AUTHOR(S): MARTIN, K., LIRET, A., WIRATUNGA, N., OWUSU, G. and KERN, M.

TITLE: Explainability through transparency and user control: a case-based recommender for engineering workers.

YEAR: 2018

Publisher citation: MARTIN, K., LIRET, A., WIRATUNGA, N., OWUSU, G. and KERN, M. 2018. Explainability through transparency and user control: a case-based recommender for engineering workers. In Minor, M. (ed.) Workshop proceedings of the 26th International conference on case-based reasoning (ICCBR 2018), 9-12 July 2018, Stockholm, Sweden. Stockholm: ICCBR [online], pages 22-31. Available from: http://iccbr18.com/wp-content/uploads/ICCBR-2018-V3.pdf#page=22

OpenAIR citation: MARTIN, K., LIRET, A., WIRATUNGA, N., OWUSU, G. and KERN, M. 2018. Explainability through transparency and user control: a case-based recommender for engineering workers. In Minor, M. (ed.) Workshop proceedings of the 26th International conference on case-based reasoning (ICCBR 2018), 9-12 July 2018, Stockholm, Sweden. Stockholm: ICCBR, pages 22-31. Held on OpenAIR [online]. Available from: https://openair.rgu.ac.uk/

Explainability through Transparency and User Control: A Case-Based Recommender for Engineering Workers

Kyle Martin[1][0000−0003−0941−3111] ✉, Anne Liret[2][0000−0003−0620−7240], and
Nirmalie Wiratunga[1][0000−0003−4040−2496], Gilbert Owusu[3], Mathias Kern[3]

[1] Robert Gordon University, Aberdeen, Scotland
{k.martin, n.wiratunga}@rgu.ac.uk
[2] BT France, Paris, France
anne.liret@bt.com
[3] British Telecommunications, United Kingdom
{gilbert.owusu, mathias.kern}@bt.com

Abstract. Within the service providing industries, field engineers can struggle to access tasks which are suited to their individual skills and experience. There is potential for a recommender system to improve access to information while being on site. However the smooth adoption of such a system is superseded by a challenge for exposing the human understandable proof of the machine reasoning. With that in mind, this paper introduces an explainable recommender system to facilitate transparent retrieval of task information for field engineers in the context of service delivery. The presented software adheres to the five goals of an explainable intelligent system and incorporates elements of both Case-Based Reasoning and heuristic techniques to develop a recommendation ranking of tasks. In addition we evaluate methods of building justifiable representations for similarity-based return on a classification task developed from engineers' notes. Our conclusion highlights the trade-off between performance and explainability.

Keywords: Case-Based Reasoning · Recommender Systems · Explainable AI · Information Retrieval · Machine Learning

1 Introduction

Within field service provisioning industries there is increasing interest into the empowerment of workers to ensure the right expert knowledge is used at the right level in the decision process. In [12], the scheduling system interactively allocates tasks with empowered engineers thanks to a personalised recommendation system that suggests tasks to an engineer based on their history of completed tasks. However, the increasing complexity of tasks lead to situations where engineers struggle to evaluate accurately the required work on tasks which are nearby and within their skill set. The amount of tasks generated every day across all business divisions can be large, which can further exacerbate the problem engineers face in finding appropriate work with the correct context information at the right time. This question becomes more critical when the type of services are inherently dynamic, such as when high priority tasks are raised that require an engineer's immediate attention, and require that he must abandon tasks which he might be unable to revisit on time. In a worst case scenario, tasks may miss their deadline.

The motivation of this study is to develop a method to access and prioritize tasks which fall within the engineers' capabilities, experience and are of relevance to the business at that point in time. A recommender system has potential to fill this gap [4], but responses to a fuzzy logic recommender [12] suggested that users resented the lack of clarity behind its recommendations. The ability to explain a system's recommendation, or display a level of transparency which allows the user to understand the reasoning behind that recommendation, encourages trust between a system and its users [13].

In [19] the authors present five goals that an explainable intelligent system must be able to satisfy - transparency, justification, conceptualisation, relevance and learning. Based on these criteria, we observe that a CBR system already achieves 3 (transparency [8], relevance and learning [1]), but does not necessarily answer the remaining 2 (justification and conceptualisation). These criteria and the improvement they may bring to interactive services scheduling motivated the contributions of this paper. A real world dataset from telecommunications services has been used for this case study. For each service request, a number of progression comments until successful or unsuccessful completion are reported by engineers. These offer a large source of unstructured data ("engineer notes"), but are difficult to exploit due to variability in content quality.

With this observation in mind, this paper presents a transparent telecommunications task case-based recommender system which has been extended to incorporate the ides of conceptualisation and justification. To assist users in understanding the necessary concepts for decision-making, we introduce a customizable modular design based upon parallel co-ordinates [9], which considers the input of various similarity assessment modules to develop a recommendation ranking. We combine related case attributes and a local similarities model to improve the conceptual level of recommended objects. To improve the system's ability to justify a decision, we evaluate several text representation learning measures to determine which is most useful to select features for justification.

We offer several contributions in this paper. We (1) present a framework for case-based recommender systems which facilitates conceptualisation inspired by parallel co-ordinates. We (2) showcase an extension to the system that allows user customisation to suit individual or business needs and improves transparency. Lastly, we (3) evaluate methods of developing representations from engineer notes for similarity-based return on the basis of their accuracy and ability to justify a decision. Though presented as a field services recommender, the concept could be adapted to fit other domains.

This paper is split into the following sections. Section 2 discusses the problem in more detail and Section 3 talks about related work. Section 4 outlines the use of conceptual parallel co-ordinates to improve conceptualisation. Section 5 discusses text representation learning methods and analyses their impact on recommendation justification. Finally, in Section 5 we offer some conclusions.

2 Learning Similarity from the Experts

As in a number of complex services provisioning organisations, the telecommunications engineering force who carry out the work in the field, gradually form a strong and concrete expertise in the field of network equipment installation and repair. To ensure service delivery, they traditionally are allocated tasks, that each is, in this scenario, a pre-defined time constrained action to perform on a specific piece of equipment. Field engineers record information about the tasks they have completed in text documents called ”notes”. These notes originally contain necessary information such as location of the task and an overview of the work to be done (Order notes) and are expected to be updated by the engineer when the task is completed with some details of how this was achieved (Closure notes). If a task cannot be completed then the cause of this is also expected to be recorded (Further notes) and lastly, the engineer can enter additional information they feel as necessary (User notes).

This work is motivated by the need to learn similarity from a user’s perspective. We believe that by using notes that have been written by engineers themselves as the information source for a similarity metric, the cases retrieved through similarity-based return will be more representative of this point of view. It will also allow greater opportunity for explainability, as it enables potentially generating post-hoc explanations for recommendations based on other engineers’ description of tasks. Finally it enables extracting and sharing implicit knowledge to proactively inform about practical work instructions, all over the service and supply chain.

3 Related Work

The main motivation behind this paper is the work presented in [19]. In that paper, the authors present the five goals of an explainable intelligent system - transparency (the system can demonstrate the reasoning behind its decision), justification (the system can justify why the proposed solution is better than other potential solutions), conceptualisation (the system can illustrate to the user the meaning of concepts required to understand the decision), relevance (the approach adopted by the system is relevant to the problem) and learning (the solution provided by the system can improve user knowledge).

Anecdotal evidence from a number of sources suggest that Case-Based Reasoning (CBR) systems can already answer 2 of the 5 goals of explainability (transparency, relevance) while facilitating another (learning) by virtue of the architecture itself [6, 17]. We can observe that the solutions offered by a CBR system are always relevant to the presented query, as they draw upon most similar problems that the system has seen before from within the same domain [1]. Furthermore, as CBR systems are based on the concept of reusing solutions to solve similar problems, it is trivial to direct users towards the original solution which led to this point, thus demonstrating transparency of decision-making [8]. We also argue that CBR systems facilitate learning through the method in which they draw on past experience to present a solution to the user. As this is similar to the way in which humans learn [14], it eases uptake of knowledge from the system. We do however acknowledge the argument that a human cannot learn from a solution they do not understand [19], so if a user cannot understand why a presented solution successfully answers a query, then their learning will be inhibited. Thus, we suggest that learning can be most easily obtained by achieving the other 4 goals. As such, we maintain focus on improving justification and conceptualisation.

We argue that a vanilla CBR system does not necessarily answer the remaining two goals - justification and conceptualisation. Although the process which led to the selection of CBR as technical approach is transparent, it is not necessarily clear to the user why the resulting recommendation would be better. Furthermore, although [10] have shown that displaying local similarity scores from known calculations can provide clarity to users regarding the justification of a decision, these do not help in understanding the underlying concepts. Hence we suggest improvements to conceptualisation (inspired by parallel co-ordinate visualisation techniques) and justification (through an evaluation of representation learning methods) aspects in the following subsections.

In this work, we propose to use parallel co-ordinates at a conceptual level with the aim of improving the conceptualisation aspect of field service recommendation systems. Parallel co-ordinates are means to visualise high-dimensional data. They are often used in CBR systems to allow simpler comprehension and comparison of local similarities in the return set [9, 10]. Previously, parallel co-ordinate visualisation methods have been used to improve the explanation of CBR solutions for pharmaceutical tablet formation [10]. In that work, pharmaceutical experts praised the clarity of the method and agreed that the visualisation could be more meaningful and easier to understand than a textual explanation.

However, this method becomes less meaningful if the user cannot understand the features or potentially even the local similarity which is being described. We propose an extension to this practice, by using parallel co-ordinate visualisation techniques at a conceptual level. Similar in essence to the work in [3], where the authors suggest a level of abstraction can be useful for classification, we propose that generalisation of concepts can be a useful tool to improve user understanding for explainability (though we do not use this as a basis to then perform induction for classification as in the original paper). By collecting related local similarities together under a meaningful heading, the user is given context which can improve their understanding of individual similarities and how each contributes to the recommendation as a whole.

In this paper we consider both distributional and distributed approaches to learning representations for text documents as a means to improve a system's ability to justify a decision. Distributional approaches, such as the statistical-based method tf-idf, produce sparse document representations that can be difficult to utilise in machine learning algorithms, but are trivial to relate back to specific features. Distributed approaches, such as document-2-vector (Doc2Vec) [7], a method derived from Word2Vec [11], produce dense representations. However, as they develop a level of abstraction it becomes more difficult to relate these to specific features.

Deep metric learners, like the Siamese Neural Network (SNN) [2], do not learn a representation directly from the text itself. They receive combinations of pre-processed representations (such as that obtained from tf-idf or Doc2Vec) as input to develop embeddings which are optimised based on an objective. This objective is defined by a matching criteria, which does not necessarily have its basis in class knowledge [5]. Deep metric learners develop representations where cases that meet these matching criteria exist close together, whilst cases that do not exist further apart. The developed space is therefore optimised for similarity-based return. Deep metric learners have shown achievements in areas such as face verification [18] and similar text retrieval [15].

4 Improving Conceptualisation with Parallel Co-ordinates

In this paper we present a case-based recommender system that uses a parallel co-ordinates approach, inspired from visualisation methods, to contextualise local similarities. Here we use the term 'concept' to describe a subset of local similarities or attributes collected under one descriptive heading. For example, in the suggested system all attributes which are heuristically evaluated to give an idea of task urgency or business priority are grouped under the Business Relevance concept module. Each concept makes use of a subset of local similarities or attributes to develop a score, before all concept scores are combined to develop a recommendation ranking. We use this method to augment the backbone of our recommender, which is based on task similarity knowledge gained from text and is described in the following section.

Collecting local similarities under a meaningful heading allows excellent opportunity for conceptualisation. Instead of being presented with only a recommendation or a general score, users can see how a recommended item is scored across several meaningful concepts. For example, an engineer presented with a top recommendation may have difficulty understanding where it has come from. However, if a score breakdown is provided such that the user can see that the recommendation scored an average of 90% relating to Business Relevance concept and 80% relating to Similarity to Previous Work concept then this could be much more meaningful. Grouping attributes into concepts can also improve understanding of the score if any of the individual attributes or local similarities are not meaningful to the user.

The basic idea is this. Let us say we have a set of concepts, C, with each individual concept represented as $c \in C$. Each concept should represent a related collection of case attributes. Therefore, if A were to represent the full set of case attributes, such that $a \in A$, then each concept represents a subset of case attributes, where $c \subset A$. We can represent the concept similarity score, c_s, between a given query q and any case x as:

$$c_s(q,x) = \frac{\sum_i^{|A|} a_i(q,x)}{|A|} \tag{1}$$

where a_i represents the local similarity calculation and $|A|$ represents the total number of local similarities or attributes that have contributed towards that concept score.

We can combine the output of all concept scores to create a global score by which to rank cases. This allows us to develop a visualisation similar to that of parallel co-ordinates [9], as in Figure 1. We argue that collecting related attributes under meaningful concept headings allows users to better understand the relationship between local similarities by giving them context. This in turn can help them understand these better at a conceptual level. Furthermore users should be better able to understand the breakdown of the decision-making process of the system, improving overall transparency.

4.1 Giving Users Control Over Recommendations

We can facilitate transparency in the presented system by giving the user control over the weighting of the individual components of the recommendation at both a local similarity and concept level. These customization options allow the system to be configured

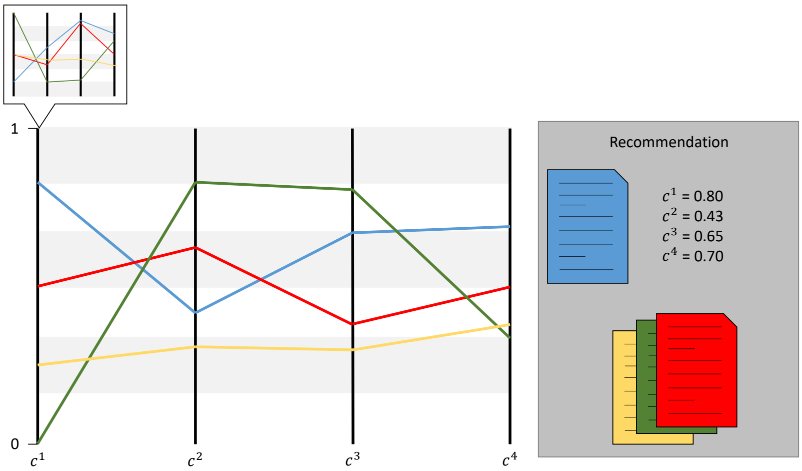

Fig. 1: The concepts of the recommender can be seen to emulate typical similarity visualisation methods. Importantly, each concept acts as a summary for the local similarity methods within it.

to meet both individual user and business needs, as well as encourage trust in the system. In addition, they ensure that the Task Recommender can be altered to meet the needs of the numerous work types within a telecommunication organisation and presents the opportunity to standardise relevant task retrieval across these disciplines. Thus the similarity calculation for any given concept becomes:

$$c_s(q,x) = \frac{\sum_i^{|A|} a_i(q,x) \cdot w_i^a}{|A| \cdot |W^a|} \tag{2}$$

While the overall similarity for a query q to any given case x becomes:

$$q \sim x = \frac{\sum_i^{|C|} c_i(q,x) \cdot w_i^g}{|C||W^g|} \tag{3}$$

By giving the user control over the impact that different concepts have upon the recommendation, we encourage the user to have better understanding of how concepts relate to each other at a summary level, one-step above that of local similarities. This can directly contribute to user learning, one of the goals of an explainable intelligent system. Furthermore the transparency of the system is again highlighted, as it allows the user a better understanding of how the recommendation was made. We plan to evaluate this claim in a subject study with telecommunication engineers in future work.

5 Learning a Text Representation for Justification

We examine several methods for learning representations for text documents in regards to their accuracy and their ability to justify a recommendation. Specifically, we consider a distributional representative (tf-idf), a distributed representative (doc2vec) and a deep metric learner (Siamese Neural Network (SNN)). Furthermore, as deep metric learners require a sample selection strategy to perform most efficiently, we also consider an SNN which uses DYNEE sample selection.

Term frequency-inverse document frequency (tf-idf) calculates a value for each term in a document by dividing the frequency of the term in said document by the percentage of documents which contain that term [16]. As such, for a document representation built from tf-idf, each feature is a value which represents an individual word. It is therefore easy to relate these features back to the raw data to form an explanation. However, tf-idf is not well-suited for extremely large corpus' or vocabulary as this can lead to sparse representations for documents. In addition, it does not explicitly handle synonyms.

Document-2-Vector (Doc2Vec) [7] is an extension of the Word2Vec algorithm [11]. Word2Vec uses contextual knowledge gained from word co-occurrence to build word embeddings for every term in a corpus and develop a metric space where words that have similar contexts exist close together. Using Doc2Vec, the word embeddings for each term in a document are averaged to produce a representation for the document itself. This allows direct similarity comparison between documents within a corpus based upon their contents, whilst avoiding sparseness of representations.

The SNN is a deep metric learner comprised of two matching sub-networks with identical weights and parameters. Examples are input to an SNN in pairs, and are labeled as either positive or negative based on whether the pair satisfies user-defined matching criteria. The metric space which is learned by the SNN is optimised based upon the stated matching criteria, such that positive examples (pairs of instances which adhere to the matching criteria) exist close together, while negative examples (pairs in which members do not fit the matching criteria) exist far apart.

As the SNNs receive pairs as input, there is an additional dimension to training in the form of a pairing strategy. Research has demonstrated that deep metric learners which incorporate a sample selection strategy can offer increased performance. We therefore also considered an SNN supported by DYNEE sample selection. DYNEE is a sample selection method which combines exploitation of the knowledge gained from training thus far and exploration of the feature space to select pairs for training.

5.1 Evaluation

In this section we evaluate the vectorial representation of notes gained through each method - tf-idf, Doc2Vec, SNN and an SNN with DYNEE sample selection - to determine which develops the best representation in terms of accuracy and explainability. For the purposes of comparison we have created a simple classification task where notes are classified according to one of four work types.

5.2 Experimental Setup

We extracted two months worth of notes generated by telecommunication engineers between March and April 2018. We used Order notes as a means of developing a representation for similarity-based return in a case-based reasoning system. Order notes have the nice property to be present in all tasks, which is not true for any other note type. We also filtered out any Order note which contained less than 50 characters, as these were judged not to contain enough information to be meaningful. This resulted in a dataset of 1610 notes split into four classes - cabling (227 notes), jointing (789 notes), overhead (503 notes) and power testing (91 notes). These classes represent the work type (i.e. the primary competence of engineer required) which is associated with each note.

The dataset was split into train and test and evaluated using 5-fold cross evaluation. Embeddings for each note were built using each of the above outlined methods[1]. We used k-nearest neighbour for similarity-based return, with k equal to 3. The Doc2Vec feature size was 300. For the SNN implementation using DYNEE, pair selection was repeated every 5 epochs. The α exploitation ratio used for DYNEE was ${}^{|P|}/_{10}$.

5.3 Results

The results can be seen in Table 1. All representations generated by deep metric learners obtain higher accuracy on the classification task. Tf-idf itself does particularly poorly - this is to be expected, as tf-idf does not consider the context of terms whereas doc2vec does. We only display results for SNNs which used input gained from the Doc2Vec model. This was simply because it achieved the best results, though both SNNs using tf-idf as input still outperformed other methods.

Figure 2 illustrates representations of the casebase using a multi-dimensional scaling scatter plot. It confirms that concepts learned by SNNs form better clusters around class boundaries. This also supports the performance gains observed with SNNs.

Architecture	Accuracy (%)
Tf-Idf$_{k-NN}$	62.24
Doc2Vec$_{k-NN}$	63.79
SNN BASE$_{k-NN}$	66.25
SNN DYNEE$_{k-NN}$	66.83

Table 1: Results of representation learning methods on a classification task.

5.4 Explainability of Results

We examined the explainability of each representation in terms of the ability to justify a classification with evidence from the feature set. Using tf-idf we can highlight exactly what terms are similar and have led to finding nearest neighbours by identifying the features that demonstrate most similarity. As each feature directly relates to a term from the documents themselves, it is trivial to find a list of correlating terms. Similarly, using the features from the doc2vec representation we can identify which concepts have seen the most activation, or are most closely related to specific features. However, the SNN is more opaque. After having been input to the network, it becomes difficult to map the meaning of the abstracted output representation back to the original input.

[1] Both SNN sub-network architectures were comprised of 3-layer perceptrons which used ReLU activations and were trained for 250 epochs.

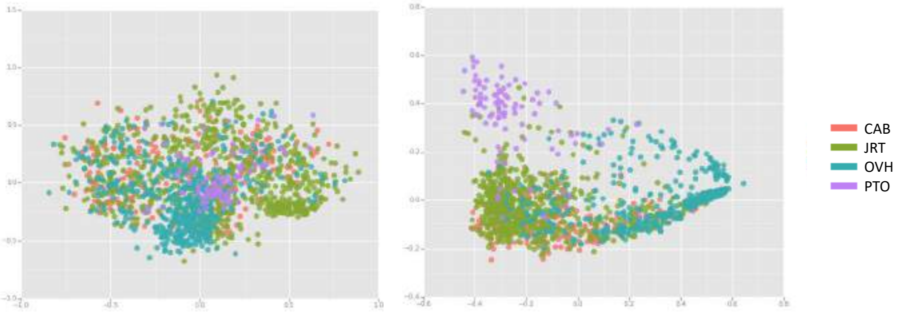

Fig. 2: Representation distribution gained from Doc2Vec (left) and SNN (right).

This leads us to the conclusion that there is a trade-off between performance and explainability. Examining this, we find that the representations built by methods which allow the easiest justification of results (such as tf-idf) are extremely close to the raw features of the data. Meanwhile, the representations built by deeper methods (such as the SNN) tend to achieve greater accuracy as they have underwent a series of abstractions. We can observe there is a trade-off between the clarity that using raw features allows and the performance boost enjoyed by architectures which can develop abstract concepts from data to produce embeddings. This leads us to the suggestion of the explainability-abstraction spectrum. In future work, we plan a subject study with telecommunication engineers to empirically evaluate the justification capacity of different text representation learning methods.

6 Conclusion

In this paper we have introduced an explainable task recommender system that is applied to telecommunications service operations flow. This system adheres to the five goals of explainable artificial intelligent systems in its design. We have also presented a method to improve conceptualisation in a CBR system using parallel co-ordinates at a conceptual level. Lastly, we have performed an evaluation of methods to produce representations for text in regards to their ability to justify a decision.

As future work, we are planning a deployment with the intent of obtaining user feedback to evaluate the system on its explainability. Moreover our study confirmed the potential of engineering notes to help drawing up some extra information about tasks. We are in particular motivated to exploit these unstructured data to classify the pieces of work according to actual required knowledge and then recommend workers with on one hand suitable task in need of action, and on the other hand conceptualised form of task context annotated with disturbance likelihood assessment.

References

1. Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Communications **7**(1), 39 – 59 (Mar 1994)

2. Bromley, J., Guyon, I., LeCun, Y.: Signature verification using a 'siamese' time delay neural network. International Journal of Pattern Recognition and Artificial Intelligence **7**(4), 669 – 688 (August 1993)
3. Drastal, G., Czako, G., Raatz, S.: Induction in an abstraction space: A form of constructive induction. In: Proc. of the 11th Int. Joint Conference on Artificial Intelligence - Volume 1. pp. 708–712. IJCAI'89, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1989)
4. Isinkaye, F., Folajimi, Y., Ojokoh, B.: Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal **16**(3), 261 – 273 (2015)
5. Koch, G., Zemel, R., Salakhutdinov, R.: Siamese neural networks for one-shot image recognition. In: Deep Learning Workshop. ICML '15 (July 2015)
6. Kofod-Petersen, A., Cassens, J., Aamodt, A.: Explanatory capabilities in the creek knowledge-intensive case-based reasoner. In: Holst, A., Kreuger, P., Funk, P. (eds.) Volume 173: Tenth Scandinavian Conference on Artificial Intelligence. pp. 28 – 35. Frontiers in Artificial Intelligence and Applications (2008)
7. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32. pp. II–1188–II–1196. ICML'14, JMLR.org (2014)
8. Leake, D.B.: Case-Based Reasoning: Experiences, Lessons and Future Directions. MIT Press, Cambridge, MA, USA (1996)
9. Lind, M., Johansson, J., Cooper, M.: Many-to-many relational parallel coordinates displays. In: 2009 13th International Conference on Information Visualisation. pp. 25 – 31 (July 2009)
10. Massie, S., Craw, S., Wiratunga, N.: Visualisation of case-base reasoning for explanation (2004)
11. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. CoRR **abs/1301.3781** (2013)
12. Mohamed, A., Bilgin, A., Liret, A., Owusu, G.: Fuzzy logic based personalized task recommendation system for field services. In: Bramer M., Petridis M. (eds) Artificial Intelligence XXXIV. SGAI 2017.Lecture Notes in Computer Science, vol 10630. pp. 300–312. Springer, Cham, Cambridge, UK (December 2017)
13. Muhammad, K., Lawlor, A., Smyth, B.: On the pros and cons of explanation-based ranking. In: Aha, D.W., Lieber, J. (eds.) Case-Based Reasoning Research and Development. pp. 227 – 241. Springer International Publishing, Cham (2017)
14. National Research Council: How people learn: Brain, mind, experience, and school: Expanded edition. National Academies Press (2000)
15. Neculoiu, P., Versteegh, M., Rotaru, M.: Learning text similarity with siamese recurrent networks. In: Rep4NLP@ACL (2016)
16. Ramos, J.: Using tf-idf to determine word relevance in document queries. In: Proceedings of the first instructional conference on machine learning. pp. 133 – 142 (2003)
17. Roth-Berghofer, T.R.: Explanations and case-based reasoning: Foundational issues. In: Funk, P., González Calero, P.A. (eds.) Advances in Case-Based Reasoning. pp. 389 – 403. Springer Berlin Heidelberg, Berlin, Heidelberg (2004)
18. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proc. of the 2015 IEEE Conf. on Computer Vision and Pattern Recognition. pp. 815 – 823. CVPR '15, IEEE Computer Society, Washington, DC, USA (June 2015). https://doi.org/doi:10.1109/cvpr.2015.7298682
19. Sørmo, F., Cassens, J., Aamodt, A.: Explanation in case-based reasoning - perspectives and goals. Artificial Intelligence Review **24**(2), 109 – 143 (2005)