Classifying and Aggregating Context Attributes for Business Service Requests - No 'One-Size-Fits-All'

Rebecca Copeland, Noel Crespi
Institute Mines-Telecom, Telecom SudParis, Paris, France

Abstract— When building decision-making models from disparate observations, there are no set rules to guide the designer on how to organize available information, how to classify vital aspects, how to emphasize important ones in the aggregation process, and how to deal with conflict and uncertainty in the aggregation procedures. This paper draws on the experience of structuring a business and risk model that evaluates service requests, which requires not only dynamic context-based decisions, but also situational and behavioral perspectives, with high uncertainty and wide variations of attribute styles. This study focuses on design issues that affect classification and aggregation options, such as corroboration, primacy and discord, and provides examples of classified key-factors that demonstrate the design issues. The paper suggests procedures and algorithms to fit the design, but shows that there is no universal method - there is no '*one-size-fits-all*'.

Keywords- Classification, Credibility, Discordant, Aggregation, Corroboration, CART, OWA, WPM, SAW, MCDM, DST, policy

I. INTRODUCTION

The ability to infer context for service requests in real-time is highly desirable for increasing number of mobile business applications. Such requests can be for person-to-person communications or for access to corporate services or Internet browsing. Business context enables organizations to decide how to prioritize delivery of services, protect their resources, and enhance the security of sensitive applications and data. If enterprises could determine business-only status to distinguish from personal use, they would change the priority for requests, cut short non-essential long-duration media, or opt for changing access modes.

Where the context is complex, there is a need for guidance in selecting attributes and sources and choosing their classification and procedures. The key factors include commonplace context characteristics that are ascertained from the physical environment (location, time), and from the digital information (media type, network type, destination), but the situational aspects (Associated-Activity, Urgency, Integrity) are less obvious and harder to establish, yet they may determine the outcome. Therefore, the choice of modelled aspects has to be guided by the designed suppositions and assertions.

Model design must consider many aspects: atomization, classification, attribute association, credibility calibration, modes of corroboration and more. Models are often portrayed as if they have common procedures and uniform algorithms, but in reality, multiple methods are required. Design influences the selection of data sources too, so that assertions are adequately supported. Too many sources raise the cost of maintaining reliable input, but too few sources risks missing out on important observations. The classification choice also influences processing: too many classes produce complex output, but too few may gloss over important aspects. Element granularity should be balanced, where flat structures are simpler but lack accuracy, while deeply hierarchical ones are well defined but perplexing to manage.

This paper explores how design issues affect modelling context and proposes classification modes and aggregation methods that deal with the main issues. In Section II, function-based design is described. In Section III, granularity is considered. In IV, primacy and precedence issues are demonstrated. In V, corroboration and correlation, and in VI discordance and uncertainty are discussed. In VII, classification issues are summarized, in VIII related work is detailed, and in IX - the conclusions.

II. FUNCTION-LED DESIGN

A. *Suppositions and Assertions*

Models are conceived with a particular function in mind, which results in specific decisions. For example, determining if the service request is for business or leisure in order to grade business priority; or to establish what level of resource usage should be granted; or to detect and grade a level of risk that merits a defensive action. Such functions must influence the model design at all levels. These functions have been described in the author's earlier studies: selecting quality and priority levels [29]; protecting resources from excessive personal usage [19,18,29]; determining business status and level of funding [19]; and optimizing the choice of access network [18].

To support the overall function, the classification structure of key-factors (classes) and attributes (atomic elements) must convey the logical function-based construct that defines 'suppositions' and 'assertions'. An attribute feature is 'asserted' when an observation is matched with filtering tables or qualifying data. By linking corresponding suppositions, assertions and qualifiers, a function-specific logic is achieved.

The choice of suppositions is closely related to the function, e.g. to determine business priority, positioning of the requesting devices are inferred as business locations or not. Choosing assertions depends on what should be tracked and prioritized by the function. While good coverage is important, excessive coverage leads to overlapping and dependencies, so the right balance has to be struck. Observations should provide sufficient coverage of any aspect that can support or conflict with the attribute. In Table I, examples of suppositions and assertions are shown.

Table I: Examples of Context Suppositions/Assertions			
	Key-Factors	***Suppositions Examples***	***Assertions Examples***
Environ	Space	Mostly on-site or on POI business location	Normal place of work Home working Partner premises
Environ	Time	During work-time, including recognized after-hours	Normal hours Busy-hour /congestion Not sick, not on holiday
Circumstantial	Activity	Concurrent/recent business-like activities	Corp. apps Scheduled duty activity Browse approved sites
Circumstantial	Urgency	Indications of urgency or emergency, critical jobs	On-duty officer Emergency event/alarms To critical system
Circumstantial	Integrity	No risky behavior, no confidentiality or data rights infringements	Undesirable Destination Implausible Location Sensitive Data/Apps
Digital	Destination	Known, appropriate or approved destinations	Internal/External P2P Corp. apps Cloud, corporate websites
Digital	Network	Suitable access mode, appropriate network usage	WLAN or Home BB 3G/4G/mobile, Roaming Hotspot
Digital	Media	Appropriate media type for the request, not unduly demanding	Special needs Interactive Voice/Video Browsing

Several logical functions can be satisfied by one model if the construct of suppositions-assertions-qualifiers is overlaid. This is achieved by the combination of customizable filtering of the same data, selectable attribute assertions, and alternative policy-based prioritization. Observations that are gathered from pre-allocated sources are qualified to fit specific assertions. The asserted attributes 'feed' a layer of that support function-driven suppositions. The 'key-factor' aspects classify attributes that support particular suppositions. This provides a layered structure of sources, attributes and key-factors that is overlaid by function-specific qualifiers (inferences), assertions (claims) and suppositions (deductions), as in *Figure 1*.

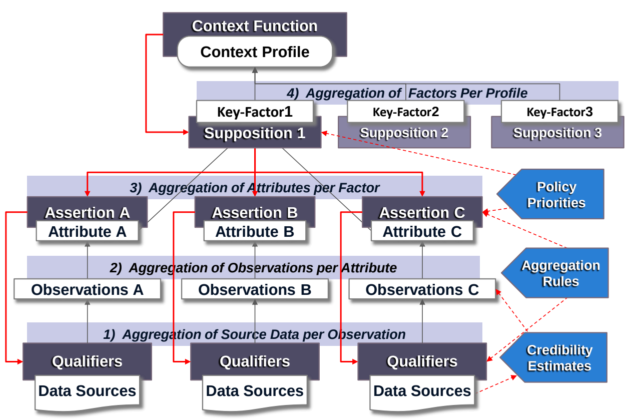

Figure 1. Assertions and Suppositions Relating to attributes and factors

B. Sources and Qualifiers

Finding the right sources that can bring important insights is vital for precise and conclusive decisions. To minimize costs, only necessary sources should be used. On the other hand, adequate alternative backup must be made available for continuous operation. In addition to data sources that discover basic facts (the 'instigators'), e.g. GPS or apps login, the qualifiers (e.g. tagged locations) are required to impute meanings. Function-specific qualifiers can infer different meaning for the same source data, e.g. interpreting the request timing as busy hour request (for optimal resourcing) compared with business hours (for business priority).

C. Classification Groups

Among the various generic classes that appear in studied communication modelling, the most often listed are spatial, temporal and destination-based characteristics. Aspects such as media and access network type are less common and situational or behavioral aspects are only chosen in relation to particular applications. *When and where* a service is requested are provided by *Environmental* factors, i.e. spatial and temporal aspects; *What and how* the session is to be delivered are derived from *Digital* key-factors, i.e. request details such as destination, network access type and media type; *Why* the service is requested is gleaned by C*ircumstance* key-factors, which represent the user's situation and behavior. They are inferred from associated activities, urgency indications and integrity risks, with information drawn obliquely from observations of current as well as recent history. Hence, circumstance key-factors are harder to establish and contain more design issues, such as discord and primacy, but they provide valuable insights for behavioral context models.

D. Profiles and Policies

Classification can be made by the type of *source input* (e.g. environment, circumstance and digital facets) and by the *consequences of the output* (e.g. 'high impact', 'time-critical' or 'essential-activity'). While most models are structured by hierarchical design, a *multi-dimensional classification* can provide further insights, relating attributes across all key-factors. This dimensional classification enables emphasizing attributes' contribution from a different viewpoint, and is used for alternative policy-prioritization.

Such cross-dimensioning is function-specific, and can also be used for 'profiling' scenarios or behavior patterns. Profiling involves analyzing the same data by alternative assertions and inferences, especially assisted by historical data analytics. For example, profiles for a Risk Model, such as '*Suspicious*', '*Intrusive*', '*Abusive*' and '*Destructive*', can be ascertained from inferred risks that are linked to the same key-factors and attributes as the business-priority function. These profiles describe common risk scenarios that are associated with defensive actions. Each profile is customized by prioritization rates of the factors and attributes as well as risks. The model computes the best-fit profile for each service request, which determines what defensive actions should be taken. At the same time, the original model, which is aggregated by the normal business key-factors and attributes (without risks), provides the best-fit business context profile. The scores of these profiles are used to combine business status with risk context, and selecting a service delivery option is governed by both.

III. GRANULARITY OF CLASSIFICATION

A. Hierarchy

Granularity level is a design issues for any model. A flat hierarchy with a handful of key-factors is easier to manage, but involves more complex elements. It is possible to increase granularity by creating an attribute for each level of 'strength', so that they are all binary (true/false). However, where there is high uncertainty, it will increase attribute numbers dramatically. Although it is desirable to build independent elements, designed as CART [2] structures, excessive splitting of members makes it "impossible to see the wood for the trees". The bottom-up approach in [26] is preferable, relating attribute granulation to their data sources. This approach is preferable, not only for avoiding excessively fragmented structures, but also because it forges strong links between the source data and the analyzed assertions.

The same granularity arguments also govern the selection of classes. Management of deep hierarchies is complex, but incorporating too many different attributes in one class hides some important features. Defining suppositions for each key-factor helps to distinguish them and confirm that the contents should be in separate classes. The need for individual prioritization is a clear indication that a separate class is required. Therefore, in the case granularity of the main key-factors, a top-down method is still required, in order to closely associate them with the model's function and suppositions.

B. *Intensity and Credibility*

Observations may be designed as binary elements, with true/false states, or as elements with a range of possible values - i.e. *Level of Intensity* (*LoI*), which is observed or estimated for each service request. This concept is not the same as the strength of the belief that the attribute assertion is proven. Hence, it is important to distinguish between plausibility, as defined in the Dempster Shafer Theory [22], and attribute Intensity. Intensity is measured according to the observed source and is interpreted by indices, e.g. spatial proximity to tagged locations that strengthens the assertion of location, and temporal proximity to recent activities, when the impact of past activities fades with the passage of time.

An observation (OB_b) becomes active when the information from the data source is matched by a Qualifier (Q_{xy}), with x lists and y items each. Binary attributes have LoI_b =1 when active, and LoI_b =0 when they are not. Depending on the type of observations and whether Intensity can be measured, the Level of Intensity (LoI_b) is gauged for the observed assertion, not per source. The measurements are converted into a numeric scale, by fuzzification of both subjective estimates and instrumental measurements, to provide a uniform unit-less numeric index.

An observation usually requires more than one data source, and the sources can have impact weights (Sw_s), to distinguish '*instigators*' from '*qualifiers*'. With the credibility approach, as in [28], the combined credibility of sources and qualifiers constitutes the observation's *inherent* trustiness value. The final observation score is the product of the Level of Intensity and the aggregated Source Credibility $Scred_b$, as in (1).

$$1 \le b \le b',\ \ 1 \le s \le s',\ Scred_{sb} = \in [1,0] \qquad \textit{Observations}\ (1)$$
$$OB_b status = \begin{cases} if source_{sb} = Q_{xy} & LoI_b = 1 \\ else & LoI_b = 0 \end{cases} \qquad \textit{Binary Status}$$
$$LoI_b = \in [1,0] \qquad \textit{Non-binary Observation Intensity}$$
$$OB_b = \left(\sum_{s=1}^{s=s'} Scred_{sb} Sw_s\right) \cdot LoI_b \qquad \textit{Observation Score}$$

The aggregation of source credibility is using SAW (Simple Weighted Aggregation) [11], which allows weighting of different sources. Qualifiers' credibility is also added, but its impact is deemed lower than the credibility of instigating data sources that discover new facts. Credibility for a data source is calculated from estimated properties, such as reliability, precision, accuracy and data integrity, as explained in [28].

C. *Nested Sub-classes - the Urgency Class*

The granularity of classification sometimes dictates that a sub-class is inserted, despite the general aim to keep the hierarchy as flat as possible. Sub-classification is justified by the need to apply different aggregation method or different business policy weighting. The example of the Urgency Key-Factor (*UKF*) in Figure 2 demonstrates a case where the class aggregation is cumulative, because the key-factor supposition is built by the corroboration of multiple signs of urgency (e.g. emergency events or alarms) that increase the overall state of urgency. One such sign is an urgent destination, when the request is destined to an emergency service, on-duty officer or a critical application. Only one destination can be valid per service request, so an operation that selects a single destination is performed in a separate sub-class, and the result of that is then aggregated as a cumulative member of the top set.

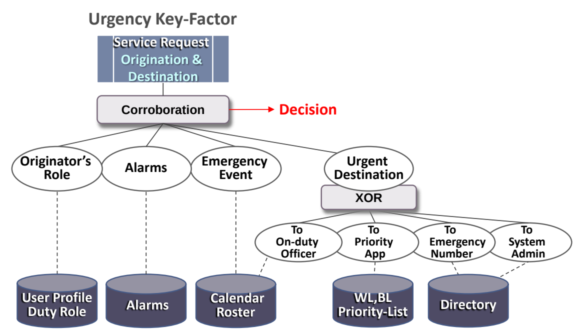

Figure 2. Urgency factor cumulative +XOR

The destination (UD_d) is selected in the sub-class as the maximum value and is considered as another contributory Urgency Attributes (UA_i). The cumulative operation needs an aggregation procedure that corroborates the key-factor's supposition that the request is urgent. Such corroboration needs to identify the main attributes (largest inherent value), and augment it with the supporting members' values, as in (2).

$$UD_{\hat{p}} = \max(UD_{d=1}, \dots UD_{d=d'}) \qquad \textit{Sub-Class \& Cedar}\ (2)$$
$$UKF = f\{UA_1, \dots (UD_{\hat{p}} = UA_a), \dots UA_{a'}\}, \qquad UA_a > UA_{a+1}$$
$$f(UA_a) = UA_{a-1} + UA_a(1 - UA_{a-1}) \qquad \textit{Recursive function}$$
$$UKF = \begin{cases} UKF = 0 & a = 0 \\ UKF = UA_{\hat{p}} & a = 1 \\ f(UA_{a-1}) + UA_a(1 - f(UA_{a-1})) & \forall (1 < a \le a') \end{cases}$$

This procedure utilizes the author's new algorithm (Cedar) that is subject to a further study. Cedar (Corroborative Evidential Diminishing Aggregation Rating) is a corroborative aggregation that increments the largest member in the set (the '*prime*') by lesser values, in proportion to their *inherent* values. The proportionality is achieved through a coefficient - the '*residual interval*', i.e. the remaining interval to the top limit of the scale, after taking out all previous contributions. The residual is = (1-previous), but since 'previous' also contains its previous, this function is recursive. The residual is reduced with each step in the proportions of the increments, so the impact of lower members is gradually diminishing.

D. Source-based Sub-classes - the Destination Class

It is useful to define a sub-class of attributes according to the type of sources, in order to streamline processing and minimize the number of sources that need to be accessed. This is demonstrated by the Destination Key-Factor (*DKF*), in Figure 3. There is no sense in splitting the key-factor into two, because there can only be one destination per service request. Instead, separate sub-classes are used to process one destination type only, P2P (Person to Person) or web URL. For example, if P2P is detected, it is qualified by the corporate telephone directory with no need to access Deep Packet Inspection, and the process of qualifying web assertions is skipped.

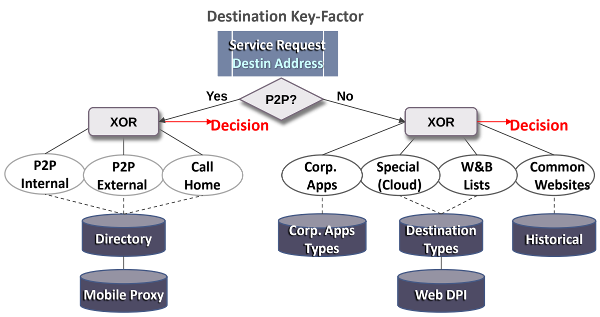

Figure 3. Destination sub-classification by source

The *DKF* is also an example of a class with attributes that are *mutually exclusive* (xor) as well as *mandatory* - i.e. at least one, and only one valid attribute must always exist, so *DKF* can never be zero (see summary in Table III). This example also shows that the same information (the destination) can be interpreted in two different classes, with different inferences, e.g. 'Internal P2P' in the Destination class may also be inferred as 'Duty Officer' in the Urgency class.

IV. PRIMACY AND PRECEDENCE PROCESSING

A. Precedence and Ranking

Key factors, such as Associated-Activities, Integrity or Urgency, which rely on cumulative effects, also need to be ranked where the higher rated members in the set have greater effect, and lesser members' impact is minimized. Most common ranking methods are the family of OWA (Ordinal Weighted Averaging) techniques, as discussed in ([6], [16], [17]). OWA computes the score for a Key-Factor (KF_k) from its ordered members attributes (A_a), as in (3).

$$0 < a \leq a' \qquad \textit{OWA Aggregation} \ (3)$$

$$A_a > A_{a+1} \quad where \ \frac{a}{a'} = \ OWA \ transposed \ ranking$$

$$KF_k = \sum_1^{a'} A_a \frac{a}{a'}$$

OWA ranking provide the desirable effect of gradually diminishing the impact of lesser contributors, but the ranking coefficients are the ranking numbers (1,2,3,4) which are not proportional to the contributions of each of the members. The OWA algorithms do not guarantee keeping to a range of 0-1, so further normalization is required. OWA is also inadequate in handling conflicting negative attributes. OWA results for just two members are erratic, due to the order-based coefficient. In Table II, comparison of bi-aggregation is given, showing the results of using SAW, OWA and WPM (Weighted Product Model) methods [15], and comparing them with Cedar.

In example 2, when aggregating a negative contributor Att2, OWA raises the score of Att1 instead of reducing it. In example 3, OWA doubles the score unreasonably with a very small contribution. In all the examples, WPM and Averaging do not augment the first attribute (the *prime*), but actually reduce it, even when there are no negative contributions. SAW aggregates scores linearly, causing the score to shrink almost to nothing in example 2. Hence, none of these methods demonstrate *proportional corroboration*, while Cedar produces the desired effects, incrementing and decrementing the scores reliably, according to the contributions.

Table II: Bi-Aggregation Comparison of Algorithms

	Input: Attributes and Weights				Aggregation Comparison				
	Att1	W1	Att2	W2	SAW	Ave	OWA	WPM	Cedar
1	0.788	0.890			0.701	0.701	0.701	0.809	0.701
	0.788	0.890	0.776	0.580	1.151	0.576	1.853	0.698	0.797
2	0.634	0.850			0.539	0.539	0.539	0.679	0.539
	0.634	0.850	-0.634	0.840	0.006	0.003	0.545	0.463	0.344
3	0.820	0.750			0.615	0.615	0.615	0.862	0.615
	0.820	0.750	0.055	0.210	0.627	0.313	1.242	0.469	0.617
4	0.634	0.850			0.539	0.539	0.539	0.679	0.539
	0.634	0.850	0.634	0.840	1.071	0.536	1.610	0.463	0.734

B. Primacy - the Spatial Class

The Spatial Key Factor (*SKF*) is an example of a class that requires determining a *prime* attribute (the highest scoring attribute). It is essential to establish an unambiguous spatial prime for clear decisions, e.g. 'Office' for on-site working policies that are different from policies for 'Home-working' or 'Roaming'. Any spatial attributes can be *prime*, or it can be *concordant* or *discordant*. Hence, spatial attributes are said to be *'role-swapping'*, and are not assigned as negative or positive by fixed design. Figure 4 shows the procedure that defines the prime, matching positioning readings with the qualifying tables (Q_{xy}). According to the selection of a prime, concordant attributes become positive and discordant attributes become negative. Discordant values decrease the prime's score, down to even near-zero, but one attribute must still be determined as the prime, to avoid indecision.

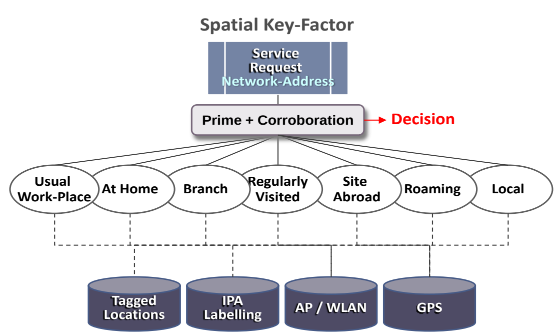

Figure 4. Spatial Key-Factor

To aggregate Spatial Attributes (*SA*), a variant of Cedar that is a combination of Cedar bi-aggregation with averaging is chosen, although full Cedar is also effective. With this aCedar (averaging Cedar), all contributions (both positive and negative) except the prime, are averaged in a single process, as in (4). The final value is always positive, because the prime is always greater than the average of lesser attributes. The advantage of aCedar is that it needs no ordering (only finding the prime as the maximum value) and no process iteration, while clearly securing the primacy of the prevailing attribute.

$$SA_{\hat{p}} = max(SA_1, \dots SA_a),\ SA_{\hat{p}} > 0 \qquad \textit{SA Prime, aCedar} \quad (4)$$

$$\begin{cases} SA_a & if\ \ Q_{xy}(SA_a) = Q_{xy}(SA_{\hat{p}}) \\ -1 \cdot SA_a & if\ \ Q_{xy}(SA_a) \neq Q_{xy}(SA_{\hat{p}}) \end{cases} \quad \textit{Set negative attributes}$$

$$SKF = SA_{\hat{p}} + \left(1 - SA_{\hat{p}}\right)\frac{1}{a-1}\left(\sum_{a=1, i\neq\hat{p}}^{a=a'} SA_a\right) \quad \textit{Residual pro-rata}$$

C. Sequential Logic - the Temporal Class

In rating-based aggregation, precedence determines the level of impact that is given to the prime and the succeeding members. Precedence order also determines which attributes are eliminated, where the logic tests attributes in a sequence of conditions. Temporal attribute processing involves a series of conditions, e.g. if the user is on 'holiday', all other options are irrelevant. Therefore, time interpretation options are eliminated sequentially, and the outcome depends on the order of precedence, as shown in Figure 5.

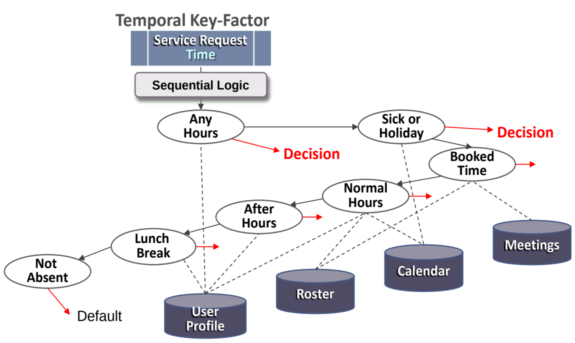

Figure 5. Temporal Precedence of Conditions

Different processing sequences may be needed for different temporal suppositions. For example, the precedence order for the resource optimization function requires that 'busy hour' is eliminated before 'normal working hours'. A *unique* Temporal Attribute (TA_a) must be selected for the Temporal Key-Factor (*TKF*), i.e. there can only be one valid attribute, and there is no logical corroboration between attributes (see summary in Table III). The design of the key factor determines whether a zero-value score is treated as neutral result (no policy attached), or it has an associated action by *default*, as in (5).

$$TA_a \in TF \qquad \textit{Temporal Key-Factor} \quad (5)$$

$$TA_a = OB_b |\, TA_a, \qquad ST_a | TA_a \in \{0,1\}$$

$$\begin{cases} ST_a |\, TA_a = 1 \wedge \sum_{a=1}^{a=a-1} ST_a |\, TA_a = 0 \rightarrow & TKF = TA_a \\ else & \rightarrow\ \ TKF = Default \end{cases}$$

D. Dealing with Exceptions - the Media Type Class

The precedence order may be activated for one member, before aggregating the rest, where factor aggregation is uniform, except for one exception. This is demonstrated by the Media Key-Factor (*MKF*) in Figure 6. For example, corporate rules may impose restrictions on video while roaming or long sessions during busy-hour. However, a 'Special Needs' attribute (e.g. audio-visual interface for disabled users) can be treated as an overriding exception, while the rest of the attributes are a homogeneous group. Exceptions are regarded as precedence-based conditions. A *mandatory* single media type is determined in a mutually exclusive process.

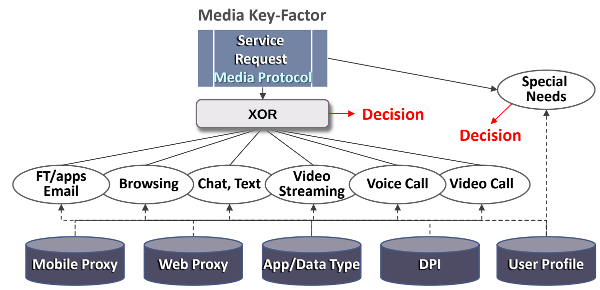

Figure 6. Media Type Key-Factor and Exceptions

V. Corroboration and Correlation

A. Corroborative Attributes - Associated Activity

Cumulative aggregation does not necessarily need to have a prime, but in models that attach particular significance to the prevailing attribute, the aggregation is corroborative. The Associated-Activity Key-Factor (*AAKF*) draws observations from application servers' logins, and the evidence is cumulative, as shown in Figure 7, i.e. the supposition of business activity is strengthened by multiple logins to various corporate apps. If there are no current or recent activities within the temporal proximity limits, the *AAKF* value is zero, which is a neutral result, with no default. While digital classes (Destination, Network, and Media) mandate a non-zero result, the *AAKF* does not regard a zero factor score as an error.

In the case of the business priority function, non-business activities contradict the key-factor's supposition. Therefore, these attributes are discordant by design, in contrast with spatial assertions that conflict with each other and can swap roles. Hence, the *AAKF* discordant assertions are deemed as '*designated'*, i.e. they are fixed for the key-factor supposition.

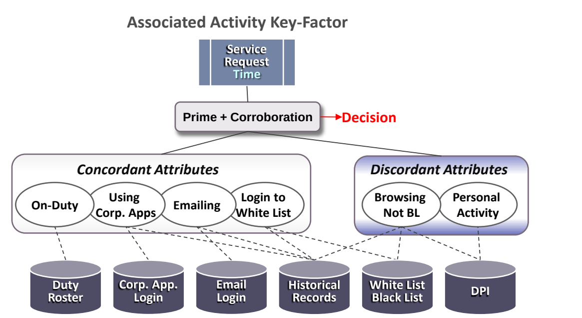

Figure 7. Associated Activity Concordant and Discordant Attributes

B. Correlation - the Network Access Key-Factor

The digital classes (Network, Destination and Media type) are characterized by assertions that are strictly non-cumulative. Their assertions must support the suppositions - or they are not valid, i.e. discordance is not possible. There must always be one *unique* true-status attribute for these digital key-factors.

However, corroboration can come from inter-class correlation: an attribute in one class may confirm or conflict with another, in a different class. Cross-factor support is not a straightforward corroboration, because it involves support for different suppositions, but such correlation assists in achieving model resilience and anomalies detection. If a digital class assertion is disputed by a credible cross-factor attribute, this will cause the request to be rejected with a serious error condition. Confirmation, on the other hand, should boost the key-factor score via corroborative aggregation. Figure 8 shows the Network Access Key-Factor (*NAKF*) with an example of cross-factor correlation by the Spatial Key-Factor.

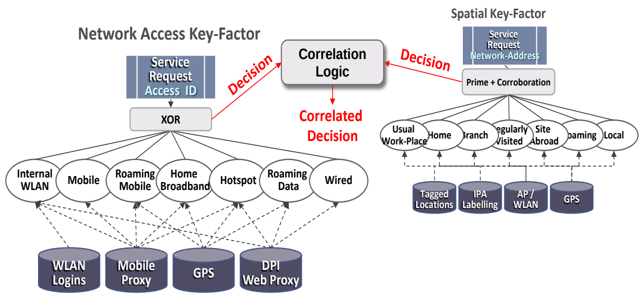

Figure 8. Network Access confirmed by Spatial Factor Location

Correlation boosting is not necessarily symmetrical, and often only one of the involved factors is incremented. For example, the spatial assertion ‘Abroad’ via GPS tagging is boosted by a ‘Roaming’ network attribute, but the correlation is not true in reverse. Similarly, correlation between ‘Home’ location and ‘Home-Broadband’ network is also asymmetric. In finding such correlations between key-factors, attributes dependencies are revealed. A design policy should aim to limit dependencies, yet they can provide important ‘clues’.

VI. Discordance and Uncertainty

A. *Uncertainty and Conflict*

Uncertainty is high on the agenda of papers dealing with context, especially behavioral models. Context semantics allow for an accuracy characteristic [12] to be recorded, but there are no indications of how to compute it or how to integrate it with the attribute scores. As mentioned above, the author has proposed in [28] a procedure to build up credibility rates from the properties of sources and integrate it in the observations’ ‘inherent’ value. This way, uncertainty is integrated in the context members’ evaluation.

The Dempster-Shafer Theory (DST) distinguishes plausibility (1-Belief) from probability, and allows for unassigned ‘mass’, which is neither discordant nor concordant, to be added to plausibility. However, it is debatable whether such ‘ignorance’ mass should be accrued to plausibility, and not to conflicting ‘disbelief’. DST’s ‘*Belief Function*’ tackles uncertainty of evidence conceptually, but their combination rule fails to account for conflict. It assigns discordant ‘beliefs’ to separate sets [25], but they are discarded instead of aggregated as negative amounts, so conflict is not incorporated in the score. In [31], uncertainty is managed via measuring sensors’ *dissimilarity* using DST, thus measuring only ‘pure’ discord, but still not addressing conflicting evidence. Hence, conflict is not well served by existing methods.

B. *Negativity*

Since behavioral context must handle conflict, aggregating attributes must cope with negativity. Negative assertions must actually decrement the scores, to reflect the raised doubt in the prime evidence. Methods that segregate negative from positive attributes, where the values are normalized by relating them to min/max in each group do not achieve such decremental effect. The algorithm in (6) cancels out the negative sign, and yields a positive number that is less than the absolute sum total, but more than a simple subtraction (see [22]). However, if the negative set $\{A_j\}$ is empty, the amount *without* conflict is *lower* than that *with* conflict, which is counter-intuitive. A signed corroborative method must ensure that the prime is always augmented by concordant corroboration, and is only decreased with discordant contributions, as accomplished by Cedar.

$$KF = f\left(\{A^+_{i=1}, \dots A^+_{i=i'}\}, \{A^-_{j=1}, \dots A^-_{j=j'}\}\right) \quad \textit{signed Attributes} \quad (6)$$
$$KF = \sum_{i=1}^{i=i'} \frac{A^+_i}{\max(A^+_i)} + \sum_{j=1}^{j=j'} \frac{\min(A^-_j)}{A^-_j}$$

Handling negative assertions raises several design issues. In particular, the Cedar proportional coefficient (the *residual*) must continue to diminish, despite negative contributions. In (7), mixed positive and negative attributes are aggregated with averagingCedar, allowing assertions to decrement or increment the total. The residual interval must be based on *absolute* values of the contributions, so that the diminishing effect remains consistent. In full Cedar, the prime, by definition, is positive, and is gradually decremented by negative assertions in proportion to their contributions. With an absolute residual, the negative contributions are decreasing, and even large negatives do not deplete the prime value entirely.

$$Prime = +A_{\hat{p}} \qquad aCedar\ for\ Negatives\ \&\ Positives \qquad (7)$$
$$KF = \{+A_1, \pm A_2 \dots \pm A_{a'}\} \ \ 0 < a \le a' \quad Negative\ Attributes$$
$$NonPrimeAtt = \pm \sum_{a=1, a\neq \hat{p}}^{a'} A_a \qquad Add\text{-}on\ Contributions$$
$$\begin{cases} AveNonPrime = \frac{1}{a-1} \sum_{a=1, a\neq \hat{p}}^{a'} A_a & a > 1 \\ AveNonPrime = 0 & a \le 1 \end{cases}$$
$$Residual = (1 - |A_{\hat{p}}| => 0 \qquad \text{‘}Absolute\text{’}\ Residual$$
$$\pm KF = +A_{\hat{p}} + \left(1 - |A_{\hat{p}}|\right) \cdot \frac{1}{a-1} \sum_{a=1, a\neq \hat{p}}^{a'} A_a$$

C. *Discordant Suppositions - the Integrity Class*

Managing discordance affects model design as well as algorithms. Signed scores require particular attention in aggregation and normalization, and in interpreting results, which must attach meaning to possible negative output. Particular issues arise in the Integrity Key-Factor (*IKF*), as in Figure 9, which contains bi-polar attributes that can swing from one extreme to another, e.g. ‘habitual’ becomes ‘unhabitual’, according to spatial proximity and historical records. Thus, the aggregation procedure cannot distinguish negative attributes.

Discordant assertion is defined as (a) *designated*, if the assertion always disagrees with the class supposition, e.g. ‘non-business activities’ in *AAKF*; (b) *swapping roles*, if it can be prime, concordant to discordant, as in *SKF*; (c) *bi-polar*, if the observation intensity indicates it, as in the Integrity Key-Factor

(*IKF*). Assertions turn negative at different processing points: (a) fixed negativity is 'hard-wired' for *designated* discordant attributes; (b) negativity is dynamically assigned to *role-swapping* attributes that are conflicting with the prime, when it is ascertained; (c) *bi-polar* assertions become negative when the Intensity is assessed.

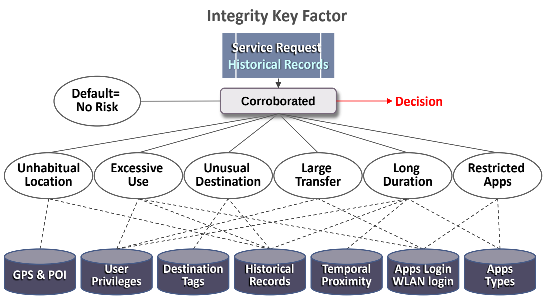

Figure 9. Integrity Key-Factor

When the *whole* set of assertions is discordant, the key-factor itself is a negative component, e.g. the *IKF* is regarded as Integrity=(1-Risk). If the *IKF* is expressed as a series of bi-polar attributes, i.e. mixed positive and negative attributes, it could produce an overall negative key-factor. Unlike *SKF*, the prime itself can be negative, if it contradicts the key-factor supposition. In this case, the negative prime should be the *minimum* value, rather than the customary *maximum*. This enables addressing the '*worst*' attribute, and applying policies to the real concern - the integrity weakness, as shown in (8).

$$-1 \le a \le +1 \qquad \textit{Negative Prime Cedar} \; (8)$$

$$IKF = \{A^{\pm}_{a=1}, \dots A^{\pm}_{a=a'}\}, \quad A^{\pm}_{a} < A^{\pm}_{a-1}$$

$$\begin{cases} A^{+}_{\hat{p}} = \max(A_1, \dots A_{a'}) & if \max(A_1, \dots A_{a'}) \ge 0 \\ A^{-}_{\hat{p}} = \min(A_1, \dots A_{a'}) & if \max(A_1, \dots A_{a'}) < 0 \end{cases} \quad \textit{Negative Min}$$

$$f(IA_a) = IA_{a-1} + IA_a\left(1 - |IA^{\pm}_{a-1}|\right) \quad \textit{Absolute Residual}$$

$$IKF = \begin{cases} IKF = 0 & a = 0 \\ IKF = \pm A_{\hat{p}} & a = 1 \\ f(IA_{a-1}) + IA_a \cdot (1 - f(IA_{a-1})) & \forall (1 < a \le a') \end{cases}$$

If the concordant values do not offset the discordant values, a negative factor is returned, which will reduce the profile's overall score. This is still a valid result, as long as the score decrements are proportional to other key-factors values. Hence, cross-class parity and consistency should be maintained.

VII. Summarizing Classification Issues

Table III summarizes the described features per key-factor. It shows the required mode of processing (by rank, corroboration, sequential logic etc.). The intensity indexing can be spatial or temporal, for some or all members, but certain attributes have specific intensity scales, e.g. data confidentiality scale for Integrity risks. The table shows which classes must have a prime attribute selected, while others must have a unique attribute selection. It also shows the appropriate type of discordance (designated, role-swapping, or bi-polar), and the requirements for mandatory results or defaults.

It is evident that some features are common within their class facets (environmental, circumstantial, digital), yet their procedures are still different. This shows that there is *no one-size-fits-all* and that the model design has to accommodate individual requirements of different components.

Table III: Summary of Class features

Group:	***Environmental***		***Situational***			***Digital***		
Class:	**Spatial**	**Temporal**	**Activity**	**Urgency**	**Integrity**	**Destin.**	**Network**	**Media**
Processing	Rank+ Corrob.	Seq. xor	Rank+ Corrob.	Rank Corrob+ xor	Rank+ Corrob.	xor+ xor	xor	Seq.+ xor
Intensity	Spatial	Temporal	Temporal	Temporal	Temporal +	no	no	no
Discordant	Swapping	no	Designated	no	Bi-polar	no	no	no
Prime	Prime	Unique	Prime	Prime	Prime	Unique	Unique	Unique
Mandatory	Default	Default	no	no	no	yes	yes	yes

VIII. Related Work

Little is said in research about how to select sources and attributes and only few studies go beyond RBAC (Role Based Access Control) for enterprise admission control. The range of context aspects is mostly limited to environmental and digital factors, but rarely circumstantial or behavioral. In [10], transport methods are deduced from environmental context. An activity factor is not uncommon, as in [1], but it has a wide range of interpretations. In [3], roles and environmental aspects are blended. In [8], OSS/BSS context relies on 'policy-continuum', i.e. retained memory of policy-based behavior. In [5], context is used for agent-based routing at the enterprise, but does not address service requests. These studies do not cover the full range of attributes and factors that is required for accurate context assessment.

Uncertainty and conflict are fundamental in context modelling. The Dempster Combination Rule, as in [24], separates and discards discordant attributes, so conflict is not taken into account. Resolving conflict in [20] uses ABAC (Attribute-Based Access Control) with rule reduction, but not in aggregation. In [22], negative and positive elements are segregated and correlated in a normalizing algorithm, but it fails when the conflict is zero. In [24] *signed* fuzzy measures are aggregated, using discrete Choquet Integral that exploits their interactions, but only for interval-based models.

Context languages, such as OWL (Web Ontology Language) and the OASIS-defined XCML (eXtensible *Context* Markup Language) [11] enable documenting context structures, but do not assist in design decisions. They do not capture source credibility, but XCML in [12] is extended to accommodate 'accuracy'. In [21], XACML (Extensible *Access Control* Markup Language) is used for policy based authorization. In [13], SAML (Security Assertion Markup Language) profile for XACML is specified, to relay authorization information for special security assertions. In [14] XACML extensions for RBAC authenticate user access requests, with spatial-temporal context. However, these semantic tools also require appropriate aggregation procedures.

Decision Trees (DT) structures are described in many studies. In [7], OWA (Ordered Weighting Averaging) is used to create hierarchical Fuzzy Operator Trees (FOT). Lack of model uniformity is reflected in [4], which decomposes OWA attribute pattern trees, to allow mixed operations (max, min, average, ordinal) for multiple algorithms in one model, acknowledging the need for complex aggregation. Data mining methods are used for mobile context in [2], including CART (Classification and Regression Trees), which atomizes attributes down to binary components in deeply hierarchical structures. In [26], a bottom-up method (using Maximum Likelihood), is proposed for building decision trees up from data, avoiding numerous 'branches', but still allowing single source branches to exist. These methods deeply affect the design of suppositions and assertions.

Several attribute aggregation methods that are regarded as MCDM (Multi-Criteria Decision Making) are listed in [9], including fuzzy algorithms. In [15], WPM (Weighted Product Model) with pairwise analysis is preferred over SAW (Simple Additive Weighting), although SAW is faster to compute. However, in [27], WPM is shown to produce extreme results, due to weights being used as exponents, while SAW is more intuitive. Aggregation by ranking is dominated by OWA, which was first proposed by Yager in 1988 [16]. In [6], 'Induced OWA' extends the concept to pairs that allow 'fusing' different types of information, e.g. linking policies to attributes. In [17], an OWA extension ranks ordinal interval vectors of spatial sources.

In this paper, the new Cedar algorithm is shown to produce appropriate aggregation results for key-factors that need corroborative methods, while the MCDM methods fail to cope with discordance, parity, proportionality and scale.

IX. CONCLUSIONS

In this study, design issues are shown to have considerable impact on context model methodology and aggregation techniques. The design affects the choice of algorithms and procedures, and significantly affects accuracy of decisions. Aspects such as atomization granularity, hierarchy, policy weighting techniques, parity between classes, aggregation modes, uncertainty and discordance, all have profound consequences in terms of the ability to produce unambiguous, conclusive decisions. A good model design strikes a balance between model simplifications to optimize real-time processing, and achieving definitive decisions, based on evidential nuances, despite conflict and uncertainty.

Several aggregation methods were proposed in this paper, to suit particular designs, including the new corroboration algorithms, Cedar, which is a recursive algorithm that meets all the corroboration requirements for proportional aggregation, diminishing impact of lesser contributions and dealing with discordance. The utility of these methods extends to similar types of behavioral-digital context models, especially for diagnostic and analytic processes.

REFERENCES

[1] D Riboni L Pareschi C Bettini "Towards the Adaptive Integration of Multiple Context Reasoners in Pervasive Computing Environments" Xplore 2010

[2] E Kahya-Özyirmidokuz, A Gezer, C Çiflikli "Characterization of Network Traffic Data" Data Analytics 2012 IARIA

[3] S Ahmad, R Ahmad "Design of Algorithm for Environment based Dynamic Access Control Model for Database Systems" International Journal of Computer Applications May 2011

[4] R Senge and E Hullermeier "Pattern Trees for Regression and Fuzzy Systems Modeling" IEEE 2011

[5] M De Choudhury H Sundaram A John Dorée D Seligmann "Context aware routing of enterprise user communications" IEEE 2007

[6] R Yager, D Filev "Induced Ordered Weighted Averaging Operators" IEEE 1999

[7] Yu Yi, T Fober, E Hullermeier "Fuzzy Operator Trees for Modeling Rating Functions" IJCIA 2009

[8] J Strassner, S v-d-Meer, D O'Sullivan, S Dobson "The Use of Context-Aware Policies and Ontologies to Facilitate Business-Aware Network Management" Springer, LLC 2009

[9] P. Meyer and M. Roubens, "Choice, ranking and sorting in fuzzy multiple criteria decision aid, in Multiple Criteria Decision Analysis: State of the Art Surveys, eds. J. Figueira, S. Greco and M. Ehrgott (Springer-Verlag, 2005), pp. 471–506

[10] L Liao D Fox H Kautz "Learning and Inferring Transportation Routines" Artif. Intell. 2007

[11] C Bettini et al "A Survey of Context Modelling and Reasoning Techniques" Elsevier 2008

[12] R robinson, K Henricksen "XCML - A runtime representation for the Context Modelling Language" IEEE 2007

[13] "SAML 2.0 profile of XACML v2.0 OASIS Standard" OASIS 2005

[14] Que Thi, Tran Dang "X–STROWL: A Generalized Extension of XACML for Context-aware Spatio-Temporal RBAC Model with OWL" IEEE 2012

[15] K.Savitha, Dr.C.Chandrasekar "Vertical Handover decision schemes using SAW and WPM for Network selection in Heterogeneous Wireless Networks" GJCST (2011)

[16] R Yager "On Ordered Weighted Averaging Aggregation Operators in ultMicriteria Decisionmaking" IEEE 1988

[17] B Ahn "The OWA Aggregation With Uncertain Descriptions on Weights and Input Arguments" IEEE 2007

[18] R Copeland, N Crespi "Selecting Access Network for BYOD enterprises with Business Context (eBC) and enterprise-centric ANDSF" Springer, Mobilware 2012

[19] R Copeland, N Crespi, "Implementing an enterprise Business Context model for defining Mobile Broadband Policy. In: IEEE CSNM (2012)

[20] C Shu, E Yang, A Arenas "Detecting Conflicts in ABAC Policies with Rule-reduction and Binary-search Techniques" IEEE 2009.

[21] A Mohan, D Blough "An Attribute-based Authorization Policy Framework with Dynamic Conflict Resolution" Internet-2, 2010

[22] J Lei, W Niu, Y Qin, H Tang and Song Ci "Aggregating User Rating and Service Context for WSN Service Ranking" 2012, Globecom

[23] P Smets "Practical Uses of Belief Functions" Universite Libre de Bruxelles, Brussels-Belgium

[24] K Huang, J Shieh, K Lee, S Wu "Applying A Generalized Choquet Integral With Signed Fuzzy Measure Based On The Complexity To Evaluate The Overall Satisfaction Of The Patients" IEEE 2010

[25] J Pearl "Reasoning with Belief Functions: An Analysis of Compatibility" 1990International Journal of Approximate Reasoning

[26] R Barros, R Cerri, P Jaskowiak, A de Carvalho "A Bottom-Up Oblique Decision Tree Induction Algorithm" IEEE2011

[27] Fred S. Azar "Multiattribute Decision-Making: Use of Three Scoring Methods to Compare the Performance of Imaging Techniques for Breast Cancer Detection" ScholarlyCommons 2000

[28] R Copeland, N Crespi "Rating Credibility of Sources for Profiling Risk and Business Context of Service Requests", in-press SSC IEEE 2014

[29] R Copeland, N Crespi "Controlling Enterprise Context-Based Session Policy and Mapping to Mobile Broadband Policy Rules" IEEE 2012.

[30] K Sentz "Combination of Evidence in Dempster-Shafer Theory" 2002

[31] M Florea, E Bosse "Dempster-Shafer Theory: combination of information using contextual knowledge" ISIF 2009.