Improving Next Location Prediction
by Using Hybrid Predictors

Jan Petzold, Faruk Bagci, Wolfgang Trumler, and Theo Ungerer

Institute of Computer Science,
University of Augsburg, Germany,
{petzold, bagci, trumler, ungerer}@informatik.uni-augsburg.de

Abstract. Neural networks, Bayesian networks, Markov models, and state predictors are different methods to predict the next context in a ubiquitous environment. For all methods a lot of parameters must be set up which differ for each user. Therefore a complex configuration must be made before a prediction method can be used. A hybrid predictor can reduce the configuration overhead. Such a hybrid predictor uses different prediction methods or configurations in parallel yielding different prediction results. A selector selects a prediction result from the result set of the base predictors. We propose and evaluate three principal hybrid predictor approaches – the warm-up predictor, the majority predictor, and the confidence predictor – with several variants. The hybrid predictors reached a higher prediction accuracy than the average of the prediction accuracies of the separately used prediction methods.

1 Introduction

A context aware application should be customized to the user’s preferences. Furthermore the application could take over decisions from humans which are dictated by human habits. Therefore next contexts should be predicted such that the system can act proactively. There are many techniques to predict the next context of a user: neural networks, Bayesian networks, Markov models, and state predictors are representatives. The problem of the techniques are the numerous parameters required to set up the methods. For every user the individual parameters must be found.

We investigated in our previous work different methods for context prediction especially next location prediction, e.g. different state predictors [11], the multi-layer perceptron [15], and a dynamic Bayesian networks [13]. We investigated for all techniques different parameter settings. The state predictors were evaluated with local and global patterns based on orders of 1 to 5. We also evaluated in this investigation the Markov predictor with the same parameters, pattern types and orders. Neural networks require several parameter settings for the structure like e.g. the number of input neurons and for the learning algorithm like e.g. the learning rate. Parameters for the Bayesian network could be the history length and different time parameters like weekday and time of day. We observed that the parameters that deliver the best prediction accuracy differ for each person.
In this paper we investigate hybrid predictors. A hybrid predictor consists of a set of base predictors which will be used to predict the next context. All base predictors deliver a prediction result and a selector determines the result of the hybrid predictor from the set of delivered results. The advantages could be a better prediction accuracy or with the same prediction accuracy a higher quantity. A disadvantage of the hybrid predictor is the additional expenses in storage and computing costs opposite to the single predictors.

Two critical choices must be made for a hybrid predictor: the choice of the set of base predictors and the selector. Our evaluations are done with a large number of base predictors from the state and Markov prediction methods (see table 2). We introduce for the selector three principal possibilities – a warm-up selector, a majority selector, and a confidence selector – and discuss several variants.

The paper is structured as follows. The next section describes the base predictors. The three proposed hybrid predictors are introduced in sections 3 to 5. Section 6 shows the evaluation results. Related work is described in section 7. The paper ends with a conclusion.

2 Base Predictors

All known prediction techniques like neural networks or Bayesian networks can be used as base predictors in the predictor set. But in our evaluation we investigated only Markov and state predictors as base predictors of hybrid predictors.

2.1 Markov Predictor

Markov models seem a good approach for the next location prediction based on location histories. A Markov model regards a pattern of the last visited locations of a user to predict the next location. The length of the regarded pattern is called the order. Thus a Markov model with order 3 uses the last three visited locations for prediction of the next visited location. For all patterns the probabilities of the next location change is stored which is calculated from the whole sequence of the locations visited by the user. A simple Markov model is the Markov predictor [3, 14]. A Markov predictor stores for every pattern the frequencies of the next visited locations. That means if a location appeared 5 times after a pattern in the whole sequence the predictor stores 5 for this location and this pattern. Figure 1 shows a Markov predictor with order 3 for the sequence

\[FSBFBSFBSFBS\]

The pattern of the last three locations is \(FBS\). The number for \(F\) is the maximum for the entry \(FBS\). Therefore the prediction is the location \(F\).

A disadvantage of the Markov predictor is that it can count up the frequency of a location to a high number, for example to 1,000. Before another location can be predicted the Markov predictor must count up the frequency for the other location again to 1,000. The state predictors prevent this problem.
2.2 State Predictor

The state predictor [11] uses a finite automaton which is called two-state predictor for every pattern thus replacing the frequency counter of the Markov predictor. Figure 2 shows a so-called global two-level two-state predictor with order 3 that uses the two-state predictor. The sequence is the same as in the case of the Markov predictor in the last subsection. Figure 3 shows a two-state predictor for three possible next locations for a fixed pattern.

The denotations of the states consist of the id of the location and a counter. We consider the pattern BSF. If a person enters for the first time the boss’ office B after the pattern BSF, the state B0 is set for this pattern. If the person moves in the pattern again, the office of the boss B is predicted as next location. If the prediction proves as correct, the predictor switches into the strong state B1 for BSF. Thus, next time the office of the boss B will be predicted again. If the person interrupts her habit once by entering a room different from the
boss’ office after BSF, the predictor changes into the state $B_0$ predicting still the office of the boss. If the person now enters the secretariat $S$ after BSF the predictor switches into the state $S_0$ independently of the room entered after BSF before, and predicts thus the secretariat as next.

2.3 Local Predictors

The Markov predictor and the state predictor above considered a global location history. A local Markov predictor or a local two-level two-state predictor uses only history of neighbor locations and disregards the movement sequences through all locations. The shift register does not contain the global history, i.e. which locations were entered before, but the local history, meaning the locations, which the user visited after the current location. For each location exists a two-level two-state predictor or a Markov predictor, i.e. the successively entered neighbor locations form the pattern for each location. Otherwise the local predictors operate like the global predictors.

2.4 Confidence Estimation

The confidence predictor uses the confidence of the base predictor to select the prediction result. In [12] several confidence estimation techniques for the state predictor were proposed. By using the confidence estimation the prediction accuracy could be significantly improved. The confidence will be considered separately for every pattern in the pattern history table of the two-state predictor and the Markov predictor respectively.

The first approach is the strong state method which can only be used with the two-level two-state predictor. The method provides two confidence levels. If the current state of a two-state predictor is a weak state the predictor is classified into the low confidence. Otherwise if a two-state predictor is in a strong state the confidence of the predictor is classified as high. Now the prediction result of the predictor will only be supplied if the confidence is high. Otherwise the result is detained.

The second confidence estimation method is the threshold method which is independent of the prediction techniques. The threshold method considers the prediction accuracy of the last predictions. If this accuracy exceeds a specified threshold the predictor’s confidence is classified as high. Otherwise the confidence of the predictor is classified low.

The third proposed method – the confidence counter method – is also independent of the used prediction algorithm. This method estimates the prediction accuracy with a saturation counter. Figure 4 shows a two-bit counter that consists of 4 states and 4 confidence levels respectively.

The initial state is state 10. Let $s$ be the current state of the confidence counter. If a prediction result is proved as correct ($c$) the counter will be incremented, that means the state graph changes from state $s$ into the state $s + 1$. If $s = 11$ the counter keeps the state $s$. Otherwise if the prediction is incorrect ($i$) the counter switches into the state $s - 1$. If $s = 00$ the counter keeps the state
Fig. 4. Confidence counter

s. If the counter is in the state 11 or in the state 10 the predictor is assumed as confident, otherwise the predictor is unconfident and the prediction result will not be supplied.

3 Warm-up Predictor

On one hand there are predictors with a short learning time and low prediction accuracy, e.g. state or Markov predictors with a small order. On the other hand predictors exist which need a long time for learning but provide a better prediction accuracy e.g. predictors with a high order. The idea of the warm-up predictor is now to combine a fast learning predictor and a slow learning predictor. In the warm-up phase, where the slow learning predictor cannot deliver a result since a context pattern occurs the first time, the fast learning predictor is used to predict the next context. Successively, the better slow learning predictor is used. This principle will be implemented by Prediction by Partial Matching (PPM) and Simple Prediction by Partial Matching (SPPM) respectively.

The Prediction by Partial Matching (PPM) [3] from the area of data compression works as follows in the case of state predictors: A maximum order $m$ is applied in the first level instead of the fixed order. Then, starting with this maximum order $m$, a pattern is searched according to the last $m$ locations. If the pattern matches, the two-state predictor of this pattern is used to predict the next location. If no pattern of the length $m$ is found, the pattern of the length $m-1$ is looked for, i.e. the last $m-1$ locations. This process can be accomplished until the order 1 is reached. If even a predictor of order 1 doesn’t generate any prediction there will be no prediction. We propose to stop the PPM with order 1 because location prediction doesn’t make sense when the current location isn’t known.

A simplification of the PPM is the Simple Prediction by Partial Matching (SPPM). Also here a maximum order is applied in the first level. If no pattern with the length of the maximum order is found, the pattern of the length 1 is looked for. If a predictor of order 1 doesn’t generate a prediction there will be no prediction.

The SPPM method reflects the idea of the warm-up predictor perfectly. The method combines a predictor with a high order with a predictor with a small order. If the predictor with maximum order cannot make a prediction because the current pattern occurs the first time, the predictor with order 1 is used to predict the next context. Figure 5 demonstrates the SPPM method using the two-level two-state predictors with order 5 and 1. The colored areas indicate that a prediction can be made. By using the predictor with order 1 the vacuity (no prediction possible) of the predictor with order 5 will be reduced. The SPPM
predictor reaches the same high prediction quantity like the predictor with order 1, however, with potential much better prediction results.

Fig. 5. Warm-up predictor

Also the PPM method is a warm-up predictor. The predictor set of a PPM predictor with maximum order 5 contains the predictors with order 1, 2, 3, 4, and 5. A disadvantage of the SPPM method is the use of the extremes - order 5 and order 1. The PPM method features a nuance between these two extremes.

4 Majority Predictor

The selector of a majority predictor considers only the predictors from the predictor set which deliver a prediction result. The prediction result of the hybrid predictor is the result which obtains the majority in the set of results delivered by the base predictors. There are three majority selection principles - the relative majority selector, the simple majority selector, and the absolute majority selector. A relative majority selector gives the result which appears most frequently in the set of results. A result is selected by the simple majority selector when more than half of the predictors deliver this result. If no result reaches the simple majority the hybrid predictor doesn’t deliver a prediction. The absolute majority selector works like the simple majority selector but it considers all base predictors including the predictors which can’t deliver a prediction result.

In the following we give a formal description of the majority predictor. Let $P$ be the set of base predictors. Let $P_r$ with $|P_r| \leq |P|$ be the set of base predictors which deliver a prediction result, and let $R$ be the set of prediction results of the predictors from $P_r$. Note that $R$ is a multi-set. Furthermore $|M|_x$ defines the number of the value $x$ in a multi-set $M$. So the majority predictor with relative majority delivers $a$ as prediction result if

$$|R|_a \geq |R|_b \quad \forall b \in R$$

Figure 6 shows a majority predictor with relative majority whose predictor set $P$ contains seven predictors. In the example five predictors deliver a result,
that means \( P_r = \{ P_1, P_3, P_4, P_6, P_7 \} \). The prediction results of these five predictors form the multi-set \( R = \{ A, C, C, D, B \} \). So it does apply \( |R|_C = 2 > 1 = |R|_x \quad \forall x \in R \setminus \{ C \} \). Thus the prediction result is \( C \).

The majority predictor with \textit{simple} majority selector delivers \( a \) as prediction result if

\[
|R|_a > \frac{|R|}{2}
\]

If there isn’t such \( a \) the hybrid predictor delivers no result. That means the predictor with simple majority makes less predictions than the predictor with relative majority. The example in figure 6 shows this case, the predictor with simple majority cannot deliver a prediction result. Figure 7 shows another example. Here the majority predictor with simple majority delivers a result.
In the example five predictors provide a prediction, that means \( P_r = \{ P_1, P_3, P_4, P_6, P_7 \} \). The prediction results of the five predictors form the multi-set \( R = \{ A, C, C, C, B \} \). Furthermore \( |R|_C = 3 > 2,5 = \frac{|R|}{2} \), so \( C \) is the prediction of the hybrid predictor. Of course the predictor with relative majority will also deliver this result.

The next step would be a predictor with absolute majority. Therefore the result of the hybrid predictor is the result which reaches the majority in the whole set of the predictors. A majority predictor with \textit{absolute} majority delivers \( \text{a} \) as prediction result if
\[
|R|_a > \frac{|P|}{2}
\]

This kind of majority predictor again provides less predictions than the other two. Figure 8 show a example where the predictor with absolute majority can make a prediction.

Let \( V_{rel} \) be the set of predictions delivered by a predictor with relative majority in a test run, \( V_{sim} \) the set of predictions delivered by a predictor with simple majority, and \( V_{abs} \) the set of predictions delivered by a predictor with absolute majority, then the following formula applies
\[
V_{abs} \subseteq V_{sim} \subseteq V_{rel}
\]

The results delivered by the predictor with absolute majority will be also delivered by the predictors with relative and simple majority. The results delivered by the predictor with simple majority will be also delivered by the predictor with relative majority. Figures 6 and 7 show examples where the majority predictors with absolute majority selection principle doesn’t deliver a result. Due to its small quantity of results reached, we omit the absolute majority selection principle in our evaluations.
5 Confidence Predictor

In [12] we proposed three confidence estimation methods: the strong state, the threshold, and the confidence counter methods. These confidence estimation methods can be applied as selector principles yielding three types of confidence hybrid predictors.

Figure 9 shows the operation of a confidence predictor. First every predictor from the base predictor set is making a prediction. The results of the predictors will be classified by one of the confidence method into confidence levels. The predictors on the highest level will be used to predict the next location. One of the proposed majority method is used to select the result from the set of results of the predictors on the highest confidence level.

In the use of a confidence predictor there are two decision steps between the predictions of the base predictors in the predictor set and the final prediction of the hybrid predictor. Primary, which confidence estimation method will be used. Secondary, a majority method must be chosen. In the following we will consider first the primary selection criterion. Then we will describe the secondary criterion and a possible extension, the tertiary selection criterion.

**Primary Selection Criterion** The confidence of the predictors can be calculated by the methods described in section 2.4, i.e. the strong state, the threshold, and the confidence counter methods. Thereby the confidence of every base predictor of the predictor set must be calculated by the same confidence method, since a comparison of the confidence level of different confidence methods isn’t possible.

The strong state method can only be used with a set of state predictors. There are only two confidence levels. A predictor is in a weak state corresponding with
a low confidence or a predictor is in a strong state corresponding with a high confidence. By using the threshold method the predictor set can contain all kind of predictors. The threshold method provides theoretically an infinite number of confidence levels since the levels correspond with the prediction accuracies of the predictors. Also by using the confidence counter method the predictor set can contain all kinds of predictors. For the confidence counter method there are a discrete number of confidence levels. Every state of the counter corresponds to a confidence level.

Secondary Selection Criterion After the classification of the prediction results in confidence levels the predictor on the highest confidence level should be selected. In most cases this won’t be a single predictor but many predictors will reach the highest level. Therefore a decision must be made by a majority method which prediction result is used for the hybrid predictor. For the majority we proposed three methods in the last section. The relative or the simple majority should be used in the confidence predictor. First of all by using the strong state method and the confidence counter method the secondary criterion will be deciding.

Tertiary Selection Criterion If the highest reached confidence level is very low, there is the problem that an unconfident predictor or a set of unconfident predictors are used for the prediction of the hybrid predictor. For example the strong state method is used and all predictors are in a weak state and thus they are classified as unconfident.

To eliminate this problem a tertiary criterion can be used which decides by a barrier whether the prediction result will be selected. By using a barrier we consider only predictors whose confidence is equal or higher than the barrier. For example, if the strong state method and the barrier is used, only predictors in the strong state are considered.

If the highest reached confidence level is equal or higher than the barrier the operation of the hybrid predictor is the same like without the use of the barrier. Otherwise if no predictor reaches a confidence level equal or higher than the barrier, the hybrid predictor detains the prediction result.

Summary Selection Criterions Table 1 summarizes the selection possibilities for using a confidence predictor. There are 12 variants of the confidence predictor.

6 Evaluation

6.1 Evaluation Methodology

We used two evaluation data sets: the Augsburg Indoor Location Tracking Benchmarks [9, 10] and the Nokia Context Data [4]. The Augsburg Benchmarks are a collection of movement data of four persons over seven months. The persons
moved through the offices and stored manually the office number and the entry
time on their PDAs. The Nokia Context Data include two location information,
the cell id and the location area code of a person who travels through different
cellular phone cells.

For the comparison of the hybrid predictor approaches we calculated the
prediction accuracy based on known patterns, that means if a pattern occurs
the first time the predictor cannot make a useful prediction. We assume that
a prediction will be requested for every location change. So the number of re-
quested predictions corresponds to the number of location changes. The number
of requested predictions \( p = p_d + p_i \) consist of deliverable predictions \( p_d \) and
impossible predictions \( p_i \) based on empty pattern history tables. The deliver-
able predictions depend on patterns which occurred at least once. Let \( c \) be the
number of correct predictions. The prediction accuracy \( a \) will be calculated as
follows:

\[
a = \frac{c}{p_d}
\]

With a higher order the number of deliverable predictions decreases \( p_d \). The
quantity \( q \) expresses this fact:

\[
q = \frac{p_d}{p}
\]

State predictors with different orders were used for the evaluation of the
warm-up predictors. To evaluate the majority and the confidence predictors a
set of predictors are needed. In the evaluation we used only state and Markov
predictors in the predictor sets. Table 2 shows the four predictor sets \( P_1 \) to \( P_4 \)
which will be used in the subsections 6.3 and 6.4.

The use of the PPM and SPPM predictors in the set \( P_1 \) is remarkable since
PPM and SPPM are hybrid predictors themselves. This shows that the predictor
set of hybrid predictors can again contain hybrid predictors.
Table 2. Predictor sets for the evaluation of the majority and the confidence predictors

<table>
<thead>
<tr>
<th>Predictor Sets</th>
<th>( P_1 )</th>
<th>( P_2 )</th>
<th>( P_3 )</th>
<th>( P_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>local one-level one-state predictor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>local one-level two-state predictor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>local two-level two-state predictor order 1 - 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>local two-level two-state predictor PPM(5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>local two-level two-state predictor SPPM(5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>local two-level Markov predictor order 1 - 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>local two-level Markov predictor PPM(5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>local two-level Markov predictor SPPM(5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>global two-level two-state predictor order 1 - 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>global two-level two-state predictor PPM(5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>global two-level two-state predictor SPPM(5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>global two-level Markov predictor order 1 - 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>global two-level Markov predictor PPM(5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>global two-level Markov predictor SPPM(5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.2 Warm-up Predictor

We evaluated the PPM and SPPM in four variants: local two-level two-state predictors, global two-level two-state predictors, local Markov predictors, and global Markov predictors all with maximum order 5. These we compared to the state
and Markov predictors with order 1, 2, 3, 4, and 5. Figure 10 shows the measurements of the location prediction of person D of the Augsburg Benchmarks and the prediction of the cell id of the Nokia Context Data.

The local base predictors work better than the corresponding local PPM and SPPM predictors. The global warm-up predictors work better than the global predictors with order 5 for person D of the Augsburg Benchmarks. In the case of the cell id of the Nokia Context Data the global warm-up predictors perform worse than the global predictor with order 5 since the global predictors with order 1 show a low prediction accuracy.

In the comparison of PPM and SPPM we see a better performance of the SPPM Markov predictor than the PPM Markov predictor for person D. For the state predictor the performance of SPPM and PPM is reversed. In the case of the cell id the PPM variant works in both cases – the state and the Markov predictor – better than the SPPM variant, since the predictors with order 1 show a low prediction accuracy.

The charts show that the warm-up predictors yield no improvement in the used scenarios.

### 6.3 Majority Predictor

Table 3 shows the prediction accuracies of the majority predictors with the Augsburg Benchmarks and the Nokia Context Data. Furthermore the quantity was measured. The measurements base on the four predictor set $P_1$ to $P_4$. The best prediction accuracy for every row is highlighted in the table.

<table>
<thead>
<tr>
<th></th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
<th>$P_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rel. sim.</td>
<td>rel. sim.</td>
<td>rel. sim.</td>
<td>rel. sim.</td>
</tr>
<tr>
<td>Person A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (in %)</td>
<td>51.17</td>
<td>53.13</td>
<td>48.83</td>
<td>55.47</td>
</tr>
<tr>
<td>Quantity (in %)</td>
<td>76.86</td>
<td>82.22</td>
<td>74.23</td>
<td>77.55</td>
</tr>
<tr>
<td>Person B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (in %)</td>
<td>50.78</td>
<td>51.07</td>
<td>48.65</td>
<td>50.92</td>
</tr>
<tr>
<td>Quantity (in %)</td>
<td>69.51</td>
<td>71.67</td>
<td>68.98</td>
<td>70.10</td>
</tr>
<tr>
<td>Person C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (in %)</td>
<td>46.23</td>
<td>46.87</td>
<td>44.62</td>
<td>48.96</td>
</tr>
<tr>
<td>Quantity (in %)</td>
<td>63.54</td>
<td>65.24</td>
<td>63.73</td>
<td>62.13</td>
</tr>
<tr>
<td>Person D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (in %)</td>
<td>52.65</td>
<td>54.62</td>
<td>50.88</td>
<td>54.03</td>
</tr>
<tr>
<td>Quantity (in %)</td>
<td>71.71</td>
<td>75.63</td>
<td>72.68</td>
<td>71.62</td>
</tr>
<tr>
<td>Cell ID</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (in %)</td>
<td>67.11</td>
<td>67.75</td>
<td>66.33</td>
<td>67.11</td>
</tr>
<tr>
<td>Quantity (in %)</td>
<td>81.12</td>
<td>82.86</td>
<td>83.33</td>
<td>82.37</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (in %)</td>
<td>77.11</td>
<td>73.63</td>
<td>73.63</td>
<td>74.13</td>
</tr>
<tr>
<td>Quantity (in %)</td>
<td>82.14</td>
<td>81.70</td>
<td>82.07</td>
<td>81.46</td>
</tr>
<tr>
<td>Area Code</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (in %)</td>
<td>97.10</td>
<td>97.10</td>
<td>97.10</td>
<td>97.10</td>
</tr>
<tr>
<td>Quantity (in %)</td>
<td>81.16</td>
<td>73.91</td>
<td>70.05</td>
<td>72.95</td>
</tr>
</tbody>
</table>
The consideration of the quantity shows that the majority predictor with relative majority reaches a high quantity, since when one predictor from the predictor set could deliver a result, the hybrid predictor is delivering a result. The predictors with simple majority reaches only a low quantity between 32\% and 50\%. For person A and C these values are higher than the quantity of the used predictors in the set which reach better prediction accuracies as the hybrid predictor.

For the Augsburg Benchmarks the predictors with the simple majority reach a better prediction accuracy than the predictors with relative majority. The best prediction accuracy for the Augsburg Benchmarks were reached with the predictor set $P_2$ – state and Markov predictors. For person D we can compare the result with the predictor from the predictor set $P_2$ (see figure 10). Here only the local two-level two-state and the local Markov predictors with order 4 and 5 work better. But the quantity of these predictors are lower than the quantity of the hybrid predictor.

For the Nokia Context Data the predictors with simple majority perform also better than the predictors with relative majority. For the prediction of the cell id the predictor with predictor set $P_3$ – the state predictors – delivers the best result which is better than all single results (see figure 10). The predictor with the set $P_1$ reaches the best result for the location area code.

These result show that the majority predictor with relative majority has no advantage opposite the single predictors. The predictor with simple majority reaches an accuracy which is clearly higher than the average of the accuracies of the single predictors. A good predictor set seems to be the set $P_2$. This set includes the state and Markov predictors as well as the local and global variants. This set doesn’t reach the best result for the Nokia Context Data, but the reached result was a little bit lower than the best result.

### 6.4 Confidence Predictor

The strong state confidence method can only be used with the predictor set $P_3$ which contains only state predictors. In the evaluation we investigated all variants of the confidence predictor shown in table 1. The threshold method and the confidence counter method were evaluated with the four sets. The comparison of the four predictor sets showed that the confidence predictors reach similar prediction accuracies. The results showed only discrepancies of one percent. For this reason we consider in the following only the predictor set $P_2$ which provides a combination of the advantages of the local and global predictors as well as the state and the Markov predictors.

Figure 11 shows the reached prediction accuracies for person D of the Augsburg Benchmarks with predictor set $P_2$ for the threshold method and the confidence counter method as well as the accuracies reached with the strong state method and predictor set $P_3$. Figure 12 shows the same measurements for the cell id of the Nokia Context Data.
Fig. 11. Confidence predictors for person D

Fig. 12. Confidence predictors for cell id
**Tertiary Selection Criterion**  In the measurements the threshold method use a barrier of 60% and the confidence counter use four states and the third state as barrier. The confidence predictor using the simple majority with and without barrier perform identically for the threshold method and the confidence counter method in most cases. Small discrepancies were found in the predictor sets $P_3$ and $P_4$. In most cases the use of the barrier and simple majority delivers the best result as expected. The identical results with simple majority with and without barrier show, that if more than half of the predictors on the highest confidence level deliver the same prediction result, then this confidence level is equal or higher than the used barrier.

For the strong state method a significant increase can be noticed from simple majority without barrier to simple majority with barrier. The reason is that this method has only two confidence levels. The probability is higher that all predictor are classified in the lower level. The use of the barrier with the relative majority improves the prediction accuracy, but opposite the simple majority the results are very bad.

**Secondary Selection Criterion**  As expected from the investigation of the majority predictors the confidence predictors with simple majority perform essentially better than the predictors with relative majority. Predictors with relative majority selector work worse than the single predictors of the predictor set using the same confidence method. But the confidence predictors with simple majority are always better than the single predictors using the same confidence method.

**Primary Selection Criterion**  When the three confidence methods are compared, there isn’t a winner. If we consider person D the predictor using the threshold method delivers the lowest prediction accuracy. The use of the confidence counter method reaches the highest accuracy. In the case of the cell id the threshold method delivers the best result and the strong state method shows the lowest accuracy. Also for the other persons in the Augsburg Benchmarks and the location area code no method of the three confidence methods was the winner. If we consider additionally the quantity the threshold method shows a disadvantage opposite the other methods.

All methods show that a higher prediction accuracy can only be reached with a lower quantity. A big difference exits between the quantity with relative majority and simple majority. Here the threshold method performs worst. The reason can be the continuous confidence level, so only few predictors from the predictor set reach highest confidence level.

A short summary for the confidence predictors follow. The use of predictor set $P_2$ delivers mostly the best results. The predictors with the simple majority and the barrier reach the highest prediction accuracy in all measurements. The threshold method shows always the lowest quantity which don’t seems acceptable.
7 Related Work

Most context prediction approaches apply only a single prediction method [7, 8, 1, 6, 2]. To our knowledge no hybrids are used in this application domain up to now.

The basic state predictor methods, the confidence estimation methods, and the hybrid predictor approach are motivated by branch prediction methods from the area of processor architecture. The idea of a warm-up predictor was proposed by Young and Smith [16] for hybrid branch predictors. During the warm-up phase of a hybrid branch predictor a static or a simple dynamic predictor should be used. This idea was improved by our warm-up predictor since our warm-up predictor can switch between the complex and the simple predictor already during the warm-up phase (see figure 5). In the processor architecture domain a static predictor is used for the first $n$ million branches. After this $n$ million branches the adaptive predictor is used only. The idea of the confidence predictor is based on the confidence estimation for branch prediction by Grunwald et al. [5]. For a hybrid branch prediction the predictor is used which reached the highest confidence level. The confidence in this case was the up to now reached prediction accuracy. This corresponds with our confidence predictor with threshold method.

8 Conclusion

For prediction methods a lot of parameters must be set up which differ for each user. Therefore a complex configuration must be made before a prediction method can be used. The paper presented three hybrid approaches which reduce the configuration overhead by automatically selecting the probably best result from differently configured base predictors and base predictors with different prediction methods respectively.

The warm-up predictor didn’t achieve the desirable improvements. Opposite this the majority and confidence predictors reached an increase of the prediction accuracy or a prediction accuracy which was higher than the average of the prediction accuracies of the single predictors of the base predictor set. The decision of the majority in both cases – the majority predictor and the confidence predictor – fall on the simple majority. For the confidence predictor there wasn’t a favorable for the confidence method. The use of the barrier achieves a little improvement for the strong state method. For the other methods there wasn’t an improvement. The confidence predictors reach always a slightly better prediction accuracies than the majority predictors. But the majority predictors show a significant better quantity than the confidence predictors in all tests. $P_2$ seems a good candidate for the predictor set. $P_2$ combines the advantages of the local and global as well as of the state and the Markov predictors.

Future investigations of hybrid approaches could select also other prediction techniques like neural networks or Bayesian networks in the base predictor set.
References