Learning Temporal Regression Models and Voronoi Tessellation for Job Offers Recommendation.

Sidahmed Benabderrahmane (a), Nedra Mellouli (a), Myriam Lamolle (a), Jean Baptiste Gabriel (b).
(a): Paris 8 University, IUT Montreuil, 140 Rue de la Nouvelle France, 93100 France.
(b): Multiposting, 3 rue Moncey, 75009 Paris - France.
s.benabderrahmane@iut.univ-paris8.fr Tel: 0033(01)48703700.

***Abstract*—Nowadays, the best ways to attract job candidates is through dedicated web-based portals, and therefore match their related data automatically using optimized algorithms. In this perspective, with the goal of sharing, at best, the job offers, many online job boards have been created, the choice of which can be sometimes very hard for the recruiters that aim at attracting the best possible candidates in the shortest amount of time. Based on these considerations, in this paper, we propose a novel jobboard recommendation system that aims at estimating the best potential job-boards for a given text job offer. Our efficient predictive model for job boards recommendation, is based on a hybrid representation, that combines semantic knowledge and time series forecasting. The semantic classification of job boards requires a textual analysis using domain knowledge. The time series analysis module is to predict the best job board for a given offer. The proposed system has been evaluated on real data, and preliminary results seem very promising.**

Index Terms—**Recommendation; Time series; Clustering; Forecasting; Data Mining; Big Data.**

I. Introduction

Since the last two decades, the use of Internet for recruitment purposes has grown considerably. This recruitment process, also known as ”*e-recruitment*” is based on the use of information technology and communications. In this context, expansion of the Web has led to an increase in the number of job diffusion web sites (also called *job board*) and consequently the number of candidates that can be contacted through these intermediary tools. However, despite the wide dissemination of existing platforms of e-recruitment, the main concern of recruiters rest of ”finding” the best profiles (i.e., the most talented potential candidates) for a given position. To better target potential candidates, some problems are to be solved such as clarity of the offer and its relevance to potential candidate profile, or adequacy of the job board in relation to the core business of the offer itself even. The search for the best candidates for a given offer returns among others target the most appropriate job board, and after that the most relevant profiles among the mass of available profiles.

Current recommendation systems process only a part of the recruitment process, concentrating on matching offers with CVs. However, the selection of the most appropriate job board regarding an offer is also very important for the optimization of this fully digital recruitment process. For this reason, various questions arise concerning the criteria for relevance of a job board over a given offer. For example, is a job board considered relevant if the number of offers are increasing? Or, if the number of visits and / or the number of clicks to view the offers by potential candidates tend to grow compared to those observed in the past?

Our main goal is to provide a tool to help recruiters to i) select the most relevant job board for a given position, ii) diffuse more effectively job ads, that is to say at the right place at the right time, iii) provide tools to connect candidates and offers automatically.

To meet the above objectives, we are also faced with problems related to the specificity of the data to be processed. We dispose from our industrial partner, a history of job advertisements on web sites, and the quantity of their visits (clicks), that are stored in a big database. The recorded data also concern the number of candidates obtained through various job boards and social networks. In this context, we propose a recommendation system of *job boards* based on a hybrid model combining modular semantic classification approaches, and time series forecasting [4]. The semantic classification of job boards and job ads requires a textual analysis of the content on the basis of business description that is given by a public french organization (ROME code [1]). The time series analysis module, aims to predict the best *job board* for a given offer, combined with textual analysis module.

The rest of this article is organized as follows: a state of the art will be discussed in section 2. The proposed model will be presented in section 3. Finally, in section 4 we discuss preliminary results and conclude the paper in section 5.

II. State of the Art

Nowadays, few automatic recommendation systems of job offers to particular users exist. These systems are generally classified into three main categories namely textual recommendation systems [1], recommendation systems based on collaborative filtering, and hybrid recommendation systems [5].

Textual-based recommendation systems analyze the content of job descriptions as well as information provided by users to identify the semantic content. To that aim, two types of semantic analysis approaches exist: approaches based on ontologies [8] and text mining approaches [10]. Whatever the approach used in the purely textual recommendation systems, weaknesses may occur. Indeed, the existing approaches require manual annotation by the recruiter and the

[1] www.pole-emploi.fr/candidat/le-code-rome-et-les-fiches-metiers-@/suarticle.jspz ? Id = 15734

candidate to describe both job offers and CVs. Therefore, the volume of processed data is quite large and require the use of highly optimized algorithms. Collaborative filtering systems are based on the analysis of the opinions of a group of users. Their opinions are considered similar to that of an active identified user. These recommendation systems can target CVs only from items related information (such as the title). The use of items certainly reduces the mass of processed data but with a loss of precision. As for hybrid systems, they combine the two previously mentioned categories.

In parallel of the semantic approaches, more formal approaches based on vectors and probabilistic models have been proposed [7] for profiling applications according to a specific offer. Although these approaches seem to be transposed to our problem since they concern the profiling of a job boards for a given job, they are unusable in our context because they only deal with text data. Finally, another approach is essentially based on the predictive linear models was proposed in [10]. This work considers the problem of recommendation as a prediction of the performance of offers on a job board regardless of the behavior of this job board in the past. The linear model was proposed in that work assumes that the model parameters are independent for simplicity. But the real data do not always check their working hypothesis since the dependencies are spatial order (depending on the job board) and also temporal (dependent of the past).

Most of these recommendation systems could be improved if the temporal dimension of information related to the job board was more taken into account in the models. We wish in this work, to consider the information relative to the temporal aspect of the dissemination of offers in the different job boards, to create a predictive model based on the values observed in the past. We propose a representation based on time series, for highlighting the trend and seasonality in the recruitment data. With these informations, decision-making and recommendation of relevant job boards for job opportunities, could contribute to long-term automating the assignment of these bidding job offers to one or more the most appropriate job board. It is this approach that we favor in our study and that we will detail starting with the presentation of our model in the following sections.

III. Description of the temporal recommendation system architecture

As reported in the precedent section, our interest in this study is to characterize the best job boards for a given job offer to automatize the the process of diffusion of vacancies (postings). Job boards of vacancies available on the web are multiple. Some may be specialized to broadcast certain categories of business or certain types of deals, for example, internships, PhD, or fixed-term contracts. Defining the intrinsic characteristics common to all job boards is a very important step for the analysis of their behavior in order to compare and evaluate them. These characteristics are related to the properties of vacancies advertised on these job boards. In particular, in this study we consider initially the business description of the offer, and number of its clicks accrued per job board over a given period. We first present the classification used for the characterization of offers based on their type of business. This classification is used later in the formalization of data relating to job boards.

A. Data formalization

For the formalization of offers, we will use two different types of textual content: jobs and job categories (business classes). An *offer* can be defined as a structured text document used to formalize an offer of an employer. It is divided (or can be divided theoretically) in various fields such as title, business description, skills, education level, etc., organized according to the publisher and / or some high standards level. Given this definition, we formalize the contents of a job offer j as a set of vectors, each representing a text field, as follows:

$$o_j = (\vec{v_{j,1}}, \vec{v_{j,2}}, ..., \vec{v_{j,k}}) \tag{1}$$

where k is the number of text fields in the offer j and v is a vector of weighted keywords representing the frequencies of words of each field (according to the method described by [9]). Specifically, considering a generic text field i (with i ¡k) of job j, we formalize its contents with a vector $v_{j,i}$

$$\vec{v_{j,i}} = \{w_{j,i,1}, w_{j,i,2}, ..., w_{j,i,n}\} \tag{2}$$

where n is the size of vocabulary of the field and $w_{j,i,k}$ is the inverse frequency (TF-IDF) of k terms in the i field in the j offer. Similarly, a job category (also known generically category in this article) can be defined as a textual description of a specific category of occupations. Its definition is generally provided by a domain expert (or any authority) and can be used effectively for classification and indexing of job offers. In our case, we used the French public ROME code categories. Thus, according to the same principle, we formalize the contents of a job class c as

$$c_i = (\vec{v_{i,1}}, \vec{v_{i,2}}, ..., \vec{v_{i,l}}) \tag{3}$$

where l is the number of terms describing the job class c, v is the key words vector representing the frequency of terms. Then we used a vector-based distance and an SVM classifier to annotate each job offer to a semantic business category.

Each day, a set of offers is deposited on one or more job board on a given date. An offer made on a job board has a finite life cycle. In this period, the number of clicks associated with each bid is incremented. Therefore, the daily number of clicks associated with an offer and job boards is available. This number is shown on other time scales: weekly, monthly, and annual midyear. We denote by T the period or the time scale associated with the number of clicks considered. To formulate such data, in particular the number of clicks, we consider a job board, noted JB, as a set of offers on a given period T:

$$JB_T = \cup_j o_j \text{ for } j = 1, ..., p \tag{4}$$

Using the vectorial classification of offers described previously, each JB become a class of offers on a period T:

$$JB_T = \cup_k c_k \text{ for } k = 1, ..., m \tag{5}$$

and a class c_j as the union of offers contained in :

$$c_j = \cup_i o_i \text{ for } i = 1, ..., n \tag{6}$$

For each class c_j, we introduce a ratio X^{c_j} calculated as the total number of relative clicks of offers in this class in a period T :

$$X^{c_j} = \frac{nb.clic}{|c_j|} \tag{7}$$

in particular, when j=1 in a JB then $X^{JB} = \frac{nb.clic}{|JB|}$.
In the following of this paper, we consider T a discrete interval $[1, N]$. Having a series of observations $X_1^{c_j}, X_2^{c_j}, .., X_N^{c_j}$ on a fixed period T, we propose the definition of previsions on a date N with a time series of observations, to estimate $\hat{X}^{c_j}(N, h)$ on future dates within a given horizon h. The objectives of temporal analysis in our study are multiple. Firstly, it concerns the prevision of future realization of a random variable X^{c_j} using the previously observed values $X_1^{c_j}, X_2^{c_j}, X_N^{c_j}$ for each class c_j and for each JB. Secondly, we are interested by estimating the trend of time series. In addition, we are interested in analyzing the variations; for example, one may ask whether an observed change in the number of visitors to a JB is the result of a seasonal fluctuation or is a reflection of a trend. Finally, evaluation of the impact of an event on a variable will measure the JB sensitivity to potential disturbances and noise. For these reasons, we will use univariate time series only, and we notice the variable X^{c_j} by x_t observed at time t. We chose to use a regression model applied to the time series using the information on the number of clicks on the offers. The figure 4 give an example of a time series of job board, where values x_t are the clicks ratios between 2008 till 2014 (1716 days).

1) Statistical Data Analysis: To analyze a time series (x_1, x_2, ..., x_n), it is useful to have statistical indexes to summarize the series. As an indicator of central tendency, we calculated firstly the average as follows: $\bar{x_n} = \frac{1}{n}\sum_{j=1}^{n} x_j$. We also calculated the index of empirical variance (or standard deviation), to rise up comprehensive information on the dispersion of temporal observations with respect to their central tendency. The variance of a set of values in a time series is defined by: $\hat{\sigma}(0) = \frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x_n})^2$. Later we calculated dependencies between two successive observations by the empirical auto covariance (of order h): $\hat{\sigma}_n(h) = \frac{1}{n-h}\sum_{t=1}^{n-h}(x_t - \bar{x_n})(x_{t+h} - \bar{x_n})$. After that we calculate for each series the empirical autocorrelation that is given by the ratio of auto-covariance and the empirical variance: $\hat{\rho}_n(h) = \frac{\hat{\sigma}_n(h)}{\hat{\sigma}_n(1)}$. These autocorrelations characterize dependencies between series values (x_1, x_2, ..., x_{n-1}) et (x_1, x_2, ..., x_n). This value can provide an overall idea about the implicit regression in the data set. Indeed, linear regression can be observed if the value of $\hat{\rho}_n(h)$ is close to +1 or -1. More auto-correlation tends to 1 in absolute value, the higher the series has a trend. The slope of the linear regression line follows the sign of $\hat{\rho}_n(h)$, while the cloud of the series values is more rounded when it is close to zero.

2) Trend and seasonality in time series: As part of our study, we seek to identify seasonality and trends in job boards time series. Seasonality is an important factor that indicates the repetition frequency of a phenomenon in a periodic manner. The trend is an attendance indicator of a stationary, increasing or decreasing job board. Therefore, it is useful to decompose a time series in order to separate the content of the trend, irregular components and the seasonal component, if it is present. In the case where the series is non-seasonal, that means it is composed of a trend and an irregular residual component. Thus, the decomposition generally requires the separation of the three (or two) sources and estimation of the trend. There are several decomposition models present in the state of the art [11]. Among these models, we opted for the special Holt-Winters (HW) probabilistic model [3], [6]. This method has the advantage of being simple to implement and can take into account both the trend and seasonality. Given the probabilistic aspect of this model, it is robust to noise and provides a single model to extract the three components. Two scenarios of decomposition by HW model are possible: additive and multiplicative. As we have no information *a priori*, we consider the two scenarios using a sliding window throughout the series. Figure 7 shows an example of decomposition. In the first part, we have the observed series of a job board (top), followed by trend (2nd), the seasonal component (3rd) and finally the noise fluctuations. This is so important because if we have a job board where the trend is decreasing hence the probability of recommending offers in it will be weak.

B. The Global Architecture of the System

The proposed recommendation system of job offers in the job boards is described in Figure 1. It is based on two main phases namely learning predictive model step and the recommendation step. During the first phase (left part of the figure), we used the previously described vectorial model and SVM classifier to annotate each job offer with the French job offers semantic vocabulary (ROME code). After that, we performed a clustering process, to regroup similar job offers on the basis of their shared semantic annotations. The clustering step is based on GLA (Generalized Lloyd Algorithm) that generates a Voronoi tessellation on the input data, by separating the job offers into convex regions (see Step 1 in Figure 1). These regions represent the classes $\zeta^{posting}$ (or clusters) of job offers (postings) obtained with GLA. Since we have more than 1 Million offers, we implemented a *Hadoop MapReduce* version of the clustering algorithm on R using *rmr2* and *rhdfs* packages, on *Hadoop 2.5 CloudEra* version. Results of the produced Voronoi tessellation with GLA clustering algorithm are displayed in figure 2.

Similarly on the same DB, we have several thousands of job boards. For each Job board and for each class of postings $\zeta^{posting}$, we construct a time series vector (see step 2 in Figure 1). Hence a time series will represent the temporal behavior of a job board by considering only job offers belonging to a similar semantic class $\zeta^{posting}$. Then, for each time series, a regression model is learned using

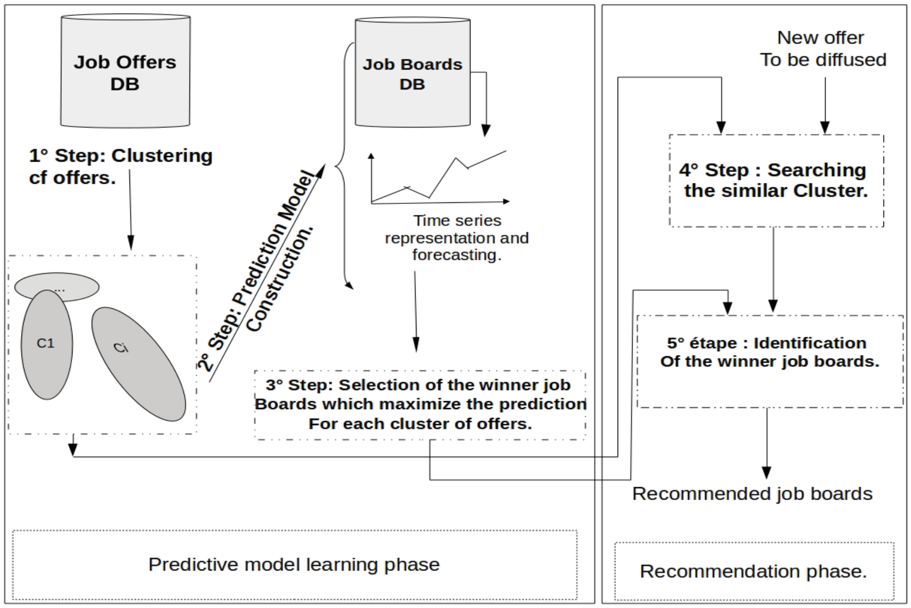


Fig. 1. Global architecture of the recommendation system.

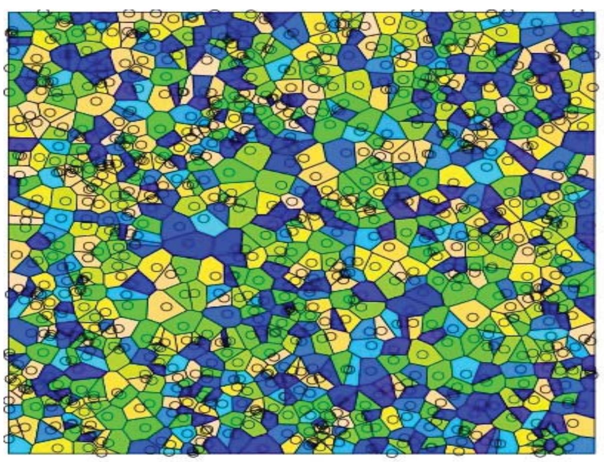


Fig. 2. Voironoi tessellation with GLA clustering algorithm.

Holt Winters probabilistic model, and future values of clicks ratios are predicted within an horizon h. After that, the job board(s) maximizing the different predicted ratio values are considered the most appropriate for the dissemination of the offers belonging to the considered class (see Step 3 in Figure). A hash table is created, containing key / values as class of offers, winner job boards.

In the second phase (the online stage), we seek to identify the job boards adapted to receive the diffusion of a new incoming job offer. The recommendation comes in two stages. Firstly, the system seeks to identify the nearest pre-generated class of offers compared to this new offer in term of its similarity with the centroid of the class (see Step 4 in the Figure). Since we already have an association for each class, the corresponding winner job board, thus we recommend the diffusion of this new offer on this job board(the fifth step on the figure).

C. The prediction algorithm

The Algorithm 1 illustrates the great steps described previously of our recommendation system. As we have shown in the previous section, this algorithm requires a list of jobs classes, regrouped by supervised manner (classification) or not (clustering). For a cluster of offers, the algorithm generates for each job board portal, a representation in time series of ratios.

IV. Experimental Results

A. Description of the dataset

To evaluate the proposed model, we used a big database of job offers and job boards. This complex DB, large and with heterogeneous information has been provided by an industrial partner (Multi Posting) in Sonar Project (http://sonar-project.com/). By considering confidentiality issues we can not present the architecture of the Data Base. However, we can attest that it represents more than a six-year follow-up containing about ten thousand of job boards and millions of job offers.

B. Results discussion

1) Quantitative analysis of the dataset: In order to quantitatively explore our data set, we firstly analyzed the variation of diffusing job offers in the different job boards, at different times. Figures 3.(a) and 3.(b) respectively show the distributions of job vacancies advertised and consulted on channels with different dates. For example, in Figure 3.(a), we can see that between 2008 and early 2011, the amount of vacancies advertised in the job boards was low. By cons, in the interval 2011-2012, offers are constant and distribution deals are around 10000 offers per day (to a maximum of about 15,000 offers). Between 2013 and early 2014, the increase is of the

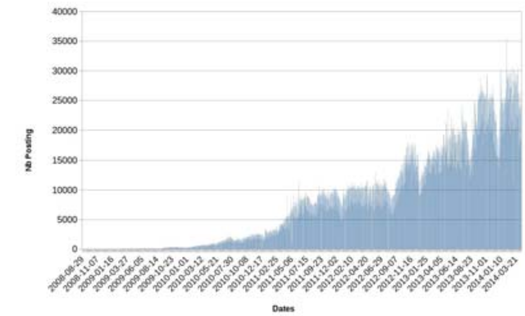
Algorithm 1 Clicks forecasting in each job board.

Require: A cluster C_{off} of similar offers, a list of job boards JB, and an horizon value h.
Ensure: The appropriate job board which maximizes the predicted value of clicks ratio.
Begin
$Maxclic = 0$
By considering all the offers $posting_j$ in C_{off}
for Each job board JB_i in DB **do**
  for each Instant $t \in \Delta_t$ **do**
    Calculate the ratio: $x(t) = \frac{|clics|}{|C_{off}|}$.
  end for
  Construct time series $X(JB_i)$=$\{x(t)|t \in \Delta_t\}$.
  Apply moving average filter on $X(JB_i)$ to reduce noises.
  Learn Holt-Winters model on $X(JB_i)$.
  Calculate $forecast(JB_i)_h$ to estimate future values of clicks in an horizon h.
  if $Maxclic \leq forecast(JB_i)_h$ **then**
    $Maxclic = forecast(JB_i)_h$
  end if
end for
return the winner JB_i having $Maxclic$.
End

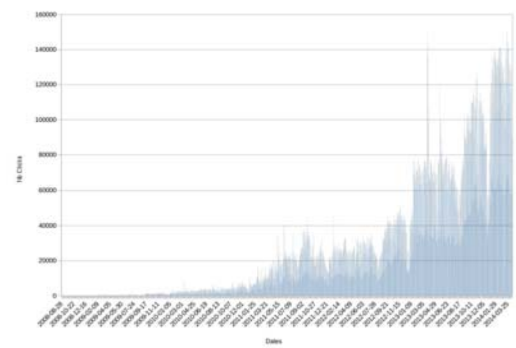
order of 25,000 to 30,000 offers per day with two points of visible change between late 2012 and early 2013. Scattering peaks are visible in the months of January, June or September which could be interpreted by the fact that certain offers are Seasonal. Referring to Figure 3.(b), we can see that there is a correlation between the number of diffusion of offers in Figure ref dist2.(a) and the number of clicks in the second figure 3.(b). The number of clicks is constant for a number of offers between July 2011 and November 2012. We also note that from November 2012 to August 2013, a sudden increase in clicks is notable for a constant number of offers. Finally, from August 2013, we can observe an extreme increase in clicks of job offers (average 120,000 hits per day). Slopes changes corroborate these observations.

2) Quantitative analysis using statistical indicators: Qualitative analysis is to explore, through statistical indicators, the properties of a series representing a job board. Figure 4 illustrates the variation of the ratio as a time series of a job board in our DB. Peak values can be seen in 2008, early 2010 and late 2012. Visually analyze a series seems difficult; for this reason, it is useful to use descriptive statistical tools to extract hidden information. For example, Figures 5 and 6 represent variations of calculating co-variances and correlations between the observed time series on different neighborhoods. We can see that for smoothing sliding windows (*Lag*) ranging from 1 to 10, the auto-correlation indexes are positive and close to 1.

3) Analysis of trends and seasonality: The information generated in the time series can look very noisy. This is due to random fluctuations intrinsic to the measures. To remedy



(a) Job offers distribution.



(b) Clicks on job offers distribution.

Fig. 3. (a) : Distribution of the diffusion of Job offers on different job boards of our DB at different dates. (b) : Distribution of the number of clicks of the offers on job boards at different dates.

this, and to visualize potential trends in the series, we used a moving average filter with different neighborhood sizes. After filtering the input time series with moving averages, it is now possible to build a predictive model. We decided initially to use the method of Holt-Winters [3], [6]. Figure 8 shows the result of Holt-Winters prediction; in Black, the original series and dark, we have a prediction of the number of clicks in a five-day horizon. We can therefore see the execution of our algorithm which, for a set of job board, calculates the prediction with the exponential model, and offers the portal that maximizes the prediction in terms of number of clicks.

4) Evaluating the predictive model: To evaluate the performances of Algorithm 1, we followed a test protocol that is to cut each time series of a job board channel into two parts. The first part of the series is used to create the regressive model with Holt-Winters. Then, we use the second part of the time series to compare the predicted values with the rest of real values. The difference is calculated using the mean square error. In Table IV-B.4 we illustrate the results obtained for a set of 22 job boards only due to space limit. We can observe the number of seasonality obtained with the additive and multiplicative decomposition, the trend and the prediction errors. A total of 11 JBs have an upward

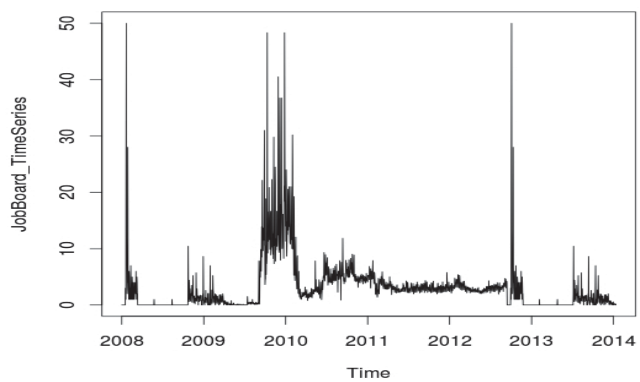


Fig. 4. Time series representation of a job board where clicks ratios values are calculated between 2008 and 2014.

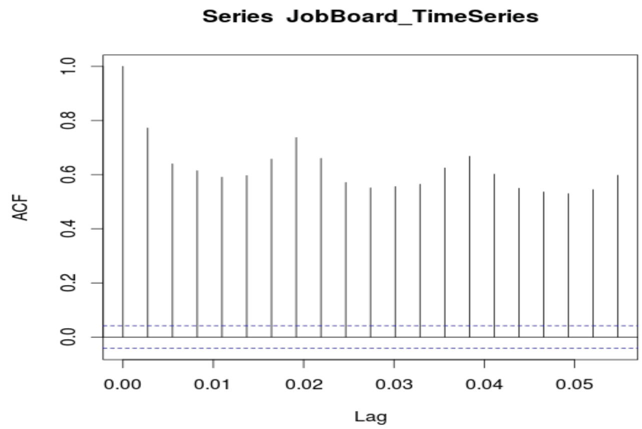


Fig. 6. Correlogram of the same series with different lag values.

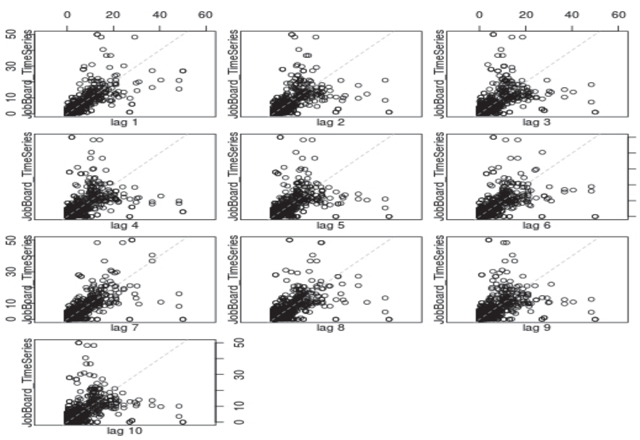


Fig. 5. Auto-correlation Analysis between observed values of the series on different neighborhoods (Lag 1 to 10).

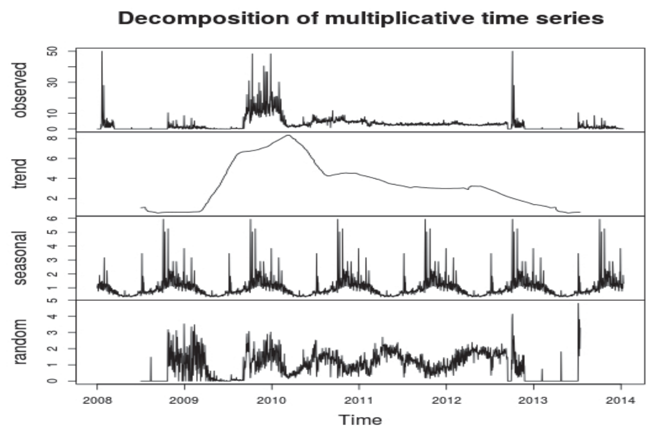


Fig. 7. Example of an additive decomposition of the original time series data with it seasonality, trends and residues.

trend (U), and we can observe that the number of additives and multiplicative seasonality are almost similar with higher values during each year. It means that these JBs not only have increasing clicks values, but also offers high diffusion frequencies in the year by their attractiveness. In addition, the prediction error in these job boards remains on average low, and thus they are highly recommendable to disseminate job offers. We can also see that there are 8 job boards with downward trends (D). These channels have a variable number of the two kind of seasonality models, averaging 5 to 6. For stationary job boards, there are 3 JBs with low seasonality.

V. Conclusion and perspectives

In this paper, we have presented a recommendation system based on semantic knowledge and temporal representation by time series. The objective is to diffuse a job offer to one or more adequate job boards. The system is based on two main stages namely learning the predictive model step and the recommendation one. We have shown the need of taking into account the seasonality for finer predictive studies, particularly in the context of the diffusion of job offers. We have integrated the probabilistic model of Holt-Winters to decompose the time series in order to identify trends, seasonality and possibly the residual noises.

Recommendation phase takes place in two stages: the identification of the most similar class of job offers regarding a new offer; then, the recommendation of job boards that maximize the ratio of the prediction of the clicks.

We presented some results of our experiments that revealed potential interesting job boards for certain class of job offers. The perspectives of this work will concern mainly taking into account other domain knowledge. On the other hand, the success of social networks has also contributed significantly to the evolution of the recruitment market on the Internet. Indeed, personal information posted by users of a social network such as LinkedIn can identify with more or less precisely their profile (academic curriculum, professional background, passion, etc). The integration of this information into the process of recommendation could refine the relevance of the proposed job board.

Processing numerical time series could represent a hard task and has the inconvenient of manipulating complex and

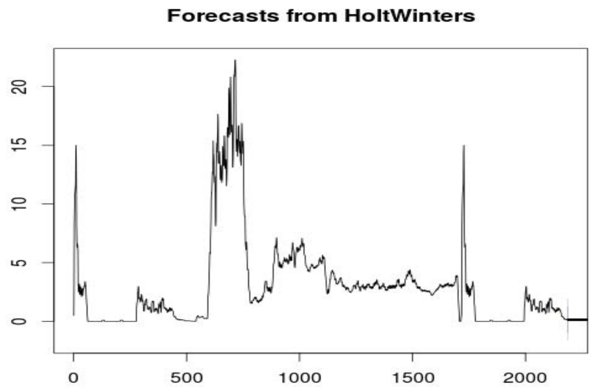


Fig. 8. Time series of the job board represented in Figure 4 with an exponential model generated by the Holt-Winters approach.

possible noisy data. We would like to use symbolic time series representation to avoid such problems and to transform the time series into symbolic sequences. Thereafter it will be possible to use symbolic data mining methods such as motifs discovery or similarity search [2]. We want also, in this context, to use other prediction algorithms on the the symbolic sequences such as Markov Models.

ACKNOWLEDGMENT

This work was supported by French Government and *Ile de France region* under a grant for FUI SONAR Project for automatic recruitment tasks. The authors would like to thank Multi Posting for data sharing.

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.*, 17(6):734–749, June 2005.

[2] Sidahmed Benabderrahmane, Rene Quiniou, and Thomas Guyet. Evaluating distance measures and times series clustering for temporal patterns retrieval. In James Joshi, Elisa Bertino, Bhavani M. Thuraisingham, and Ling Liu, editors, *Proceedings of the 15th IEEE International Conference on Information Reuse and Integration, IRI 2014, Redwood City, CA, USA, August 13-15, 2014*, pages 434–441. IEEE, 2014.

[3] Chris Chatfield and Mohammad Yar. Holt-winters forecasting: Some practical issues. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 37(2):pp. 129–140, 1988.

[4] Tak chung Fu. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1):164 – 181, 2011.

[5] Mamadou Diaby and Emmanuel Viennet. Développement d'une application de recommandation d'offres demploi aux utilisateurs de facebook et linkedin. In *Atelier Fouille de Données Complexes de la 14e Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances (EGC'14)*, Rennes, jan 2014.

[6] Sarah Gelper, Roland Fried, and Christophe Croux. Robust forecasting with exponential and Holt-Winters smoothing. *Journal of Forecasting*, 29(3):285–300, 2010.

[7] Joseph S. Kong, Kyle Teague, and Justin Kessler. The love-hate square counting method for recommender systems. In Gideon Dror, Yehuda Koren, and Markus Weimer, editors, *KDD Cup*, volume 18 of *JMLR Proceedings*, pages 249–261. JMLR.org, 2012.

[8] V. Radevski, Z. Dika, and F. Trichet. Common: A framework for developing knowledge-based systems dedicated to competency-based management. In *Information Technology Interfaces, 2006. 28th International Conference on*, pages 419–424, 2006.

Job board Id	Nb Seasonality (Additive model)	Nb Seasonality (Multiplicative model)	Trend	Error of prediction
1	5	5	D	0.1
4570	9	9	S	0.0022
4962	36	36	S	0.4
4630	15	16	U	0.016
6922	20	24	U	0.26
6605	10	10	U	0.4
919	5	5	U	0.1
4499	10	11	U	0.1
757	5	5	S	0.03
4497	5	5	D	0.16
5638	5	5	D	0.1
139	5	5	U	0.041
434	5	5	U	0.2
446	6	5	D	0.3
9	6	5	U	0.08
2769	6	6	D	0.1
32	6	6	U	0.031
16	6	6	U	0.11
2843	6	6	U	0.044
14	5	5	D	0.07
5	5	5	D	0.16
8	12	11	U	0.05
7	6	6	D	0.7
6	6	6	D	0.02
18	6	7	D	0.22
2843	6	6	S	0.2
447	11	11	U	0.007
25	6	6	U	0.48
11	6	7	U	0.8
12	12	12	D	0.12
13	6	6	S	0.18
15	6	6	D	0.043
17	7	7	S	0.05
47	6	6	S	0.2
125	6	7	U	0.03
136	6	6	S	0.1
166	7	7	S	0.046
169	5	6	S	0.083
174	6	6	U	0.026
263	20	23	S	0.1
272	7	7	S	0.05
505	18	18	S	0.1
422	6	6	S	0.25
594	6	6	S	0.01
667	5	5	S	0.01
675	5	5	S	0.1
680	12	12	S	0.04
30	6	6	U	0
798	6	6	U	0.001
815	6	6	S	0.65
1375	12	12	S	0.02
773	6	6	D	0.2

TABLE I

EXAMPLE OF A SOME JOB BOARDS AVAILABLE IN THE DB. FOR EACH RANDOMLY SELECTED JOB BOARD, WE HAVE THE NUMBER OF SEASONALITY OBTAINED BY ADDITIVE AND MULTIPLICATIVE DECOMPOSITION, AND THE TREND (U: UPWARD, D: DOWNWARD, S: STATIONARY), AND THE PREDICTION ERROR OF THE PREDICTIVE MODEL FOR TIME SERIES.

[9] Ian Ruthven and Mounia Lalmas. A survey on the use of relevance feedback for information access systems. *Knowl. Eng. Rev.*, 18(2):95–145, June 2003.

[10] J. Sgula. *Fouille de donnes textuelles et systmes de recommandation appliqus aux offres d'emploi diffuses sur le web*. PhD thesis, CEDRIC Laboratory, Paris, France, 2012. Directeur: Gilbert Saporta Rapporteur 1: Ludovic Lebart Rapporteur 2: Emmanuel Viennet Examinateur 1: Yves Lechevallier Examinateur 2: Thierry Artieres Examinateur 3: Michel Crucianu Examinateur 4: Yannick Fondeur Examinateur 5: Stphane Le Viet .

[11] Robert A. Yaffee and Monnie McGee. *Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS*. Academic Press, Inc., Orlando, FL, USA, 1st edition, 2000.