User Modeling of Skills and Expertise from Resumes

Hua Li, Daniel J. T. Powell, Mark Clark, Tifani O'Brien and Rafael Alonso
Leidos, Inc., Virginia, U.S.A.

Keywords: User Modeling, Expertise Modeling, Resume, Profile, Skill.

Abstract: Job applicants describe their skills and expertise in resumes and curriculum vitaes (CVs). These biographic data are often evaluated by human resource personnel or a search committee. This manual approach works well when the number of resumes is small. However, in this information age, the volume of available resumes can be overwhelming and there is a need for automatic evaluation of applicant skills and expertise. In this paper, we describe a user modeling algorithm to quantitatively identify skills and expertise from biographic data. This algorithm is called REMA (Resume Expertise Modeling Algorithm). REMA takes data from a resume document as input and produces an expertise model. The expertise model details the expertise topics for which the resume owner has claimed competency. Each topic carries a weight indicating the level of competency. There are two key insights for this algorithm. First, one's expertise is the cumulative result of the various "learning events" in one's career. These learning events are mentioned in various sections of the resume, such as earning a degree, writing a paper, or getting a patent. Second, one's knowledge and skills can become outdated or forgotten over time if not reinforced by learning. We have developed a prototype resume evaluation system based on REMA and are in the process of evaluating REMA's performance.

1 INTRODUCTION

A resume is a written summary of one's education, work experience, skills, credentials, and accomplishments that is often used to apply for jobs. One key function of a resume is to provide information regarding one's skills and expertise. The resume evaluator is required to manually judge the competence of a skill, i.e., the level of mastery of that skill, and/or the expertise in a skill domain, i.e., the level of mastery of all skills in that domain. Expertise modeling is used to automatically produce a quantitative assessment of the competence of various skills and the expertise of relevant skill domains from a resume.

Expertise modeling is valuable for evaluators when faced with a large number of resumes to review. Given a position description, expertise modeling can be used to find suitable candidates by matching a set of skills between a position and a candidate's resume. There are many other applications of expertise modeling, some of which are listed below.

- Expertise finder: given a skill, find experts with that skill, i.e., find resumes with a high expertise level for that skill.
- Virtual expertise group finder: given a set of resumes, cluster them based on skill set similarities.
- Job finder: given a candidate resume, find matching positions within a career market.
- Next skill to learn: suggest skills for career development. Given a user resume, find career-conducive skills (CCS) that the user should learn. The CCS are skills that frequently co-occur with a user's skill set in expert resumes.

Even though there's extensive literature on expertise recommendation, to our knowledge, expertise modeling from resumes is still an open research area. Our model is called Resume Expertise Modeling Algorithm (REMA). It takes a resume, CV or biographical document as input and produces an expertise model. The algorithm focuses on the concept of "expertise mention" in the resume, e.g., earning a Ph.D. in psychology in 2000 or publishing a paper on psychology in 2001. Such mentions imply a certain level of education on a given topic (i.e., psychology) at a particular point in time (i.e., 2000 and 2001, respectively). The greater the number of learning events on a specific topic, the more competent that person is with the topic. This process is referred to as "reinforcement".

Li, H., Powell, D., Clark, M., O'Brien, T. and Alonso, R..
User Modeling of Skills and Expertise from Resumes.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 3: KMIS*, pages 229-233
ISBN: 978-989-758-158-8

Conversely, skills can get rusty over time if a person stops learning. This process is referred to as “forgetting”. The expertise modeling algorithm processes the expertise mentions incrementally, in a chronological order. New expertise topics are inserted into the expertise model and their weights are increased by reinforcement and decreased by forgetting.

2 RELATED WORK

The related work is mainly in the area of expertise recommendation or expert finding which attempts to find the right person with the appropriate skills and knowledge. This is useful for many purposes including problem solving, question answering, and collaboration. A significant amount of research has been generated in the Information Retrieval community (Smirnova1 and Balog, 2011, Balog et al. 2007; Balog et al. 2009; Liebregts and Bogers, 2009). This line of research focuses on content-based algorithms, similar to document search. These algorithms identify experts based on the content of documents that they are associated with (Liebregts and Bogers, 2009; Serdyukov et al. 2007). While these approaches have been very effective in finding the most knowledgeable people on a given topic based on a large collection of documents from an enterprise or the internet, it’s not clear how they can be used to assess the expertise on multiple topics based on a single resume.

3 THE REMA ALGORITHM

The REMA algorithm is shown in Figure 1. An input resume is first parsed into expertise mentions. The mentions are evaluated by Natural Language Processing (NLP) tools to extract expertise topics.

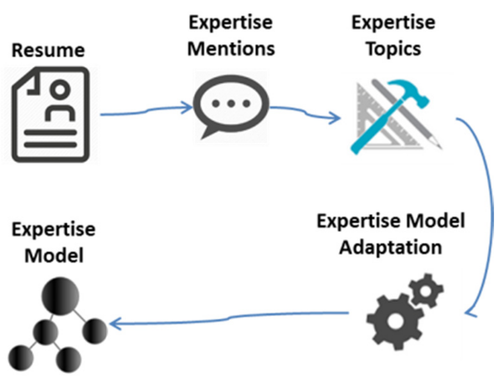


Figure 1: REMA algorithm diagram.

These topics are then processed by REMA’s expertise model adaptation component to generate the expertise model. This algorithm is an extension of our user modelling algorithm RAMA (Reinforcement and Aging Modeling Algorithm) described in detail elsewhere (Li and Alonso, 2014; Li and Alonso, 2012; Alonso et al., 2010).

3.1 Expertise Mentions

Expertise mentions are phrases or statements in the resume that indicate significant learning events. For example, a resume may mention a paper on databases in a certain year in the publication section. When parsing the expertise mentions, the associated resume section and the date of the event are captured because they are important indicators of level of expertise. REMA uses a source relevance parameter to register the fact that expertise mentions in different parts of a resume carry different significance. For example, a mention in a patent and publication section should indicate more expertise than one in an experience and education section. Even within the same section, mentions originated from different sources may carry different significances. For example, within the publication section, mentions of a book or journal paper are more indicative of expertise than those of a conference paper. The date of event mentioned reflects the recency of the learning. In other words, skills or expertise acquired more recently are more up-to-date and less likely to be forgotten. We use regular expression and GATE to extract the date and time information from the expertise mention.

3.2 Expertise Topics

Expertise topics are terms indicating skills or expertise such as database or machine learning. They are extracted from expertise mentions using NLP tools. In particular, Apache Lucene®[1] is used to extract simple terms from text. WordNet®[2] is used to identify noun words. GATE is used to extract noun chunks and named entities. OpenCalais web service is used to extract expertise related tags including "Industry Term", "Technology", and "Programming Language". Relationships between

[1] Apache Lucene is a registered trademark of the Apache Software Foundation within the United States and/or other countries.

[2] WordNet is a registered trademark of the Trustees of Princeton University within the United States and/or other countries.

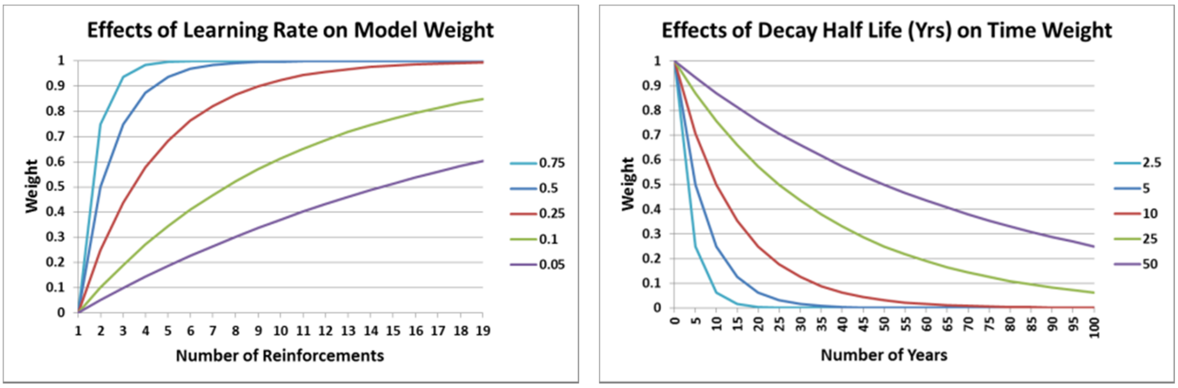


Figure 2: Left: Effects of Reinforcement Factor (Learning Rate) on Weight; Right: Effects of decay half-life on time (Forgetting Factor) weight.

expertise topics are also extracted from expertise mentions. For example, WordNet's semantic relation "hyponym" is useful to map a broader expertise to a narrower one. We can also use Wikipedia's®[3] categories to establish a similar kind of relationship between expertise topics.

3.3 Expertise Model Adaptation

This component has two subcomponents: content adaptation and weight adaptation. The former refers to the addition of new expertise topics in the expertise model. As expertise mentions are being processed in chronological order, unseen expertise topics will be automatically inserted into the expertise model with a default weight. The size of the default weight is equal to the weight change from the reinforcement of one expertise mention (see below).

Weight adaptation refers to the dynamic adjustment of the weights of expertise topics in the model by reinforcement and forgetting mechanisms. Reinforcement increases the topic weight. The more expertise mentions on a given topic, the more life learning events the person has on that topic. As a result, the more competent that person is with the topic. This process is termed "reinforcement". The parameter that controls the rate of learning is called the reinforcement factor. The effects of this factor on the learning rate in simulation experiments are shown in the left panel of Figure 2.

Conversely, naturally forgetting will decrease the topic weight. In general, a person's skills will stagnate over time if the person stops learning. There are at least two contributing factors to this stagnation. The first is our memory decay as described in decay theory[4]. This theory proposes that memory fades and information becomes less available for later retrieval as time passes. The second factor is the technological knowledge depreciation, or decay (Nemet, 2012, Park, Shin, and Park, 2006). The average decay rate is estimated at 13.3%, which corresponds to a half-life of 4.86 years (Park, Shin, and Park, 2006). The effects of five different decay half-life values over time are shown in the right panel of Figure 2.

3.4 Expertise Model

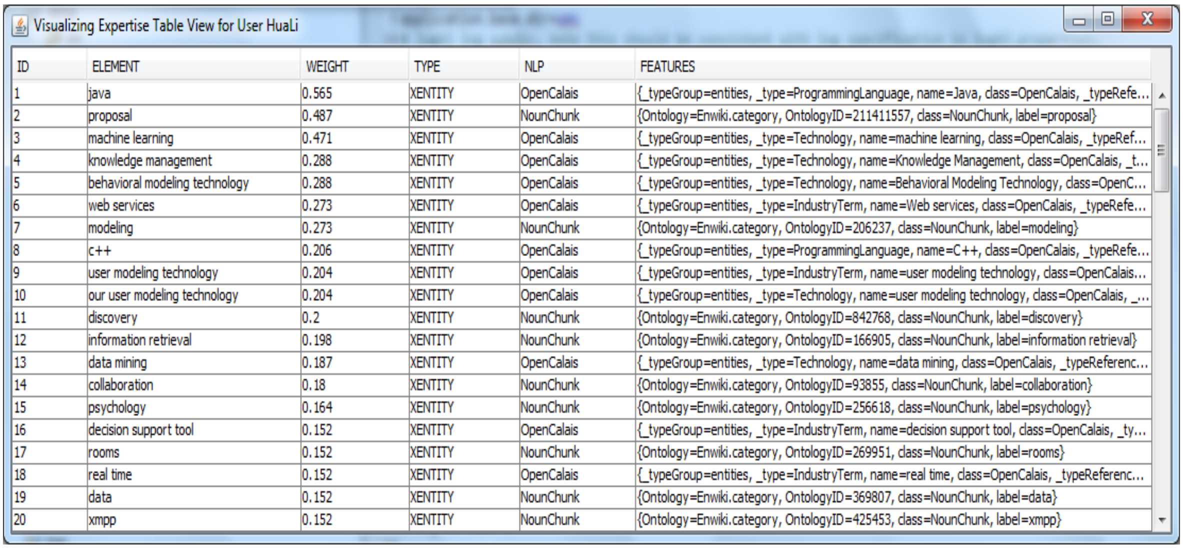
The expertise model consists of weighted expertise topics and their relationships. Each topic carries an authority label that denotes the origin of the expertise term, such as a WordNet, Wikipedia, or a web service like OpenCalais. The collection of expertise topics represents the breadth of the person's skills and expertise. The topic weights indicate the level of expertise and have a range from zero to one. The larger the weight, the more competent the person is with that skill or expertise.

4 REMA RESULTS

We implemented a prototype REMA system in Java. It has a simple Graphical User Interface (GUI) that allows the user to process a directory of resumes and then display the resulting expertise models. This prototype will be used to conduct evaluation experiments in the near future to assess the

[3] Wikipedia is a registered trademark of the Wikimedia Foundation, Inc., in the United States and/or other countries.

[4] http://en.wikipedia.org/wiki/Decay_theory



Visualizing Expertise Table View for User HuaLi

ID	ELEMENT	WEIGHT	TYPE	NLP	FEATURES
1	java	0.565	XENTITY	OpenCalais	{_typeGroup=entities, _type=ProgrammingLanguage, name=Java, class=OpenCalais, _typeRefe...
2	proposal	0.487	XENTITY	NounChunk	{Ontology=Enwiki.category, OntologyID=211411557, class=NounChunk, label=proposal}
3	machine learning	0.471	XENTITY	OpenCalais	{_typeGroup=entities, _type=Technology, name=machine learning, class=OpenCalais, _typeRef...
4	knowledge management	0.288	XENTITY	OpenCalais	{_typeGroup=entities, _type=Technology, name=Knowledge Management, class=OpenCalais, _t...
5	behavioral modeling technology	0.288	XENTITY	OpenCalais	{_typeGroup=entities, _type=Technology, name=Behavioral Modeling Technology, class=OpenC...
6	web services	0.273	XENTITY	OpenCalais	{_typeGroup=entities, _type=IndustryTerm, name=Web services, class=OpenCalais, _typeRefe...
7	modeling	0.273	XENTITY	NounChunk	{Ontology=Enwiki.category, OntologyID=206237, class=NounChunk, label=modeling}
8	c++	0.206	XENTITY	OpenCalais	{_typeGroup=entities, _type=ProgrammingLanguage, name=C++, class=OpenCalais, _typeRefe...
9	user modeling technology	0.204	XENTITY	OpenCalais	{_typeGroup=entities, _type=IndustryTerm, name=user modeling technology, class=OpenCalais...
10	our user modeling technology	0.204	XENTITY	OpenCalais	{_typeGroup=entities, _type=Technology, name=user modeling technology, class=OpenCalais, _...
11	discovery	0.2	XENTITY	NounChunk	{Ontology=Enwiki.category, OntologyID=842768, class=NounChunk, label=discovery}
12	information retrieval	0.198	XENTITY	NounChunk	{Ontology=Enwiki.category, OntologyID=166905, class=NounChunk, label=information retrieval}
13	data mining	0.187	XENTITY	OpenCalais	{_typeGroup=entities, _type=Technology, name=data mining, class=OpenCalais, _typeReferenc...
14	collaboration	0.18	XENTITY	NounChunk	{Ontology=Enwiki.category, OntologyID=93855, class=NounChunk, label=collaboration}
15	psychology	0.164	XENTITY	NounChunk	{Ontology=Enwiki.category, OntologyID=256618, class=NounChunk, label=psychology}
16	decision support tool	0.152	XENTITY	OpenCalais	{_typeGroup=entities, _type=IndustryTerm, name=decision support tool, class=OpenCalais, _ty...
17	rooms	0.152	XENTITY	NounChunk	{Ontology=Enwiki.category, OntologyID=269951, class=NounChunk, label=rooms}
18	real time	0.152	XENTITY	OpenCalais	{_typeGroup=entities, _type=IndustryTerm, name=real time, class=OpenCalais, _typeReferenc...
19	data	0.152	XENTITY	NounChunk	{Ontology=Enwiki.category, OntologyID=369807, class=NounChunk, label=data}
20	xmpp	0.152	XENTITY	NounChunk	{Ontology=Enwiki.category, OntologyID=425453, class=NounChunk, label=xmpp}

Figure 3: Table view of an example expertise model.

performance of the REMA algorithm. In this section we show some preliminary results.

4.1 Topic Cloud View for Expertise Model

An expertise model is shown as a topic cloud where the larger the font size and the redder the color indicate higher expertise levels (Figure 3).

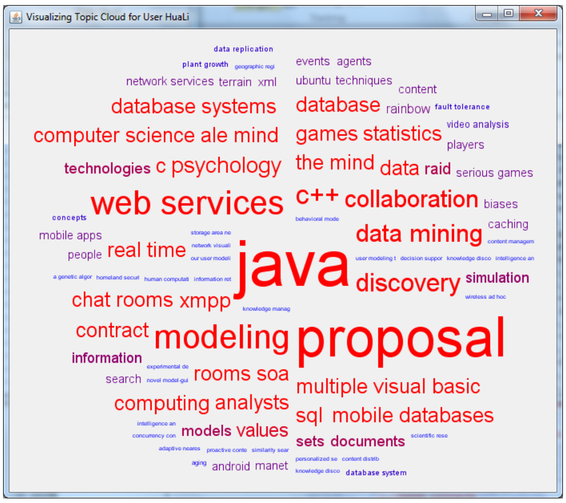


Figure 4: Topic Cloud view of an example expertise model.

4.2 Table View for Expertise Model

An example expertise model is shown as a table with the following six columns (Figure 4):

- ID – the sequence number of each model element (expertise topic)
- ELEMENT – the expertise topic
- WEIGHT – the level of expertise
- TYPE – the type of topic, XENTITY denotes an entity topic, XRELATION denotes a synset (set of synonyms) -> hypernym relationship defined in WordNet.
- NLP – Natural language processing tool used for extracting this element
- Features – features associated with this element.

5 CONCLUSIONS

We developed REMA, a user modeling algorithm that quantitatively identifies skills and expertise from biographic data. REMA takes data from a resume document as input and produces an expertise model. There are two key concepts for this algorithm. First, one's expertise is the cumulative result of the various "learning events" in one's career. Second, one's knowledge and skills can become outdated or forgotten over time if not reinforced by learning. We have developed a prototype resume evaluation system based on REMA and are in the process of evaluating REMA's performance.

REFERENCES

Alonso, R., P. Bramsen, and H. Li, 2010. Incremental user modeling with heterogeneous user behaviors, International Conference on Knowledge Management and Information Sharing, *International Conference on Knowledge Management and Information Sharing (KMIS 2010)*.

Li, H. and R. Alonso. User Modeling for Contextual Suggestion. In *Proceedings of 21st Text REtrieval Conference (TREC 2014)*. NIST, 2014. http://trec.nist.gov/pubs/trec23/papers/pro-RAMA_cs.pdf

Li, H. and R. Alonso, Managing Analysis Context, ESAIR'12: *Fifth International Workshop on Exploiting Semantic Annotations in Information Retrieval*, 2012.

Nemet, G., 2012. Historical Case Studies of Energy *Technology Innovation*, 1–11.

Park, G., Shin, J., & Park, Y., 2006. Measurement of depreciation rate of technological knowledge : Technology cycle time approach. *Journal of scientific and industrial research* 65 (2), 121–127.

Balog, K., Bogers, T., Azzopardi, L., de Rijke, M., van den Bosch, A.: Broad expertise retrieval in sparse data environments. In*: SIGIR* 2007, pp. 551–558 (2007)

Balog, K., Soboroff, I., Thomas, P., Craswell, N., de Vries, A.P., Bailey, P. Overview of the TREC 2008 enterprise track. In: TREC 2008 (2009).

Smirnova1 E. and K. Balog. A User-Oriented Model for Expert Finding. P. Clough et al. (Eds.): ECIR 2011, *LNCS* 6611, pp. 580–592, Springer-Verlag Berlin Heidelberg, 2011.

Liebregts, R. and Bogers, T.: Design and evaluation of a university-wide expert search engine. In: Boughanem, M., Berrut, C., Mothe, J., Soule-Dupuy, C. (eds.) ECIR 2009. *LNCS*, vol. 5478, pp. 587–594. Springer, Heidelberg (2009).

Serdyukov, P., Hiemstra, D., Fokkinga, M.M., Apers, P.M.G.: Generative modelling of persons and documents for expert search. In: *SIGIR* 2007, pp. 827–828 (2007).