Personalised Automated Assessments

Patricia Gutierrez and **Nardine Osman** and **Carles Sierra**
Artificial Intelligence Research Institute (IIIA-CSIC), Barcelona, Spain
{patricia, nardine, sierra}@iiia.csic.es

Abstract

Consider an evaluator, or an assessor, who needs to assess a large amount of information. For instance, think of a tutor in a massive open online course with thousands of enrolled students, a senior program committee member in a large peer review process who needs to decide what are the final marks of reviewed papers, or a user in an e-commerce scenario where the user needs to build up its opinion about products evaluated by others. When assessing a large number of objects, sometimes it is simply unfeasible to evaluate them all and often one may need to rely on the opinions of others. In this paper we provide a model that uses peer assessments to generate expected assessments and tune them for a particular assessor. Furthermore, we are able to provide a measure of the uncertainty of our computed assessments and a ranking of the objects that should be assessed next in order to decrease the overall uncertainty of the calculated assessments.

1 Introduction

Consider an assessor who needs to assess a large amount of information. For instance, think of a tutor in a massive open online course with thousands of enrolled students, a senior program committee member in a large peer review process who needs to decide what are the final marks of reviewed papers, or a user in an e-commerce scenario where the user needs to build up its opinion about products evaluated by others. When assessing a large number of objects, sometimes it is simply unfeasible to evaluate them all and often one may need to rely on the opinions of others. In the process of building up our opinion, some questions need to be answered, such as: How much should I trust the opinion of a peer? What should I believe given a peer's opinion? What should I believe when many peers give their different opinions? Which objects should be assessed next, such that the certainty of my belief improves?

This paper addresses these questions through the Personalised Automated ASsessment model (PAAS). PAAS uses peer assessment to calculate and predict assessments. However, what is fundamentally different from many previous works [Piech *et al.*, 2013; de Alfaro and Shavlovsky, 2013; Walsh, 2014; Wu *et al.*, 2015] is that the computed peer-based assessment is tuned to the perspective of a specific community member. PAAS aggregates peer assessments giving more weight to those peers that are trusted by the specific community member whom the automated assessments are computed for. How much this specific member trusts a peer is then based on the similarity or evaluation rate between his (past) assessments and the peer's (past) assessments over the same assignments. To compute such a trust measure, we build a trust network conformed of direct and indirect trust values among community members. Direct trust values are derived from common assessments while indirect trust is based in the notion of transitivity. We clarify that our target is not consensus building, but to accurately estimate unknown assessments from a specific member's point of view, based on the peers' assessments and reliability.

Finally, we are also able to provide a measure of the uncertainty of our calculated assessments and a ranking of the objects that should be assessed next in order to decrease the overall uncertainty of those calculated assessments.

2 The PAAS Model

2.1 Notation and Problem Definition

Let ϵ represent an assessor who needs to assess a large set of objects $\mathcal{I}$, and let $\mathcal{P}$ be a set of peers that are able to assess objects in $\mathcal{I}$.

We understand assessments as probability distributions over an evaluation space $\mathcal{E}$ at a given moment in time. For example, one can define a set of elements for the evaluation space for the quality of an English classroom homework as $\mathcal{E} = \{poor, good, excellent\}$. The assessment $\{poor \mapsto 0, good \mapsto 0, excellent \mapsto 1\}$ would represent the highest assessment possible, whereas the assessment $\{poor \mapsto 0, good \mapsto 1/2, excellent \mapsto 1/2\}$ would represent that the quality of the homework is most probably between good and excellent, and so on.

We define an assessment e_i^α (also referred to as evaluation or opinion) as a probability distribution over the evaluation space $\mathcal{E}$, where $\alpha \in \mathcal{I}$ is the object being evaluated and $i \in \{\epsilon \cup \mathcal{P}\}$ is the evaluator. We say $e_i^\alpha = \{x_1 \mapsto v_1, \ldots, x_n \mapsto v_n\}$, where $\{x_1, \ldots, x_n\} = \mathcal{E}$ and $v_i \in [0,1]$ represents the value assigned to each element $x_i \in \mathcal{E}$, with the condition that $\sum_{i\in|\mathcal{E}|} v_i{=}1$.

Finally, we define $\mathcal{L}$ as the history of all assessments performed, and $\mathcal{O}_\alpha \subset \mathcal{L}$ as the set of past peer assessments over the object α.

The ultimate goal of our work is to compute the probability distribution of ϵ's evaluation over a certain object α, given the evaluations of several peers over that same object α. In other words, what is the probability that ϵ's evaluation is x given the set of peers' evaluations O_α? Such expectation can be formalized with the conditional probability as follows:

$$p(X{=}x \mid \mathcal{O}_\alpha)$$

.

To calculate the above conditional probability, we take into account every particular evaluation in O_α. In other words, expectations (or probabilities) are calculated for each individual evaluation in O_α, before those expectations are aggregated into $p(X{=}x \mid \mathcal{O}_\alpha)$. The probability that ϵ's assessment is x given a particular evaluation $e_\mu^\alpha \in O_\alpha$ is formalized as follows:

$$p(X{=}x \mid e_\mu^\alpha)$$

.

The more general probability $p(X{=}x \mid \mathcal{O}_\alpha)$ is then defined as an aggregation of the individual probabilities:

$$p(X{=}x \mid \mathcal{O}_\alpha){=}\overline{p(X{=}x \mid e_\mu^\alpha)}$$

where the exact definition of the aggregation is presented later on in Section 2.4.

We strongly base the intuition behind the computation of the individual conditional probabilities on the notion of *trust* between peers based on previous experiences, where trust is understood in this context as the expected similarity between the assessments given by those peers. In other words, our intuition is that we expect ϵ will tend to agree with μ's assessments if his trust on μ is high. Otherwise, ϵ's evaluation will probably be different. We perform then a sort of analogical reasoning: if in the past μ gave opinions that were a certain degree dissimilar from ϵ's opinions, then this will probably happen again now.

The remainder of this section is divided accordingly. We first describe in detail how the measure of trust between peers is calculated (Section 2.2). Then, we illustrate how to calculate ϵ's assessment on an object α given μ's assessment over α and ϵ's trust in μ's assessments (Section 2.3). In other words, we present an approach for calculating the individual probability $p(X{=}x \mid e_\mu^\alpha)$. We then illustrate how to combine those probabilities to build the probability distribution of ϵ's assessments given the assessments of several peers (Section 2.4). In other words, we present an approach for calculating the probability $p(X{=}x \mid \mathcal{O}_\alpha)$. Finally, we provide a measure of the uncertainty of the computed assessments and a ranking of the objects that should be assessed next by ϵ in order to decrease that uncertainty (Section 2.5).

2.2 Step 1. How much should I trust a peer?

ϵ needs to decide how much can he or she trust the assessment of a peer μ. We define this trust measure based on the following two intuitions. Our first intuition states that if ϵ and μ have both assessed the same object, then the similarity of their assessments can give a hint of how close their judgments are. However, cases may arise where there are simply no objects evaluated by both ϵ and μ. In such a case, one may think of simply neglecting μ's assessment, as ϵ would not know how much to trust μ's assessment. Our second intuition, however, proposes an alternative approach for such cases, where we approximate that unknown trust between ϵ and μ by looking into a chain of trust between ϵ and μ through other peers. Roughly speaking, we relay on the transitive notion: "if ϵ trusts μ, and μ trusts μ', then ϵ will likely trust μ'". In the following, we define these two intuitions through two different types of trust relations: direct trust and indirect trust.

Direct Trust

Direct trust is the trust relation that emerges between evaluators that have assessed one or more objects in common. One possible approach is to measure such relation as aggregations of their evaluations' similarity over those objects assessed in common. For instance, let the set $A_{i,j}{=}\{\alpha \mid e_i^\alpha, e_j^\alpha \in \mathcal{L}\}$ be the set of objects that have been assessed by both evaluators i and j. Then different definitions for the direct trust between i and j based on the similarity between two assessments ($sim(e_j^\alpha, e_j^\alpha)$) may be adopted, such as as:

- The average of the similarities for all commonly assessed objects:

$$T_D(i,j){=}\frac{\sum_{\alpha\in A_{i,j}} sim(e_i^\alpha, e_j^\alpha)}{|A_{i,j}|}$$

- The conjunction of the similarities for all commonly assessed objects:

$$T_D(i,j){=}\bigwedge_{\alpha\in A_{i,j}} sim(e_i^\alpha, e_j^\alpha)$$

- The Pearson coefficient [Upton and Cook, 2008], or linear correlation between i and j, for all commonly assessed objects:

$$T_D(i,j){=}\frac{\sum_{\alpha\in A_{i,j}} sim(e_i^\alpha, \bar{e}_i)\cdot sim(e_j^\alpha, \bar{e}_j)}{\sqrt{\sum_{\alpha\in A_{i,j}} sim(e_i^\alpha, \bar{e}_i)^2}\sqrt{\sum_{\alpha\in A_{i,j}} sim(e_j^\alpha, \bar{e}_j)^2}}$$

  where $\bar{e}_i$, $\bar{e}_j$ are the means of the evaluations performed over the set $A_{i,j}$ by i and j respectively.

However when we calculate such aggregations we loose relevant information. For instance, we are not able to tell if j usually under rates with respect to i, if it usually over rates, or neither. We are also not able to tell if the dissimilarities between i and j's evaluations are highly variable or not.

To cope with such loss of information, we define the direct trust between two peers i and j as a probability distribution

$\mathbb{T}_{\mathbb{D}i,j} : [0,1] \rightarrow [0,1]$ built from the historical data of previous evaluations performed by i and j. This probability distribution describes, as we will explain shortly, the *expected similarity* or the *expected evaluation rate* between i and j's assessments. The support of the distribution is $[0,1]$ since both the expected similarity and the expected evaluation rate are in the range $[0,1]$, as we will see shortly, and the range of the distribution is $[0,1]$ as this is a probability distribution and the range of any probability is $[0,1]$. Note that we do not consider here any summarizing measure for trust that would translate that distribution into a single value, although a number of measures could be used, such as the average similarity (as the center of gravity of the distribution) or entropy (as a measure of the uncertainty of the distribution).

When defining $\mathbb{T}_{\mathbb{D}i,j}$ we distinguish two cases: (1) a first case with a non-ordered evaluation space, such as $\mathcal{E} = \{visionary, original, sound\}$; and (2) a second case with an ordered evaluation space, such as$=\{bad, good, excellent\}$. In the second case, we are interested in maintaining information about whether a peer under rates or over rates with respect to another peer, therefore we are interested in the *expected evaluation rate* between i and j. In the first case, this is not an issue as assessments cannot be ordered and therefore the notion of under/over rating does not exist, therefore we are rather interested in the *expected similarity* between i and j's assessments. Next we detail the trust probability distributions $\mathbb{T}_{\mathbb{D}i,j}$ built for both cases.

- *Non-Ordered Case.*

  In the non-ordered case, we are interested in the similarity between i and j's assessments. As such, the support of the distribution representing i's direct trust on j (i.e. the x-axis of $\mathbb{T}_{\mathbb{D}i,j}$) consists of the possible degrees of similarity between i and j's assessments.

  Trust distribution $\mathbb{T}_{\mathbb{D}i,j}(x)$ then describes the probability that peers i and j evaluate an object with a similarity x (or the probability that the similarity of their evaluations is x).

- *Ordered Case.*

  In the ordered case, we are interested in the evaluation rate ${}^{e_j}/_{e_i}$ between evaluations made by peers i and j. If ${}^{e_j}/_{e_i} = 1$, this means that i and j provide the same evaluation. If ${}^{e_j}/_{e_i} > 1$, this meas that j over rates with respect to i. If ${}^{e_j}/_{e_i} < 1$, this means that j under rates with respect to i.

  We normalize the evaluation rate to values between 0 and 1. To do so, we require a non decreasing function $r : \mathcal{R} \rightarrow [0,1]$ such that $\lim_{x\rightarrow\infty} r(x)$=1, and for convenience we constraint $r(1)$=0.5. We adopt the following normalized evaluation rate function that satisfies these properties:

  $$r(x){=}e^{\ln {}^{1/2}/_{x}} \qquad (1)$$

  As such, the support of the distribution representing i's direct trust on j (i.e. the x-axis of $\mathbb{T}_{\mathbb{D}i,j}$) consists of the possible normalized evaluation rates between i and j. Trust distribution $\mathbb{T}_{\mathbb{D}i,j}(x)$ then describes the probability that i and j would assess an object with a normalized evaluation rate x.

In what follows, we explain how we build direct trust distributions computationally, based on previous experiences.

Initially, the direct trust distribution between any two peers is the uniform distribution $\mathbb{F}{=}\{{}^{1}/_{n}, \ldots, {}^{1}/_{n}\}$ (describing ignorance), where n is the size of the distribution's support. Every new assessment made would then update the trust distributions accordingly. Consider a new assessment e_i^α. The distribution $\mathbb{T}_{\mathbb{D}i,j} \forall j$ s.t. $A_{i,j} \neq \emptyset$ is updated as follows:

1. We find the element x in $\mathbb{T}_{\mathbb{D}i,j}$'s support whose probability needs to be adjusted. So we calculate $x{=}sim(e_j^\alpha, e_i^\alpha)$ in the ordered case (where the definition of sim is domain dependent and outside the scope of this paper, although we do note that several approaches may be adopted, such as using semantic similarity measures [Li *et al.*, 2003]), or $x = r({}^{e_j^\alpha}/_{e_i^\alpha})$ in the non-ordered case (Equation 1).

2. We update the probability of the *single expectation* x in $\mathbb{T}_{\mathbb{D}i,j}$ accordingly:

   $$p(X{=}x) = p(X{=}x) + \gamma \cdot (1 - p(X{=}x)) \qquad (2)$$

   The update is based on increasing the latest probability $p(X = x)$ by a fraction $\gamma \in [0,1]$ of the total potential increase $(1 - p(X = x))$. For instance, if the probability of x is 0.6 and γ is 0.1, then the new probability of x becomes $0.6 + 0.1 \cdot (1 - 0.6) = 0.64$. We note that the ideal value of γ should be closer to 0 than to 1 so that one single experience does not result in considerable changes in the distribution. In other words, a *single* assessment cannot result in *considerable* change in the probability distribution. Considerable changes can only be the result of information learned from the accumulation of many assessments.

3. We normalize $\mathbb{T}_{\mathbb{D}i,j}$ by updating *several expectations* following the entropy based approach of [Sierra and Debenham, 2005]. The entropy-based approach updates $\mathbb{T}_{\mathbb{D}i,j}$ such that: (1) the value $p(X{=}x)$ is maintained and (2) the resulting distribution has a minimal relative entropy with respect to the previous one. In other words, we look for a distribution that contains the updated probability value $p(X{=}x)$ and that is at a minimal distance from the original $\mathbb{T}_{\mathbb{D}i,j}$ (as the relative entropy is a measure of the difference between two probability distributions). Following this approach, we update $\mathbb{T}_{\mathbb{D}i,j}(X)$ as follows:

   $$\begin{aligned} \mathbb{T}_{\mathbb{D}i,j}(X) = \arg\min_{\mathbb{P}'(X)} \sum_{x'} p(X{=}x') \log \frac{p(X{=}x')}{p'(X{=}x')} \\ \text{such that} \quad \{p(X{=}x) = p'(X{=}x)\} \end{aligned} \qquad (3)$$

   where $p(X{=}x')$ is a probability value in $\mathbb{T}_{\mathbb{D}i,j}$, $p'(X{=}x')$ is a probability value in $\mathbb{P}'$, and $\{p(X{=}x) = p'(X{=}x)\}$ specifies the constraint that needs to be satisfied by the resulting distribution.

Indirect Trust

Given a direct trust relation between peers i and j and between peers j and k, the question now is: What can we say about the indirect trust between peers i and k when i and k

have no objects assessed in common? In other words, given the direct trust distributions $\mathbb{T}_{\mathbb{D}i,j}$ and $\mathbb{T}_{\mathbb{D}j,k}$, what can we say about the indirect trust distribution $\mathbb{T}_{\mathbb{I}i,k}$?

As with direct trust distributions, we distinguish two cases: a first case where assessments cannot be ordered and thus trust is based on a similarity measure sim; and a second case where assessments can be ordered and thus trust is based on a normalized evaluation rate function $r(x)=e^{\ln 1/2/x}$.

- *Non-Ordered Case.*

  In this case, we want to preserve the fundamental triangular inequality property of similarity functions that says that: T-norm$(sim(a,b), sim(b,c)) \leq sim(a,c)$. As with $\mathbb{T}_{\mathbb{D}i,k}$, the support (or the x-axis) of $\mathbb{T}_{\mathbb{I}i,k}$ consists of the possible degrees of similarity between i and k's assessments. But since these degrees of similarity should satisfy the T-norm, the support is defined as the set:

  $$\text{supp}(\mathbb{T}_{\mathbb{I}i,k})=\{x_{ik}=\text{T-norm}(x_{ij}, x_{jk}) \mid x_{ij} \in \text{supp}(\mathbb{T}_{\mathbb{D}i,j}) \wedge x_{jk} \in \text{supp}(\mathbb{T}_{\mathbb{D}j,k})\}$$

  where supp represents the support of a distribution.

  We then compute the probabilities of the expectations of $\mathbb{T}_{\mathbb{I}i,k}$ as follows:

  $$\{p(X=x_{ik}=\text{T-norm}(x_{ij}, x_{jk}))=\mathbb{T}_{\mathbb{D}i,j}(x_{ij}) * \mathbb{T}_{\mathbb{D}j,k}(x_{jk}) \mid x_{ij} \in \text{supp}(\mathbb{T}_{\mathbb{D}i,j}) \wedge x_{jk} \in \text{supp}(\mathbb{T}_{\mathbb{D}j,k})\} \quad (4)$$

  This could result in more than one probability computed for the same expectation x_{ik}. As such, we then add up all the probabilities that correspond to the same expectation x_{ik}.

  We note that we follow a conservative approach by adopting the product operator (Equation 4), which is a T-norm that gives the smallest possible values, as we prefer not to overrate indirect trust values since they are not inferred directly from historical data. Of course, other operators could also be used, such as the min function.

- *Ordered Case.*

  In this case, we want to preserve the property: $e_j/e_i * e_k/e_j = e_k/e_i$ with respect to the evaluations performed by i, j and k. For instance, if the evaluation rate between e_j and e_i is 0.5 (j under rates a 50% with respect to i) and the evaluation rate between e_k and e_j is 0.5 (k under rates a 50 % with respect to j) then the evaluation rate between e_k and e_i should be 0.25 (then k under rates a 75 % with respect to i).

  As above, the support (or the x-axis) of $\mathbb{T}_{\mathbb{I}i,k}$ consists of the possible degrees of similarity between i and k's assessments. The support us then defined as the set:

  $$\text{supp}(\mathbb{T}_{\mathbb{I}i,k}) = \{x_{ik}=x_{ij} * x_{jk} \mid x_{ij} \in \text{supp}(\mathbb{T}_{\mathbb{D}i,j}) \wedge x_{jk} \in \text{supp}(\mathbb{T}_{\mathbb{D}j,k})\}$$

  We then compute the probabilities of the expectations of $\mathbb{T}_{\mathbb{I}i,k}$ as follows:

  $$\{p(X=x_{ik}=x_{ij} * x_{jk}) = \mathbb{T}_{\mathbb{D}i,j}(x_{ij}) * \mathbb{T}_{\mathbb{D}j,k}(x_{jk}) \mid x_{ij} \in \text{supp}(\mathbb{T}_{\mathbb{D}i,j}) \wedge x_{jk} \in \text{supp}(\mathbb{T}_{\mathbb{D}j,k})\} \quad (5)$$

  Again, this could result in more than one probability computed for the same expectation x_{ik}. As such, we then add up all the probabilities that correspond to the same expectation x_{ik}.

The calculations presented above provide an approach for calculating indirect trust between two peers i and k when those peers are linked through a direct trust chain passing through only one intermediate peer j. For direct trust chains of increasing length between i and k, the previous process is iterated. For instance, if there is a direct trust chain linking i to j, j to m, and m to k, then we first compute the indirect trust distribution $\mathbb{T}_{\mathbb{I}i,m}$ from the direct trust distributions $\mathbb{T}_{\mathbb{D}i,j}$ and $\mathbb{T}_{\mathbb{D}j,m}$, and then we compute the indirect trust distribution $\mathbb{T}_{\mathbb{I}i,k}$ from the direct/indirect trust distributions $\mathbb{T}_{\mathbb{I}i,m}$ and $\mathbb{T}_{\mathbb{D}m,k}$, following the same approach as above.

When multiple chains of direct trust connect two peers (e.g. say a chain linking i to j and j to k, and another chain linking i to m and m to k), we obtain multiple indirect trust distributions (one from every chain). In those cases, we pick the resulting distribution which is most optimistic. In other words, while our approach to calculate the indirect trust follows the pessimistic approach (through our choice of the product operator in Equations 4 and 5), we now choose the most optimistic of the pessimistic outcomes. To do that, we choose the distribution that is closest to the *equivalence* distribution, which is a distribution that describes that the evaluations of two peers are equivalent. In the non-ordered case, the equivalence distribution is $\mathbb{P}_{\mathbb{E}}(1)=1$; that is, the similarity between two peers is maximum. In the non-ordered case, the equivalence distribution is $\mathbb{P}_{\mathbb{E}}(0.5) = 1$; that is, the normalized evaluation rate between two peers is 0.5, which implies that they always provide the same evaluation. The distance between an indirect trust distribution $\mathbb{T}_{\mathbb{I}i,k}$ and the equivalence distribution $\mathbb{P}_{\mathbb{E}}$ can be calculated as:

$$emd(\mathbb{T}_{\mathbb{I}i,k}, \mathbb{P}_{\mathbb{E}}) \quad (6)$$

where emd is the earth mover's distance which calculates the distance between two probability distributions [Rubner *et al.*, 1998].[1] We note that the range of emd is [0,1], where 0 represents the minimum distance and 1 represents the maximum possible distance.

In the remainder of this paper, when we refer explicitly to a direct or indirect trust distribution between peers i and j, we refer to such distribution as $\mathbb{T}_{\mathbb{D}i,j}$ or $\mathbb{T}_{\mathbb{I}i,j}$, respectively. Whereas when we refer generically to a trust distribution that could either be the direct or indirect trust distribution, we refer to such a distribution as $\mathbb{T}_{i,j}$.

Trust Graph

Direct and indirect trust relations in a community can be represented by a weighted directed graph. We define a community's *trust graph* as:

$$G=\langle N, E, w\rangle$$

[1] If probability distributions are viewed as piles of dirt, then the earth mover's distance measures the minimum cost for transforming one pile into the other. This cost is equivalent to the 'amount of dirt' times the distance by which it is moved, or the distance between elements of the probability distribution's support.

where the set of nodes N is the set of evaluators in $\{\epsilon \cup \mathcal{P}\}$, $E \subseteq N \times N$ are edges between evaluators with direct or indirect trust relations, and $w : E \mapsto [0,1]^n$ is the weight of an edge, described as a trust probability distribution.

$D \subset E$ is the set of edges that link evaluators with direct trust relations: $D = \{(i,j) \in E \mid \mathbb{T}_{\mathbb{D}i,j} \neq \bot\}$. Similarly, $I \subset E$ is the set of edges that connect evaluators with indirect trust relations: $I = \{(i,j) \in E \mid \mathbb{T}_{\mathbb{I}i,j} \neq \bot\} \setminus D$. We note that the set of edges E is then composed of the union of the set of direct and indirect edges: $E = D \cup I$. Weights in w describe direct and indirect trust probability distributions and are defined as follows:

$$w(i,j) = \begin{cases} \mathbb{T}_{\mathbb{D}i,j} & \text{, if } (i,j) \in D \\ \mathbb{T}_{\mathbb{I}i,j} & \text{, if } (i,j) \in I \end{cases}$$

Our goal is to determine how much a particular evaluator ϵ can trust a peer μ. So the trust graph is constructed with respect to ϵ's point of view only. Therefore, we maintain a trust graph of the whole community containing all the *direct* edges between peers (as they are needed to calculate indirect trust relations), but we only maintain the *indirect* edges that connect ϵ with the rest of the peers.

Information Decay

An important notion in our proposal is the notion of the *decay* of information. We say the integrity of information decreases with time. In other words, the information provided by a trust probability distribution should lose its value over time and decay towards a default value. We refer to this default value as the *decay limit distribution* $\mathbb{D}$. For instance, $\mathbb{D}$ may be the uniform distribution, which describes that trust information learned from past experiences tends to ignorance over time.

To implement such a decay mechanism, we need to:

1. Record all evaluations $e_i^\alpha \in \mathcal{L}$ made at time t with a timestamp t, noted $e_i^{\alpha^t}$.
2. Record all direct trust distributions $\mathbb{T}_{\mathbb{D}i,j}$ with a timestamp t, noted $\mathbb{T}_{\mathbb{D}i,j}^{\ t}$, where t is the timestamp of the last evaluation that modified the trust distribution. The first time $\mathbb{T}_{\mathbb{D}i,j}$ is calculated, t is the timestamp of the latest evaluation amongst the two evaluations leading to this calculation. (Recall that it is the similarity between two evaluations or the evaluation rate that updates the probability distribution.) Then, every time a new evaluation with timestamp $t' > t$ is considered to update $\mathbb{T}_{\mathbb{D}i,j}^{\ t}$, $\mathbb{T}_{\mathbb{D}i,j}^{\ t}$ is first decayed from t to t' before the distribution is updated.
3. Record all indirect trust distributions $\mathbb{T}_{\mathbb{I}i,j}$ with a timestamp t, noted $\mathbb{T}_{\mathbb{I}i,j}^{\ t}$, where t is the time the distribution is calculated. Every time $\mathbb{T}_{\mathbb{I}i,j}$ is calculated, all probability distributions involved in this calculation will first need to be decayed to the time of calculation t. The time of calculation is usually the latest timestamp amongst the timestamps of the distributions involved in this calculation.

Information in a trust probability distribution $\mathbb{T}_{i,j}$ decays from t to t' (where $t' > t$) as follows:

$$\mathbb{T}_{i,j}^{t \rightsquigarrow t'} = \Lambda(\mathbb{D}, \mathbb{T}_{i,j}^t) \tag{7}$$

where Λ is the *decay function* satisfying the property: $\lim_{t' \to \infty} \mathbb{T}_{i,j}^{t \rightsquigarrow t'} = \mathbb{D}$. One possible definition for Λ could be:

$$\mathbb{T}_{i,j}^{t \rightsquigarrow t'} = \nu^{\Delta_{t,t'}} \cdot \mathbb{T}_{i,j}^t + (1 - \nu^{\Delta_{t,t'}})\mathbb{D} \tag{8}$$

where ν is the decay rate, and:

$$\Delta_{t,t'} = \begin{cases} 0 & \text{, if } t' - t < \omega \\ 1 + \dfrac{t' - t}{t_{max}} & \text{, otherwise} \end{cases}$$

The definition of $\Delta_{t,t'}$ above serves the purpose of establishing a minimum *grace* period, determined by the parameter ω, during which the information does not decay, and that once reached the information starts decaying. The parameter t_{max}, which may be defined in terms of multiples of ω, controls the *pace of decay*. The main idea behind this is that after the grace period, the decay happens very slowly; in other words, $\Delta_{t,t'}$ decreases very slowly.

2.3 Step 2: What to belief when a peer gives an opinion?

Given a peer assessment e_μ^α, the question now is how to compute the probability distribution of ϵ's evaluation. In other words, what is the probability that ϵ's evaluation of α is x given that μ evaluated α with e_μ^α. As illustrated earlier, this is expressed as the conditional probability:

$$\mathbb{P}(X{=}x \mid e_\mu^\alpha)$$

To calculate this conditional probability, the intuition is that ϵ would tend to agree with μ's evaluation if his trust on μ (that is, the expected similarity between their assessments or the expected evaluation rate between their assessments) is high. Otherwise, ϵ's evaluation would probably be different. We perform then a sort of analogical reasoning: if in the past μ gave assessments that were a certain degree dissimilar from ϵ's opinions, or with a certain evaluation rate with respect to ϵ, then this will probably happen again now.

We then calculate the above conditional probability based on the following desired properties:

- If $\mathbb{T}_{\epsilon,\mu}$ is a flat distribution (i.e. a distribution representing ignorance), then $\mathbb{P}(X \mid e_\mu^\alpha)$ should also be a flat distribution. That is, the closer ϵ's trust on μ is to ignorance, the less information μ is giving to ϵ with his/her assessment.
- The degree of belief $e_\epsilon^\alpha = x$ should increase for those points x whose similarity (or evaluation rate, in the case of the ordered case) to e_μ^α is high (i.e. for higher values of $\mathbb{T}_{\epsilon,\mu}$).
- The degree of belief $e_\epsilon^\alpha = x$ should decrease for those points x whose similarity (or evaluation rate, in the case of the ordered case) to e_μ^α is low trust (i.e. for lower values of $\mathbb{T}_{\epsilon,\mu}$).

Formally, these properties are achieved by defining the probabilities accordingly (where the denominator of the following two equations, Equations 9 and 10, is used for normalisation to ensure that the resulting distribution is a probability distribution):

- *Non-Ordered Case.*

$$p(X{=}x \mid e^{\alpha}_{\mu}) = \frac{e^{\mathbb{T}_{\epsilon,\mu}(sim(e^{\alpha}_{\mu},x))\cdot\mathbb{I}(\mathbb{T}_{\epsilon,\mu})}}{\sum\limits_{x'\in\mathcal{E}} e^{\mathbb{T}_{\epsilon,\mu}(sim(e^{\alpha}_{\mu},x'))\cdot\mathbb{I}(\mathbb{T}_{\epsilon,\mu})}} \tag{9}$$

- *Ordered Case.*

$$p(X{=}x \mid e^{\alpha}_{\mu}) = \frac{e^{\mathbb{T}_{\epsilon,\mu}(r(e^{\alpha}_{\mu}/x))\cdot\mathbb{I}(\mathbb{T}_{\epsilon,\mu})}}{\sum\limits_{x'\in\mathcal{E}} e^{\mathbb{T}_{\epsilon,\mu}(r(e^{\alpha}_{\mu}/x'))\cdot\mathbb{I}(\mathbb{T}_{\epsilon,\mu})}} \tag{10}$$

where $\mathbb{I}(\mathbb{T}_{\epsilon,\mu})$ is a measure of how informative the probability distribution $\mathbb{T}_{\epsilon,\mu}$ is. We calculate $\mathbb{I}(\mathbb{T}_{\epsilon,\mu})$ as:

$$\mathbb{I}(\mathbb{T}_{\epsilon,\mu}) = 1 - \mathbb{H}(\mathbb{T}_{\epsilon,\mu}) \tag{11}$$

where $\mathbb{H}$ describes the entropy of a probability distribution. In other words, the lower the entropy of the distributions then the more informative it is, and vice versa.

We finally define the probability distribution of ϵ's expected evaluation given μ's opinion accordingly: $\mathbb{P}(X \mid e^{\alpha}_{\mu})$, where X varies over the evaluation space $\mathcal{E}$.

2.4 Step 3: What to belief when many give opinions?

In the previous section we computed $\mathbb{P}(X \mid e^{\alpha}_{\mu})$. That is, the probability distribution of ϵ's evaluation on α given the evaluation of a peer μ on α. But what does ϵ do when there is more than one peer assessing α?

Given the set of opinions $\mathcal{O}_{\alpha}$ describing a set of peer evaluations over the object α, we define the probability of ϵ's assessment being x as follows:

$$p(X{=}x \mid \mathcal{O}_{\alpha})) = \frac{\prod\limits_{\mu\in\mathcal{O}_{\alpha}} p(X{=}x \mid e^{\alpha}_{\mu})}{\sum\limits_{x'\in\mathcal{E}} \prod\limits_{\mu\in\mathcal{O}_{\alpha}} p(X{=}x' \mid e^{\alpha}_{\mu})} \tag{12}$$

In other words, the probability of ϵ's assessment on α being x given the set of opinions over α is an aggregation (a product in this case) of the probabilities of ϵ's assessment on α being x given each evaluation $e^{\alpha}_{\mu} \in \mathcal{O}_{\alpha}$.

We then define the probability distribution of ϵ's expected evaluation given all opinions in $\mathcal{O}_{\alpha}$ as $\mathbb{P}(X \mid \mathcal{O}_{\alpha})$, where X varies over the evaluation space $\mathcal{E}$.

We note that instead of the product operator $\prod$ other connectives could be used, for instance the $\min$ operator might be used. However, we note that using the minimum operator does not take into account the number of assessments made. That is, having assessments of 20 peers could be equivalent to having the assessment of just one peer. In fact, the proposed aggregation of Equation 12 ensures that:

- The larger the number of identical opinions, the less uncertain the final probability distribution is, and
- The more trusted the opinions, the less uncertain the final probability distribution is.

Finally, to translate the final assessment from a probability distribution $\mathbb{P}(X \mid \mathcal{O}_{\alpha})$ into a single value, we calculate the mean (average) of the distribution and select the closest mark to that mean.

2.5 Step 4: What should be evaluated next?

The previous three steps have provided a model to calculate automated assessments of objects that have not been assessed by ϵ, based on peers opinions. The level of uncertainty of the automated assessments generated by our model can be calculated as the uncertainty of the probability distribution of ϵ's expected evaluation based on those peers opinions $\mathbb{P}(X \mid \mathcal{O}_{\alpha})$. This level of uncertainty is measured by the distribution's entropy:

$$\mathbb{H}(\mathbb{P}(X \mid \mathcal{O}_{\alpha}))$$

The question that naturally arises then is what objects can be assessed next by ϵ to decrease such uncertainties? For example, how many more assignments should a tutor evaluate so that the uncertainty of the calculated assessments becomes *acceptable*. We suggest ϵ to evaluate objects with maximum uncertainty, or maximum entropic value. The ranking of objects with respect to their entropic value is then defined as follows:

$$\begin{aligned} Rank(\alpha) &= 1 - \mathbb{H}(\mathbb{P}(X \mid \mathcal{O}_{\alpha})) \\ &= 1 + \sum_{x\in X} p(X{=}x \mid \mathcal{O}_{\alpha}) \ln p(x \mid \mathcal{O}_{\alpha}) \end{aligned} \tag{13}$$

ϵ can then continue to evaluate objects one by one until the uncertainty of the automated assessments becomes less than some predefined *acceptable uncertainty threshold*.

3 Conclusions and Future Work

In this paper we have presented the personalised automated assessments model (PAAS), a trust-based assessment service that helps compute group assessments from the perspective of a specific community member. This computation essentially aggregates peer assessments, giving more weight to those peers that are trusted by the specific community member whom the automated assessments are computed for. How much this specific member trusts a peer is then based on the similarity or evaluation rate between his (past) assessments and the peer's (past) assessments over the same assignments.

The proposed work is an extension of the work carried out in [Gutierrez *et al.*, submitted for publication]. In fact, the COMAS model is a much more simplified model of the non-ordered case. It is much more simplified as it assumes that the probability of the similarity between two assessors is 1 for the aggregation of the similarities of past evaluations over the same objects. PAAS' use of probability distribution makes it a richer and more informative model as much more information is preserved in the calculations. Furthermore, PAAS computes the uncertainty of the automated assessments, helping suggesting which objects should be evaluated next in order to decrease the overall uncertainty of PAAS' calculations.

In COMAS, experimental results were conducted on a real classroom datasets as well as simulated data that considers different social network topologies (where we say students assess some assignments of socially connected students). Results show that the COMAS method 1) is sound, i.e. the error of the suggested assessments decreases for increasing numbers of tutor assessments; and 2) scales for large numbers of students.

Future work on PAAS should follow a similar approach for evaluation, where the same real classroom datasets can be used as the groundtruth of marks, and we can then compare PAAS' automated assessments to that groundtruth.

Additionally, we could also test the ranking of marks (Section 2.5) by running experiments in a real classroom where we ask the tutor to evaluate assignments once in a random order and another time following the suggested ranking. This could help us check whether the error decreases faster in the latter case. Also, we expect to find that for a given acceptable uncertainty threshold, the tutor should evaluate less assignments in order to reach that threshold than evaluating randomly.

Acknowledgments

This work is supported by the CollectiveMind project (funded by the Spanish Ministry of Economy and Competitiveness, under grant number TEC2013-49430-EXP) and the PRAISE project (funded by the European Commission, under grant number 388770).

References

[de Alfaro and Shavlovsky, 2013] L. de Alfaro and M. Shavlovsky. Crowdgrader: Crowdsourcing the evaluation of homework assignments. *Thech. Report 1308.5273, arXiv.org*, 2013.

[Gutierrez *et al.*, submitted for publication] Patricia Gutierrez, Nardine Osman, and Carles Sierra. Trust-based community assessment. *Pattern Recognition Letters*, submitted for publication.

[Li *et al.*, 2003] Yuhua Li, Zuhair A. Bandar, and David McLean. An approach for measuring semantic similarity between words using multiple information sources. *IEEE Trans. on Knowl. and Data Eng.*, 15(4):871–882, July 2003.

[Piech *et al.*, 2013] Chris Piech, Jonathan Huang, Zhenghao Chen, Chuong Do, Andrew Ng, and Daphne Koller. Tuned models of peer assessment in moocs. *Proc. of the 6th International Conference on Educational Data Mining (EDM 2013)*, 2013.

[Rubner *et al.*, 1998] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. A metric for distributions with applications to image databases. In *Proceedings of the Sixth International Conference on Computer Vision (ICCV 1998)*, ICCV '98, pages 59–, Washington, DC, USA, 1998. IEEE Computer Society.

[Sierra and Debenham, 2005] Carles Sierra and John Debenham. An information-based model for trust. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '05, pages 497–504, New York, NY, USA, 2005. ACM.

[Upton and Cook, 2008] G. Upton and I. Cook. *A Dictionary of Statistics*. Oxford Paperback Reference. OUP Oxford, 2008.

[Walsh, 2014] Toby Walsh. The peerrank method for peer assessment. In Torsten Schaub, Gerhard Friedrich, and Barry O'Sullivan, editors, *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 909–914. IOS Press, 2014.

[Wu *et al.*, 2015] J. Wu, F. Chiclana, and E. Herrera-Viedma. Trust based consensus model for social network in an incompletelinguistic information context. *Applied Soft Computing*, 2015.