2014 IEEE 27th International Symposium on Computer-Based Medical Systems

FRAT-up, a Rule-based System Evaluating Fall Risk in the Elderly

Luca Cattelani*, Federico Chesani†, Pierpaolo Palumbo*, Luca Palmerini*,
Stefania Bandinelli‡, Clemens Becker§, and Lorenzo Chiari*
*Department of Electrical, Electronic, and Information Engineering
University of Bologna, Italy
Email: {luca.cattelani, pierpaolo.palumbo, luca.palmerini, lorenzo.chiari}@unibo.it
†Department of Computer Science and Engineering
University of Bologna, Italy
Email: federico.chesani@unibo.it
‡Geriatric Unit, Azienda Sanitaria Firenze, Italy
Email: stefania.bandinelli@asf.toscana.it
§Robert-Bosch-Krankenhaus Geriatric Rehabilitation Clinic, Germany
Email: clemens.becker@rbk.de

***Abstract*—About one-third of persons over 65 are subject to at least one fall during a year, and many of them are subjected to health, psychological and financial consequences. A requirement to improve the effectiveness of preventive interventions is to timely identify subjects at higher risk.**

In this work we introduce the Farseeing Fall Risk Assessment Tool (FRAT-up), a software tool for evaluating the fall risk of a subject, based on known risk factors. The tool is based on probabilistic rules, generated automatically from a light ontology capturing the scientific findings about risk factors. FRAT-up has been tested on the InCHIANTI dataset, showing performances comparable with state-of-the-art tools.

I. INTRODUCTION

Falls in the elderly are a common and burdensome problem. About one-third of the population aged 65 or more fall at least once a year and about 20% of the falls result in injuries that necessitate medical intervention. Many preventive interventions have been proposed [1]. Screening tools can help identifying subjects at increased risk, that could potentially benefit the most from these interventions. They are generally developed to be used in specific settings, namely the community, hospitals, and nursing and residential care facilities.

The EU project Farseeing[1] aims to promote better prediction, identification and prevention of falls with a focus on ICT devices, tools and software. In this work we introduce the *Farseeing Fall Risk Assessment Tool* (*FRAT-up*), a software component that evaluates the fall risk of a subject within a year time span. FRAT-up is based on the reasonable assumption that the total fall risk of a subject is given by the contributions of the risk factors related to falls, and to which the specific subject is exposed. Hence, our tool takes as input the characteristics of a subject in terms of the list of known risk factors affecting her/him. As output FRAT-up provides an estimation of the fall risk.

A vast scientific literature is available on risk factors. FRAT-up exploits such scientific knowledge to compute the fall risk: our approach directly uses the *odds ratio* computed within epidemiological statistical papers. From the technical viewpoint, FRAT-up is based on one (or more) rule for each risk factor. Starting from a light ontology of risk factors, we generate a set of probabilistic rules which are expressed as a Logic Program with Annotated Disjunction (LPAD, [2]). The subject's health profile is added to the LPAD program, that computes the overall fall probability.

We tested FRAT-up using the data collected within the InCHIANTI study [3], used in a prospective manner. To ease the use of the tool, we also implemented a web-based graphical user interface allowing a user to enter information about exposure to risk factors and to get the estimated risk factor.

II. RISK FACTORS, ESTIMATORS AND UNKNOWN DATA

Risk factors are variables correlated with an increased risk. In order to identify the presence of a risk factor in a subject it is necessary to start from an assessment method, and more than one method is possible. We call the variables produced by these estimation methods estimators. Usually there is no standard way to relate estimator values to risk factor ones, so one of the possible practices needs to be chosen. More than one estimator may contribute to a single risk factor.

We assume that a typical user has access to a set of estimators, coming from direct observation, self report from the subject or clinical tests, and inputs them to the tool in order to obtain a fall risk assessment. It is possible than not all supported estimators are known to the user. To provide nonetheless a reasonable answer, FRAT-up accepts a special, "unknown", value to fill an estimator variable. When a risk factor cannot be derived because one or more required estimators are unknown, it assumes a distribution on the risk factor according to its prevalence.

Some risk factors are either present or not present in a given subject. We call these "Boolean Risk Factors". To take also uncertainty into account we use three-valued logic where the variable may be either "true", "false" or "unknown". Other

[1]http://farseeingresearch.eu/

1063-7125/14 $31.00 © 2014 IEEE
DOI 10.1109/CBMS.2014.35

risk factors are commonly seen as scalar. We call these "Scalar Risk Factors", the variables may be assigned a number or the special value "unknown". Some risk factors act synergistically in contributing to the fall probability [4]. We introduce a class of factors called "Synergistic Risk Factors"; a Synergistic Risk Factor derives from a set of other risk factors and has a value equal to the number of these factors that are present in the subject. Estimators may be represented with three-valued logic and scalars with "unknown" too. Boolean Risk Factors may be function of scalar estimators.

The LPAD program assessing risk is automatically produced starting from risk factors and estimator names and types, risk factor odds ratios and prevalences, and functions from estimators to risk factors. This input source is made transparent accessing through a wrapper; an ideal one may be a quantitative risk ontology in electronic format since, as stated in [5], ontologies are important for epidemiology. Our prototypical ontology has been designed in the OWL language, containing risk factors classification along various dimensions, and scientific references for every included information. The main dimensions are "kind" ("environmental", "behavioral", and "endogenous"), "setting" ("supportive housing", "long term care", "acute care", and "community"), and "reversibility" ("irreversible", "surely reversible", and "subject specific").

III. Fall Risk Evaluation based on Rules

Statistical findings can be summarized as sentences of the type "subjects affected by a risk factor X_i show an odds ratio x_i of the fall probability w.r.t. not affected subjects": the knowledge is in form of rules that specify also a probability/odds about the chance of fall. A rule-based approach then appears to be the closest solution to the knowledge provided in the literature: for each risk factor one or more rules would be used to determine the contribution of that risk factor to the overall fall risk. The information provided by the knowledge base is of two types: (a) that a specific factor is a risk factor for falls, and (b) that the risk factor contributes to the risk in a way that can be estimated through the odds ratio values. Beside being a natural choice, the use of rules also allows other advantages. For these reasons and the following ones we oriented towards the LPAD language, a rule based system supporting probabilistic reasoning.

i) As new risk factors can be added by extending the LPAD with new clauses for each factor, similarly, risk factors can be ignored by the tool by simply deleting the corresponding clauses.
ii) Adjusting the assessment tool for a specific population or setting can be done by modifying the LPAD parameters. Adjustments can be the consequence of selecting certain published works rather than others, or simply because the tool needs to be adapted to a specific population/setting etc.
iii) The fall probability is computed on the basis of the known facts about a subject, and there is not a minimum number of mandatory "questions" that must be answered to provide such a probability.
iv) Optimized inference mechanisms have been published [6], [7], and implementations are freely available[2].
v) Learning algorithms for both the *parameters* [8] and the *structure* [9] of an LPAD are available. E.g., a specific long-term care facility could be interested in evaluating the fall risk on the basis of its own historical records about falls: it would suffice to learn the LPADs parameters from the available data.

[2]http://www.ing.unife.it/software/cplint/

A. Introduction to LPAD

FRAT-up uses LPAD to assess the risk starting from the estimators. LPADs are logic programs [10] where the head of a clause is a disjunction of annotated atoms. The clauses are of the form:

$$h_1 : p_1 \vee \ldots \vee h_n : p_n \leftarrow b_1 \wedge \ldots \wedge b_m \wedge \neg c_1 \wedge \ldots \wedge \neg c_l$$

where $h_1, \ldots, h_n$ are the atoms ,and $p_1, \ldots, p_n$ are the probabilities related to each disjunct, b_i and c_i are predicates. The intended meaning of such a clause is that each atom h_i has probability p_i if the body is true, and the atom does not appear in the head of any other clause; when it does, the intended semantics is the distribution semantics as in [11]. The probabilities $p_1, \ldots, p_n$ should sum up to 1, with an implicit "null" atom when the explicit probabilities sum up to less than 1. From the operational viewpoint an LPAD program is treated as follows: for each clause containing a disjunct, different logic program *instances* are generated, each containing the clauses with exactly one disjunct.

We adopt the syntax of the `cplint`[2] implementation. Note that the disjunction in the head of the clauses is indicated with the symbol "`;`", while the conjunction is indicated as usual in Prolog with "`,`" .

B. Structure of our Rules

In FRAT-up we model the risk as a probability of observing the adverse event in a given time span, given by the composition of a base probability (the probability of the adverse event in absence of any risk factor) and contributions from each of the risk factors that are present in the subject. We will follow an example with just one risk factor:

```
fall(S) : 0.085.
fall(S) : 0.017 :-
     factorB(S, 'physical activity limitation', t).
```

where `S` is the subject, and `factorB` stands for "boolean factor", a variable accepting only `t` (true) or `f` (false). The rule captures the following meaning: the probability to fall of a subject described by the profile S due to the relation to the risk factor "physical activity limitation" is 1,7%. This is different than just stating that the probability to fall is 1,7%, because the system takes into account all the rules that may lead to a fall, and computes the probability of having at least one rule causing a fall assuming the categorical distributions in the head of the rules as independent. Since each factor may be "unknown", we need rules to handle that case using the

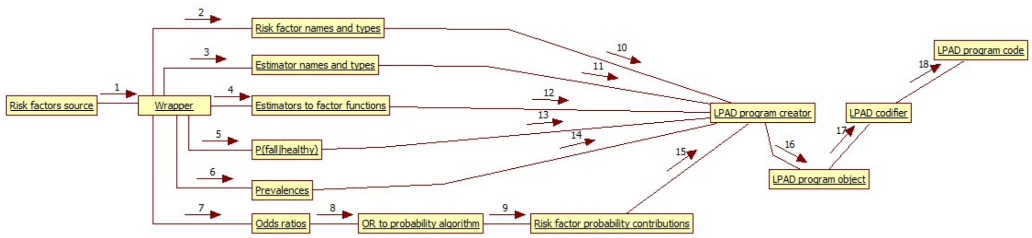


Fig. 1. UML Collaboration Diagram of LPAD production steps.

prevalence.

```
factorB(S, 'physical activity limitation', t) :-
     factor3(S, 'physical activity limitation', t).
factorB(S, 'physical activity limitation', f) :-
     factor3(S, 'physical activity limitation', f).
factorB(S, 'physical activity limitation', t) : 0.56 ;
factorB(S, 'physical activity limitation', f) : 0.44 :-
     factor3(S, 'physical activity limitation', u).
```

where `factor3` means "risk factor in three-valued logic" and `u` means "unknown". In this case we have a three-valued logic risk factor that is function of the scalar estimator "Physical activity level". The following rules, where `estimatorSU` means "estimator accepting scalars and unknown", handle the discretization.

```
factor3(S, 'physical activity limitation', t) :-
     member(estimatorSU('Physical activity level', N), S),
     number(N), N =< 2.
factor3(S, 'physical activity limitation', f) :-
     member(estimatorSU('Physical activity level', N), S),
     number(N), N > 2.
factor3(S, 'physical activity limitation', u) :-
     member(estimatorSU('Physical activity level', u), S).
```

C. Automatic Generation of the Rules

The rules are automatically generated from a Java program. The process of producing and writing the LPAD program goes through a sequence of steps as depicted in the Collaboration Diagram of Figure 1. First a data source, like a quantitative fall ontology, is read through a wrapper, and data are extracted; they include the names and types of the risk factors and the estimators, the prevalences of the risk factors, the probability to fall in absence of risk factors, the odds ratios associated to the risk factors, and for each risk factor a function that specifies how it is computed starting from one or more estimators.

Applying a mathematical model, the probability contribution associated to each risk factor is computed starting from the probability to fall in absence of risk factors and the odds ratio of the risk factor. This is required because the LPAD language has a native support for probabilities, while odds ratios may be handled only indirectly.

In the next step an `LPAD program creator` algorithm builds an `LPAD program object`, that is a representation in a composite Java object of an LPAD program, a tree that mirrors the syntactical tree of a well-formed LPAD program. Finally in the last step an LPAD codifier writes the actual LPAD program according to customizable formatting rules.

IV. FRAT-up Architecture and Web Interface

The health professional interacts directly with a web-based Graphic User Interface, by filling a form regarding subject exposure to risk factors. Rather than specifying the risk factor, the User is requested to indicate a number of estimators. Current FRAT-up version (v. 3.3) supports 15 boolean estimators and 13 scalar estimators, linked to 26 different risk factors. Accepted values are `yes`, `no`, and `Use prevalence` for boolean estimators, while for scalar estimators an input field is provided, together with the `Use prevalence` option. The option meaning is related to the use of the distribution over the population of the presence of the risk factor.

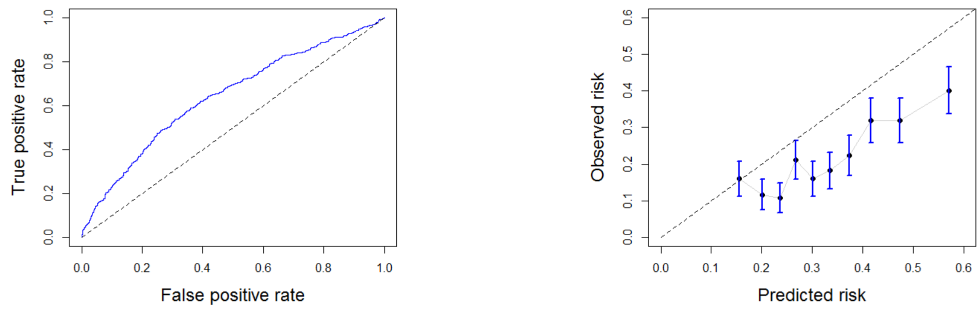
The application is built as a Java Servlet. On the client side a Javascript program intercepts the inputting events: every time a data is inserted, an Ajax query is executed against the server. The server invokes the Prolog engine[3], computing the score. The whole process requires few milliseconds on a desktop machine with moderate computation power.

V. Validation

The validity of the tool was tested on the database of the InCHIANTI study [3]. The study consists of 4 waves, covering globally a 9-year follow-up. At each wave, subjects were asked about the occurrence of any fall in the previous 12 months. In addition, information were collected on fall risk factors considered in our tool. We extracted a total of 2319 observations from 977 subjects aged 65 or more. At each wave, the risk factors of each subject were used to calculate their risk of falling at the subsequent wave.
Figure 2a shows the Receiver Operating Characteristic (ROC) curve. The area under the ROC curve (AUC), expressing the discriminatory ability of the tool, is 0.640 (95% C.I. 0.612-0.667). Even though this value of AUC signifies modest discrimination, it is superior or comparable with what found for other tools that received proper validation. In two validation studies the Tinetti Balance Scale showed an AUC between 0.62 and 0.64 [12], [13] while the Functional Reach an AUC of 0.509 [13]; a recent systematic review and meta-analysis estimated for the Timed Up and Go Test (TUG) an AUC of 0.57 (95% C.I. 0.54-0.59) [14]. Some other studies reported better results but without any kind of validation (e.g. [15]).

Figure 2b shows the calibration plot. Subjects have been grouped in deciles according to their estimated fall risk. For

[3]An optimized version of YAP (http://www.dcc.fc.up.pt/∼vsc/Yap/), based on the inference mechanisms in [6], [7], freely available at https://sites.google.com/a/unife.it/ml/cplint.



(a) Receiver Operating Characteristic curve

(b) Calibration plot.

Fig. 2. Validation graphs

each decile, the observed risk and its 95% C.I. are shown as a function of the decile-averaged, estimated fall risk. The Hosmer-Lemeshow test [16] produces a very low p-value indicating statistical significance of miscalibration, due to risk over-estimation consistent over the risk strata.

VI. Conclusion

An innovative web-based methodology for fall risk estimation in the elderly was presented, that combines scientifically validated sources of fall risks already present in the literature and transforms them in a single risk factor value by using statistical methods and logic programming. There is no use of data sets for training purposes: all the model parameters are generated starting from widely available scientific references. The methodology was validated on 977 subjects obtaining satisfactory accuracies (comparable to other traditional tools routinely used in the clinics like the Tinetti Balance Scale or the Timed Up and Go Test [13]).

Future work will focus on improving the presented methodology to obtain higher performance in identifying a persons fall risk. Moreover, additions to this method on supporting other settings and on supporting more risk factors will be developed. The versatility of the presented solution will allow to combine clinical information with other sources of data, like wearable sensors recording physical activity.

Acknowledgment

This work has been partially supported by the EU Seventh Framework Programme under grant agreement FARSEEING n. 288940, by Sovvenzione Globale Spinner 2013, and by Health Sciences and Technologies - Interdepartmental Center for Industrial Research (HST-ICIR) - University of Bologna. Special thanks to Fabrizio Riguzzi for the precious support.

References

[1] L. D. Gillespie, M. C. Robertson, W. J. Gillespie, C. Sherrington, S. Gates, L. M. Clemson, and S. E. Lamb, "Interventions for preventing falls in older people living in the community," *Cochrane database of systematic reviews*, no. 9, 2012.

[2] J. Vennekens, S. Verbaeten, and M. Bruynooghe, "Logic programs with annotated disjunctions," in *ICLP*, ser. LNCS, B. Demoen and V. Lifschitz, Eds., vol. 3132. Springer, 2004, pp. 431–445.

[3] L. Ferrucci, S. Bandinelli, E. Benvenuti, A. Di Iorio, C. Macchi, T. B. Harris, and J. M. Guralnik, "Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study," *Journal of the American Geriatrics Society*, vol. 48, no. 12, pp. 1618–25, Dec. 2000.

[4] J. E. Lee and D. S. Stokic, "Risk factors for falls during inpatient rehabilitation." *American journal of physical medicine & rehabilitation / Association of Academic Physiatrists*, vol. 87, no. 5, pp. 341–50; quiz 351, 422, May 2008.

[5] J. a. D. Ferreira, D. Paolotti, F. M. Couto, and M. J. Silva, "On the usefulness of ontologies in epidemiology research and practice." *J. of epidem. and community health*, vol. 67, no. 5, pp. 385–8, May 2013.

[6] F. Riguzzi, "Extended semantics and inference for the Independent Choice Logic," *Logic Journal of the IGPL*, vol. 17, no. 6, pp. 589–629, 2009.

[7] F. Riguzzi and T. Swift, "Well-definedness and efficient inference for probabilistic logic programming under the distribution semantics," *Theory and Practice of Logic Programming*, vol. 13, no. Special Issue 02, pp. 279–302, March 2013.

[8] E. Bellodi and F. Riguzzi, "Expectation Maximization over binary decision diagrams for probabilistic logic programs," *Intelligent Data Analysis*, vol. 17, no. 2, pp. 343–363, 2013.

[9] ——, "Learning the structure of probabilistic logic programs," in *ILP 21st Intl. Conf., 2011. Revised Papers*, ser. LNCS, vol. 7207. Heidelberg, Germany: Springer, 2012, pp. 61–75.

[10] J. W. Lloyd, *Foundations of Logic Programming, 2nd Edition*. Springer Verlag, 1987.

[11] T. Sato, "A statistical learning method for logic programs with distribution semantics," in *ICLP*, 1995, pp. 715–729.

[12] M. Raîche, R. Hébert, F. Prince, and H. Corriveau, "Screening older adults at risk of falling with the Tinetti balance scale," *Lancet*, vol. 356, no. 9234, pp. 1001–2, Sep. 2000.

[13] M.-R. Lin, H.-F. Hwang, M.-H. Hu, H.-D. I. Wu, Y.-W. Wang, and F.-C. Huang, "Psychometric comparisons of the timed up and go, one-leg stand, functional reach, and Tinetti balance measures in community-dwelling older people." *J. of the Am. Geriatrics Society*, vol. 52, no. 8, pp. 1343–8, Aug. 2004.

[14] E. Barry, R. Galvin, C. Keogh, F. Horgan, and T. Fahey, "Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta- analysis." *BMC geriatrics*, vol. 14, no. 1, p. 14, Jan. 2014.

[15] P. Stalenhoef, J. Diederiks, J. Knottnerus, A. Kester, and H. Crebolder, "A risk model for the prediction of recurrent falls in community-dwelling elderly," *Journal of Clinical Epidemiology*, vol. 55, no. 11, pp. 1088–1094, Nov. 2002.

[16] D. W. Hosmer and S. Lemeshow, *Applied logistic regression*, 2nd ed. New York: John Wiley & Sons, 2000.