Complex Adaptive Systems Modeling

REVIEW

Open Access

Lifelong aspect extraction from big data: knowledge engineering

M. Taimoor Khan[1*], Mehr Durrani[2], Shehzad Khalid[1] and Furqan Aziz[3]

*Correspondence: taimoor.muhammad@gmail.com
[1] Bahria University, Islamabad, Pakistan
Full list of author information is available at the end of the article

Abstract

Traditional machine learning techniques follow a single shot learning approach. It includes all supervised, semi-supervised, transfer learning, hybrid and unsupervised techniques having a single target domain known prior to analysis. Learning from one task is not carried to the next task, therefore, they cannot scale up to big data having many unknown domains. Lifelong learning models are tailored for big data having a knowledge module that is maintained automatically. The knowledge-base grows with experience where knowledge from previous tasks helps in current task. This paper surveys topic models leading the discussion to knowledge-based topic models and lifelong learning models. The issues and challenges in learning knowledge, its abstraction, retention and transfer are elaborated. The state-of-the art models store word pairs as knowledge having positive or negative co-relations called must-links and cannot-links. The need for innovative ideas from other research fields is stressed to learn more varieties of knowledge to improve accuracy and reveal more semantic structures from within the data.

Keywords: Knowledge-based topic models, Lifelong topic models, Aspect extraction, Automatic knowledge-based models, Knowledge engineering, Big textual data analysis

Background

Probabilistic topic models perform statistical evaluations on words co-occurrence to extract popular words and group them in topics. A topic can be considered as a concept represented through its top words. In aspect based sentiment analysis (ABSA), topics are used to represent product aspects or sentiment category. Due to the amount of content produced online, there is rich information available that can be processed for decision making. It can help to identify public trends, e.g., popular products and their features. In social sciences it is used to aggregate opinions of a group of people as opinion mining or sentiment analysis. Its sub-domains are hazard analysis, threat analysis, bias analysis, etc. Government bodies can use it to make policies that address the concern of majority of the people and can even plan their speeches and official statements accordingly. Sentiment analysis in Khan et al. (2015) stress on user centered health care facilities. The importance of ABSA can be realized from the fact that its practical applications are available while it is far from mature. Sentiment analysis aggregated at aspect level is more informative than product level analysis, as shown in Fig. 1.

Springer Open

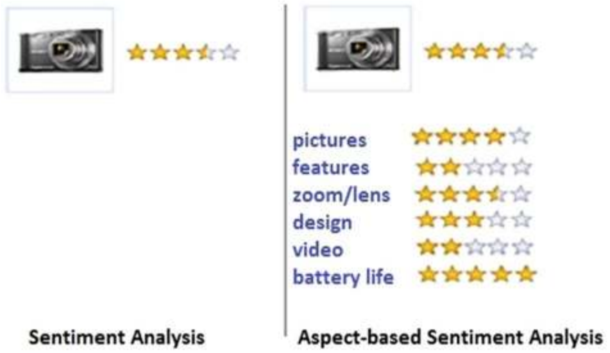

Fig. 1 Comparison of information provided through sentiment analysis (at product level) and aspect-based sentiment analysis

Machine learning topic models

Topic models are preferred for aspect extraction as coherent topics without requiring any manual support. Along with coherent topics, it also tends to find incoherent topics that doesn't make much sense, however, it shows better results for large datasets. It makes topic models a suitable choice for big data analysis. The supervised techniques require labeled data, while semi-supervised techniques need guidance from domain experts. Both supervised and semi-supervised techniques are least effective to large-scale data and are not scalable for being domain specific. Among traditional unsupervised techniques, the dictionary based approach fails to address the domain specific relevance due to the use of general purpose dictionary. For example, *screen* and *resolution* generally have no relevance but are strongly co-related in domains of electronic devices. Frequency-relation based techniques struggle at the hands of multiple forms in which a word can exist. Aspect extraction is the most challenging part of ABSA, having the extraction mechanism shown in Fig. 2. Topic models hold the advantage of identifying aspect terms and grouping them together under topics as a single step.

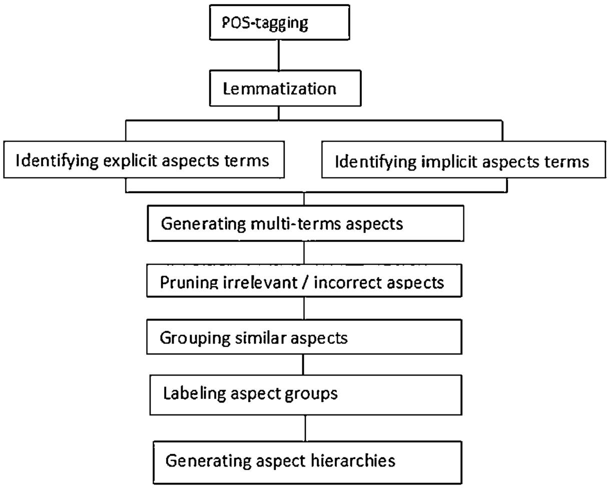
POS-tagging
Lemmatization
Identifying explicit aspects terms
Identifying implicit aspects terms
Generating multi-terms aspects
Pruning irrelevant / incorrect aspects
Grouping similar aspects
Labeling aspect groups
Generating aspect hierarchies

Fig. 2 The process of extracting product aspects which are used as opinion targets in aspect-based sentiment analysis

Topic models have evolved over the past decade, having many extensions, to meet the challenges of real world applications. Knowledge-base topic models are provided with domain specific knowledge rules instead of seed aspects, following a semi-supervised approach. A paradigm shift was taken in Chen et al. (2014) to introduce automatic lifelong learning model that has a knowledge-base maintained automatically. The model learns by itself and apply that knowledge to improve results. The quality of knowledge improves with experience. It benefits from the grey (overlap) region among various domains in big data to learn popular patterns as knowledge and transfer it to the task in hand. Being unsupervised, the model is expected to learn wrong and irrelevant knowledge as well. But with experience they are pushed down the priority queue and are eventually filtered out. Lifelong learning models have four main components, i.e., knowledge representation, extraction, transfer, and maintenance/retention. With innovative ideas for knowledge engineering, more types of knowledge can be explored.

Big subjective data

There is big subjective data ever growing on various online sources that is produced by amateur authors (Katz et al. 2015). It can be used for aspect extraction by analyzing the data at user, relationships and content levels (Tang et al. 2014; Guellil and Boukhalfa 2015). Aspects are extracted in relation to the product it belongs and sentiment word used for it. Sentiments are to be contextualized and conceptualized for specialized domains (Gangemi et al. 2014; Weichselbraun et al. 2014). Preserving the flow of association can be extended to explore reasons for the given sentiments. Online data is in large volume, covering a variety of domains, on various platforms having different formats and is being poured in continuously. It exists in abundance, covering a variety of topics and is freely available on blogs, forums, social media and review websites. It is being produced by amateur authors and is expected to have all sorts of inconsistencies including under or over use of capitalization, spelling mistakes, shortened words, slang, swearing, etc. The data available is huge and there is little known about it. Separating out a single domain from big data is a cumbersome task and it is unwise not to benefit from it for generalization and high accuracy. Considering the massive content produced online, only the techniques that can scale up to support big data can survive. That is how real world problems of data analysis and information extraction can be addressed.

Domain in big subjective data is a collection of documents about a single subject usually a product in commercial data. A document is a user review submitted for that product using an online platform. The review documents can be processed at different levels where phrase level analysis is preferred for topic extraction. It yields better results to extract product aspects as topics, which is a collection of co-related phrases. Therefore, phrase level analysis is also known as aspect level analysis. Document or sentence level analysis can be used at pre-processing for ABSA to identify language, spam, subjectivity, slang, etc. Transcribed text has many words that doesn't have semantic meaning, e.g., Hmm, Aah and need to be filtered (Katz et al. 2015; Takeuchi and Yamaguchi 2014; Cailliau and Cavet 2013). Datasets from transcribed and micro-blogging platforms have low accuracy due to noise (Ofek et al. 2014). Structural (meta) information can be used in support of content for improving accuracy (Katz et al. 2015), e.g., likes, shares, tweets, retweets, etc. Product aspects can exist in various forms, i.e., known or unknown based on prior knowledge of the domain. To increase data for stable results (Xia et al. 2013) generated inverse of selected documents. Known aspects are provided as seeds to identify unknown aspects in a semi-supervised approach. They can be either explicitly mentioned or implied in a situation. It can consist of one or multiple terms, e.g., "*battery life*". The aspects and sentiments extracted are used for aspect-based sentiment analysis (Katz et al. 2014; Hai et al. 2014; Poria et al. 2013; Machova and Marhefka 2014; Medhat et al. 2014).

Complex network analysis

The study of complex network analysis gained popularity to observe the nature of a system through the properties of its constituents. For example, Internet is a complex network of routers and computers having physical or wireless links. In social networks people have edges based on their relationships with other people. There are many complex networks available in real world as biological network, neural network, sports, hobbies, movies, religion, etc. These networks are used for a variety of purposes, e.g., constructing marketing strategies, tracking criminal organizations, analyzing social and psychological behavior, etc. There are different online social networks having complex nature, where World Wide Web is the largest complex network with known topological information having billions of nodes. Complex network falls in category of graph theory which initially focused on simple graphs. The use of complex networks has dramatically increased due to computerization of data acquisition leading to large datasets. The increased computing power has allowed to process millions of nodes at high performance. Most of the latest research is focused on combining different disciplines with complex networks. Complex networks have its roots in mathematics and statistics and are successfully used to give efficient solutions for problems in sociology, psychology, business, computer science and other disciplines.

Complex networks are highly resourceful where a system can be analyzed from a number of perspectives. It offers a range of properties to study the nature of a problem. Some of the commonly used complex network characteristics discussed in Albert and Barabási (2002), Javarone et al. (2013) and Newman (2003) are briefly explained. Clustering, component or communities refers to many small clusters detected in a network. Clustering coefficient is the measure of degree to which nodes in a graph tends to cluster together. Most of the technological and real world complex networks have clustering coefficient higher than that of random clusters, which is an obvious proof for presence of patterns. Deeper analysis of these systems can help to identify the hidden patterns and use them for future tasks. Average path length is the mean length between any two randomly picked nodes. The concept of small-world property is based on the fact that most real world complex networks have short average path length. Diameter or connectedness is the maximal distance across the network. Diameter of a disconnected graph is infinite in which case maximum diameter across all clusters is considered. Degree is the number of edges of a node while degree distribution shows the average spread of node degrees across the network. Scale free networks, network resilience, network navigation, etc. are some other properties of networks. Complex networks are constructed for natural language can introduce these diverse properties to Machine Learning for Natural Language Processing tasks. It has words as nodes associated through co-occurrence or semantic co-relation. Like many other real world complex networks, the Natural Language networks also obeys the small world property and showed high clustering co-efficient.

Complex network analysis is used to address problems in a variety of domains. It has relevance to text analysis for being rich in information, preserving order and hierarchy, etc. Complex networks are used for various NLP tasks in the past. Complex networks can offer a number of properties discussed in previous section. Complex networks are used for text categorization and summarization in literature (Chang and Kim 2013; Cancho et al. 2007; Hassan et al. 2007). Constructing a complex network of the big textual data, the global properties of the network reveals the nature of dataset in general. Networks having more randomness with low clustering coefficient and longer average path lengths are assumed to have more information with low confidence. Biased networks tends to show popular patterns with higher confidence. Complex networks helps to compartmentalize different NLP tasks and provide visualization to help analyze the problem. However, it can also be used as a knowledge-base in support with a machine learning technique. An important aspect of complex networks is that the effect of a community within a network can be identified through its neighbors. This concept is also used for NLP tasks in machine learning where the context of a word is better understood by its surrounding words. A complex network of big textual data is shown in Fig. 3 where (a) gives a view of the complex network of big data. The colors represent different domains in dataset while the size of the nodes show the popular terms in each domain. Figure 5b shows how communities exist within a domain of the network and their impact on each other through neighboring nodes. Communities with high clustering coefficient and modularity gives a popular pattern with greater support from different review authors in the dataset.

Rough sets

Rough set theory is an independent discipline and has substantial progress in different application domains. Probabilistic rough sets relax the rigid thresholds of Pawlak rough sets (Pawlak 1982) to overcome its limitation and raise its applicability. The

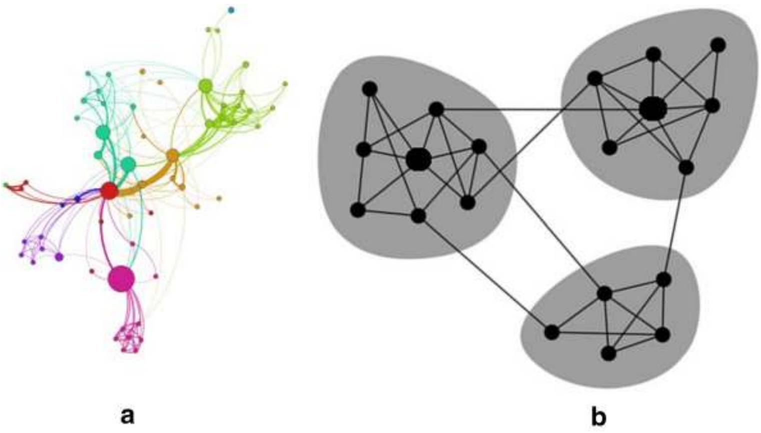

Fig. 3 Knowledge as complex network with **a** using *colors* as different domains while sizes vary based on popularity (as degree). **b** shows identifying communities, within a domain, as knowledge rules, based on clustering coefficient

suitable probabilistic thresholds having some level of error tolerance are evaluated in relevance to the data by using decision-theoretic rough set (DTRS) model, Bayesian rough set (BRS) model, variable precision rough set (VPRS) model, game-theoretic rough set (GTRS) model (Yao 2007; Śle et al. 2005; Ziarko 1993; Herbert and Yao 2011; Azam and Yao 2014). Jaccard Index is used in Clark et al. (2013) to explore similarity measure among regions. Rough sets based analysis helps in situations of decision making when there is limitation of information, multi-decision criteria, feature selection, rule mining, business prediction and fault diagnosis. The decision rules are analyzed for robustness to classify on partially matched data (Azam and Yao 2013). LML topic models can benefit from rough sets to decide learning patterns due to multi-decision criteria as the information available is insufficient.

Rough sets separate the problem space into three disjoint regions has Positive, negative and boundary region by following certain criteria. The objects (data instances) at the boundary region partially accept the criteria applied on it. A satisfaction function is used to measure the relevance of an object to the conditional criteria. To address classification problems with rough sets, three-way decisions with probabilistic rough set (Yao 2010) are introduced. Three-way decision detect communities with relationships as completely belong, completely not belong and incompletely belong (Liu et al. 2013). It achieved higher accuracy as compared to two-way decision, which were forcefully classifying objects even if there isn't enough information available for it. Three-way decision use defer state to store objects that are not classified to any of the available classes. Sequential three-way decisions (Yao 2013; Li et al. 2013a, b; Su et al. 2013) utilize a hierarchical structure of multi-view descriptions to discover objects that are less known by increasing the amount of information at each level. The objective is to use only the information required to classify an object for scenarios where acquiring information is expensive. Only the objects with defer state are further investigated. A cluster ensemble technique is used in Yu and Zhou (2013), using three-way agreement among multiple clustering models.

Aspect extraction review

Unsupervised techniques are preferred for aspect extraction because they are efficient and domain independent. Therefore, they can be easily scaled up to big data consisting of aspects from a variety of domains. Since no training data is required, they are inexpensive and can be directly applied to fresh content. The accuracy of unsupervised techniques is low as it generate incoherent topics that does not have semantic relationship with other terms in the topic. Supervised and semi-supervised techniques require experts and domain specific guidance and are therefore discouraged for real world big data analysis. Recent work on aspect extraction is focused on improving the accuracy of unsupervised techniques by enabling them to improve through a learning mechanism. Parts-of-speech (POS) tagging and stop words removal is performed at pre-processing, so that the efficiency can be improved by applying aspect extraction on potential aspects only, also called candidate aspects. Taxonomy of unsupervised topic extraction techniques covered in the study is given in Fig. 4.

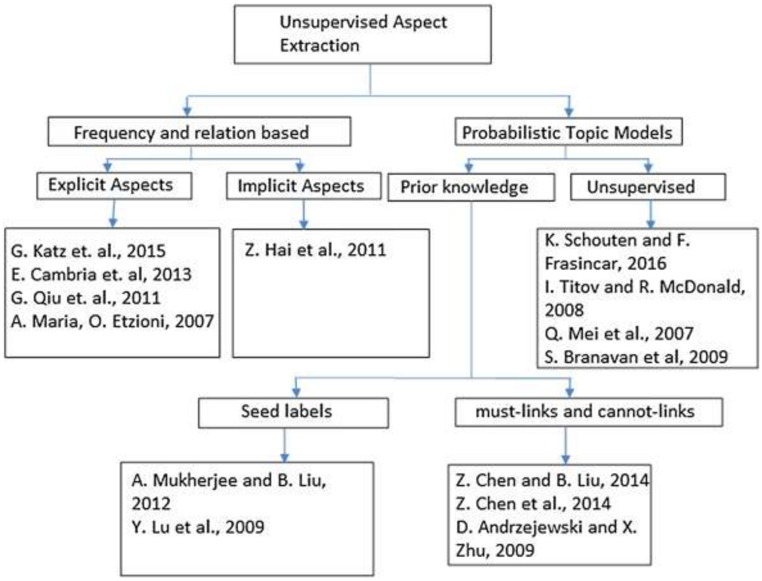
Unsupervised Aspect Extraction
Frequency and relation based
Probabilistic Topic Models
Explicit Aspects
Implicit Aspects
Prior knowledge
Unsupervised
G. Katz et. al., 2015
E. Cambria et. al, 2013
G. Qiu et. al., 2011
A. Maria, O. Etzioni, 2007
Z. Hai et al., 2011
K. Schouten and F. Frasincar, 2016
I. Titov and R. McDonald, 2008
Q. Mei et al., 2007
S. Branavan et al, 2009
Seed labels
must-links and cannot-links
A. Mukherjee and B. Liu, 2012
Y. Lu et al., 2009
Z. Chen and B. Liu, 2014
Z. Chen et al., 2014
D. Andrzejewski and X. Zhu, 2009

Fig. 4 Taxonomy of literature studied for unsupervised aspect extraction

Frequency and relation-based techniques

Aspect-based SA was introduced in Hu and Liu (2004) using frequency-relation based technique for unsupervised aspect extraction. They generated a list of candidate aspects above a frequency threshold and then removed candidate aspects that do not exhibit aspects-like behavior. An aspect is expected to belong to an entity and is associated to a sentiment word. Some rarely mentioned candidate aspects are added to the list for their strong aspects-like nature. It did not perform well with multi-term aspects and completely ignored implicit aspects. The precision of this technique was improved in Popescu and Etzioni (2007) by verifying aspect-entity relationship for candidate aspects through point-wise mutual information (PMI). Some pre-defined discriminators were used to mine patterns like "___ of an entity", "entity having ___", etc. in which an aspect may exist with its entity. They are applied at lexicons and therefore avoid over-fitting but miss the context (Katz et al. 2015; Cambria et al. 2013). The online lexicon sources has many irrelevant words that are not contributing effectively (Tsai et al. 2013).

$$PMI(a, d) = \frac{p(a, d)}{p(a)p(d)} \tag{1}$$

The later work with frequency-relation based approach is focused on applying more variety of relational filters to reduce the long list of frequent candidate aspect. Additional filters are applied at document and paragraph level. They associated high rating with aspects given in pros and cons format and used them to explore more aspects. Majority of the aspects happen to be domain specific thus candidate aspects having high frequency in generic corpus are least likely to be an aspect. To extract multi-term aspects with improved accuracy the C_{value} measure (Frantzi 1998) resolve term distance. Frequent candidate aspects are used as core aspects by using the distance measure in Cilibrasi and Vitanyi (2007) to explore more aspects. Relational dependency parsers are used in Qiu et al. (2011) to drop out irrelevant candidate aspect terms. The aspects extracted are mapped to their associated sentiments (Tuveri and Angioni 2014; Zhang and Liu 2014).

Limited work has been done in finding implicit aspects which are complex to find as compared to explicit aspects. They can be found in the form of adjectives and adverbs. They are used in linguistic patterns to explore implicit aspects. An implicit aspect is not used but is implied in a situation. A clustering based approach is used to map implicit aspects to their respective explicit aspects through mutual reinforcement relationship among clusters (Su et al. 2008). The associations among clusters were weighted based on frequency to explore implicit aspects through reliable links. Two-phase association rule mining (ARM) is used to map implicit aspects to their corresponding explicit aspects using sentiment words as condition (Hai et al. 2011). The rule consequents (explicit aspects) are used to generate more conditions for ARM, where implicit aspects could be mined. Bag-of-Nouns scheme is used to extract candidate aspects from each cluster and is provided to self-organizing maps (Kohonen and Somervuo 1998) to reduce the feature set. The aspects extracted are domain specific while keeping the model independent of domain and language. Clustering based approaches have been used previously in weakly supervised approaches. Multi-level aspect extraction has been a new concept introduced in aspect extraction as agglomerative clustering over the clusters of aspects. These techniques has suffered from the richness of natural language. Some of the major NLP challenges faced by machine learning techniques are discussed in Khan et al. (2016).

Probabilistic topic models

Probabilistic topic models have been extensively used in information extraction to great effect. Although it doesn't mine the semantics in content but uses co-relations between words through heavy computations for calculating probabilities of words co-existence. It follows BOW approach by considering the importance of words and their positions only. This idea doesn't seem convincing, however, it produced reliable results when applied on a large volume of data. The initial parameters required can be extracted from the corpus or provided as constant. The probabilistic topic models extract topics from documents or generate documents from topics. They are either based on Probabilistic Latent Semantic Analysis pLSA (Hofmann 1999) or Latent Dirichlet Allocation LDA (Blei et al. 2003). Candidate aspects are mostly found in the form of nouns and noun phrases. Topic models are extended differently to focus on candidate aspect terms only. One of the advantages of using topic models, is that they identify and aggregate aspect terms at the same time. Topic models extract hidden thematic patterns in large collection of documents (Schouten and Frasincar 2016).

Aspect-sentiment joint model extract topics and distribute them further among aspect, positive-sentiment and negative-sentiment models (Mei et al. 2007). It is based on pLSA and requires user guidance to define the separation among the given models. The topic models are found to converge to global topics which are products and brand, while ignore aspects due to homogeneity. MG-LDA (Multi-grain LDA) (Titov and McDonald 2008) addresses this problem by running LDA globally to extract entities while locally (through a sliding window) to extract aspects. The model proposed in Branavan et al. (2009) considers a special type of review format, i.e., having pros and cons section separate from the review document. The key phrases from the pros and cons section are used to extract more candidate aspects, based on distributional similarity between them. They claimed to extract aspects even from incomplete reviews, however, did not provide the separation between aspects and sentiment words. The sentiment words extracted through topic models as candidate aspects are pruned by considering them as non-aspect adjectives. Two joint models, that are sentiment-LDA and dependency-sentiment-LDA are used to extract aspects through their dependency on sentiment words, however, the separation between aspects and sentiment words was inconclusive.

MaxEnt-LDA (Zhao et al. 2010), a hybrid model is used to discover aspects and sentiments and the separation is performed through syntactic patterns. Multinomial distributions are performed for the extracted word to indicate whether it is aspect or sentiment word, while it learns initial values from training data. More hybrid models that are used for reliable separation between aspects and sentiments by training the model with a labeled dataset (Griffiths et al. 2004). The semi-supervised topic models required user guidance as initial seeds in order to extract more aspects. They provide high accuracy with guided inference to produce aspect distribution confirming to user's needs. They are incorporated differently in Titov and McDonald (2008) where the extracted aspects are also supported with ratings and evidence to claim their authenticity.

Knowledge-based topic models

In order to reduce the gap in accuracy of supervised and semi-supervised topic models, several knowledge-based models were proposed that improved results with manually provided knowledge rules instead of seed aspects. DF-LDA (Andrzejewski and Zhu 2009) was introduced as a knowledge-based topic model, to which knowledge rules are provided in the form of must-links and cannot-links. The earlier semi-supervised models required experts to provide seed aspects. It changed the perspective towards aspect extraction as the accuracy was related to the quality and usefulness of knowledge. The knowledge rules guided the model to decide which aspect terms to put together under the same topic. Intra-topic terms possess high must-link probability while inter-topic aspect terms hold high cannot-link probabilities. It empowered the model to decide where to place an extracted aspect term by incorporating its knowledge. A problem identified with this model is that of being two strict for its rules, limiting its utility to certain scenarios.

In spite of the benefits of knowledge-based models, the knowledge is domain specific and is provided manually. Therefore, the model lack scalability to process big data. Secondly, it always believe the user support to be accurate, where as a mediocre domain expert might want to provide some intuition but is not exactly sure of the accuracy of the support provided. The model does not have any mechanism of verifying the knowledge provided and rectifying it through the actual data. The issue was addressed through Automatic knowledge-based model that learns and apply knowledge without any external support. It can process many unknown domains in big data and can extract and process knowledge from within, by exploiting the large volume of data. The results of the model improves, as it grows with experience.

Automatic knowledge-based topic models learn knowledge automatically without any user intervention proposed in Chen et al. (2014). The mechanism of aspect extraction, learning knowledge and its transfer in current task is depicted in Fig. 5. The model learns

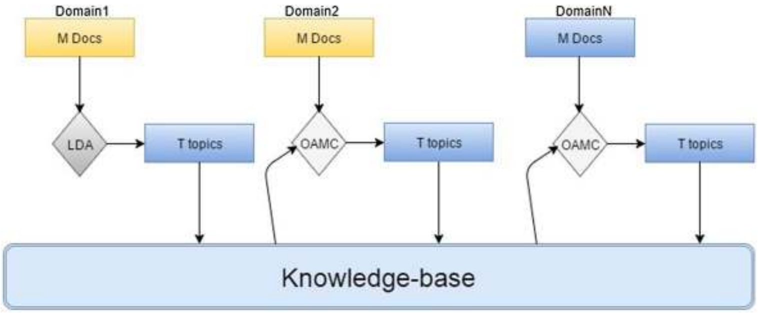

Fig. 5 LML model learning from each task to grow knowledge-base, which is consulted for each future task

adaptively, is independent of the nature of domains that it processes and can be tuned for big data analysis. The model exploits the huge size of data and its variety of formats to its own advantage for knowledge verification. It benefits from the grey region, overlapping among various domains. But an automatic learning model is expected to learn wrong knowledge that can be identified at the later stages as the model grows in experience. Topics extracted with baseline-LDA are clustered to ensure good quality knowledge. The knowledge rules learnt are of equal importance, which make it inconclusive towards conflicting knowledge rules. The model in Chen and Liu (2014) learn through multi-support frequent itemset mining (MS-FIM) (Liu et al. 1999). Knowledge is transferred into the model through multi-generalized poya urn (M-GPU) model (Mahmoud 2008; Mimno et al. 2011) to adjust the position of knowledge terms in current domain. Transitivity problem is addressed by connecting knowledge rules that share a word and have common context. LML models make results highly dependent on the knowledge-base to guide the inference in Gibbs sampling. With knowledge engineering, the quality of knowledge and their contribution can be enhanced. New varieties of knowledge can help with more useful insights. knowledge-based content recommendation systems are analyzed in Burke (2002, 2007) and Tang et al. (2013).

Discussion

Lifelong machine learning (LML) models can only keep up with the processing needs of data produced online. It requires minimal pre-processing and can be directly applied to mine popular topics for exploring big data. For aspect-based sentiment analysis, the extracted topics are considered as aspects while the words in the topic are the different forms of referring to it. Automatic knowledge-based models are inspired from semi-supervised models to which domain specific knowledge is manually provided. LML models manage a knowledge-base automatically. It may learn wrong and irrelevant knowledge as well. Wrong knowledge can resonate for few future tasks, lowering their accuracy, until filtered out. Despite of that, LML models guide the big data analysis problem in the right direction, where new ideas from different domains can be incorporated to improve the quality and contribution of the knowledge-base. LML models require minimal effort in pre-processing. It does not require data from different domains to be separated. It can also be applied to data with noise and inconsistencies. However, in that case the model takes longer to converge to coherent topics. The learning curve is heavily dependent on the nature and structure of the knowledge-base. LML models are specifically designed for big textual data and therefore performs poor with single domain. It would therefore, be inappropriate to compare its topic coherence with that of the traditional ML techniques. LML models exploit the large volume of data to its advantage, to identify popular patterns and use them towards improving the accuracy of the machine learning model. Table 1 shows comparison in the approach of various knowledge based topic models. Table 2 shows that how optimum learning is critical for the accuracy of the model. At low error tolerance the model is not learning enough and there isn't much improvement in results. Whereas by raising the error tolerance below suitable values, noise is introduced to the knowledge-base and the results deteriorate. The value with (*) is the highest accuracy and shows suitable thresholds for error tolerance. However, these values vary with the nature of data and therefore, suitable thresholds for learning must-links and cannot-links are to be evaluated for each dataset.

Table 1 Comparison of knowledge-based topic models

Features	DF-LDA	SAS	ME-SAS	MC-LDA	AKL	LTM
Model approach	Semi-superivsed	Semi-supervised	Hybrid	Semi-supervised	Automatic	Automatic
Knowledge type	1–1 mapping	Rule sets	Rule sets	Rule sets	Rule sets	Rule sets
User support	Must-link/cannot-link	Seed aspects	Seed aspects	Must-set/cannot-set	Must-set/cannot-set	1–1 mapping
Handling big data	No	No	No	No	Yes	Yes
Learning criteria	Nil	Nil	Nil	Nil	Knowledge clusters	PMI
Losing wrong knowledge	Nil	Nil	Nil	Nil	Yes	Yes
Transitivity issue addressed	No	No	No	Yes	Yes	Nil
Aspect/sentiment separation	No	Yes	Yes	Yes	Sentiments pre-processed	Sentiments pre-processed

Table 2 Learning of LML models at different levels of error tolerance and its effect on accuracy of the model

$t_{mustlink}$	$t_{cannotlink}$	Total rules used	Topic coherence
−1	1	3	−855
−1	0.9	5	−851
−0.9	0.9	6	−842
−0.9	0.8	9	−839*
−0.8	0.8	11	−843
−0.8	0.7	14	−852

Topic models stand on strong conceptual and mathematical base and are very flexible; therefore, more high level ideas from different domains can be used to incorporate knowledge into it. Unlike must-link and cannot-link word pairs, knowledge as complex network is highly resourceful and can help on multiple fronts. It can reveal the general nature of data based on its global properties and decide how much to learn from it. While based on the local properties, particular clusters or communities are preserved as knowledge. It shares some similarities with the data as network (DAN) model (Armano and Javarone 2013). Like DAN model a normalized weighted network is generated for the given data. Elements are represented as nodes sharing edges based on similarity, as co-occurrence. However, communities with low clustering coefficient may not be preferred as knowledge. Dense communities can be trusted for having high confidence. Complex network have provided useful solutions in different disciplines and can be effective with knowledge-based models. Popular nodes can be given high priority in decision making introduced in Javarone et al. (2013) where node with high degree are given preference over others in decision making. The percolation theory discussed in Albert and Barabási (2002) is the presence of a critical probability at which giant cluster is formed. The concept is known as percolation transition in mathematics and statistical mechanics. At a value smaller than critical probability the number of communities are identified for the given data as probable number of knowledge rules.

LML models are far from where they can be used for many purposes to recommend items, when applied on big data. The knowledge-based models, i.e., DF-LDA , MC-LDA used must-links and cannot-links manually provided by domain experts. GK-LDA and LTM used must-links only that were verified from the data through probability and clustering based filters. AKL used both automatically mined must-links and cannot-links, however, the clustering method used for knowledge extraction proves to be the performance bottleneck. The knowledge extraction and transfer mechanism was improved in AMC (Chen and Liu 2014) using FIM with multiple support. It gives high accuracy and performance as compared to the previous models. All of the existing LML topic models use machine learning techniques to extract knowledge that has proved to be limited in scope. It is therefore desired to introduce ideas from different domains so that the learning mechanism of LML models can be enhanced. The existing models lack knowledge maintenance and retention module, due to which the knowledge extraction process is repeated each time a task is performed. It adversely affect the performance of the model. Secondly the quality of a knowledge rule and its contribution can never be monitored. Therefore, a knowledge retention module is required to store knowledge learnt. More variety of knowledge is required that would relate aspects to their respective products. Product-aspect association is important to be resolved when multiple products and their aspects are discussed in comparison to each other. Similarly an aspect can be complex enough to have lower order of aspects forming a hierarchy. The machine learning knowledge extraction mechanisms used, i.e., point-wise mutual information (PMI), clustering, frequent itemset mining (FIM) do not have the depth to address the discussed issues and limitations.

There are different content recommender systems developed that incorporates some degree of knowledge by processing content along with meta information (Bobadilla et al. 2013; Ricci et al. 2011; Lops et al. 2011). The learning patterns of LML models can benefit from rough sets as very few rules can be learnt as knowledge. The others having boundary level satisfaction are hard to decide. Unfortunately most of the relations lie in this region and therefore, discarding them will leave the learning module ineffective. For optimum learning the model needs to analyze data and then decide what to learn from each domain. In spite of aspect extraction and sentiment analysis based models, there is still a need of standardizing the dataset and evaluation criteria (Schouten and Frasincar 2016). With the availability of ConceptNet as a more precise semantic lexicon resource, Bag-of-concepts should be used for analysis instead of Bag-of-words. Influence analysis are the extensions of aspect-based sentiment analysis identifying the influence of an entity on another (Nguyen et al. 2015; Rabade et al. 2014). These newly found sub-fields are modifying the process towards building problem specific real world application.

Conclusion

LML models are recently used for NLP tasks. Complex networks can be used as a learning module for LML models. In order to have LML models with improved accuracy, the contextual semantics of domains are to be investigated. With a good learning model, accuracy of the model improves with experience. Since LML models can learn wrong knowledge as well, therefore, the model needs to have a strong filtering mechanism. The learning curve of the model is dependent on sequence of tasks as well. Tasks having more relevance will help the model learn better. To extract hierarchical topics, major modifications are to be made to the learning and transfer mechanism. To support streaming data, the knowledge module needs to grow adaptively. LML models have a wide range of applications as they suit the processing needs of the real world data produced online. With social media the model can monitor the online activities of people for sentiment analysis. It can help to dig deeper for hazard analysis, threat analysis, detecting criminal organization and identifying sparks for potential chaos. With the cost effective availability of hardware, LML models can best serve the processing needs of big data.

Authors' contributions

MTK performed critical analysis on the data, drafted the manuscript and concluded the concepts presented. MD helped with data acquisition about knowledge-based systems and drawing important interpretations from it. FA identified the key challenges in knowledge-based systems and provided overview on complex networks and how they can be used with knowledge-based models. SK supervised the whole activity, helped in drafting and revised the whole document critically for proof of concepts. All authors read and approved the final manuscript.

Author details
[1] Bahria University, Islamabad, Pakistan. [2] COMSATS IIT, Attock, Pakistan. [3] IMSciences, Peshawar, Pakistan.

Acknowledgements
We are thankful to Bahria University, Islambad for providing the necessary environment and support to carry out this work.

Competing interests

The authors declare that they have no competing interests.

Received: 6 October 2015 Accepted: 21 March 2016
Published online: 31 March 2016

References

Andrzejewski D, Zhu X (2009) Latent Dirichlet allocation with topic-in-set knowledge. In: Proceedings of the NAACL HLT 2009 workshop on semi-supervised learning for Natural Language processing, Association for Computational Linguistics, pp 43–48

Armano G, Javarone MA (2013) Clustering datasets by complex networks analysis. Complex Adaptive Syst Model 1(1):1–10

Azam N, Yao J (2013) Formulating game strategies in game-theoretic rough sets. In: Rough sets and knowledge technology. Springer, Heidelberg, pp 145–153

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022

Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Recommender systems survey. Knowl Based Syst 46:109–132

Branavan SRK, Chen H, Eisenstein J, Barzilay R (2009) Learning document-level semantic properties from free-text annotations. J Artif Intell Res 34:569–603. doi:10.1613/jair.2633

Burke R (2002) Hybrid recommender systems: survey and experiments. User Model User Adapt Interact 12(4):331–370

Burke R (2007) Hybrid web recommender systems. In: The adaptive web. Springer, Heidelberg, pp 377–408

Cailliau F, Cavet A (2013) Mining automatic speech transcripts for the retrieval of problematic calls. In: Computational linguistics and intelligent text processing. Springer, Heidelberg, pp 83–95

Cambria E, Schuller B, Xia Y, Havasi C (2013) New avenues in opinion mining and sentiment analysis. IEEE Intell Syst 2:15–21

Chang J, Kim I (2013) Analysis and evaluation of current graph-based text mining researches. Adv Sci Technol Lett 42:100–103

Chen Z, Liu B (2014) Mining topics in documents: standing on the shoulders of big data. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 1116–1125

Chen Z, Mukherjee A, Liu B (2014) Aspect extraction with automated prior knowledge learning. In: Proceedings of ACL. pp 347–358

Cilibrasi RL, Vitanyi PMB (2007) The google similarity distance. Knowl Data Eng IEEE Trans 19(3):370–383

Clark PG, Grzymaa-Busse JW, Rzasa W (2013) Generalizations of approximations. In: Rough sets and knowledge technology. Springer, Heidelberg, pp 41–52

FERRER I CANCHO R, Capocci A, Caldarelli G (2007) Spectral methods cluster words of the same class in a syntactic dependency network. Int J Bifurc Chaos 17(07):2453–2463

Frantzi KT (1998) Automatic recognition of multi-word terms. Ph.D. thesis. Manchester Metropolitan University, Manchester

Gangemi A, Presutti V, Reforgiato RD (2014) Frame-based detection of opinion holders and topics: a model and a tool. Comput Intell Mag IEEE 9(1):20–30

Griffiths TL, Steyvers M, Blei DM, Tenenbaum JB (2004) Integrating topics and syntax. In: Advances in neural information processing systems. pp 537–544

Guellil I, Boukhalfa K (2015) Social big data mining: a survey focused on opinion mining and sentiments analysis. In: 2015 12th international symposium on programming and systems (ISPS). IEEE, New York, pp 1–10

Hai Z, Chang K, Kim J (2011) Implicit feature identification via co-occurrence association rule mining. In: Computational linguistics and intelligent text processing. Springer, Heidelberg, pp 393–404

Hai Z, Chang K, Kim J, Yang CC (2014) Identifying features in opinion mining via intrinsic and extrinsic domain relevance. Knowl Data Eng IEEE Trans 26(3):623–634

Hassan S, Mihalcea R, Banea C (2007) Random walk term weighting for improved text classification. Int J Semant Comput 1(04):421–439

Herbert JP, Yao J (2011) Game-theoretic rough sets. Fundamenta Informaticae 108(3–4):267–286

Hofmann T (1999) Probabilistic latent semantic indexing. In: Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval. ACM, New York, pp 50–57

Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, New York, pp 168–177

Javarone MA, Armano G (2013) Perception of similarity: a model for social network dynamics. J Phys A Math Theor 46(45):455102

Katz G, Elovici Y, Shapira B (2014) Coban: a context based model for data leakage prevention. Inf Sci 262:137–158

Katz G, Ofek N (2015) Consent. Knowl Based Syst 84(C):162–178

Khan MT, Durrani M, Ali A, Inayat I, Khalid S, Khan KH (2016) Sentiment analysis and the complex natural language. Complex Adaptive Syst Model 4(1):1–19

Khan MT, Khalid S (2015) Sentiment analysis for health care. Int J Priv Health Inf Manag (IJPHIM) 3(2):78–91

Kohonen T, Somervuo P (1998) Self-organizing maps of symbol strings. Neurocomputing 21(1):19–30

Li H, Zhou X, Huang B, Liu D (2013a) Cost-sensitive three-way decision: a sequential strategy. In: Rough sets and knowledge technology. Springer, Heidelberg, pp 325–337

Li J, Deng X, Yao Y (2013b) Multistage email spam filtering based on three-way decisions. In: Rough sets and knowledge technology. Springer, Heidelberg, pp 313–324

Liu B, Hsu W, Ma Y (1999) Mining association rules with multiple minimum supports. In: Proceedings of the fifth ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 337–341

Liu Y, Pan L, Jia X, Wang C, Xie J (2013) Three-way decision based overlapping community detection. In: Rough sets and knowledge technology. Springer, Heidelberg, pp 279–290

Lops P, De Gemmis M, Semeraro G (2011) Content-based recommender systems: state of the art and trends. In: Recommender systems handbook. Springer, Heidelberg, pp 73–105

Machova K, Marhefka L (2014) Opinion classification in conversational content using n-grams. In: Recent developments in computational collective intelligence. Springer, Heidelberg, pp 177–186

Mahmoud H (2008) Pólya urn models. CRC Press, Boca Raton

Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J 5(4):1093–1113

Mei Q, Ling X, Wondra M, Su H, Zhai C (2007) Topic sentiment mixture: modeling facets and opinions in weblogs. In: Proceedings of the 16th international conference on World Wide Web. ACM, New York, pp 171–180

Mimno D, Wallach HM, Talley E, Leenders M, McCallum A (2011) Optimizing semantic coherence in topic models. In: Proceedings of the conference on empirical methods in Natural Language processing. Association for Computational Linguistics, pp 262–272

Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45(2):167–256

Nguyen DT, Hwang D, Jung JJ (2015) Time-frequency social data analytics for understanding social big data. In: Intelligent distributed computing VIII. Springer, Heidelberg, pp 223–228
Nouman A, Jingtao Y (2014) Analyzing uncertainties of probabilistic rough set regions with game-theoretic rough sets. Int J Approx Reason 55(1):142–155
Ofek N, Rokach L (2014) Methodology for connecting nouns to their modifying adjectives. In: Computational linguistics and intelligent text processing. Springer, Heidelberg, pp 271–284
Pawlak Z (1982) Rough sets. Int J Comput Inf Sci 11(5):341–356
Popescu A, Etzioni O (2007) Extracting product features and opinions from reviews. In: Natural language processing and text mining. Springer, Heidelberg, pp 9–28
Poria S, Gelbukh A, Hussain A, Howard N, Das D, Bandyopadhyay S (2013) Enhanced senticnet with affective labels for concept-based opinion mining. IEEE Intell Syst 2:31–38
Qiu G, Liu B, Bu J, Chen C (2011) Opinion word expansion and target extraction through double propagation. Comput linguist 37(1):9–27
Rabade R, Mishra N, Sharma S (2014) Survey of influential user identification techniques in online social networks. In: Recent advances in intelligent informatics. Springer, Heidelberg, pp 359–370
Réka A, Albert-László B (2002) Statistical mechanics of complex networks. Rev Modern Phys 74(1):47
Ricci F, Rokach L, Shapira B (2011) Introduction to recommender systems handbook. Springer, Heidelberg
Schouten K, Frasincar F (2016) Survey on aspect-level sentiment analysis. IEEE Trans Knowl Data Eng 28:813–830
Śle D, Ziarko W et al (2005) The investigation of the bayesian rough set model. Int J Approx Reason 40(1):81–91
Su F, Markert K (2008) From words to senses: a case study of subjectivity recognition. In: Proceedings of the 22nd international conference on computational linguistics, vol 1. Association for Computational Linguistics, pp 825–832
Su W, Ziou D, Bouguila N (2013) A hierarchical statistical framework for the extraction of semantically related words in textual documents. In: Rough sets and knowledge technology. Springer, Heidelberg, pp 354–363
Takeuchi H, Yamaguchi T (2014) Text mining of business-oriented conversations at a call center. In: Data mining for service. Springer, Heidelberg, pp 111–129
Tang J, Chang Y, Liu H (2014) Mining social media with social theories: a survey. ACM SIGKDD Explor Newsl 15(2):20–29
Tang J, Xia H, Liu H (2013) Social recommendation: a review. Soc Netw Anal Min 3(4):1113–1133
Titov I, McDonald R (2008) Modeling online reviews with multi-grain topic models. In: Proceedings of the 17th international conference on World Wide Web. ACM, New York, pp 111–120
Titov I, McDonald RT (2008) A joint model of text and aspect ratings for sentiment summarization. In: ACL, vol 8. Citeseer, pp 308–316
Tuveri F, Angioni M (2014) An opinion mining model for generic domains. In: Distributed systems and applications of information filtering and retrieval. Springer, Heidelberg, pp 51–64
Weichselbraun A, Gindl S, Scharl A (2014) Enriching semantic knowledge bases for opinion mining in big data applications. Knowl Based Syst 69:78–85
Wojciech Z (1993) Variable precision rough set model. J Comput Syst Sci 46(1):39–59
Wu C, Tsai RT, Hsu JY (2013) Building a concept-level sentiment dictionary based on commonsense knowledge. IEEE Intell Syst 2:22–30
Xia R, Zong C, Hu X, Cambria E (2013) Feature ensemble plus sample selection: domain adaptation for sentiment classification. Intell Syst IEEE 28(3):10–18
Yao Y (2007) Decision-theoretic rough set models. In: Rough sets and knowledge technology. Springer, Heidelberg, pp 1–12
Yao Y (2010) Three-way decisions with probabilistic rough sets. Inf Sci 180(3):341–353
Yao Y (2013) Granular computing and sequential three-way decisions. In: Rough sets and knowledge technology. Springer, Heidelberg, pp 16–27
Yu H, Zhou Q (2013) A cluster ensemble framework based on three-way decisions. In: Rough sets and knowledge technology. Springer, Heidelberg, pp 302–312
Zhang L (2014) Data mining and knowledge discovery for big data. In: Aspect and entity extraction for opinion mining. Springer, Heidelberg, pp 1–40
Zhao WX, Jiang J, Yan H, Li X (2010) Jointly modeling aspects and opinions with a maxent-lda hybrid. In: Proceedings of the 2010 conference on empirical methods in Natural Language processing. Association for Computational Linguistics, pp 56–65