors such as high frequency non-discriminating words such as "cancer". Likewise, the setting of a significant probability threshold for confidently assigning a code remains a problem. All of these issues offer future avenues of research.

The influence of the size of training datasets, as has been proven by most authors working with these techniques, poses an important obstacle to obtaining high degrees of accuracy. The authors fully agree with Wilcox and Hripcsak [1999] that manual preparation through manual physician review may be prohibitive in production scenarios. Nevertheless, it is important to note that free-text lists of diagnoses are commonly encountered in real world settings under the form of lists of diagnoses in discharge abstracts. For public health providers these lists are required by law in Argentina. In the private sector, requirements for provision of these lists is becoming increasingly common due to the necessity of performance analysis in the face of rising healthcare costs.

Although techniques exist for working with small datasets in the case of Bayesian Machine Learning, it is nonetheless true that the sheer number of possible ICD9-CM codes (more than 15.000) might render any of these techniques finally unusable. Notwithstanding this fact, it is none less true that a non-negligible number of these diagnoses belong into low prevalence diagnoses and could thus be predicted to have a low impact on real world settings, as they would reasonably find themselves excluded from training sets until they find their way into the discharge diagnoses databases. The manner in which the first appearance of a low prevalence diagnoses should be dealt with remains a question to be settled but it might reasonably be predicted that the first candidate code should render a low probability which might be dealt with using thresholds. For the present dataset, an arbitrarily set threshold lowered the correct code detection in about a 10%, but more refined techniques which remain to be developed and tested might improve this result.

Aside from coding errors in the original datasets and the above mentioned low prevalence diagnoses, the most frequent cause of erroneous assignations was the appearance of general morphological (ie: "cancer of", "fracture of") modifiers in diagnostic phrases. For diagnoses appearing on only few occasions, the resulting probability weight of other instances of the modifier shifted the assignation to the most frequently occurring site (ie: "cancer of the esophagus", which only appeared twice in our dataset, was assigned to the section code of the most frequently occurring cancer). This problem should be neutralized with the use of greater data sets, or by defining the whole concept ("cancer of the esophagus") as an attribute rather than considering the separate attributes ("cancer", "of", "the", "esophagus"). Likewise, the elimination of stop words (ie: "of", "the") could further enhance the accuracy by removing words devoid of intrinsic meaning, but this last factor must be further explored as these words occur in a uniform fashion throughout the whole dataset.

With regard to the medical domain, Bayesian techniques, and possibly most statistical text analysis methods, may be particularly well adapted to complex classification schemes such as ICD as opposed to controlled terminology-based coding methods such as SNOMED. Assigning an ICD code to a particular set of discharge diagnoses is a complex rule-based method more akin to an expert system rather than a natural language processor. The existence of exclusions, inclusions and secondary coding may pose NLPs with too complex requirements and a statistical technique based on co-ocurrence of words and/or phrases impresses us as a simpler approach. Nevertheless, some elementary form of NLP might serve to improve the rate of correct code assignments in what would finally configure a hybrid approach. As has been showed above, a particularly confounding factor for correct classification is imposed by the existence of general morphological terms which apply to many topographies, as is the case of "cancer of" or "fracture of". The authors have employed a simple bag-of-words approach to generate the training sets. The usage of more sophisticated forms of parsing, such as a grammar-based method which may be able to detect compound word concepts ("cancer of larynx" instead of "cancer" and "larynx" as separate attributes) may improve our results. The fact that diagnoses usually come under the form of simple nominal syntagmas suggest that this approach might be tenable, not requiring sophisticated natural language parsers.

Conclusion

Bayesian text classification may be applied to automated ICD-9CM "section codes" assignation with a reasonable accuracy (86%) using small sets of diagnoses. It may be safely predicted, considering the requirements for Bayesian learning, that give larger training sets the percent of correct diagnoses will increase and coding at a greater level of detail (3, 4 or 5-digit) may be procured.

The application of these techniques may greatly reduce the workload of coding departments with an acceptable degree of error, not substantially greater than the errors commonly incurred by manual coding.

References

Carlin B, Louis T (1996): Bayes and Empirical Bayes Methods for Data Analysis. Chapman and Hall.

Dempster AP, Laird NM, Rubin DB (1977): Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society. B 39, 1-38.

James W, Stein C (1960): Estimation with Quadratic Loss, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (vol. 1), Berkeley, CA: University of California Press, pp. 361-379.
Koller D, Sahami M (1997): Hierarchically classifying documents using very few words. In ICML-97: Proceedings of the Fourteenth International Conference on Machine Learning, pp 170--178.
Mitchell T (1997), *Machine Learning*, McGraw-Hill.
Ramoni M, Sebastiani P (1999): Bayesian Methods for Intelligent Data Analysis. In: Berthold M and Hand DJ (Ed.): *Intelligent Data Analysis: An Introduction*, Springer, New York, NY
Rish I (2001): An empirical study of the naive Bayes classifier. In IJCAI-01 workshop on Empirical Methods in AI.
Sebastiani F (2002): Machine Learning in Automated Text Categorization. ACM Computing Surveys 34 (1): 1-47
Wilcox A, Hripcsak G (1999): Classification Algorithms Applied to Narrative Reports. Proceedings of the 1999 AMIA Symposium

Author's Address:
Conceptum SA, Vuelta de Obligado 2596 Piso 5, 1428 Ciudad Autónoma de Buenos Aires, Argentna; e-mail: amarch@conceptum.com.ar.

Appendix A: List of "section codes"

"Section Code"	Title
1	ENFERMEDADES INFECCIOSAS INTESTINALES (001-009)
2	TUBERCULOSIS (010-018)
3	ENFERMEDADES BACTERIANAS ZOONOSICAS (020-027)
4	OTRAS ENFERMEDADES BACTERIANAS (030-041)
5	INFECCION DEL VIRUS DE LA INMUNODEFICIENCIA HUMANA (VIH) (042)
6	POLIOMIELITIS Y OTRAS ENFERMEDADES VIRALES DEL SNC NO TRANSMITIDAS POR ARTROPODOS (045-049)
7	ENFERMEDADES VIRALES ACOMPAÑADAS DE EXANTEMA (050-057)
8	ENFERMEDADES VIRALES PORTADAS POR ARTROPODOS (060-066)
9	OTRAS ENFERMEDADES DEBIDAS A VIRUS Y A CHLAMYDIAE (070-079)
10	RICKETTSIOSIS Y OTRAS ENFERMEDADES PORTADAS POR ARTROPODOS (080-088)
11	SIFILIS Y OTRAS ENFERMEDADES VENEREAS (090-099)
12	OTRAS ENFERMEDADES ESPIROQUETALES (100-104)
13	MICOSIS (110-118)
14	HELMINTIASIS (120-129)
15	OTRAS ENFERMEDADES INFECCIOSAS Y PARASITARIAS (130-136)
16	EFECTOS TARDIOS DE LAS ENFERMEDADES INFECCIOSAS Y PARASITARIAS (137-139)
17	NEOPLASIA MALIGNA DE LABIO, CAVIDAD ORAL Y FARINGE (140-149)
18	NEOPLASIAS MALIGNAS DE LOS ORGANOS DIGESTIVOS Y DEL PERITONEO (150-159)
19	NEOPLASIA MALIGNA DE LOS ORGANOS RESPIRATORIOS E INTRATORACICOS (160-165)
20	NEOPLASIA MALIGNA DE HUESO, TEJIDO CONECTIVO, PIEL Y MAMA (170-176)
21	NEOPLASIA MALIGNA DE ORGANOS GENITOURINARIOS (179-189)
22	NEOPLASIA MALIGNA DE OTRAS LOCALIZACIONES Y DE LOCALIZACIONES NO ESPECIFICADAS (190-199)
23	NEOPLASIA MALIGNA DE TEJIDOS LINFATICOS Y HEMATOPOYETICOS (200-208)
24	NEOPLASIAS BENIGNAS (210-229)
25	CARCINOMA IN SITU (230-234)
26	NEOPLASIAS DE EVOLUCION INCIERTA (235-238) (L)
27	NEOPLASIAS DE NATURALEZA NO ESPECIFICADA (239)
28	TRASTORNOS DE LA GLANDULA TIROIDEA (240-246)
29	ENFERMEDADES DE OTRAS GLANDULAS ENDOCRINAS (250-259)
30	DEFICIENCIAS NUTRITIVAS (260-269)
31	OTROS TRANSTORNOS METABOLICOS Y DE INMUNIDAD (270-279)
32	ENFERMEDADES DE LA SANGRE Y DE LOS ORGANOS HEMATOPOYETICOS (280-289)
33	PSICOSIS ORGANICAS (290-294)
34	OTRAS PSICOSIS (295-299)
35	TRASTORNOS NEUROTICOS, DE LA PERSONALIDAD Y OTROS TRASTORNOS MENTALES NO PSICOTICOS (300-316)
36	RETRASO MENTAL (317-319)
37	ENFERMEDADES INFLAMATORIAS DEL SISTEMA NERVIOSO CENTRAL (320-326)
38	ENFERMEDADES HEREDITARIAS Y DEGENERATIVAS DEL SISTEMA NERVIOSO CENTRAL (330-337)
39	OTROS TRASTORNOS DEL SISTEMA NERVIOSO CENTRAL (340-349)
40	TRASTORNOS DEL SISTEMA NERVIOSO PERIFERICO (350-359)
41	TRASTORNOS DEL OJO Y DE LOS ANEXOS (360-379)
42	ENFERMEDADES DEL OIDO Y PROCESO MASTOIDEO (380-389)
43	FIEBRE REUMATICA AGUDA (390-392)

44	ENFERMEDAD CARDIACA REUMATICA CRONICA (393-398)
45	ENFERMEDAD HIPERTENSIVA (401-405)
46	CARDIOPATIA ISQUEMICA (410-414)
47	ENFERMEDADES DE LA CIRCULACION PULMONAR (415-417)
48	OTRAS FORMAS DE ENFERMEDAD CARDIACA (420-429)
49	ENFERMEDAD CEREBROVASCULAR (430-438)
50	ENFERMEDADES DE LAS ARTERIAS, ARTERIOLAS Y CAPILARES (440-448)
51	ENFERMEDADES DE VENAS Y LINFATICOS, Y OTRAS ENFERMEDADES DEL APARATO CIRCULATORIO (451-459)
52	INFECCIONES RESPIRATORIAS AGUDAS (460-466)
53	OTRAS ENFERMEDADES DEL TRACTO RESPIRATORIO SUPERIOR (470-478)
54	NEUMONIA Y GRIPE (480-487)
55	ENFERMEDAD PULMONAR OBSTRUCTIVA CRONICA Y ENFERMEDADES ASOCIADAS (490-496)
56	NEUMOCONIOSIS Y OTRAS ENFERMEDADES PULMONARES OCASIONADAS POR AGENTES EXTERNOS (500-508)
57	OTRAS ENFERMEDADES DEL APARATO RESPIRATORIO (510-519)
58	ENFERMEDADES DE LA CAVIDAD ORAL, GLANDULAS SALIVARES Y MAXILARES (520-529)
59	ENFERMEDADES DEL ESOFAGO, ESTOMAGO Y DUODENO (530-537)
60	APENDICITIS (540-543)
61	HERNIA DE LA CAVIDAD ABDOMINAL (550-553)
62	ENTERITIS Y COLITIS NO INFECCIOSA (555-558)
63	OTRAS ENFERMEDADES DEL INTESTINO Y DEL PERITONEO (560-569)
64	OTRAS ENFERMEDADES DEL APARATO DIGESTIVO (570-579)
65	NEFRITIS, SINDROME NEFROTICO Y NEFROSIS (580-589)
66	OTRAS ENFERMEDADES DEL APARATO URINARIO (590-599)
67	ENFERMEDADES DE ORGANOS GENITALES MASCULINOS (600-608)
68	TRASTORNOS DE MAMA (610-611)
69	ENFERMEDAD INFLAMATORIA DE LOS ORGANOS PELVICOS FEMENINOS (614-616)
70	OTROS TRASTORNOS DEL TRACTO GENITAL FEMENINO (617-629)
71	EMBARAZO ECTOPICO Y MOLAR (630-633)
72	OTRO EMBARAZO CON RESULTADO ABORTIVO (634-639)
73	COMPLICACIONES PRINCIPALMENTE RELACIONADAS CON EL EMBARAZO (640-648)
74	PARTO NORMAL Y OTRAS INDICACIONES PARA CUIDADOS DURANTE EL EMBARAZO, TRABAJO DE PARTO Y PARTO (650-659)
75	COMPLICACIONES QUE SE PRESENTAN PRINCIPALMENTE DURANTE EL CURSO DEL PARTO (660-669)
76	COMPLICACIONES DEL PUERPERIO (670-676)
77	INFECCIONES DE LA PIEL Y DEL TEJIDO CELULAR SUBCUTANEO (680-686)
78	OTROS ESTADOS INFLAMATORIOS DE LA PIEL Y DE LOS TEJIDOS SUBCUTANEOS (690-698)
79	OTRAS ENFERMEDADES DE LA PIEL Y DEL TEJIDO SUBCUTANEO (700-709)
80	ARTROPATIAS Y TRASTORNOS RELACIONADOS (710-719)
81	DORSOPATIAS (720-724)
82	REUMATISMO, SALVO DE LA ESPALDA (725-729)
83	OSTEOPATIAS, CONDROPATIAS Y DEFORMIDADES MUSCULOESQUELETICAS ADQUIRIDAS (730-739)
84	ANOMALIAS CONGENITAS (740-759)
85	CAUSAS MATERNAS DE MORBILIDAD Y MORTALIDAD PERINATALES (760-763)
86	OTRAS ENFERMEDADES CON ORIGEN EN EL PERIODO PERINATAL (764-779)
87	SINTOMAS (780-789)
88	HALLAZGOS ANORMALES NO ESPECIFICOS (790-796)
89	CAUSAS DE MORBILIDAD Y MORTANDAD DESCONOCIDAS Y MAL DEFINIDAS (797-799)

90	FRACTURA DE CRANEO (800-804)
91	FRACTURAS DE CUELLO Y TRONCO (805-829)
92	FRACTURA DE MIEMBRO SUPERIOR (810-819)
93	FRACTURAS DE MIEMBRO INFERIOR (820-829)
94	LUXACION (830-839)
95	ESGUINCES Y TORCEDURAS DE ARTICULACIONES Y MUSCULOS ADYACENTES (840-848)
96	LESION INTRACRANEAL, SALVO AQUELLAS CON FRACTURA DEL CRANEO (850-854)
97	LESION INTERNA DE TORAX, ABDOMEN Y PELVIS (860-869)
98	HERIDAS ABIERTAS DEL MIEMBRO SUPERIOR (880-887)
99	HERIDAS ABIERTAS DEL MIEMBRO INFERIOR (890-897)
100	LESION DE VASOS SANGUINEOS (900-904)
101	EFECTOS TARDIOS DE LESIONES,ENVENENAMIENTOS,EFECTOS TOXICOS Y OTRAS CAUSAS EXTERNAS (905-909)
102	LESION SUPERFICIAL (910-919)
103	CONTUSION CON SUPERFICIE CUTANEA INTACTA (920-924)
104	LESION POR APLASTAMIENTO (925-929)
105	EFECTOS DE CUERPO EXTRAÑO QUE ENTRA A TRAVES DE ORIFICIO (930-939)
106	QUEMADURAS (940-949)
107	LESION DE NERVIOS Y MEDULA ESPINAL (950-957)
108	CIERTAS COMPLICACIONES TRAUMATICAS Y LESIONES NO ESPECIFICADAS (958-959)
109	ENVENENAMIENTO POR DROGAS, SUSTANCIAS MEDICAMENTOSAS Y SUSTANCIAS BIOLOGICAS (960-979)
110	EFECTOS TOXICOS DE SUSTANCIAS PRIMORDIALMENTE NO MEDICAMENTOSAS CON RESPECTO A SU ORIGEN (980-989)
111	OTROS EFECTOS Y EFECTOS NO ESPECIFICADOS DE CAUSAS EXTERNAS (990-995)
112	COMPLICACIONES DE CUIDADOS QUIRURGICOS Y MEDICOS NO CLASIFICADOS BAJO OTROS CONCEPTOS (996-999)