Terrorist Threat Assessment with Formal Concept Analysis

Paul Elzinga

Department of Business Information
Police Amsterdam-Amstelland
Amsterdam, The Netherlands
Paul.Elzinga@amsterdam.politie.nl

Jonas Poelmans

Faculty of Business and Economics
Katholieke Universiteit Leuven
Leuven, Belgium
Jonas.Poelmans@econ.kuleuven.be

Stijn Viaene

Vlerick Leuven Gent Management School
Leuven, Belgium
Stijn.Viaene@vlerick.be

Guido Dedene

Faculty of Business and Economics
Katholieke Universiteit Leuven
Leuven, Belgium
Guido.Dedene@econ.kuleuven.be

Shanti Morsing

Department IPOL
National Police Service Agency
Zoetermeer, The Netherlands
Shanti.Morsing@klpd.politie.nl

Abstract— **The National Police Service Agency of the Netherlands developed a model to classify (potential) jihadists in four sequential phases of radicalism. The goal of the model is to signal the potential jihadist as early as possible to prevent him or her to enter the next phase. This model has up till now, never been used to actively find new subjects. In this paper, we use Formal Concept Analysis to extract and visualize potential jihadists in the different phases of radicalism from a large set of reports describing police observations. We employ Temporal Concept Analysis to visualize how a possible jihadist radicalizes over time. The combination of these instruments allows for easy decision-making on where and when to act.**

Keywords-component; Formal Concept Analysis, Temporal Concept Analysis, Contextual Attribute Logic, Text mining, Terrorist threat assesment.

I.   INTRODUCTION

In the modern day globalized world, the ease of terrorist network information exchange is characterized by contact moments through the internet and an absence of time and location restrictions. The amount of information available to police forces is continuously increasing and many police forces are not ready for handling data amounts of this size. As a consequence, pro-actively observing potential threats to our national security becomes increasingly difficult. The National Police Service Agency (KLPD) of the Netherlands started a new Intelligence Led Policing (ILP) project with the aim of collecting terrorist-related information in visually appealing models (Knowledge in Models, KiM, project) to ease the extraction and sharing of actionable knowledge. The KiM project is part of the Program Improvement by Information Security Awareness (VIA). This program is a partnership between the National Coordinator of Counterterrorism (NCTb), the National Forensic Institute (NFI), the General Intelligence and Security Service (AIVD) and the KLPD. Shortly described, the program includes research on and implementation of methods and techniques for supporting police services in their fight against terrorism.

One of the results of this project was the development of a model describing the radicalization process a potential jihadist may pass through before committing attacks. This model consists of 4 phases and its feasibility and practical usefulness have been validated by members of the intelligence services on known suspects. After the validation of this model on known jihadists, the next and probably most important, step consists of finding unknown potential suspects by applying the model to the large amounts of structured and unstructured text available in the police databases.

In this paper, we make use of the techniques known as Formal Concept Analysis (FCA) [4-6] and Temporal Concept Analysis (TCA) [7,8]. Contextual attribute logic [13] is used to group and transform the terrorism indicators available into new attributes for generating the concept lattices. After extracting the potential suspects for each phase of the model, a detailed profile based on TCA is constructed giving the history of the suspect and his current level of threat to national security.

The remainder of this paper is composed as follows. In section 2, we give some background on Jihadism in the Netherlands and the four phase model of radicalism used by the KLPD. In section 3, we elaborate on the dataset used. In section 4, the essentials of FCA and TCA theories are introduced. In section 5, the research methodology is explained and the results of the analysis are presented. Finally, section 6 concludes the paper.

II. BACKGROUNDER

A. Home-grown terrorism

In November 2004 the Dutch society was confronted for the first time with an act of terrorism, namely the brute murder of the Dutch film maker Theo van Gogh. The people suddenly realized that the ideology of violent jihad against the West had also established a foothold in the Netherlands and that the Netherlands as well had become a scene of terrorist violence. It ensued that the murderer, and most other members of the extremist network to which he belonged, were young Muslims born and bred in the Netherlands [2,3].

The latter fact has been seen as a confirmation of the new phase in Islamist terrorism, the phase in which the threat emanates principally from extremist European Muslims who are prepared to commit attacks in their own country, also known as the *European jihad*. The AIVD formulated four general trends in the development of jihadism [1]. The first and most important is the evolvement from exogenous foreign terrorist threat to indigenous *home-grown* terrorism. This threat has led to the project VIA, Information Security Awareness coordinated by the National Coordinator of Counterterrorism (NCTb). One of the results of this project is the development of a four phase model of Muslim radicalization by the National Police Service Agency. This model will be discussed in detail in the next section.

B. The four phase model of radicalism

The four phase model of radicalism, displayed in figure 1, developed by the National Police Service Agency, is based on the idea a jihadist might pass through several phases before he or she might commit serious acts of terrorism. Several indicators (i.e. words and/or sentences) are associated with each phase which are used to decide based on automated text analysis, to which phase a subject belongs. Due to National Security reasons the indicators can not be published. Interested and authorized intelligence services can contact the National Police Service Agency (www.politie.nl/klpd). An exception is made for the indicator "change of behavior" from type 2-A. Some of the keywords belonging to this indicator are the phrases "not shaking hands with women", "wearing traditional clothes", "suddenly let grow a beard" and "Islamistic marriage".

The model should be interpreted in a bottom up fashion. If 4 or more indicators of type 1 become true or 2 or more of type 2-A, than the subject enters the preliminary phase. But if the number of type 2-A comes below 2 or the number of type 1 comes below 4, then the subject will leave the preliminary phase. If 5 or more indicators of type 1 becomes true and 6 of type 2-B then the subject will enter the Social alienation phase, etc.
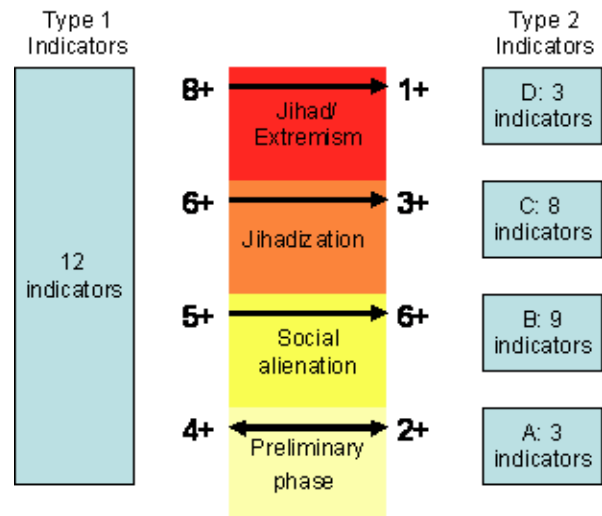

Figure 1. The four phase model of radicalism

In the preliminary phase the subject experiences a crisis of confidence; confidence in the authorities is undermined. At this point it is not so much a matter of an ideological rift, but certainly of distrust. Many young Muslims, especially those who have grown up in the West, turn to Islam in search of their identity and place in Western society. Often their parents still live in accordance with the strict traditional norms and values of Islam, which the young people can no longer relate to. They seek a new place for themselves within Dutch society, where their ethnic, religious and national identity can be a balanced part of the whole.

In the social alienation phase a small minority of these young Muslims cannot handle this situation. On the one hand, the first generation of migrants look down on them for becoming too 'Dutchified'; on the other hand, they do not really fit in with their Dutch peers because they are viewed in terms of their origins. This is generally where a shift occurs from the desire for a national identity to the desire for a religious identity. Strict Islam, as a guiding principle in their lives, provides certainty and stability because it tells them precisely what to do and what not to do. This increases their susceptibility for the ideology of strict, extremist religion, and makes them feel alienated from the rest of society. This alienation finds expression in increasing rejection of Dutch society.

In the Jihadization phase the subjects are characterized by strong radical Islamic convictions and the fact that they condone violence. Strong involvement in a radical group may ultimately lead to a willingness to support terrorists in the Netherlands or elsewhere in the world. This may include all sorts of support (e.g. funding). This phase entails further alienation from society and an even greater readiness to make an active contribution to the Jihad. The subjects' firm belief in the rightness of their radical ideology and of radical Islam may lead to recruitment activities to convince others of radical Islamic beliefs and possibly also of the necessity of the Jihad. Isolation from the rest of the world is part of a gradual process.

The last phase, Jihad/Extremism, is a phase of total isolation. The subjects' entire lives are governed by their radical Islamic beliefs. This is the last step before carrying out Islamist terrorist acts. In this phase, subjects are prepared to use violence themselves to achieve their objectives. In most cases, the definitive preparation for perpetrating an attack takes the form of physical training, often at a training camp abroad. The final step is actually carrying out violent activities.

C. Current situation

All police forces in the Netherlands (25 police regions and the National Police Service Agency) have a monitoring task of collecting information about potential jihadists.

Due to the nature of their activities, potential jihadists will avoid contact with the police and other legal authorities as much as possible. The consequence is that finding new potential jihadists is like finding a needle in the haystack. Attempts were made to search the national police database BlueView containing over 50 million documents. Unfortunately this turned out to be a laborious task .

The four phase model is not used yet as an instrument for finding new potential jihadists from large datasets, but as a checklist. To apply the model on large datasets, the KLPD has started a cooperation with the police Amsterdam-Amstelland who is investigating intelligent text mining applications, like the classification system for domestic violence [10]. This application has been used to evaluate the first version of the four phase model on the police dataset of Amsterdam. The results of this investigation has led to fine tuning the conditions imposed by the model to maximize the recall and to find as many potential jihadists as soon as possible.

III. DATASET

Our dataset consists of 166577 general police reports from the years 2006 (41990), 2007 (54799) and 2008 (69788) from the region Amsterdam-Amstelland, which holds the communities Amsterdam, Amstelveen, Uithoorn, OuderAmstel and Diemen. These general reports contain observations made by police officers during motor vehicle inspections, during a police patrol, when a known subject was seen at a certain place, etc. Next to general reports there are incident reports like car accidents, burglary, violence cases, etc. There are two reasons why we have chosen for analyzing the general reports. Since the implementation of an Intelligence Led Policing program at the Amsterdam-Amstelland police, the number of general reports has been growing rapidly over the years. The unstructured text describing the observations made by police officers has a lot of underexploited potential for finding potential extremists or radicalizing subjects. The challenge is to find new potential jihadists within the huge amount of general reports.

An example of a report is displayed in Figure 2 where two police officers asked two repeat offenders information about a third subject, called C. The reason of the inquiries is that the officers might have an indication that C. might be a recruiter.
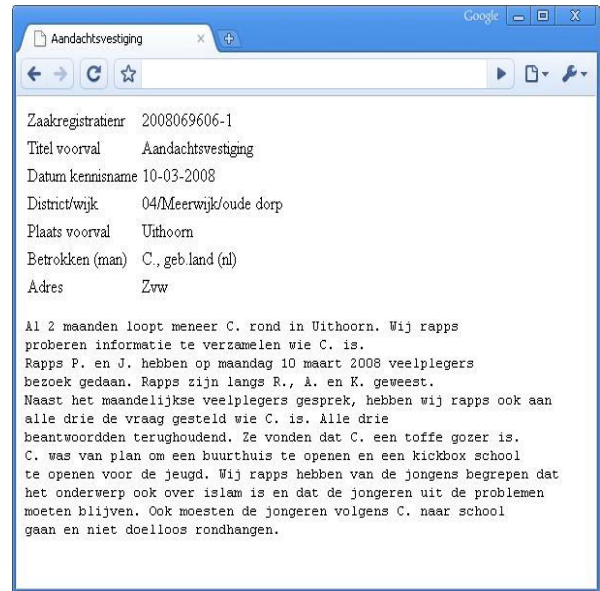

Figure 2. Example police report

IV. FORMAL CONCEPT ANALYSIS

FCA arose twenty-five years ago as a mathematical theory [4]. It has over the years grown into a powerful framework for data analysis, data visualization and information retrieval [4-6]. FCA is particularly suited for exploratory data analysis because of its human-centeredness and can be used as an unsupervised clustering technique [11,12]. Police reports containing terms from the same term clusters are grouped in concepts.

The starting point of the analysis is a database table consisting of rows M (i.e. objects), columns F (i.e. attributes) and crosses $T \subseteq M \times F$ (i.e. relationships between objects and attributes). The mathematical structure used to represent such a cross table is called a formal context (T, M, F). FCA concept lattices are derived from this table and used to describe the conceptual structures inherent in these data tables without loss of information by means of line diagrams yielding valuable visualizations of real data [9]. In a previous paper, we analysed the concept of domestic violence, where FCA was used to support human actors in their information discovery and knowledge exercise [10]. The main difference with domestic violence are the possible combinations and the frequencies of the indicators which trigger a subject in a phase together with the time dimension describing the subjects radicalization process.

A. Temporal Concept Analysis

Temporal Concept Analysis (TCA) is an extension of traditional FCA that was introduced in scientific literature about nine years ago [7,8]. TCA addresses the problem of conceptually representing time and is particularly suited for the visual representation of discrete temporal phenomena.

The pivotal notion of TCA theory is that of a conceptual time system [8]. In the visualization of the data, we express the "natural temporal ordering" of the observations using a time

relation *R* on the set *G* of time granules of a conceptual time system. We also use the notions of transitions and life tracks. The basic idea of a transition is a "step from one point to another" and a life track is a sequence of transitions.

V. RESEARCH METHOD

The method we propose is summarized in figure 4. First, we extract all subjects who have at least one attribute from the large set of observations with FCA. Second, we construct lattices for each phase of jihadism. Third, we use TCA to profile the selected subjects and their evolution over time.

We used a toolset which was developed for text mining in large sets of documents, extracting entities from these sets and generating cross tables in various formats. It has been used to develop and to apply knowledge models for amongst others detecting domestic violence [10]. The toolset uses a thesaurus where indicators can be defined with specific properties. For the purpose of this investigation, the thesaurus and toolset are extended with the property of range of numbers of different occurrences of an indicator which must true.

A. Extracting potential jihadists with FCA

For detecting potential jihadists from the large amount of observations, we make use of an FCA lattice. The subjects mentioned in the reports are the objects of the lattice. The indicators observed during the observations for these subjects are combined in one feature vector. This results in an FCA lattice as displayed in figure 3.
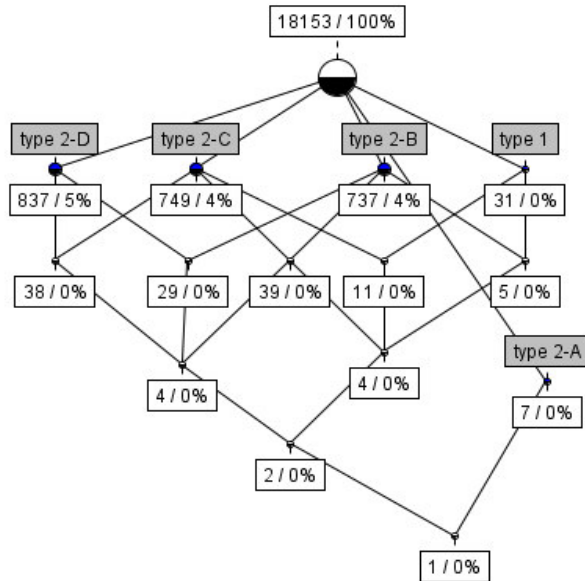
Figure 3. FCA lattice used for extracting potential jihadists

Out of 166578 documents 18153 subjects are selected into the cross table of FCA. Each of the subjects meets at least one of the 35 indicators of the original model. These indicators are grouped together based on the four phase model of the KLPD.

TABLE I. RESULTS OF EXTRACTION OF SUBJECTS

Attribute	# of subjects applied to
1	31
2-A	7
2-B	737
2-C	749
2-D	837

Table 1 shows the number of subjects who meet the requirements of the attributes of the four phase model. In the next section we will showcase how the combinations of type 1 and type 2 indicators will reveal the subjects in the different phases. One of the advantages of FCA theory is the ability to zoom in and out on the data and to create smaller lattices by amongst others deselecting the attributes from the main lattice.

B. Constructing Jihadism phases with FCA

The next step consists of constructing a lattice for each phase of jihadism and showing subjects. The FCA lattice serves as an intuitive knowledge browser making the interaction between the police officer and data more efficient. Based on this lattice, police officers can easily extract subjects for in-depth investigation. Figure 4 shows the process model of finding potential jihadists. The four lattices will be discussed from left to right.

The first lattice shows the preliminary phase where 38 subjects are detected. An in depth search after these 38 subjects revealed that 19 were highlighted correctly. The other 19 subjects were mostly persons of the domestic sphere of the subject and therefore frequently reported in the same documents with the potential jihadists. Out of the 19 correctly highlighted subjects, 3 were previously unknown by the police Amsterdam-Amstelland, but known by the National Police Agency Service. The second lattice shows 5 subjects for the social alienation phase which were all highlighted correctly. The third lattice shows 11 subjects for the Jihadization phase where 8 subjects are highlighted correctly. The fourth and last lattice shows 2 potential jihadists, who both are highlighted correctly.

C. Build detailed TCA lattice profiles for subjects

To show how the selected subject radicalizes over time a TCA lattice is constructed. C. from the example report is a subject who satisfies the conditions of all phases. Figure 4 shows the TCA lattice of C.

We clearly see his radicalization process over time in action (black arrow). There were 8 observations of C. that did not trigger sufficient conditions for entering one of the four terrorism threat phases. In 29/9/2006, C. for the first time appeared under the preliminary phase and 13 months later again he was observed and again fulfilled the requirements of the preliminary phase. 5 months later, C. for the first time had all the properties of the social alienation phase and climbed from the fourth to the third phase of alert. Afterwards he was categorized 6 times under the second phase of alert: jihadism. In 17/6/2008 he reached the highest point of alert: Jihad extremism (red oval). Afterwards he was spotted 2 times by the police, once in the Jihadism phase and once outside any phases (2 arrows).
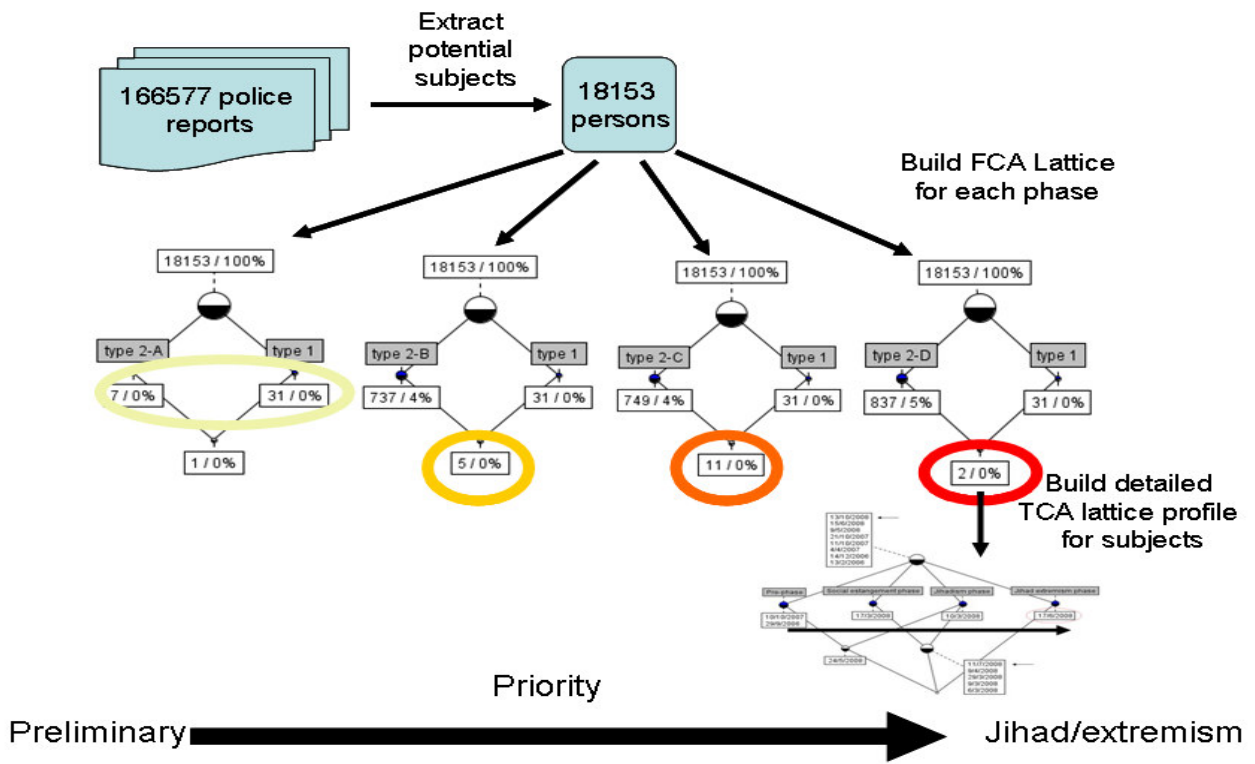
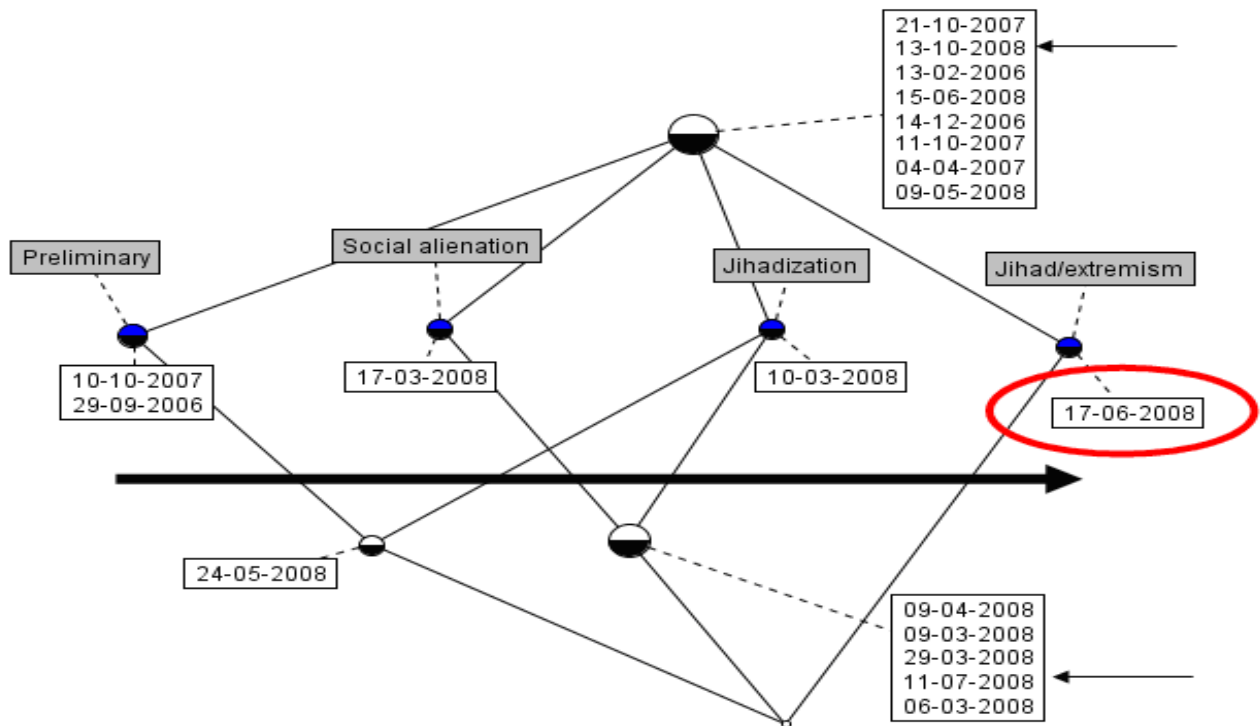Figure 4.   The process model of extracting and profiling potential jihadists

Figure 5.   TCA lattice for subject C.

VI. CONCLUSIONS

In this paper, we showed that the combination of techniques known as Formal Concept Analysis and Temporal Concept Analysis provides the user with a powerful method for identifying and profiling potential jihadists. We built a set of attributes based on the original knowledge model of radicalism which is used when searching the police reports. Out of 166577 police reports we distilled and visualized 38 potential jihadism suspects using FCA. TCA is used to analyze the radicalisation over time of the potential jihadists.

Avenues for future research include the embedding of this sandbox discovery model into operational policing practice and applying.