Quality of Life through Quality of Information
J. Mantas et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-101-4-240

Information extraction from SMS text related to a reminder service for outpatients

Stefania RUBRICHI[a,1], Stéphanie EKU NDAM[a], Andrea BATTISTOTTI[b], and Silvana QUAGLINI[a]
[a] *Dipartimento di Ingegneria Industriale e dell'Informazione, University of Pavia, Italy,*
[b] *Azienda Ospedaliera Fatebenefratelli e Oftalmico – Milano, Italy*

Abstract. This work evaluates the users' satisfaction with an SMS-based reminder system that is being used since about six years by an Italian healthcare organization. The system was implemented for reducing dropouts. This goal has been achieved, as dropout decreased from 8% to 4%. During these years, a number of reminded citizens, even not required, sent an SMS message back, with comments about the service, further requirements, etc.. We collected some thousands of them. Their analysis may represent a useful feedback to the healthcare organization. We used conditional random fields as the information extraction method for classifying messages into *appreciation*, *critique*, *inappropriateness*, etc. The classification system achieved a very good overall performance (F_1-measure of 94%), thus it can be used from here on to monitor the users' satisfaction in time.

Keywords. NLP, Information Extraction, Conditional Random Fields, SMS

Introduction

A dropout is a missed visit, i.e. a visit that has been scheduled but not attended by a patient, without any notice from him. This phenomenon, diffused in every healthcare organization delivering services on a scheduling basis, may have different causes, but the main one is due to citizens, that simply forget the appointment, or that cannot attend for any reason, and forget to notify the doctors. Dropout is a serious problem for healthcare organizations because every such event represents an economic damage. A planned visit, while not executed, still implies a fixed cost and is not reimbursed by the national healthcare system. Moreover, it represents a social damage, because it extends the waiting lists. In 2005, a reminder system has been implemented at the healthcare organization "Azienda Ospedaliera di Pavia" [1]. After 1.5 years of usage, the first system evaluation showed that dropout decreased from 8% to 4.2%. During the next years, this rate was maintained. The system sends this message three days before the scheduled date: "*The Healthcare Company of Pavia reminds you the visit of dd/mm. If you want to cancel it, call free 800448800. Thank you for the cooperation*". The citizen is not required to answer, but someone does, for different reasons: to thank for the service, to notify a possible error (e.g., he does not remember to have booked any visit), to ask for further details, etc. After collecting a significant number of answers, we decided to analyze them, in order to evaluate not only the satisfaction of the citizens

[1]Corresponding Author. Stefania RUBRICHI, stefania.rubrichi@unipv.it

with the system, but also to individuate system pitfalls and consequently take corrective actions. As a matter of fact, any SMS message a citizen perceives as an error, reveals an organizational bug or malpractice. Thus, we developed a classification system that not only takes a picture of the actual situation, but can be used also in the future to constantly monitor the outpatient service quality.

1. Materials and Methods

1.1. The SMS message corpus

Among the about 6000 SMS messages received from July 2005 to July 2011, 2300 have been used for the analysis. The remaining ones were not usable because either they were empty (due to a lot of people that try calling back), or they were clearly uncorrelated to the service (probably sent by error). First of all we start reading the messages, in order to understand the citizens' motivations to send them. Examples of received SMS messages are:

"Very good initiative, I confirm the visit, thank you for your efficiency"
"The reservation is for the hospital of Voghera, not the hospital of Pavia"
"Don't send any other messages and save public money!"
"Every week I receive this message but I repeat I NEVER BOOKED ANY VISIT".

Our first step lay in identifying the underlying semantic classes of concepts that are relevant to the concepts, relations and events described in the SMS messages and that we intend to extract. The derived classification is the following:

Appreciation	**Misunderstanding**	Not living in Pavia province
Confirmation (e.g., message #1)	about the town where the visit has	Visit already done
Thanks (e.g., message #1)	been reserved (e.g., message #2)	Visit already shifted
Further requests	**Error notification**	Visit already cancelled
Cancellation	Error	**Disappoint**
Shifting	Repeated error (e.g., message #4)	Scorn (e.g., message #3)
Details request: Ward/Time	Surprise	**Problem with toll free number**

1.2. Information Extraction

Nowadays, with the increasing amount of the biomedical information contained in large repositories of text documents, the need to effectively and readily access and query them becomes more and more important. Natural language processing (NLP), and more precisely information extraction (IE), can solve this problem since they attempt to use automated means to process text and deduce its syntactic and semantic structure. IE is the task of extracting explicitly stated facts from text documents by selectively identifying mentions of entities and their relationships, and then representing them in a consistent structured format, e.g., a database or ontology.

Recently, substantial resources have been allocated for applying NLP to biomedical documents. Excellent efforts have been documented in the literature on IE in the biomedical domain [2-6], and its subsequent applications. Approaches to IE span a broad range, from rule-based systems to machine learning (ML). Rule-based systems make decisions on sets of disambiguation rules that play an important role in discovering a named entity or relation, which specify, for example, that an ambiguous word belongs to a particular named entity rather than to another one if it follows another specific named entity. Rule-based IE systems can be very effective, but cannot easily adapt to domain changes. To address this problem of portability, recent research has focused on applying empirical methods to the IE process. Indeed automatic IE task is an attractive field for ML applications and is receiving increasing attention from the ML. The advantages of ML approaches are that they do not require human intuition and can be retrained without reprogramming for any domain. For the purpose of this work, we employed the Conditional Random Fields (CRFs) algorithm, which belongs to the family of *supervised ML* approaches.

1.2.1. Conditional Random Field

We formulate the problem as a sequence-labeling task. This task consists in predicting a sequence of labels y = $<y_1, \ldots, y_t>$ given an observed sequence of tokens x = $<x_1, \ldots, x_t>$. For example, x might be a natural language sentence, i.e., SMS text, and y the semantic labels of the word of this sentence. CRFs [7] are the state-of-the-art. A CRF is probabilistic framework for labeling and segmenting sequential data. In what follows, X is a random variable over data sequences to be labeled, and Y is a random variable over corresponding label sequences. The underlying idea is to define a conditional probability p(y|x) distribution over label sequences, given a particular observation sequence x rather than the joint distribution over both label and observation sequences. Based on this formulation, let x* be a novel observation sequence, the understanding process select the label sequence y* which maximizes the a posteriori probability:

$$y^* = \arg max_{y \in Y} \; p(y|x^*)$$

Linear-chain CRFs define the conditional probability of a label sequence given an input sequence to be a normalized product of potential functions, each of the form:

$$\exp\left(\sum_j \lambda_j t_j(y_{i-1};\, y_i;\, x;\, i) + \sum_k \mu_k s_k(y_i;\, x;\, i)\right)$$

where $t_j(y_{i-1}; y_i; x; i)$ is a transition feature function of the entire observation sequence and the labels at positions i and i-1 in the label sequence; $s_k(y_i; x; i)$ is a state feature function of the label at position i and the observation sequence; and λ_j and μ_k are parameters to be estimated from training data. Therefore, CRFs predict the labels of words by using large number of interdependent descriptive characteristics (the feature functions precisely) of the input by assigning real-value weight to these features. This can be seen as a way to "capture" the hidden patterns of labels and features, and "learn" what the likely output might be, given these patterns. For this purpose, we wish to exploit all available information in x to infer the right y, by incorporating both semantic and syntactic characteristics. Our system uses the MALLET[2] implementation of CRFs.

1.2.2. The Extraction Framework

For our problem, we followed these steps:

[2] http://mallet.cs.umass.edu

a. Pre-elaboration of the SMS text: this has been necessary because they usually contains abbreviations (e.g., for the Italian language: ‹‹ki›› → ‹‹chi››, ‹‹cmq›› → ‹‹comunque››, ‹‹xk›› → ‹‹perché››)
b. Annotation of the SMS text, according to the previous classification (manual task). A result of manual annotation is for example:
   <Very good service>**Thanks** <I confirm>**Confirmation** the visit of 26th gen 2011 could I have more information about <ward and time>**WardTime** ?
c. Choice of the classification features. We compiled three types of features to describe the data: (i) lexical, (ii) morphosyntactic and (iii) semantic features.

Lexical Features Covers word-related features: (1) the words identity that is the vocabulary from the training data, (2) the identity of the 3 words before, (3) the identity of the 3 words after, and (4) the lemmas of these words.
Morphosyntactic Features Includes: (1) parts-of-speech of all the words, (2) negation feature which indicates if a negation is detected (e.g., not, no) at a limited distance ([-4; 4]) of words befor, and after the considered one. We used TreeTagger [8] to perform POS tagging of the corpus sentences.
Semantic Features To improve the system performance we incorporated more advanced features related to this task that use external semantic resources. The first type relies on some trigger words and indicates whether they occur in the text (e.g., the word "toll" may be useful for identifying problems with the toll free number). The trigger words were collected manually from the training corpus. The second class relies on a particular semantic resource that is the lists of all Italian cities and regions. Various SMS messages mention Italian cities or regions to report an error (e.g., You definitely got the wrong number because I live in Sicily), or in case of misunderstanding on where the visit will be held (e.g., message #2, par 1.1). We found that this type of feature can provide good clues for identifying such error notifications.

2. Results

We split the 2300 SMS messages into two sets uniformly at random: one for training (about 80%) and one for testing (about 20%) purposes. The performance of the classifier is shown in tables 1 and 2. The resulting overall accuracy was around 94%.

Table 1. Overall experimental results (in %) of CRFs

Micro-average			Macro-average			Overall
Precision	Recall	F_1-measure	Precision	Recall	F_1-measure	accuracy
94.04	94	94.87	93.94	77.64	83.69	94

Table 2. Performance results (in %) of the classifiers on individual labels

Label	N_{train}	N_{test}	Precision	Recall	F_1-measure
Cancellation	27	157	88.88	88.88	88.88
Thanks	257	1074	98.81	97.66	98.23
None	1743	7138	93.73	97.07	95.37
Confirmation	268	1058	95.71	91.79	93.71
Surprise	281	1090	87.54	82.56	84.98
LiveElsewhere	51	178	100	92.15	94.91
Shifting	4	27	100	25	40
AlreadyShifted	4	25	100	25	40
AlreadyDone	5	21	100	80	88.88
CityMisunderstanding	5	29	100	100	100
RipeatedError	19	43	100	73.68	84.84
TollNumber	8	28	100	75	85.71
AlreadyCancelled	20	44	89.47	85	87.17

Scorn	40	161	100	82.50	90.41
Ward/Time	19	74	100	68.42	81.25

Table 2 shows the performance results on the individual labels for all features. Overall, labels whose training examples are scarce suffers from relatively low performance, see *Shifting* and *AlreadyShifted*. On the other hand, some other labels, see *AlreadyDone* and *CityMisunderstanding*, although rare, perform better. Such labels can rely on a more dedicated and informative feature that facilitates good performance.

3. Discussion

The work presented in this paper is an example of how technology can be used to improve the relationship between citizens and institutions, in our case a local healthcare agency of the National Health Service. First of all, analysis of feedback from citizens/patients can be considered to improve the reminder system. For example a significant number of citizens do not understand from the message that "Pavia" is part of the name of the local healthcare agency, and not the location where the visit will be done. Others would be happier with more details in the message, like ward and time. The text of the message could be "re-negotiated" with the AO Pavia responsible for the privacy, who defined the message text in such a way that no sensible information could be derived from it. Second, the analysis discovered anomalies in the outpatient management workflow, which should be corrected to increase the company organizational efficiency. Actual anomalies are related to visit booking performed by healthcare employees who (a) had no *official* role to do it, or (b) did it without following the correct protocol, resulting in non-traceability of the reservation. As regards the results obtained, the system developed can identify the type of message with very good accuracy. For the future we would like to make all the steps described in this paper as much automatic as possible, to provide continuous control for the hospital administrators on the feedback received from citizens, with the ultimate goal of evaluating the effectiveness of corrective actions taken to improve the system and avoid the malpractices described above.

References

Battistotti A, Quaglini S, Cuoco E. Reducing dropouts in outpatient care through an SMS-based system. Stud Health Technol Inform. 2006; 124:935-940.

Rubrichi S, Quaglini S. Summary of Product Characteristics content extraction for a safe drugs usage. J Biomed Inform. 2012; 45(2):231-239.

Friedman C, Johnson SB, Forman B, Stanner J. Architectural requirements for a multipurpose natural language processor in the clinical environment. In: Proceedings of the Annu Symp Comput Appl Med Care. 1995; 347–351.

Meystre S, Haug PJ. Natural language processing to extract medical problems from electronic clinical documents: performance evaluation. J Biomed Inform. 2006; 589–599.

Aronson A. Effective mapping of biomedical text to the UMLS metathesaurus: the MetaMap program. In: AMIA Annu Symp Proc. 2001; 17–21.

Tanabe L, Wilbur WJ. Tagging gene and protein names in biomedical text. Bioinformatics. 2002; 18: 1124–1132.

Lafferty J, McCallum A, Pereira F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: Proc Int Conf Mach Learn, 2001.

Schmid, H. Probabilistic part-of-speech tagging using decision trees. In: Proceedings of the International Conference on New Methods in Language Processing. 1994; 44–49.